The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong

Department of Computing

An Accurate and Robust Indoor Localization System

By Chan Chun Lun, Eddie

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

MARCH 2010

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

___________________________________ (Signed)

_________Chan Chun Lun, Eddie___________ (Name of student)

Abstract

Abstract of dissertation entitled: An Accurate and Robust Indoor Localization System, Submitted by Chan Chun Lun, Eddie for the degree of Doctor of Philosophy in the Department of Computing at The Hong Kong Polytechnic University in April 2010.

A wireless tracking system such as the Global Positioning System (GPS) is the most effective in relatively open and flat outdoor environments but is much less effective in non-line-of-sight (NLOS) environments such as hilly, mountainous, or built-up areas. In recent years, IEEE 802.11 Wireless Local Area Networks (WLANs) have been widely deployed and are now fuelling a wide range of location-aware computing applications. The localization of devices in these networks (i.e., estimating their location) makes use of the Wi-Fi signal strength. There are two location-sensing techniques for indoor environments, propagation-based and location-fingerprinting techniques.

Although these two techniques can locate WLAN-enabled devices, recent research on indoor localization system is still not satisfactory. There are five major problems that lead to inaccurate and low efficiency localization systems. First, signal overlaps and interferences seriously worsen the performance of localization system. Second, existing positioning algorithms suffer from high computational complexity of location estimation as a result of the burden of having to carry out many signal sampling, weighting, and

filtering tasks. Third, existing WLAN infrastructures are often deployed in an ad-hoc, empirical and non-optimal configurations. Such unstructured approaches lead to poor resource utilization and inaccurate localization due to signal overlap. Fourth, there is a lack of analytical models that can be used as a framework for designing and deploying the positioning systems. Most existing models ignore radio signal properties and some assume the distribution of the RSS is Gaussian and pair wise. Such assumptions may ignore or distort the real behavior of RSS and lead to RSS analysis that is not accurate enough. Finally, a localization system should provide location-aware information according to the user's needs. However, it is very difficult to retrieve location-aware information that matches the behavior of users and at the same time suits current conditions. An accurate, robust, scalable and cost-effective localization solution should have (1) a stable WLAN signal transmission, (2) an accurate positioning algorithm, (3) a structural WLAN infrastructure that can support intensive tracking, (4) a model that can visualize the WLAN signal distribution to prevent the occurrence of signal black spots and interference and finally (5) a query interpretation and information retrieval system that can provide location-aware information suitable to user behaviors and preferences.

This thesis proposes an indoor localization framework that tackles each of these problems. First, it proposes a cell-based installation of wireless access points and suggests a channel assignment scheme to have a stable WLAN signal transmission. Second, it makes use of the Newton Trust-Region algorithm and Kalman Filter to improve the accuracy of positioning algorithms. Third, it suggests the use of Fuzzy Logic and Topographic modeling to visualize the signal distribution in a 2-D and 3-D manner. Finally, it develops an agent-based module for retrieving location-aware information that can improve the speed of retrieval while maintaining or even improving the accuracy by making use of semantic information in the data to develop smaller training sets.

A series of experiments was carried out to measure the performance of the proposed framework and contrast the result with existing approaches. The major findings are that the proposed framework could help engineers to save 50% of WLAN infrastructure resources (access points) while at the same time increasing the accuracy of localization by 20%. Experimental analysis also shows that channel interference usually obeys a right-skewed distribution and positioning accuracy is greatly affected by channel interference between access points. Finally, the proposed framework provides a quick reference and efficient analytical tool for improving the design of WLAN infrastructure.

List of Publications

Journal Papers

[1] Eddie C. L. Chan, George Baciu and S.C. Mak, "Properties of Channel Interference for Wi-Fi Location Fingerprinting", IEEE Journal of Communications Software and Systems, vol. 6, no. 2, 2010 (JCOMSS)

[2] Eddie C. L. Chan, George Baciu and S.C. Mak, "Cognitive Location-Aware Information Retrieval by Agent-based Semantic Matching", International Journal of Software Science and Computational Intelligence (IJCiNi), vol. 2, no. 3, pp.21-31, July-September, 2010

[3] Eddie C. L. Chan, George Baciu and S.C. Mak, "Using a Cell-based WLAN Infrastructure Design for Resource-effective and Accurate Positioning", IEEE Journal of Communications Software and Systems, vol. 5, no. 4, pp.117-127, 2009 (JCOMSS)

[4] Eddie C. L. Chan, George Baciu and S.C. Mak, "Newton Trust Region Method for Wi-Fi Tracking", submitted to IEEE Transactions on Mobile Computing

[5] Eddie C. L. Chan, George Baciu and S.C. Mak, “Wireless Signal and Information Tracking Using Fuzzy Logic”, submitted to Spring-Verlag Studies in Computational Intelligence.

[6] Rong-Hua, Li Zhiping Luo, Eddie C. L. Chan, George Baciu, Zheng-Jun Zha, Jinhui

Tang, Tat-seng Chua, “Equivalence Between Direct LDA and LDA/QR”, submitted to Pattern Recognition Journal

Conference Papers

[7] Eddie C. L. Chan, George Baciu and S.C. Mak, "Effect of Channel Interference on Positioning in Wireless Local Area Network", accepted by 6th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. IEEE WiMOB 2010. Niagara Falls, Canada.

[8] Eddie C. L. Chan, George Baciu and S.C. Mak, "Orientation-based Wi-Fi Positioning on the Google Nexus One", accepted by 6th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. IEEE WiMOB 2010. Niagara Falls, Canada.

[9] Rong-hua Li, Eddie, C. L. Chan, George Baciu, “DLDA and LDA/QR Equivalence Framework for Human Face Recognition”, accepted by 9th IEEE International Conference on Cognitive Informatics, IEEE ICCI 2010. Beijing, China.

[10] Shuang Liang, Eddie C. L. Chan, George Baciu, Rong-Hua Li, “Cognitive Garment Design Interface Using User Behavior Tree Model”, accepted by 9th IEEE International Conference on Cognitive Informatics, IEEE ICCI 2010. Beijing, China.

[11] Eddie C. L. Chan, George Baciu and S.C. Mak, "Wi-Fi Positioning based on Fourier Descriptors", IEEE International Conference on Communications and Mobile

Computing 2010, vol. 3, pp.545-551. IEEE CMC 2010. Shenzhen, China

[12] Shuang Liang, Rong-Hua Li, George Baciu, Eddie C. L. Chan and D. Zheng, "Partial Matching of Garment Panel Shapes with Dynamic Sketching Design", 1st ACM Augmented Human International Conference, article 16, pp. 1-6. AH 2010. Megève, France

[13] Eddie C. L. Chan, George Baciu and S.C. Mak, "Using the Newton Trust-Region Method to Localize in WLAN Environment", 5th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pp.363-369. IEEE WiMOB 2009. Marrakech, Morocco.

[14] Eddie C. L. Chan, George Baciu and S.C. Mak, "Fuzzy Topographic Modeling in WLAN Tracking Analysis", International Conference on Fuzzy Computation, pp. 17-24. ICFC 2009. Madeira, Portugal. **

[15] Eddie C. L. Chan, George Baciu and S.C. Mak, "Resource-effective and Accurate WLAN Infrastructure Design and Localization Using a Cell-structure Framework", 5th IEEE International Conference on Wireless Communications, Network and Mobile Computing, volume 6, pp.9-15, 2009. Beijing, China.

[16] Eddie C. L. Chan, George Baciu and S.C. Mak, "Cognitive Location-Aware Information Retrieval by Agent-based Semantic Matching", 8th IEEE International Conference on Cognitive Informatics, pp. 435-440 IEEE ICCI 2009, Hong Kong.

[17] Eddie C. L. Chan, George Baciu and S.C. Mak, "Using Wi-Fi Signal Strength to Localize in Wireless Sensor Networks", IEEE International Conference on Communications and Mobile Computing, pp. 538-542. IEEE CMC 2009. Kunming, Yunnan, China.

[18] Eddie C. L. Chan, George Baciu and S.C. Mak, "Wireless Tracking Analysis in Location Fingerprint", 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pp.214-220. IEEE WiMOB 2008. Avignon, France.

[19] Raymond S.T. Lee, Eddie C. L. Chan and Raymond Y. W. Mak, "iJADE Content Management System (CMS) - An Intelligent Multi-agent Based Content Management System with Chaotic Copyright Protection Scheme, Knowledge-Based Intelligent Information and Engineering Systems", pp. 652-658, Springer-Verlag Berlin Heidelberg. 2006. Bournemouth, United Kingdom.

[20] Eddie C. L. Chan and Raymond S.T. Lee, "iJADE reporter: An intelligent multi-agent based context- aware news reporting system", Knowledge-Based Intelligent Information and Engineering Systems, pp. 693-699, Springer-Verlag Berlin Heidelberg. 2005. Melbourne, Australia.

Other Papers

[21] George Baciu, Eddie C. L. Chan, "Where Virtual Mobility and Real Location-aware

Information Merge Using Apple's iPhone", HKSAR Information Technology Innovation Section e-Newsletter 2010, Issue 7, pp.1-4

[22] George Baciu, Eddie C. L. Chan, "Face Recognition Using DLDA and LDA/QR Equivalence Framework", HKSAR Information Technology Innovation Section e-Newsletter 2010, Issue 6, pp.5-8

[23] George Baciu, Eddie C. L. Chan, "Channel Interference Effects in Wi-Fi Positioning", HKSAR Information Technology Innovation Section e-Newsletter, Issue 5, pp.5-7, 2010

[24] George Baciu, Eddie C. L. Chan, "Accurate Positioning Using Newton Trust-Region Method in Wi-Fi Environment", HKSAR Information Technology Innovation Section e-Newsletter, Issue 4, pp.1-4, 2009

[25] George Baciu, Eddie C. L. Chan, "Agent-based Location-Aware Information Retrieval System", HKSAR Information Technology Innovation Section e-Newsletter, Issue 3, pp.5-8, 2009

[26] George Baciu, Eddie C. L. Chan, "Resource-effective and Accurate WLAN Infrastructure Design and Localization Using a Cell-structure Framework", HKSAR Information Technology Innovation Section e-Newsletter, Issue 2, pp.4-7, 2009

[27] George Baciu, Eddie C. L. Chan, "Using Wi-Fi Signal Strength to Localize in Wireless Sensor Networks", HKSAR Information Technology Innovation Section

e-Newsletter, Issue 1, pp.4-6, 2009

[28] George Baciu, Eddie C. L. Chan, "Wireless Tracking Analysis on Wi-Fi Signal Strength", HKSAR Information Technology Solution Centre e-Newsletter, Issue 2, pp. 5-8, 2008.

[29] George Baciu, Eddie C. L. Chan, "Where Virtual Meets Real: Virtual Mobility in the Real World", HKSAR Information Technology Solution Centre e-Newsletter, Issue 1, pp. 2-4, 2008

** Best Student Paper Award

Acknowledgements

First, I would like to express my heartfelt thanks to my supervisor, Prof. George Baciu, for his dedicated teaching and nurturing. I had my knowledge enriched, my skills polished, my creativity enhanced, and my view of science developed. His encouragement and valuable advice have helped me to complete my PhD studies. It has been my pleasure to work on this study under his guidance.

I wish to take this opportunity to express my sincere gratitude to Mr. Martin Kyle for his kind and professional proofreading of this thesis. I am pleased to acknowledge the co-operation and technical support of the staff of the GAMA Lab, which was founded by Prof. Baciu. They provided me with a wealth of ideas to help improve the quality of my research.

I would like to thank the Information Technology Service Center and Department of Computing of The Hong Kong Polytechnic University for providing information about the wireless infrastructure of The Hong Kong Polytechnic University.

Finally, I would like to thank my colleague, Mr. Mak Sin Cheung, who under my instruction helped me to carry out a wireless coverage survey of the PolyU Campus.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

IEEE 802.11 Wireless Local Area Networks (WLANs) are now widely deployed and are fuelling a wide range of location-aware computing applications. Cities all over the world, such as Hong Kong, Taipei, Singapore, Tokyo and New York have adopted "Wi-Fi city" concepts in which users can access the Internet through existing Wi-Fi infrastructure and use location sensing in both indoor and outdoor environments. One kind of indoor location sensing technique is Wi-Fi Positioning Technology, which makes use of Wi-Fi signal strength in a wireless localization system. Outdoor positioning systems such as GPS, and acoustic and light-based approaches, are most effective in relatively open and flat outdoor environments but are much less effective in non-line-of-sight (NLOS) environments such as hilly, mountainous, or built-up areas. Acoustic localization applications have two particular disadvantages: first, they require the sound source to have a high intensity and to be continuously propagated and, second, they can localize only within the area covered by the sound. Wi-Fi positioning may be a useful supplement to any of these localization approaches.

1.1. Research Background

Global Positioning System (GPS) is a fully functional Global Navigation Satellite System (GNSS) developed by the United States Department of Defense. In the early days it was used as a tool for map-making, land surveying, and scientific uses. Nowadays it is more widely used. Some individuals may own a pocket PC or palm phone with GPS functions. However, GPS is limited in that it requires dedicated hardware. It is also very expensive in terms of labor, spectrum and capital costs to implement a specialized infrastructure in indoor areas solely for position location.

Recently, many public places and campuses have deployed the WLANs such as Wi-Fi - IEEE 802.11b. Wi-Fi is a wireless technology brand owned by the Wi-Fi Alliance intended to improve the interoperability of wireless local area network products based on the IEEE 802.11 standards. Wi-Fi networks have become widely deployed and are fuelling a wide range of location-aware computing applications. Accurate user location information enables a wide range of location-dependent applications. A software-only solution can be integrated as a location-sensing module of a larger context-aware application on infrastructure wireless LANs.

The most widely used techniques that Wi-Fi positioning uses to locate Wi-Fi enabled devices are the propagation-based and location-fingerprinting-based (LF-based) techniques. Propagation-based techniques measure a Wi-Fi transmission's received signal strength (RSS), angle of arrival (AOA), or time difference of arrival (TDOA). Propagation-based techniques use mathematical models to determine the location of the device. LF-based techniques locate devices by accessing a database containing the fingerprint (i.e., the RSSs and coordinates) of other devices within the Wi-Fi footprint.

They then calculate their own coordinates by comparing them with those contained in the relevant LF database. Yet, neither of these two techniques is able to achieve an entirely accurate and robust localization.

1.2. Problem Statement

Although there has been a large amount of research on indoor localization systems, the accuracy and robustness of these systems are still not entirely satisfactory and many open problems have not yet been solved. There are five major problems: (1) unstable wireless signal transmission, (2) computational-intensive and inaccurate positioning algorithms, (3) interference caused by unstructured WLAN infrastructure, (4) lack of models for visualizing the signal distribution used to deploy WLAN infrastructure, and (5) a failure to understand user's preferences so as to provide useful location-aware information. In the following, I will briefly describe each of these problems. Figure 1 depicts the five major localization problems.

1. Unstable wireless signal transmission

Unstable wireless signal transmissions are usually caused by interference and by wave propagation problems. Signal overlaps and interferences can seriously worsen the performance of localization systems.

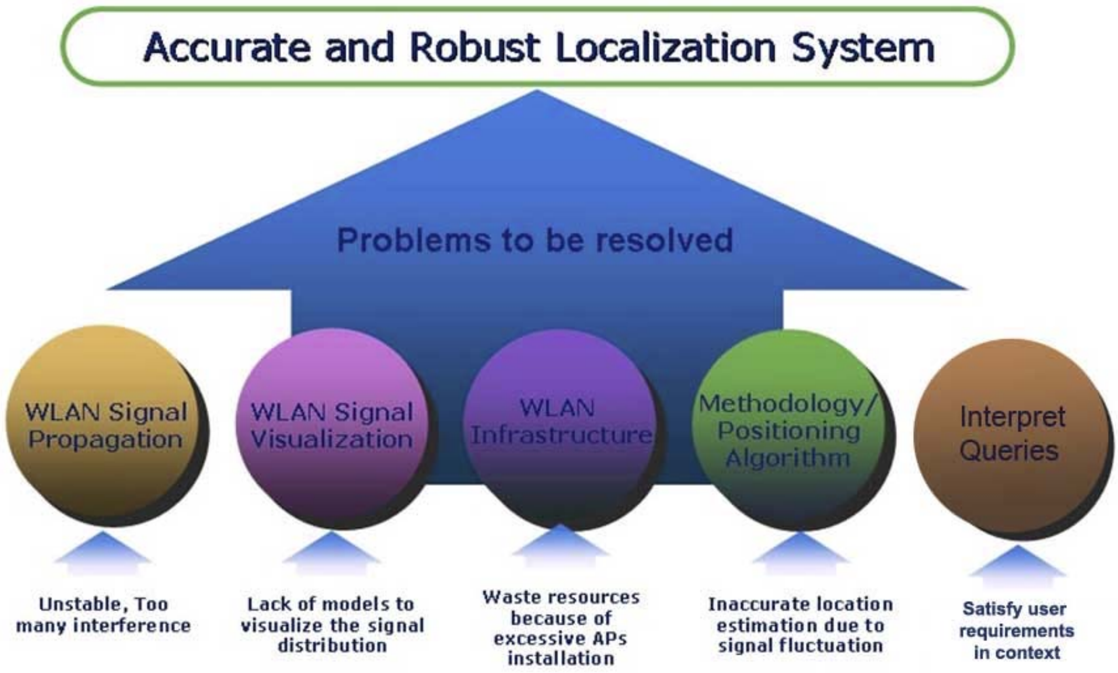


Figure 1 Five major problems of WLAN localization system

A stable and accurate localization depends on a stable wireless signal transmission yet neither the propagation techniques nor the LF-based techniques can guarantee that. LF-based localization techniques [Kaemarungsi and Krishnamurthy 2004a; Taheri, Singh et al. 2004; Li, Wang et al. 2005; Fang, Lin et al. 2008; Kj rgaard and Munk 2008; Swangmuang and Krishnamurthy 2008] require an initial survey with a very large training dataset and each signal sampling is very sensitive to signal fluctuation. Propagation-based localization techniques [Prasithsangaree, Krishnamurthy et al. 2002; Jan and Lee 2003; Kwon, Dundar et al. 2004] must compute every condition that can cause a wave signal to blend. Both techniques are strongly dependent on stable wireless signal transmissions. At the same time, the common unsystematic approach is to place more APs to improve coverage. However, this may still leave blind spots where there are too many access points packed too closely together. This can lead to signal fluctuation and overlap, which is both wasteful and can cause interference [Budianu, Ben-David et al. 2006]

2. Computation-intensive and inaccurate positioning algorithms

Existing positioning algorithms [Kaemarungsi and Krishnamurthy 2004a; Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Li, Wang et al. 2005] suffer from the high computational complexity of location estimation. These algorithms perform signal sampling, weighting, and filtering. Some algorithms [Tiemann, Martin et al. 2000; Chen, Hou et al. 2004; Wang, Zha et al. 2006] make assumptions, such as assuming that WLAN signals obey a Gaussian distribution, even while it has been shown that WLAN signals very often obey a left-skewed distribution [Bahl and Padmanabhan 2000]. Another unsatisfactory assumption is that no loss of energy arises from signals being reflected by obstacles.

Some filter signal algorithms [Liu and Chen 2004; Chang, Hong et al. 2007] use a convergence factor of a trajectory to eliminate the noise from signal strength. Sequential Monte Carlo (SMC) approaches [Hu and Evans 2004; Dil, Dulman et al. 2006] estimate locations by calculating the weighted average of all signal samples. However, these are not effective and have high computational costs in sensor networks. An optimization or filtering algorithms may help to improve the dependency problem of wireless signal transmission.

3. Unstructured WLAN infrastructure

WLAN infrastructure is a crucial part of wireless transmission and localization. WLANs are typically made up of many access points (APs) or nodes. These access points (APs) are manually placed and positioned on the basis of measurements of RSS (received signal strength) taken by engineers empirically. An unstructured approach to WLAN infrastructure design implies poor resource utilization. [Budianu, Ben-David et al. 2006]

WLAN infrastructures are installed and operated in two configurations. The first configuration is in the absence of wireless access points, communicating directly with each other in a peer-to-peer style. The second configuration does make use of wireless access points where all devices on the wireless network communicate with each other and services provided by the network operate through these access points. This configuration is dominant and used by all network providers in a structured environment. However, peer-to-peer connections constantly change as nodes are removed and thus coverage cannot be guaranteed. One part of this study focuses on this second configuration. To provide wireless coverage in a particular area several wireless access points are placed in strategic positions and emit a signal that the clients use to communicate. One of the main issues concerning this second configuration is the placement of the base stations so as to ensure that optimal coverage is provided, meaning that the number or extent of ‘flat spots’ (where no signal is present) is minimal. Incorrect placement of these access points can lead to two problems.

First, when wireless access points are placed too close together, there is signal overlap. Not only does this cause interference, it also poses a potential security risk and increases the cost of the installation since in some cases fewer access points could be used to achieve optimal coverage. The second problem arises when access points are placed too far apart. This potentially increases the number or the extent of flat spots or weak signal areas, leading to unusable connections and poor localization.

4. Lack of signal visualization models

The task of localization is not limited to location estimation but is also carried out using analytical models. For example, WLAN signal visualization models can be used to visualize the distribution of signals and help to improve the design of positioning systems by eliminating WLAN access points (APs), shortening the sampling time of WLAN received signal strength (RSS) in location estimation, and ensuring that all vital areas of a building have wireless coverage.

Currently there are very few support tools available for planning the installation of APs, or to visualize and monitor signal coverage of the installed network. Currently, the available tools are designed only for outside installations, where the wireless signal strength and Global Positioning Service information can be combined to provide feedback.

Most existing methods for modeling location fingerprinting [Kaemarungsi and Krishnamurthy 2004a; Swangmuang and Krishnamurthy 2008] depend on the accurate performance of positioning systems and proximity graphs, such as Voronoi diagrams and clustering graphs. They usually make use of the Euclidean distance to determine positions. Some researchers [Fang, Lin et al. 2008; Kj rgaard and Munk 2008; Swangmuang and Krishnamurthy 2008] have ignored radio signal properties and others have assumed the distribution of the RSS is Gaussian and pair wise. Yet in [Bahl and Padmanabhan 2000], it was shown that the distribution of the RSS is not usually Gaussian but rather left-skewed and the standard deviation varies according to the signal level.

These models should have spatial elements for visualizing the RSS distribution, evaluating and predicting the precise performance of indoor positioning systems based on location fingerprinting. Such an analytical model would support the placement of Wi-Fi APs so as to achieve the maximum throughput.

5. Difficulty of providing useful location-aware information that suits user needs

There are many applications used in a WLAN environment that depend upon accurate and efficient localization. For example, wireless tracking applications nowadays help people navigate using current location-aware information. However, very often the relevant search engines retrieve too much information and present it in a way that is not structured to be optimally useful to users. One of the central purposes of retrieving location-aware information is to provide users with information that is relevant to their current location.

In recent years, researchers have focused on how to provide higher accuracy and faster retrieval based on keyword and textual semantics. However, the results have not been satisfactory. Location-aware information is distributed through different locations. Information, such as traffic information may quickly become out-of-date and it is difficult to update it frequently. Some approaches, such as Naïve-Bayes [Danesh, Moshiri et al. 2007] and K-Nearest Neighbors (K-NN) classifiers [Weiss, Kasif et al. 1996] machine learning approaches ignore semantics when classifying text, leading to unmatched information.

1.3. Objectives and Motivations

The main objective of the present work is to create an accurate and robust indoor localization framework. This objective encompasses five sub-goals. The framework should have (1) a stable WLAN signal transmission, (2) an accurate positioning algorithm, (3) a structural WLAN infrastructure that could support intensive tracking, (4) a model that can visualize the WLAN signal distribution to prevent signal black spots and interference and, finally (5) a retrieval system that can provide location-aware information that matches or responds to user needs. In the following, I will briefly describe each sub-goal. Further details will be discussed in later chapters.

1. A stable WLAN signal transmission

Interference is a major source of signal fluctuation so it is necessary to reduce channel interference between access points. Channel interference can occur because adjacent channels overlap in the frequency spectrum in IEEE 802.11b/g WLAN. Recent research has focused on either creating a new channel assignment scheme [Chiu, Yeung et al. 2006; Haidar, Ghimire et al. 2008; Subramanian, Gupta et al. 2008] or enhancing existing MAC protocols [Dutta, Jaiswal et al. 2007; Zeng, Wang et al. 2007] to reduce channel interference. The objective of this recent work has been to improve the data transmission through wireless networks. However, none of existing work on channel interference has addressed accurate performance and the improvement of location estimation algorithms. In some cases it has even been wrongly suggested that interference might increase the accuracy of positioning.

In Chapter 3, I investigate the influence of channel interference in a location fingerprinting approach and describe localization experiments and simulations on the

IEEE 802.11 test-bed. In this chapter, I also investigate AP channel assignment, the distribution of received signal strength (RSS) values, and variations in covered areas and the distances between APs. This analysis will provide guidance as to 1) how to assign channels, 2) space APs to reduce interference, and 3) calculate how many access points are required to uniquely identify a location at a given accuracy and precision. The relevant findings should be of assistance to engineers in designing and better understanding a WLAN channel assignment specifically for positioning. This idea has been previously published in [Chan, Baciu et al. 2009c].

2. An accurate positioning algorithm/approach

An accurate positioning approach is crucial for effective indoor localization. Chapter 5 discusses applications of the Newton Trust-Region method and the use of the convergence of a trajectory to remove the noise from the received signal strength. This method has been previously published in [Chan, Baciu et al. 2009e] and applies a Newton Trust-Region (TR) algorithm to trajectory estimation based on the traditional Location Fingerprinting/Localization approach. The Newton Trust-Region method optimizes Location Fingerprinting iteratively because each point in a trajectory normally falls into a region and converges in the same direction.

3. A structural WLAN infrastructure

Large-scale WLAN infrastructures contain thousands of access points (APs) that are often deployed in ad-hoc, empirical, and non-optimal configurations. Such an unstructured approach leads to poor resource utilization and poor localization due to signal overlap and black spots. As alternatives, Chapter 7 proposes three structured approaches to WLAN infrastructure deployment with the goal of achieving high localization accuracy and optimal coverage. Chapter 7 first proposes, analyzes, and discusses triangular, square and hexagonal distributions. Finally, I present experiment results that show that the ad-hoc deployment of APs is less effective than any of the three structured approaches. Overall, the hexagonal approach is the most effective for localization operations. A surprising result of this work is the discovery that the center of a square distribution produces a localization accuracy that is 30% higher than either of the other two structured approaches. The worst performer is the ad-hoc distribution, which requires 50% more APs than the hexagonal distribution to achieve effective localization. The proposed structured approaches allow WLAN infrastructure designers to achieve optimal localization in a cost-effective, resource-effective, and accurate manner. This investigation has been published in [Chan, Baciu et al. 2009d].

4. A graphical fuzzy signal visualization model

Fuzzy logic modeling can be applied to evaluate the behavior of the received signal strength (RSS) in Wireless Local Area Networks (WLANs), which is a central part of

WLAN tracking analysis. Previous analytical models have considered in depth how the WLAN infrastructure affects the accuracy of tracking. Chapter 8 proposes a novel fuzzy spatio-temporal topographic model implemented in a large (9.34 hectare), physical university campus where the WLAN received signal strength (RSS) was surveyed from more than 2,000 access points. The Nelder-Mead (NM) method was applied to simplify this authors' previous work on fuzzy color maps into a topographic (line-based) map. The new model provides a detailed, quantitative representation of WLAN RSS. This model was published in [Chan, Baciu et al. 2008; Chan, Baciu et al. 2009b] and [Chan, Baciu et al. 2009b] and received the best paper award.

5. A location-aware information retrieval system

One of the most challenging problems in retrieving location-aware information is to understand user queries in such a way that it is optimally suitable to the user's current location. The speed and accuracy of retrieval and the usefulness of the retrieved data depends on a number of factors including constant or frequent changes in its content or status, the effects of environmental factors such as the weather and traffic and the techniques that are used to categorize the relevance of the retrieved data.

In Chapter 9, I preliminarily attempt to deal with this task that makes use of agents operating in both wired and wireless networks to find and retrieve location-aware information. In the proposed system, agents are endowed with a wide range of human-like abilities, including perception, the use of natural language, learning, and the ability to understand user queries. More specifically, this work proposes *semantic TFIDF*, an agent-based system for retrieving location-aware information that improves the speed of retrieval while maintaining or even improving the accuracy by making use of semantic information in the data to develop smaller training sets. The method assigns intelligent agents to first gather location-aware data and then, using semantic graphs from the WordNet English dictionary, the agents classify, match and organize the information to find a best match for a user query. Experiments compare the proposed system with three other commonly used systems and show it to be significantly faster and more accurate. This agent system has been published in [Chan, Baciu et al. 2009a].

1.4. Contributions of this thesis

This thesis makes four major contributions. First, the proposed Newton Trust-Region method is more robust than the traditional localization approaches and 20% more accurate. Second, the proposed WLAN infrastructure can remove up to 50% of access points while achieving the same localization accuracy. In other words, it reduces the cost of infrastructure installation. Third, the proposed comprehensive 3D wireless signal visualization model can be used as a quick reference and efficient analytical tool for use in improving the design of WLAN infrastructure. Fourth, this work refutes the erroneous belief that interference can make positioning more accurate. On the contrary, as later experiments show, channel interference between APs using the same frequency significantly degrades the accuracy of a positioning system.

1.5. Organization of this Thesis

This thesis consists of eight chapters, organized as follows:

Chapter 1 introduces the contents of this thesis.

Chapter 2 describes the Global Positioning System, propagation-based and location-fingerprinting-based indoor localization system, existing WLAN infrastructure and visualization models and, finally, the location-aware agent system.

Chapter 3 proposes a methodology for maintaining a stable WLAN signal transmission for localization. This chapter provides the background to IEEE 802.11b/g channel interference and discusses how to reduce the channel interference between access points so as to stabilize wireless transmission. This chapter also presents the experimental setup and the results of the methodology. These ideas were previously proposed in [Chan, Baciu et al. 2009c].

Chapter 4 proposes the use of a Kalman Filter method to estimate locations. This chapter describes how to filter the noise and presents the results of experiments using this positioning methodology. This method has been published in [Chan, Baciu et al. 2009f].

Chapter 5 proposes the use of the Newton Trust-Region method to estimate locations. This chapter describes and compares other positioning methods as well as an experiment conducted in an indoor environment and evaluations of the positioning methodology. This method has been published in [Chan, Baciu et al. 2009e].

Chapter 6 proposes the use of a Fourier descriptor to solve the problem of the compatibility of the absolute received signal strength of different mobile platforms.

Chapter 7 investigates whether the existing empirical approach to the installation of WLAN infrastructure is appropriate. This chapter describes and provides the results for three structural installations and carries out simulations in a testing area of 150m x 200m. This investigation has been published in [Chan, Baciu et al. 2009d].

Chapter 8 proposes a model for visualizing the WLAN signal distribution so as to indicate black spots. This chapter introduces the background to fuzzy logic and topographical modeling. The author carried out a survey of Wi-Fi signals in a 9.34 hectares university campus. A discussion is provided analyzing models of the underlying features of WLAN signal distribution. The model has been published in [Chan, Baciu et al. 2009b] and [Chan, Baciu et al. 2008].

Chapter 9 proposes an agent-based location-aware information retrieval system. This chapter provides some background to agent technology and information retrieval techniques. The result of this agent-based system is presented at the end of this chapter. This agent-based system has been described in [Chan, Baciu et al. 2009a].

Chapter 10 concludes this thesis, summarizing its contributions, discussing limitations of the proposed models, and suggesting directions for further work.

Chapter 2: Literature Review

This chapter provides some basic background and a literature review relevant to localization systems. It is organized in six major sections. Section 2.1 introduces location-sensing techniques. Section 2.2 introduces and discusses the limitations of existing WLAN indoor localization approaches including propagation-based and location-fingerprinting-based approaches. Section 2.3 describes the factors that give rise to channel interference and how it affects localization. Section 2.4 describes the existing WLAN infrastructure and Section 2.5 describes existing visualization.

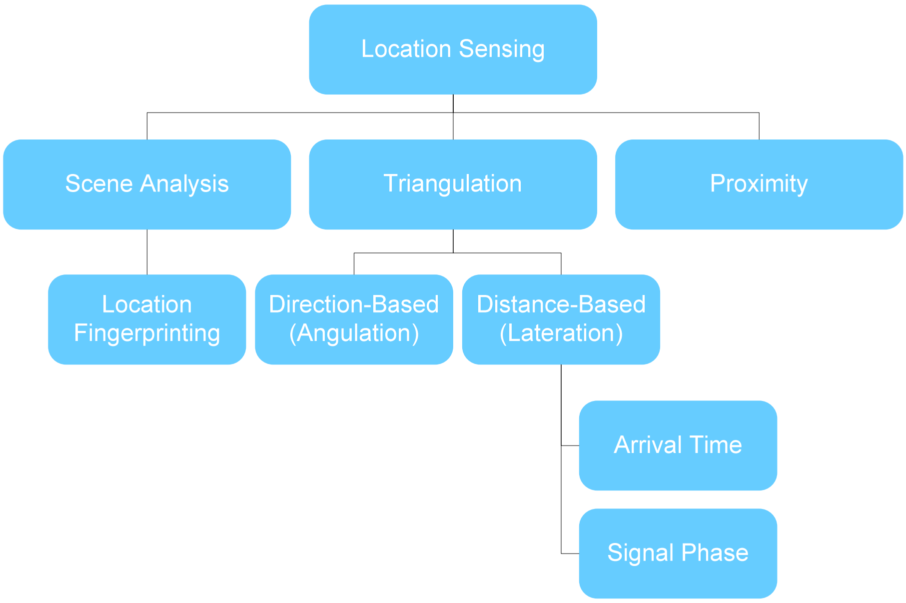


Figure 2 Location Sensing Taxonomy

2.1. Location Sensing Techniques

Location sensing techniques can be categorized into three general types: scene analysis [Hightower and Borriello 2001], triangulation [Goor], and proximity [Hightower and

Borriello 2001]. Figure 2 provides taxonomy of the three different types of location-sensing techniques.

2.1.1. Scene Analysis

The scene analysis location-sensing method collects and extracts features from an observed scene. The observed features (these are known as “fingerprints” but should not be confused with the “fingerprints” of location fingerprinting) are usually specific and unique and are used to estimate the location of the observer or of observed objects in the scene. We can estimate the distance by matching the similarity of features. The scene could be radio frequency waves, acoustic sound, visual images or any other measurable physical phenomena which can be detected near a target object.

Static scene analysis searches for observed features in a predefined dataset that maps them to object locations. Differential scene analysis estimates location by tracking the difference between successive scenes. Differences in the scenes correspond to movements of the observer and if the features in the scene are known to be at specific positions, the observer can compute their relative positions. [Metsälä 2003]

2.1.2. Triangulation

Triangulation positioning algorithm uses trigonometry and geometry to compute the locations of objects. In a 2D environment, this requires the location of three satellites. The locations of these three satellites location are denoted as (x_1, y_1), (x_2, y_2), (x_3, y_3) and the object location is denoted (*x, y*). This same concept also works in 3-D space but deals with spheres rather than circles and requires four spheres rather than three circles to find an exact location. The heart of a GPS receiver is the ability to find the receiver's distance from four (or more) GPS satellites. It can be divided into the subcategories of lateration, using distance measurements, and angulations, using primarily angle or bearing measurements.

Lateration

In the lateration method, receivers calculate their positions by measuring the distance to satellites and the location of satellites. The heart of a GPS receiver is the ability to find the receiver's distance from four (or more) GPS satellites. Once it determines its distance from the four satellites, the receiver can calculate a location on Earth. If the receiver can find only three satellites, it uses an imaginary sphere to represent the earth, giving location information but no altitude information (Figure 3).

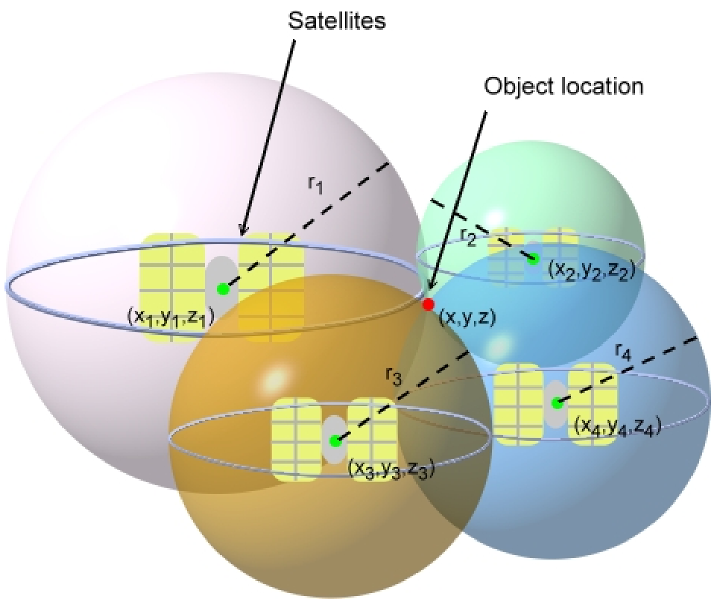


Figure 3 3D Localization using lateration

GPS was developed by the United States Department of Defense and is the only fully functional Global Navigation Satellite System (GNSS). It has been deployed commercially and has become an indispensable service in large urban areas and over large geographical locations. It is low-cost, easy to use for recreational navigation as well as maritime logistics and transportation. It is becoming an important part of the global communications infrastructure. A minimum of 24 GPS satellites in orbit above Earth are tracking GPS signals and at any point on Earth users can receive at least six satellites signals. These satellites continuously broadcast information for use in positioning. A basic GPS makes use of a tri-lateration method to locate an object, first collecting signals from satellites and then calculating a position on Earth. Satellite signal information includes the satellite send-time, orbit information, and a satellite's status. The GPS receivers use this information to calculate the distance to the satellite by lateration.

The lateration technique uses measure three different metrics: direct measurement, time-of-flight, and attenuation measurements.

The direct measurement of distance exploits physical actions or movements, which are somewhat difficult to obtain automatically because of complexities involved in coordinating autonomous physical movement. [Metsälä 2003]

Time-of-flight measurement makes use of the directly proportional relationship of the propagation time and distance between the observer and the object. The object itself may either be moving at a known velocity for a given time interval or it may be approximately stationary. In either case, the difference between transmission and arrival time of an emitted signal is observed. However, this measurement suffers when the radio waves cannot pass though solid objects such as mountains and buildings and are deflected. Signal deflection leads to multi-path interference. If the objects are large reflective surfaces, then the error can be more than 15 meters.

Attenuation measurement [Metsälä 2003] is similar to lateration except angles are used instead of distances in determining the position of an object. Two-dimensional angulations require two angle measurements and one length measurement such as the distance between the reference points as shown in Figure 4.

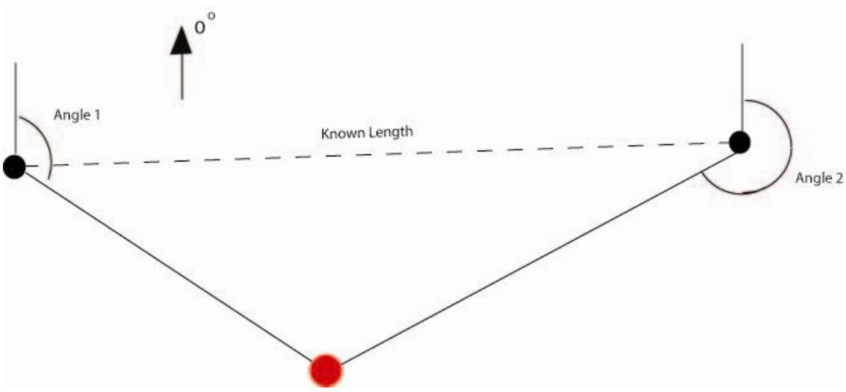


Figure 4 Determining 2D position of the object using angulations.

2.1.3. Proximity

The proximity location-sensing technique usually provides symbolic relative location information. This technique identifies an object with a tag and the transmitter in a known location detects the tag. If the tag is detected, we can say that the object is “near” the transmitter. The presence of the object is sensed using a mechanism with a limited range, for example, RFID, pressure and touch sensors, and capacitive field detectors. Monitoring is possible when a mobile device or tag is in the range of one or more access points and antennas.

2.2. Existing WLAN indoor localization approaches

Two positioning methodologies are typically applied in WLANs, propagation based approaches and location fingerprinting (LF). In previous work [Chan, Baciu et al. 2008; Chan, Baciu et al. 2009d] we made use of LF to track a WLAN-enabled device. In the simulations described later in this work, only the LF approach will be used as, fortunately, the use of a propagation-based approach gives the same accuracy as is obtained using an LF approach. For completeness, however, in the following I briefly describe both.

2.2.1. Propagation-based Approach

Propagation-based approaches measure the received signal strength (RSS), angle of arrival (AOA), or time difference of arrival (TDOA) of received signals and apply mathematical models to a set of triangulation algorithms to determine the location of the device.

The triangular positioning algorithm uses trigonometry and geometry to compute the locations of objects. In a 2D environment, this requires three access points (APs). The locations of these three APs’ location are denoted as (x_1, y_1), (x_2, y_2), (x_3, y_3) and the object location is (x, y). Using the propagation-based theorem, I denote the distance between the access points and object location as d_1, d_2, and d_3, where d_0 is the initial RSS at the reference distance.

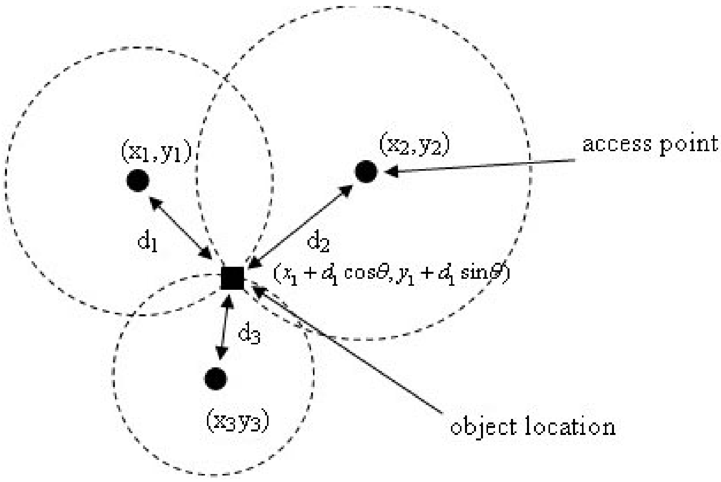


Figure 5 Triangular Algorithm

To estimate the location of the object, this work uses the tri-lateration method as follows:

$$\begin{cases} d_1 = d_0 10^{\frac{r_0 - r_1}{10 \cdot \alpha}} \\ d_2 = d_0 10^{\frac{r_0 - r_2}{10 \cdot \alpha}} \\ d_3 = d_0 10^{\frac{r_0 - r_3}{10 \cdot \alpha}} \end{cases} \tag{1}$$

After calculating the distance, the angle θ between the object location and APs is found and it is then possible to calculate the location of the object as follows:

$$(x, y) = \begin{cases} x_1 + d_1 \cos\theta_1, y_1 + d_1 \sin\theta_1 \\ x_2 + d_2 \cos\theta_2, y_2 + d_2 \sin\theta_2 \\ x_3 + d_3 \cos\theta_3, y_3 + d_3 \sin\theta_3 \end{cases} \tag{2}$$

2.2.2. Location Fingerprinting Approach

Location Fingerprinting (LF) [Kaemarungsi and Krishnamurthy 2004a; Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Li, Wang et al. 2005] locates a device by accessing a pre-recorded database containing the location fingerprint (i.e., the RSSs and coordinates) and then calculating its own coordinates compared with those in an LF database. More specifically, the LF approach requires the collection of data $\{(\mathbf{LF}_n, D_i), i = 1...N\}$, for **N** locations in the site, where D_i is the known location of the i'th measurement and $\mathbf{LF}_i$ = ($\mathbf{LF}_i$, ..., $\mathbf{LF}_{iN}$) is the RSS vector when the signal receiver is at D_i. The vector $\mathbf{LF}_i$ is the "fingerprint" of the location D_i. When a receiver is in an unknown location A, the user can locate A by searching for the fingerprint $\mathbf{LF}_i$ that is closest to **LF** and then estimates the location.

There are two alternative Location Fingerprinting Approaches, the K-Nearest Neighbor (K-NN) and the probabilistic approach.

The K-Nearest Neighbor Location Fingerprinting Approach

The K-Nearest Neighbor (K-NN) algorithm requires two sets of data. The first set of data is the samples of RSS from N APs in the area, which this paper refers to as the sampling vector. This vector is denoted as $R = [r_1, r_2, r_3...r_N]$. Each element in a vector is an independent RSS (in dBm) collected from APs in the location. The second set of data contains all of the average RSS from N APs at a particular location. This second dataset forms the location-fingerprinting database, which this paper refers to as the Location Fingerprint Vector. This vector is denoted as $F = [f_1, f_2, f_3...f_N]$ at the position $D = [d_1, d_2,$

$d_3 \ldots d_N$].

The location d is estimated by clustering the Euclidean distance $|r-f_i|$ between sampling fingerprint vector r and location fingerprint vector f_i with position d_i as follows:

$$d = \frac{\sum_{i=0}^{n} \frac{d_i}{|r - f_i|}}{\sum_{i=0}^{n} \frac{1}{|r - f_i|}} \tag{3}$$

The Probabilistic Location Fingerprinting Approach

Probabilistic LF applies Baye's Theorem to calculate the most probable location out of the pre-recorded LF database, $F = [f_1, f_2, f_3 \ldots f_N]$. I can estimate d by

$$\arg\max{}_d \left[P(d/F)\right] = \arg\max{}_d \left[\frac{P(F/d)P(d)}{P(F)}\right] \tag{4}$$

Since $P(F)$ is constant for all d, the algorithm can be rewritten as

$$\arg\max{}_d \left[P(d/F)\right] = \arg\max{}_d \left[P(F/d)P(d)\right] \tag{5}$$

As *P(d)* can be factored out from the maximization process, the probabilistic localization algorithm is as follows:

$$P(F/d) = \prod_{i=1}^{N} P(f_i/d) \tag{6}$$

2.3. Channel Interference in WLAN indoor localization

The IEEE 802.11 standard establishes several requirements for radio frequency transmission. These include the canalization schemes and the spectrum radiation of the signal. [Sun, Qin et al. 2008]. IEEE 802.11 b/g WLAN mandates 14 channels but in the FCC or North American domain, the 2.4GHz frequency ISM band is divided into 11 channels. Each channel is spread over 22 MHz in line with the Direct Sequence Spread Spectrum (DSSS) technique employed by IEEE 802.11b/g. These channels have only five MHz of center frequency separation, so channel interference may occur in adjacent channels which share the frequency spectrum.

The bandwidth of a wireless network is limited because wireless networks and stations have to share limited bandwidth. IEEE 802.11b/g has 14 overlapping frequency channels. Channels 1, 6 and 11 are non-overlapping channels.

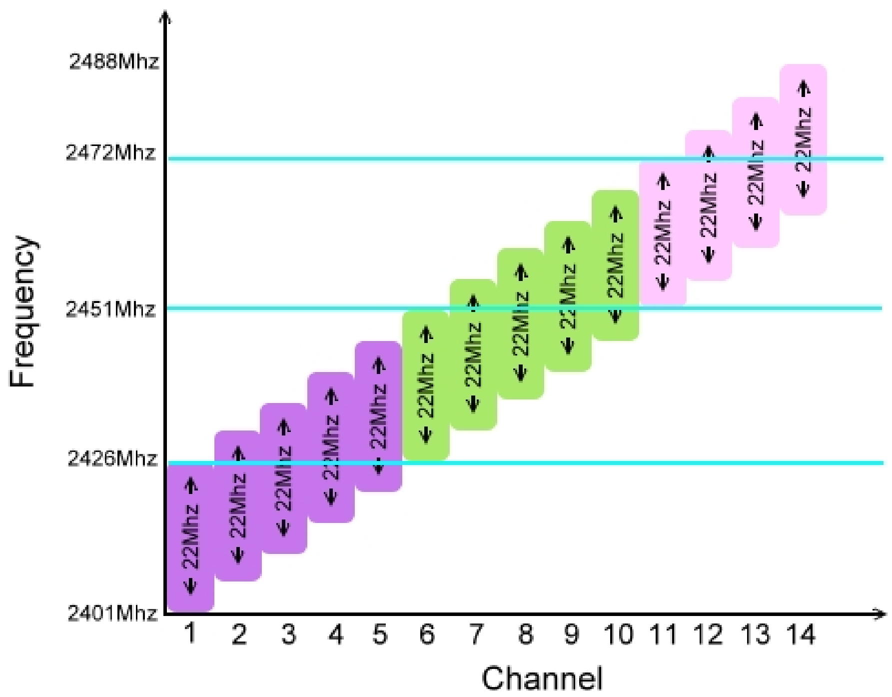


Figure 6 IEEE 802.11b/g frequency spectrum to channel number

As shown in Figure 6, IEEE 802.11 b/g spreads through 2,401 MHz to 2,483 MHz. Each channel spreads over 22 MHz. Two adjacent channels are separated only by 5 MHz such that most of the existing channels overlap.

2.3.1. The Interference-level function

The interference-level function γ is defined as follows:

$$\gamma(\Delta c) = \max(0, 1 - k\Delta c) \tag{7}$$

where Δc is the absolute channel difference (e.g, for channel 1 and 6, Δc = **5**) and k is the non-overlapping ratio of two channels. γ and Δc are in Db units. When Δc increases, γ decreases. In reality, if APs are installed far enough away from each other, γ should be at least equal to the calculated threshold.

2.3.2. The Signal Propagation Algorithm

A signal propagation algorithm [Prasithsangaree, Krishnamurthy et al. 2002; Jan and Lee 2003; Li, Wang et al. 2005] calculates the received signal strength (RSS) with path loss as follows:

$$R = r - 10\alpha \log_{10}(d) - wallLoss \tag{8}$$

where *r* is initial RSS, *d* is a distance from APs to a location, α is the path loss exponent (clutter density factor) and wallLoss is the sum of the losses introduced by each wall on the line segment drawn at Euclidean distance *d*.

Initially, *r* is the initial RSS at the reference distance of d_0 (typically 1 meter). This is 41.5 dBm for LOS propagation and for 37.3 dBm NLOS propagation in some report measurements. The path loss exponent α at a carrier frequency of 2.4 GHz is reported to be 2 for LOS propagation and 3.3 for NLOS propagation [Chan, Baciu et al.]. Under other circumstances α can be any value between 1 and 6.

2.3.3. The Signal-to-Interference-plus-Noise-Ratio

The Signal-to-Interference-plus-Noise-Ratio (SINR) is a very common indicator for measuring interference. The SINR is defined as follows:

$$SINR = \frac{R_b}{\gamma(\Delta c)\sum R + n} \tag{9}$$

where R_b is the highest RSS after path loss calculation. *R* is the remaining set of RSS after path loss calculation and n is the noise signal strength. R_b, R, n are expressed in dBm units. Usually, n should have the value of -100dbm. Again, SINR should be at least equal to the calculated threshold (Equation 9), which depends on the distances between APs, the transmission rate, the modulation scheme, and the required bit-error rate.

2.4. Existing WLAN infrastructure

Wireless Local Area Networks (WLANs) are nowadays often deployed on a large scale and in a wide range of urban environments. Very often, these WLANS are made up of many access points (APs) or nodes deployed across extensive, topographically varied, heavily built up, and constantly changing environments that may carry heavy traffic [Huang, Wang et al. 2005]. The basic requirement of an effective WLAN is that it should provide adequate coverage so that when users wish to access location-aware (e.g., pervasive computing-enabled) applications and services, the WLAN will permit mobile devices to be accurately positioned and that it be possible to deploy the network in a cost-effective and resource-efficient way, including in indoor environments. Current approaches to WLAN infrastructure design and deployment, however, apply an unstructured approach. This implies poor resource utilization, placing and positioning APs manually on the basis of empirical measurements of RSS (received signal strength) taken by engineers.[Budianu, Ben-David et al. 2006]. Such an approach may use many APs but still leave blind spots or places where there are too many access points packed closely together, producing wasteful signal overlap and interference and reducing positioning accuracy.

Although mobile phone networks do not require highly accurate positioning, they do address a set of problems very similar to those of WLAN AP positioning. In particular they must efficiently and cost-effectively cover an area with a signal of adequate strength. The current mobile network approach is to establish a regularly-placed infrastructure of mobile stations (MS) [Rubin and Choi 1997]. Each MS is placed at the centre of a hexagon that lies within the radius of the signal of the MS and that is contiguous with other similar hexagons so as to form a honeycomb-celled tessellation over the network area.

This is an effective framework but while the requirements and the constraints of the mobile cell network and an AP network are similar, they are not identical in at least five ways. First, black spots and interference are more tolerable in cell phone networks because mobile users can very quickly move into areas of more acceptable signal strength or quality. Second, MS apparatus are much more powerful than AP devices. Therefore, MS devices are less limited in their ability to pass through walls and are less influenced by traffic or by other signaling devices. Third, MS devices may be licensed to override competing signals. Fourth, a very large mobile area can be covered by a relatively small number of MS devices, while a WLAN (Wi-Fi) network will require a proportionally much larger number of APs. Finally, and perhaps most distinctively, positioning is qualitatively different in the two technologies. In a mobile network, devices can be located merely as being within the radius of an MS. In a Wi-Fi network, we can locate a device much more accurately, ie close to a point in a particular coordinate grid. Currently, this requires that the device be in communication with two or more APs so as to be able to calculate data from them. However, in certain real-world scenarios, a device will not be simultaneously within the range of more than two APs. Consequently, while it might be able to recognize a device in its radius, it may not be able to calculate its coordinates.

2.5. Existing WLAN visualization model

Current research on the visual representation of WLAN signals [Kaemarungsi and Krishnamurthy 2004a; Swangmuang and Krishnamurthy 2008] is based on the accuracy of positioning systems and proximity graphs, such as the Voronoi diagram and clustering graphs. These assume the distribution of the RSS is Gaussian and pair wise. Some researchers [Fang, Lin et al. 2008; Kj rgaard and Munk 2008; Swangmuang and Krishnamurthy 2008] have ignored the radio signal properties. In [Bahl and Padmanabhan 2000], the distribution of the RSS has been shown not to be Gaussian; it is often left-skewed and the standard deviation varies according to the signal level. It is clear from their measurements that signals from multiple access points (APs) are mostly independent and that interference comes from other APs. In such a case, the use of a simple Euclidean distance to determine the location may wrongly classify some patterns. In this way, incorrect assumptions may ignore or distort the real behavior of RSS and provide inadequate and inaccurate RSS analysis. There are thus two particular drawbacks to these representations. First, none of them provide any way to visualize location uncertainty, for example, by showing where blind spots may be located on a 2-D or 3-D display. Second, neither provides any way to estimate how many APs might be required for optimal coverage nor do they provide any guidance as to where they should be placed.

Fuzzy visualization map concepts have been widely applied in other fields to phenomena such as temperature, rainfall and atmosphere but have not been applied in the modeling of a wireless positioning system. This chapter studies the problem of modeling relationships between wireless infrastructure, large obstacles, and the spatial distribution of RSS using two concepts: (1) multi-layer fuzzy model and (2) 2D propagation-based simulation model. The fuzzy map models are used to find the regions where the RSS is denser and cluster different RSS in different colored layers. Topographic mapping has been widely recognized as a comprehensive method for visualizing geographical information, such as the reflectance of slope and terrain but could also be applied to modeling in a Wi-Fi RSS analytical model.

Chapter 3: Properties of Channel Interference for WLAN Location Fingerprinting

Localization systems for indoor areas have recently been suggested that make use of existing wireless local area network (WLAN) infrastructure and location fingerprinting approach. However, most existing research work ignores channel interference between wireless infrastructures and this could affect accurate and precise positioning. A better understanding of the properties of channel interference could assist in improving the positioning accuracy while saving significant amounts of resources in the location-aware infrastructure. This chapter investigates to what extent the positioning accuracy is affected by channel interference between access points. Two sets of experiments compare how the positioning accuracy is affected in three different channel assignment schemes, with ad-hoc, sequential, and orthogonal data is analyzed to understand what features of channel interference affect positioning accuracy. The results show that choosing an appropriate channel assignment scheme could make localization 10% more accurate and reduces the number of access points that are required by 15%. The experimental analysis also indicates that the channel interference usually obeys a right-skewed distribution and positioning accuracy is heavily dependent on channel interference between access points (APs).

The rest of this chapter is organized as follows: Section 3.1 introduces the background to channel interference. Section 3.2 describes related work on the deployment of modeling of wireless signal strength properties and infrastructure design of Wi-Fi Networks. Section 3.3 defines the characteristics of channel interference metrics. Section 3.4 defines location uncertainty. Section 3.5 presents experiments and results on channel interference metrics. Finally, Section 3.6 presents experiments and results on the effect of channel interference on positioning accuracy.

3.1. Introduction

Wireless Local Area Networks (WLANs) are often deployed on a large-scale in a wide range of urban environments. Covering a very large urban environment requires that thousands of access points be placed and installed properly, without interference. The basic requirements of an effective WLAN are, first, adequate coverage where users wish to access location-aware (e.g., pervasive computing-enabled) applications and services. Second, the WLAN should allow accurate localization of mobile devices. The deployment of the network should reduce the interference as much as possible so as to achieve these functions in a cost-effective and resource-efficient manner.

Unfortunately, access points (APs) are usually deployed in an empirical way, manually placed and positioned on the basis of measurements of RSS (received signal strength) taken by engineers. Such an unstructured approach to WLAN infrastructure design implies poor resource utilization and strong channel interference [Budianu, Ben-David et al. 2006]. For example, more APs may be used to improve coverage but this may still leave blind spots or places where there are too many access points packed too closely together. This can lead to signal overlap, which is wasteful and causes interference.

Location-Fingerprinting-based approaches [Kaemarungsi and Krishnamurthy 2004a; Taheri, Singh et al. 2004; Li, Wang et al. 2005; Fang, Lin et al. 2008; Kj rgaard and Munk 2008; Swangmuang and Krishnamurthy 2008] locate a device by comparing its coordinates with the received signal strengths (RSSs) and coordinates of other devices within the Wi-Fi footprint as held in an LF database. More specifically, the LF approach requires the collection of data $\{(\mathbf{Y}_n,C_i)$, i = 1...$N\}$, for N locations in an area, where C_i is the known location of the i'th measurement and $\mathbf{Y}_i = (\mathbf{Y}_i, ...,\mathbf{Y}_{iN})$ is the received signal strength (RSS) vector when the transmitter is at C_i. The vector $\mathbf{Y}_i$ is the "fingerprint" of the location C_i. When a new fingerprint $\mathbf{Y}$ is derived from a transmitter at an unknown location A, I can locate A by searching for the fingerprint $\mathbf{Y}_i$ that is closest to Y in say d distance and estimate the location with the corresponding C_i. The drawbacks of the LF approach are (1) LF requires an initially survey with a very large training dataset and (2) LF is very sensitive to signal fluctuation due to the changes of infrastructure of buildings and channel interference among APs leading to inaccurate positioning.

The IEEE 802.11 standard establishes several requirements for radio frequency transmission, including the canalization schemes and the spectrum radiation of the signal.[Sun, Qin et al. 2008]. In IEEE 802.11 b/g WLAN, there are 14 channels. In North America, the 2.4GHz frequency ISM band is divided into 11 channels. Each channel is spread over 22 MHz due to the Direct Sequence Spread Spectrum (DSSS) technique employed by IEEE 802.11b/g. These channels have only five MHz of center frequency separation. Channel interference occurs because frequency spectrum is shared with each adjacent channel. Recent research has focused on reducing channel interference by either creating a new channel assignment scheme [Chiu, Yeung et al. 2006; Haidar, Ghimire et al. 2008; Subramanian, Gupta et al. 2008] or enhancing existing MAC protocols [Dutta, Jaiswal et al. 2007; Zeng, Wang et al. 2007]. The goal is to improve the data transmission through wireless networks. However, none of existing work on channel interference has focused on the accuracy of the location estimation algorithm. Some researchers have even maintained that interference could increase positioning accuracy.

This chapter investigates the influence of channel interference in a location fingerprinting approach. The study of channel interference is essential for accurate indoor positioning system. This chapter also describes localization experiments and simulations on the IEEE 802.11 test-bed and investigates the channel assignment of APs, the distribution of received signal strength (RSS) values, the variation of coverage, and distances between APs. The analysis of these features provide insights into how to assign channels, how to space APs so as to reduce interference, and how many access points are required to uniquely identify a location at a given accuracy and precision. The results would be of interest and assistance to engineers designing WLAN channel assignments specifically for positioning.

3.2. Related Work

This section summarizes current research on modeling wireless signal strength properties and on infrastructure design for Wi-Fi networks.

3.2.1. Modeling of Wireless Signal Strength Properties

The modeling of wireless signal strength properties [Kaemarungsi and Krishnamurthy 2004a; Swangmuang and Krishnamurthy 2008] is crucial to deploying efficient indoor positioning systems. Analytical models can be used to improve the design of positioning systems. For example, by eliminating the installation of Wi-Fi access points and shortening the sampling time of Wi-Fi received signal strength (RSS) in location estimation. Yet, we currently lack an analytical model that might be used as a framework for designing and deploying positioning systems. Such an analytical model should have spatial elements to visualize the RSS distribution and to evaluate and predict "precision" performance of indoor positioning systems based on location fingerprinting. Such a model could be used to improve the design of positioning systems, for example by eliminating some fingerprints and reducing the size of the location fingerprint database. Our previous work [Chan, Baciu et al. 2008] and [Chan, Baciu et al. 2009b] made use of fuzzy logic and topographic mapping techniques to visualize the received signal strength (RSS) but did not investigate how channel interference affects positioning accuracy. Most existing research in channel interference has focused on improving wireless data transmission [Dutta, Jaiswal et al. 2007; Zeng, Wang et al. 2007] and channel assignment [Chiu, Yeung et al. 2006; Haidar, Ghimire et al. 2008; Subramanian, Gupta et al. 2008]. The study on the relationship between channel interference and positioning accuracy has been insufficient.

3.2.2. Infrastructure Design of Wi-Fi Network

WLANs are made up of many access points (APs) or nodes. These access points (APs) are manually placed and positioned on the basis of measurements of RSS (received signal strength) taken by engineers empirically. An unstructured approach to WLAN infrastructure design implies poor resource utilization. [Budianu, Ben-David et al. 2006] as extra APs may be used to improve coverage but this may still leave blind spots or places where there are too many access points packed too closely together. This can lead to signal overlap, which is both wasteful and causes interference. As long as an ad-hoc approach is used, it is not possible to estimate in advance how many APs are optimally required for localization or where they should be placed. In our previous work [Chan, Baciu et al. 2009d], we addressed this issue and proposed a more structured approach which would produce economies of scale and efficiencies as well as improved localization capabilities. However, different Wi-Fi infrastructure designs could cause different patterns of channel interference and to date there has been no in-depth study of how channel interference might affect positioning accuracy.

3.3. Characterization of Channel Interference Metric

This section investigates whether interference is normally or log normally distributed. The signal difference between the preset (maximum) RSS value of AP and sample RSS of an AP could be seen as interference strength value (in dB). Two APs are placed and interfere with each other. We assume that there is a zero (or very short) distance between receiver and APs, In other words, the signal should not be reduced by propagation-loss. The preset RSS value is denoted by ρ and the interference strength value by $F = \{f_1, f_2, ..., f_n\}$. The interference could then be by:

$$Y = \sqrt{\sum_{i=1}^{n}(\rho - r_i)^2} = \sqrt{\sum_{i=1}^{n} f_i^2} \tag{10}$$

where n represents the number of collected samples. In equation (10), the difference between a preset RSS value and measured RSS, r_i is the signal interference strength. In fact, the random variable of interference, f_i, should have a zero mean if and only if the wireless RSS obeys a normal distribution. In other words, the random variable of interference f_i has a non-zero mean when the wireless RSS does not obey a normal distribution.

It is assumed that the random variable $X = Y^2$ where the random variable X is the square of the difference between the sample RSS and the preset RSS. Assuming that the RSS is normally distributed, the random variable $X = Y^2$ has a central chi-squared distribution with n degrees of freedom, i.e. $\mathrm{E}\{r_i\}=\rho$ or the mean of sample RSS is equal to a preset RSS value. Thus, the difference-squared f_i obeys a zero mean Gaussian distribution. A probability density function (PDF) of X will be the chi-square distribution:

$$P_{X_n^2}(x) = \frac{1}{2^{\frac{n}{2}}\sigma^n\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2\sigma^2}} \tag{11}$$

where the variance of each Gaussian component in X is σ^2, and Γ denotes Gamma function which has closed-form values at the half-integers.

However, in some real-world scenarios, the distribution of the RSS is not usually Gaussian and it is often left-skewed. The standard deviation for this real distribution varies according to the signal level. This has been verified in [Bahl and Padmanabhan 2000]. Actually, the distribution of the RSS is a non-central chi-squared distribution. In this case, the random variable of interference f_i will have a non-zero mean equal to $\mu = \rho - E\{r_i\}$.

Here, λ is defined as a non-centrality parameter of the non-central chi-squared distribution. Parameter λ could be defined as $\lambda = \sum_{i=1}^{n} \mu_i^2$. A larger value of λ indicates that some regions are experiencing higher signal interference. The PDF of non-central chi-square distribution is seen to be a Poisson-weight mixture of central chi-squared distribution. It could be defined by

$$P_X(x;n,\lambda) = \sum_{i=0}^{\infty} \frac{e^{-\lambda/2\sigma^2}(\lambda/2\sigma^2)^i}{i!} P_{P_{X_n^2}(n+2i)}(x) \tag{12}$$

Alternatively, the PDF can be written as

$$P_X(x;n,\lambda) = \frac{1}{2\sigma^2} e^{-\frac{x+\lambda}{2\sigma^2}} \left(\frac{x}{\lambda}\right)^{\frac{n-2}{4}} I_{\frac{n-2}{2}}\left(\frac{\sqrt{\lambda x}}{\sigma^2}\right) \tag{13}$$

where $I_k(x)$ is the k^{th}-order modified Bessel function of the first kind given by:

$$I_k(x) = \left(\frac{x}{2}\right)^k \sum_{i=0}^{\infty} \frac{x^{2i}}{4^i i! \Gamma(k+i+1)} \tag{14}$$

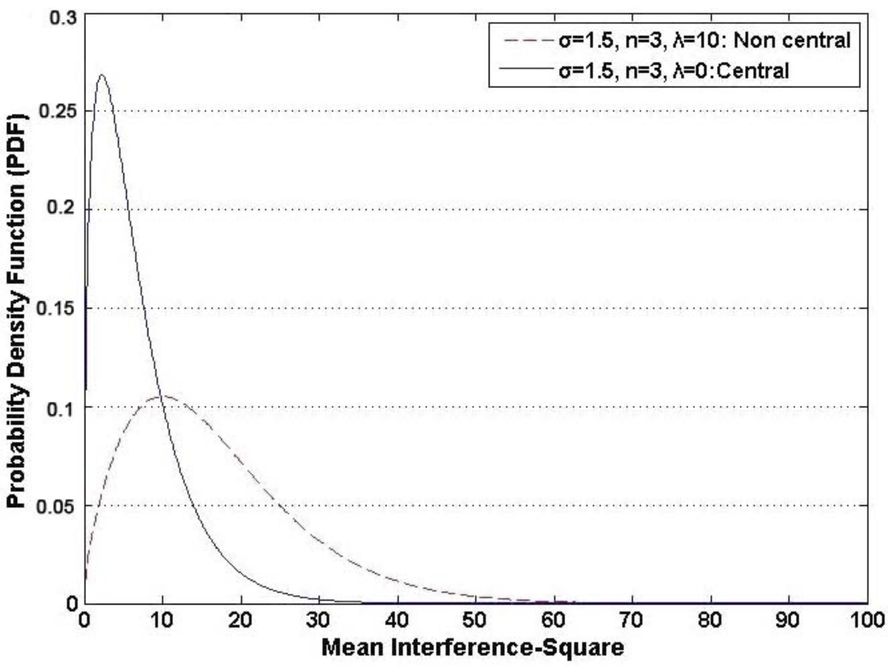


Figure 7 Theoretical comparison of the PDFs of central and non-central chi-squared distribution of interference under σ = 1.5, n = 3

In Figure 7 we have the theoretical distribution formed by (11) and (14). The dotted purple line represents the central chi-squared distribution used by (1). The blue line represents a non-central chi-squared distribution used by (14). As can be seen, the use of a central chi-square distribution produces a more evenly distributed interference-square. In other words, the strength of the interference fluctuates widely. When σ = 1.5, n = 3, the interference-square of the non-central distribution would be mostly around 10. When a non-central distribution is used, the distribution is skewed more to the right and interference strength is more likely to be a smaller value. When σ = 1.5, n = 3 and λ = 10, the interference-square of non-central distribution would be mostly around 4.

Figure 8 shows the PDF versus mean interference-square under σ = 20, n = 3. The non-central and central chi-square becomes very similar when the standard deviation of the received signal strength has a larger value. This indicates that the larger standard deviation of received signal strength causes the mean interference-square of central and non-central chi-distribution to move closer together.

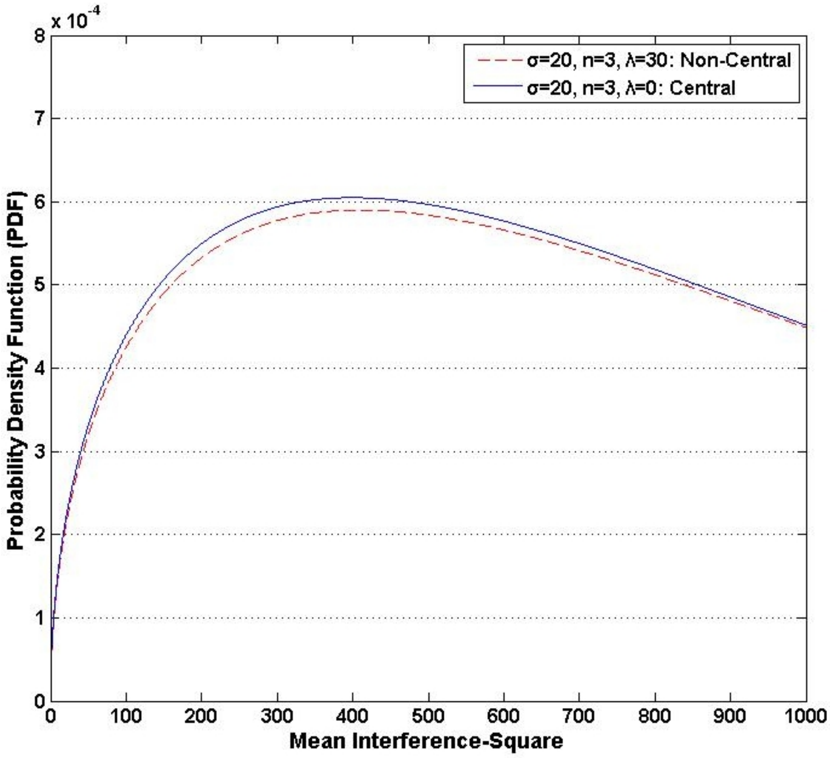


Figure 8 Theoretical comparison of the PDFs of central and non-central chi-squared distribution of interference under σ = 20, n = 3

In conclusion, depending on whether RSS is a central or non-central normal distribution, the interference distribution could be defined by either (11) or (14). Having said this, it should also be noted that experiments with RSS distributions could vary, with some experiment results showing that RSS obeys a normal distribution and some showing otherwise. To my knowledge, however, the distribution of interference has not yet been studied. The later experimental section will show that interference usually follows the distribution described in (14).

3.4. Problem Definition of Location Uncertainty

Location uncertainty occurs when it is not possible to identify the location of a co-ordinate in 2-D or 3-D space. This chapter uses existing WLAN infrastructure to estimate the position of a WLAN-enabled device. There are two fundamental positioning approaches. They are propagation-based [Prasithsangaree, Krishnamurthy et al. 2002; Jan and Lee 2003; Kwon, Dundar et al. 2004] and location fingerprinting (LF) [Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Li, Wang et al. 2005]. This work will use only an LF approach as accuracy obtained using an LF approach will also be obtained using a propagation-based approach.

Given an LF approach and WLAN infrastructure, location uncertainty could be specifically defined as a locus - a line (a series of points) having the same RSS measured within the overlap of two or more AP signals. For example, Figure 9 shows one AP with circular signal propagation and points A, B and C all having the same received signal strength (location fingerprint). If a user is located at point B, an LF approach could not estimate whether the user is at point A, B, or C. If there is just one AP, the location uncertainty happens on the circumference of signal propagation.

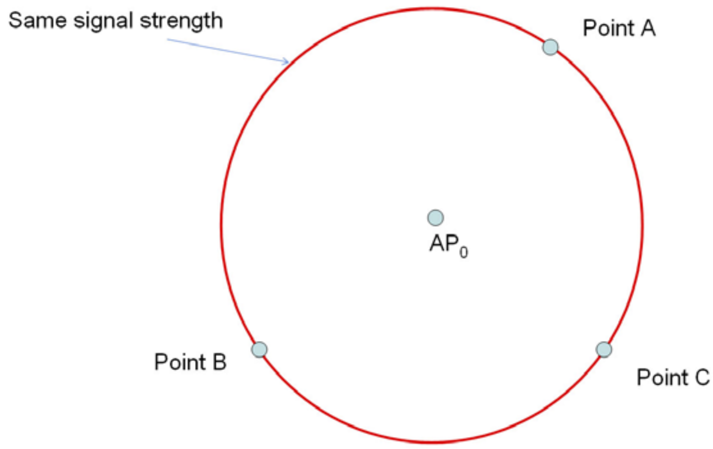


Figure 9 One AP with a circular signal propagation and points A, B and C all having the same received signal strength (location fingerprint).

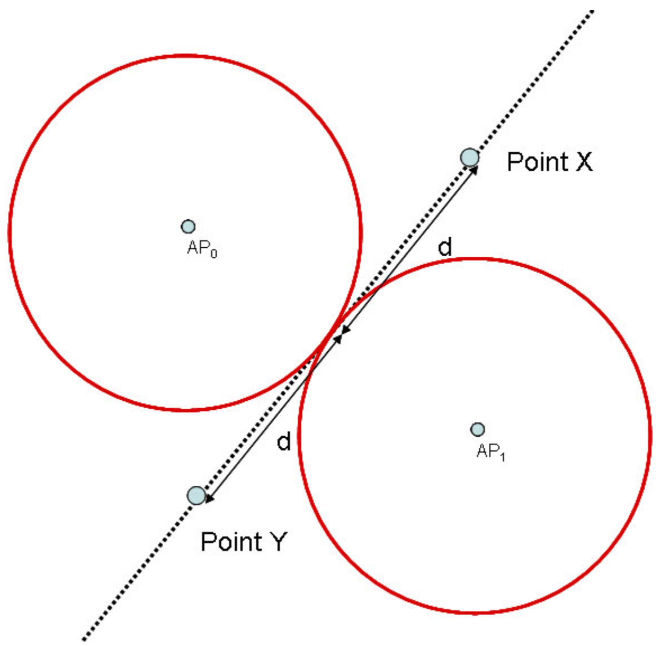


Figure 10 Location uncertainty with 2 access points' case is on the tangent of circular signal propagation where they share same RSS.

Another example is given in Figure 10, showing two APs with circular signal propagation and points X and Z having the same received signal strength. It is not possible to locate only pairs of point in a tangent where they share the same RSS. Increasing the number of APs in an area reduces the zone of most common overlap and thereby shortens the locus (seen as a line). Given enough APs, the localization uncertainty can be eliminated. This would result in perfect localization. But as long as an ad-hoc approach is used, it is not possible to estimate in advance how many APs are optimally required for localization or where they should be placed.

3.5. Experiment & Result on Channel Interference Metric

The following section describes an experiment on channel interference metrics and discusses the experiment results. The purpose of the experiment is to determine whether interference distribution obeys a normal, mean chi-square, or non-mean chi-square distribution.

Item	**Description**
Number of Aps	2 APs
Sampling time	2 hours
Samples of Signal Strength	1,000
Wi-Fi coverage from each APs	80 meters
Range of signal strength	-70dBm to -30dBm

Table 1 EXPERIMENT SETTINGS (SET A)

The experiment places two APs within a short distance of an RSS receiver. The assumption is that all signal fluctuations are caused by interference between two APs. Two APs were set to the same channel and emitted a WLAN signal at the maximum strength -70dBm. A receiver recorded 1000 samples of signal strength from two APs over two hours. The sample result was used to form a distribution and the theoretical distribution was compared using (11) and (14). Table 1 summarizes the experimental settings of Set A. Two parameters, σ and λ are input to adjust the shape of the distribution curve. σ is the standard distribution of interference. λ is a non-centrality parameter.

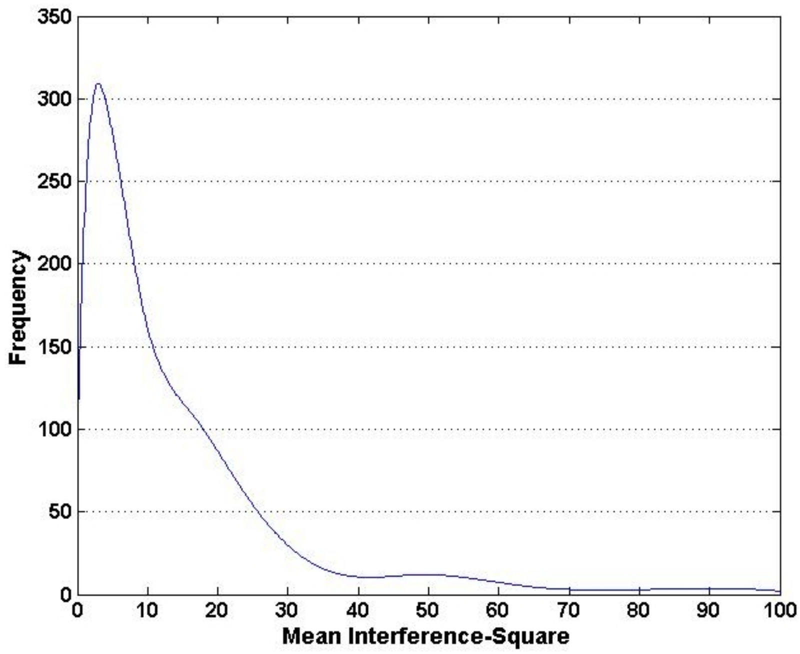


Figure 11 Result of Frequency to Mean Interference-Square experiment

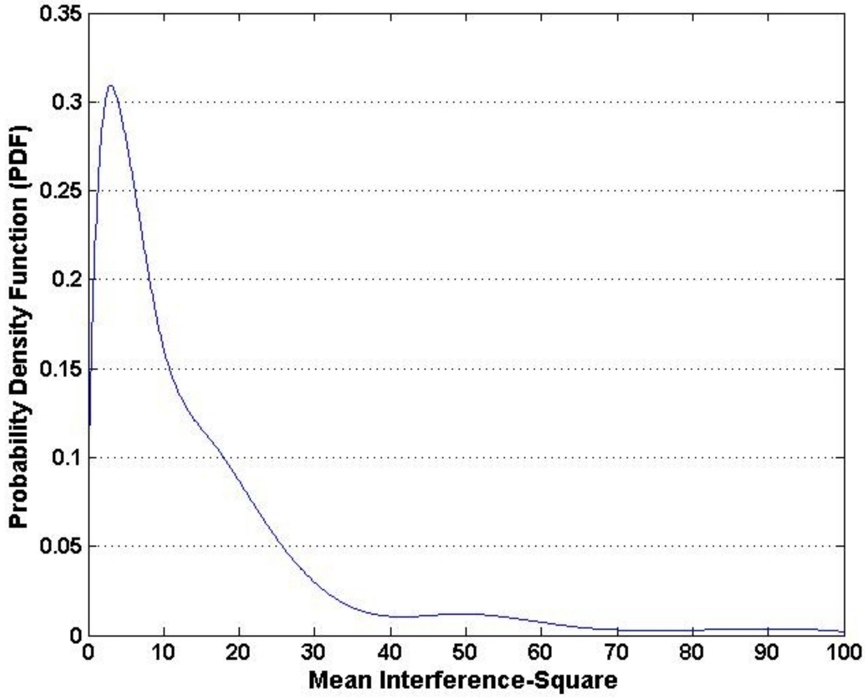


Figure 12 Result of PDF to Mean Interference-Square experiment

Section 3 discussed the theoretical distribution of interference. Here, it is compared with the actual experimental result. Figure 11 shows the relationship of the mean interference-square to frequency. Along the Y-axis, frequency represents the number of occurrences of a particular value of the interference-square. The interference-square is mostly at 7 and is a right-skewed distribution.

Figure 12 shows the relationship of the mean interference-square to the probability density function (pdf). As can be seen in Figure 4, it is again a right-skewed distribution. A smaller value for the mean interference-square means a larger value of the pdf. When the mean interference-square is 4, the pdf has the largest value of 0.31.

The visual results so far suggest that interference is a random variable if the received signal strengths from APs are also all random variables. Intuitively this makes sense. The results in these experiments indicate that the interference mostly occurs in a right-skewed distribution, which implies a smaller value for the mean interference-square. It also shows that a non-central chi-distribution (14) could represent the distribution of interference. It would be ideal to have less interference and a smaller fluctuation.

The following sections will continue to look at some factors that affect interference: channel assignment, number of APs, the SINR value, and distribution of location uncertainty. It also carries out a 2D simulation of location uncertainty in a space based on the proposed interference distribution (14).

3.6. The Effect of Channel Interference on Positioning Accuracy: Experiment and Results

This section describes experiments on the effect of channel interference on positioning accuracy. The experiments were conducted in a 150m X 100m testing area. Accuracy was measured using only the probabilistic LF approach, which is defined by (6). The radius of coverage of each AP is 80m. The signal strength ranges between -85dBm and -30dBm. The positioning resolution is set to three meters. Figure 13 shows the actual measurement setup and Table 2 summarizes the experiment settings of Set B.

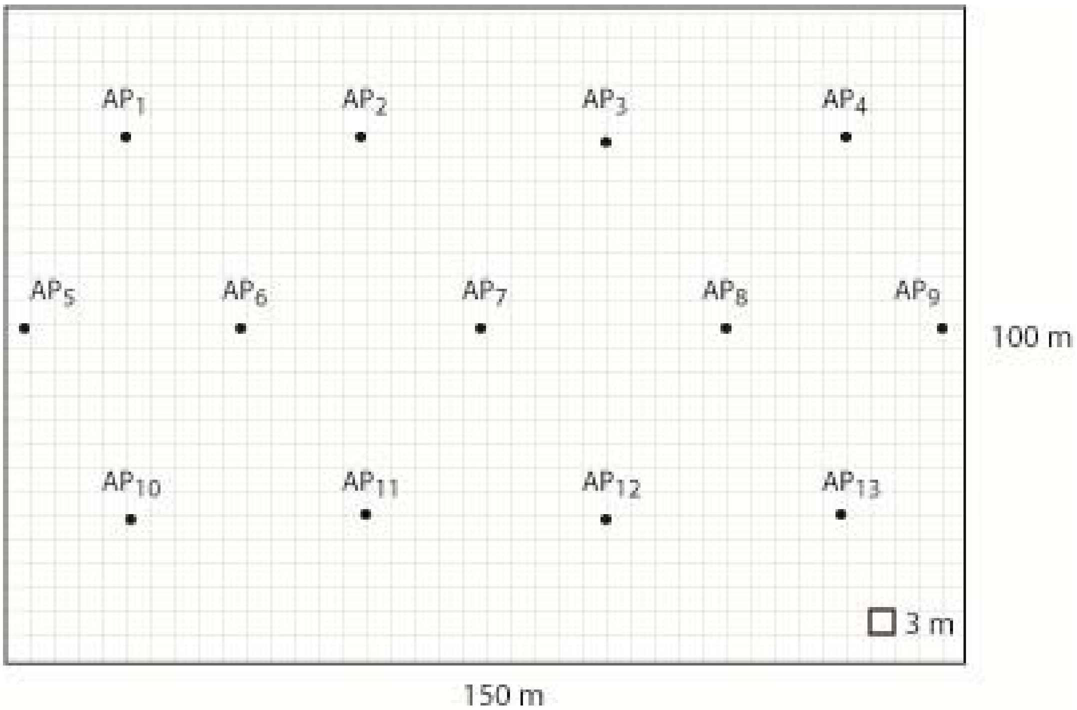


Figure 13 Testing area with 3 X 3 m2 grid

Item	**Description**
Total area	150m x 100m
Number of APs	13 Aps
Positioning resolution	3 meters
Wi-Fi coverage from each APs	80 meters
Range of signal strength	-85dBm to -30dBm

Table 2 EXPERIMENTAL SETTINGS (SET B)

The following discusses the results and compares the positioning accuracy under three typical channel allocation schemes. Subsection 3.6.1 looks at the impact of channel interference on positioning accuracy. Subsection 3.6.2 investigates the positioning accuracy by varying the number of access points under different channel allocations. Subsections 3.6.3 and 3.6.4 look at how channel interference is affected by varying the number of APs and SINR values. Subsection 3.6.5 illustrates how location uncertainty is distributed under different channel allocations.

3.6.1. Effect of Interference on the positioning accuracy

Figure 14 shows the relationship of channel interference to positioning accuracy. In order to see how channel interference affects positioning accuracy specifically, the number of APs is set to 13. The channel interference is varied from 0 to 25 dBm and the accuracy is in a scale from 0 to 1 (1 represents 100% accuracy). The three different channel allocations do not have major difference of positioning accuracy when the interference value is still small. When the interference strength increases above 15dBm, positioning accuracy deteriorates seriously. It is thus clear that when channel interference increases, the positioning accuracy decreases. However, this interference value is difficult to control because it depends on the environment. One way of improving this is to take more iteration. As can be seen in Figure 14, the positioning performance of orthogonal channel allocation is the most accurate. This result indicates that orthogonal channel allocation is 10% more accurate when the system is burdened with high channel interference.

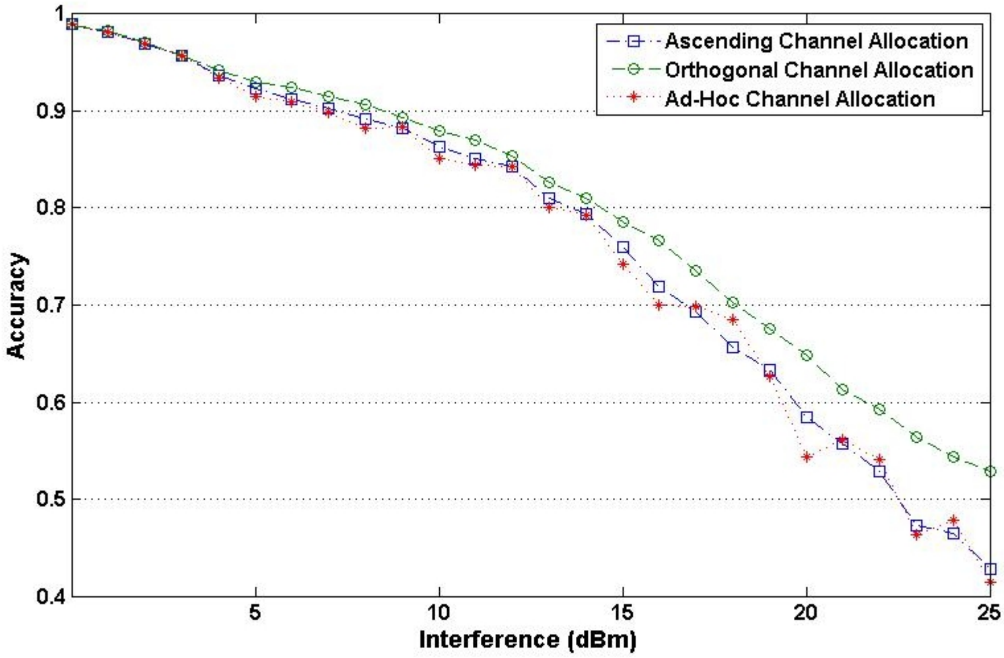


Figure 14 Relationship of interference to accuracy under ascending, orthogonal and ad-hoc channel allocations

3.6.2. Effect of number of APs on the positioning accuracy

This section considers the impact of the number of APs. Figure 15 shows the relationship of the number of access points to accuracy using each of the three allocation schemes. The resolution is 2m. A higher number of APs improves the precision dramatically up to the point that nine APs are used. If more than nine APs are used, the accuracy does not increase significantly due to the interference between them.

The channel interference between APs increases when the number of APs increases. Figure 15 shows that orthogonal channel allocation with only nine APs achieve 90% accuracy. Perhaps the most important point to note is that orthogonal channel allocation could require 15% fewer APs than either ascending channel allocation or ad-hoc channel allocation. Again, the smaller the interference is, the more accurate the positioning is. This is because orthogonal channel allocation provides less interference than the other two channel allocations. The next subsection further investigates this issue.

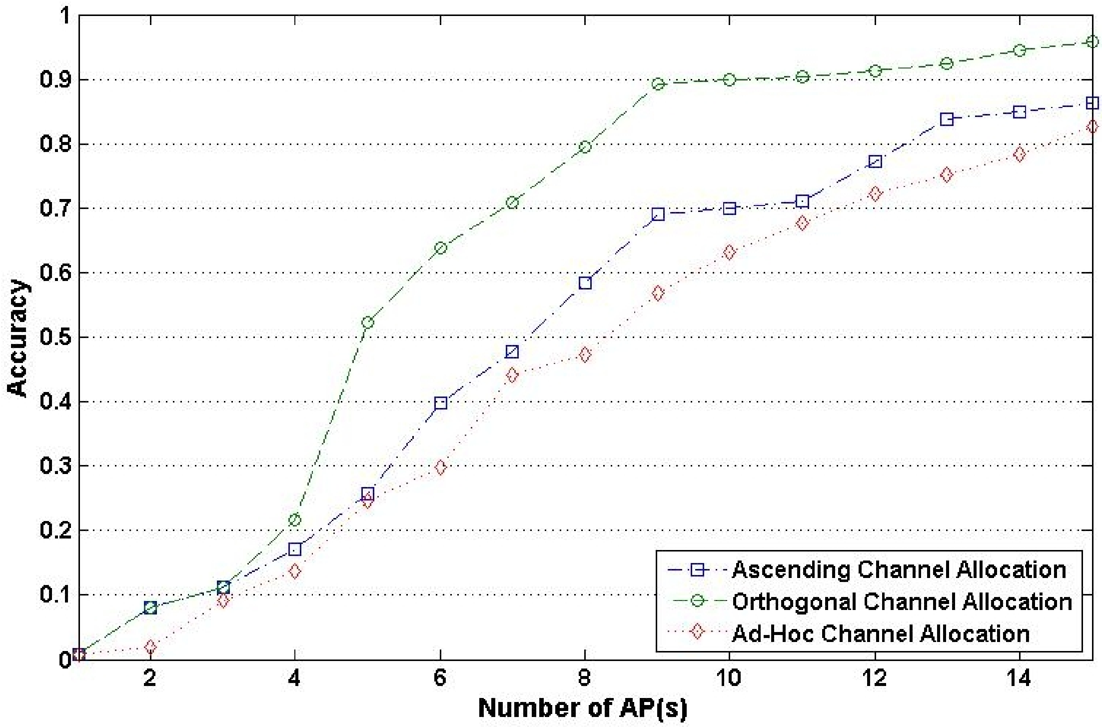


Figure 15 Relationship of number of access points to accuracy given ascending, orthogonal and ad-hoc channel allocations

3.6.3. Effect of number of APs on the interference

Figure 16 shows the relationship of the number of APs to interference. The result suggests that more APs cause in more interference. Orthogonal channel allocation is associated with less interference in any case, an average of 10.9 dBm, whereas ascending channel allocation and ad-hoc channel allocation respectively average of 13.1 dBm and 14.24 dBm of interference.

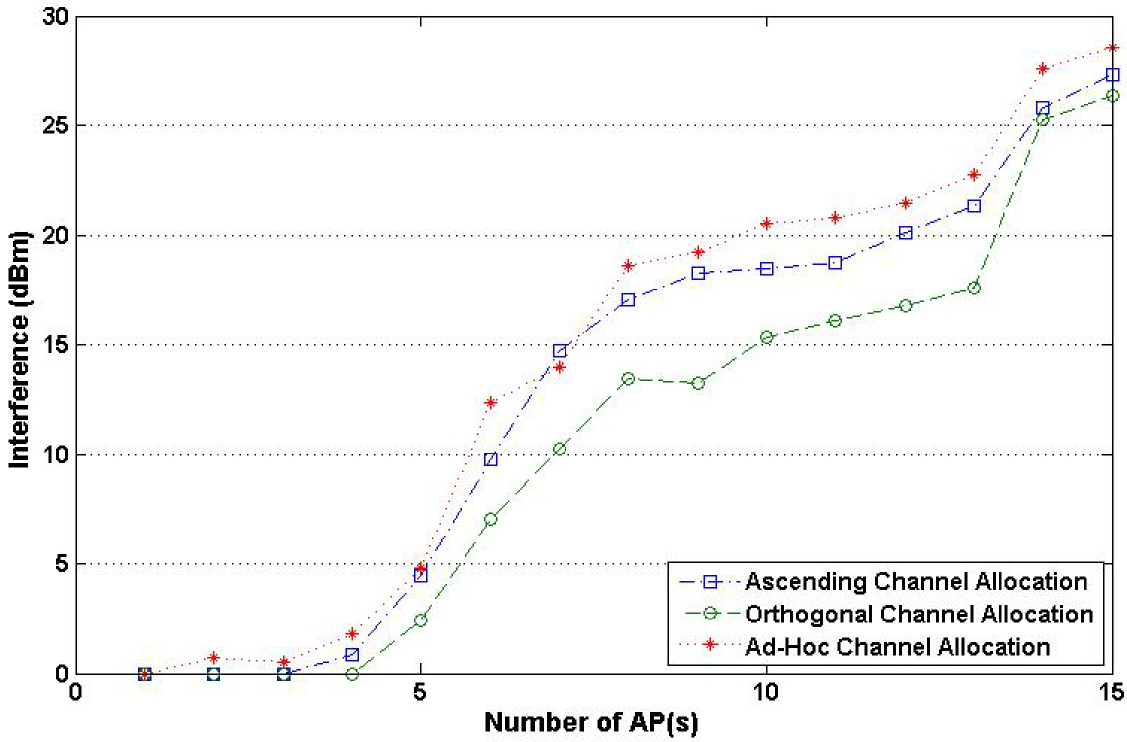


Figure 16 Relationship of number of APs to interference under ascending, orthogonal and ad-hoc channel allocations

3.6.4. Effect of Signal-to-Interference-plus-Noise-Ratio on the positioning accuracy

In order to see how SINR affects the positioning accuracy, the number of APs is set to 13. The SINR is varied from 0 to 1 and the accuracy was in a scale from 0 to 1 (1 represents 100% accuracy). Figure 17 shows the relationship of SINR to positioning accuracy. The result indicates that the higher the value of SINR, the more accurate the positioning. Orthogonal channel allocation has the best positioning performance and was 8% more accurate than the other two channel allocations.

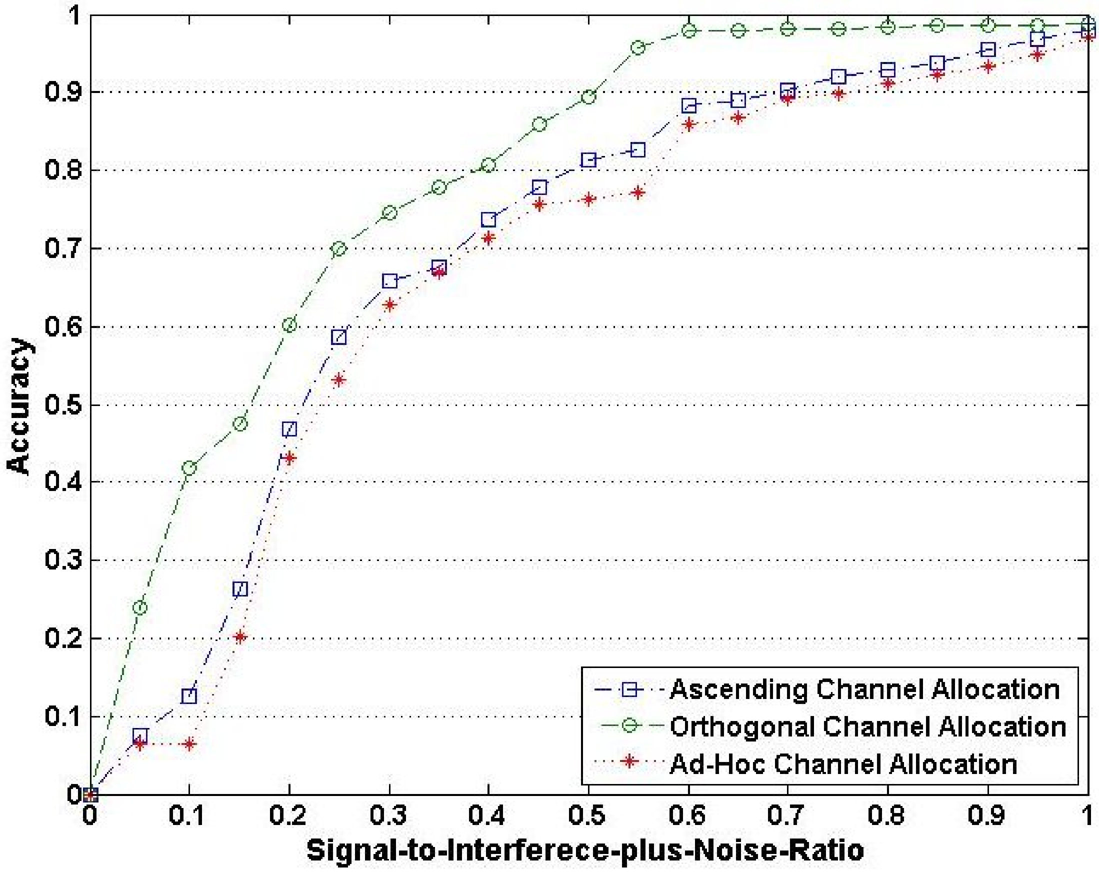


Figure 17 Relationship of Signal-to-Interference-plus-Noise-Ratio to accuracy under ascending, orthogonal, and ad-hoc channel allocations

3.6.5. Result of 2D Simulation of Location Uncertainty under ascending, orthogonal and ad-hoc allocation

The purpose of 2D simulation is to investigate how different channel allocations might affect points of location uncertainty and to allow a visual analysis of the effect of channel interference.

Figure 18, 19, 20 show 2D simulation results of location under ascending, orthogonal, and ad-hoc allocation. Each blue arc represents the coverage of an AP. The colored dots represent positions of localization uncertainty. Each pair of colored dots represents where localization returned the same RSS reading for two coordinates. In other words, more colored dots means the positioning accuracy is worse. The number of APs is fixed at 13 and each AP is distributed hexagonally.

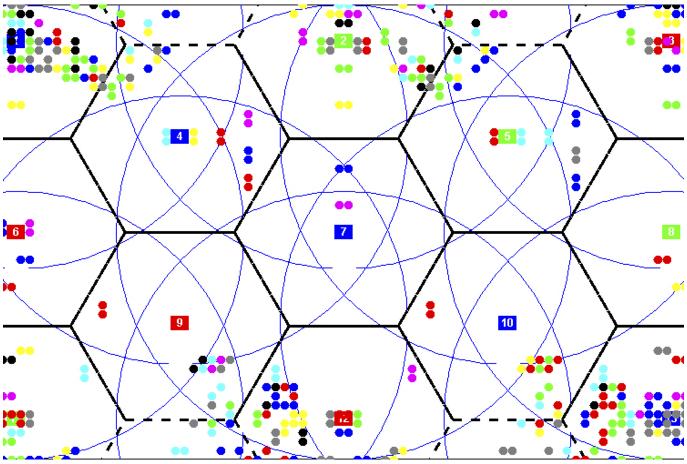


Figure 18 The distribution of location uncertainty using Ascending Channel Allocation

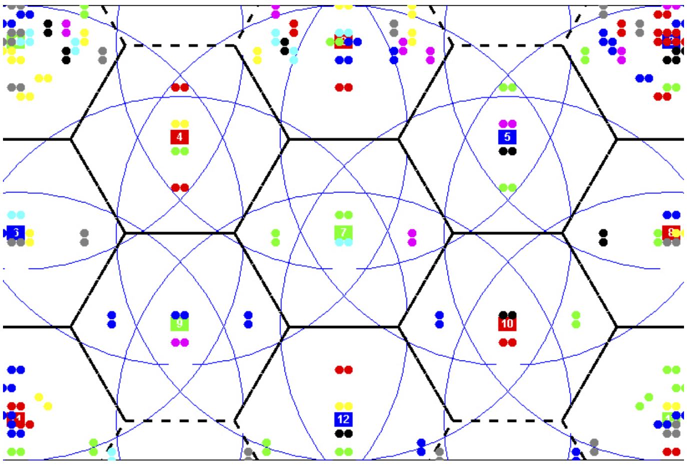


Figure 19 The distribution of location uncertainty using Orthogonal Channel Allocation

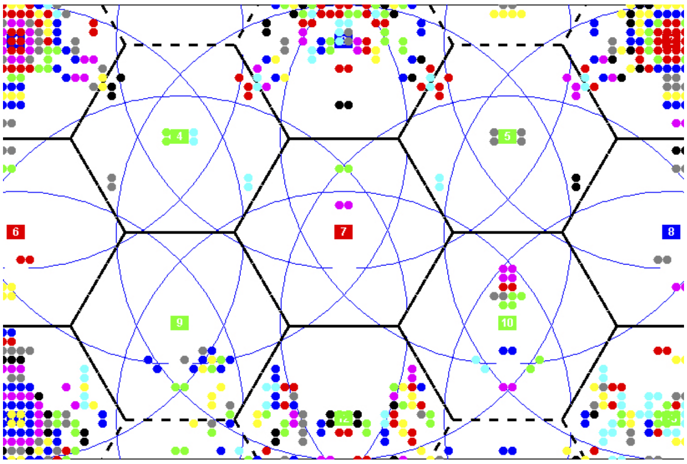


Figure 20 The distribution of location uncertainty using Ad-Hoc Channel Allocation

The colored dot are distributed in corners where there are too few APs to achieve positioning accuracy. Figure 18 shows 3648 colored points and the positioning accuracy is approximately 94.11%. Figure 19 shows 3758 colored points and the positioning accuracy is approximately 96.96%. Figure 20 shows 3558 colored points and the positioning accuracy is 91.8%. Orthogonal channel allocation outperforms the other two channel allocations at every setting.

Chapter 4: Using Wi-Fi Signal Strength to Localize in Wireless Sensor Networks

Wireless sensor networks (WSNs) are widely used in many applications such as localization and real-time tracking systems. Previous research commonly suffered from the line-of-sight (LOS) problem and dependence on contrast of the background light intensity. Location Fingerprinting (LF) methods use a training dataset of received signal strength (RSS) at different locations to track the target. The drawback of the LF method is the need to have an extensive training dataset based on site surveys and that it is greatly affected by changes in the internal building infrastructure. This chapter implements a sensor-based LF method to replace extensive site surveying. A Kalman Filter is used to trace multiple points to characterize a trajectory. Our experimental result shows that the proposed method leads to more accurate and effective tracking.

The rest of this chapter is organized as follows: Section 4.1 describes how the use of a Kalman Filter can improve WLAN tracking. Section 4.2 describes the design of the proposed WSN localization approach. Section 4.3 presents trajectory accuracy improvement using a Kalman Filter. Finally, Section 4.4 discusses any improvements in location estimation accuracy obtained using this method.

4.1. Introduction

WSNs are increasingly ubiquitous in public places, in for example, airports, malls, cafes, campuses, and public squares. They are fuelling a wide range of location-aware computing applications. Currently, Wi-Fi enabled devices can be located by applying one of two location-sensing techniques, propagation-based [Prasithsangaree, Krishnamurthy et al. 2002; Jan and Lee 2003; Kwon, Dundar et al. 2004] and location fingerprinting (LF) [Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Li, Wang et al. 2005]. Propagation-based techniques measure the received signal strength (RSS), angle of arrival (AOA), or time difference of arrival (TDOA) of received signals and apply mathematical models to determine the location of a device. The drawback of propagation-based methods is that in order to achieve accurate localization they need to compute every condition that can cause wave signals to blend. Location fingerprinting allows a device to locate itself by using a device to access a database containing the fingerprint (i.e., the RSSs and coordinates) of other devices within the Wi-Fi footprint and then calculate its own coordinates by comparing them with the LF database. The drawbacks of the LF method include the need to have an extensive training dataset based on surveys and that the method is much affected by changes in internal building infrastructure.

This chapter describes the implementation of a localization approach that makes use of the increasingly ubiquitous Wi-Fi network by using a WSN that estimates the location of sensors using an LF methodology. The approach uses LF-based techniques and sensors in two phases. The first phase detects IEEE 802.11b Wi-Fi signal strength and then uses a set of static location fingerprint sensors to collect the location fingerprints for a training database. In the second phase, the location fingerprints are retrieved by the mobile Wi-Fi enabled device and the location is estimated by applying the k-nearest neighbor algorithm to the LF training database. A Kalman Filter is used to track multiple points to characterize a trajectory. This WSN-based localization approach offers a number of benefits. First, it obviates the need for extensive manual site surveying. Second, it is potentially suitable for every environment, indoor or outdoor, notwithstanding topography, the presence of man-made structures, or environmental conditions. Finally, it is accurate and cost-effective.

4.2. The design of the WSN Localization Approach

A test bed was established in a laboratory on 7th floor of the PQ building, in Department of Computing, at The Hong Kong Polytechnic University. The layout of the laboratory is shown in Figure 1. The room is approximately 10m by 4m.

The radio frequency channels of IEEE 802.11b are in the 2.4GHz band. There are three non-overlapping channels for 802.11b. The received signal sensitivity also limits the range of the RSS to be -90 dBm and -30 dBm. Nevertheless, the highest typical value of the RSS is approximately -40 dBm at one meter from any AP. Samples at 33 locations are shown in Figure 1. Each data set of samplings consists of 20 samplings collected at approximately 0.25-second sampling interval. In our case, four access points were distributed in the room. Figure 21 illustrates the position of the access points on the grid at (2, 4), (1, 0), (8, 4) and (10, 1).

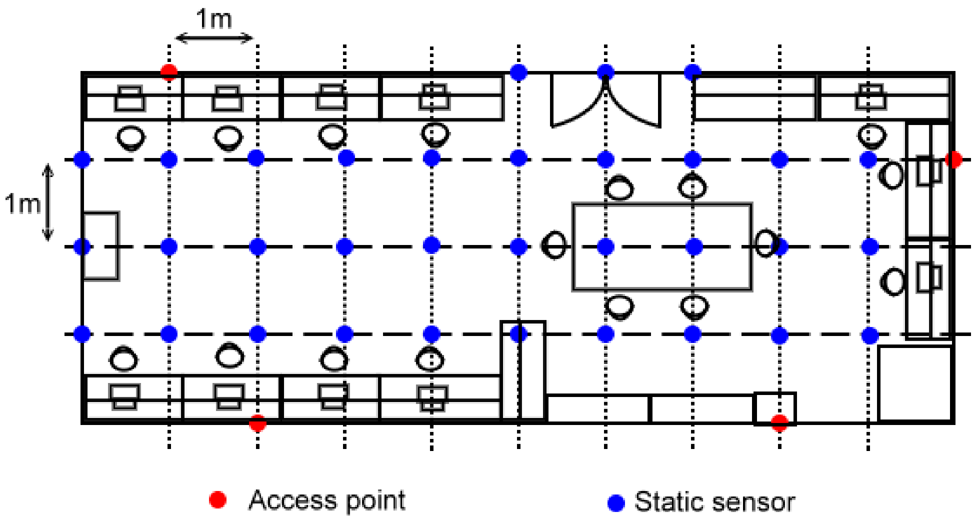


Figure 21 Test bed environment with 33 grid points

4.3. Trajectory Estimation with K-Nearest Neighbor approach

The K-Nearest Neighbor (K-NN) algorithm requires two sets of data. The first set of data is the offline samples of RSS from *m* APs in the area. Each element in a vector is an independent RSS (in dBm) collected from APs in the location. $\mathrm{S}=\left\{\mathrm{s}_1 \ldots s_2 \mid s_i \in \mathbb{R}^n\right\}$ is a set of online sampling LF vectors in database. The second set of data contains online RSSs, $\mathrm{R}=\left\{\mathrm{r}_1 \ldots r_2 \mid r_i \in \mathbb{R}^n\right\}$ from *n* APs at a particular location.

The K-NN algorithm requires the collection of data $\left\{(s_i, d_q), i, q \in \mathbb{N}\right\}$, for *n* locations in the site, where d_q is the known location of the *q*th measurement and the vector s_i is the "fingerprint" of the location d_i. When a receiver in unknown location A becomes aware of a new fingerprint *r*, it searches for the fingerprint s_i that is closest to *r* and then estimates the location. The unknown location for r is decided by a majority vote from the K shortest distance fingerprints.

We can estimate the location d_q by clustering the distance between online received LF vector r and offline sampling LF vector s_i as

$$\mathrm{d_q}(r,f)=\left(\sum_{i=1}^{n}\left|r-s_i\right|^q\right)^{\frac{1}{q}} \tag{15}$$

d_q is called Manhattan distance if $q=1$ and Euclidean distance if $q=2$ the accuracy does not necessarily higher as q increases.

Let s_{ij} be jth sample RSS in the ith access points. m is the number of access points. n is the number of sample data. The distance between s_i and s_{ij} is defined as

$$d_j=\sqrt{\sum_{i=1}^{m} f_{ij}-s_i} \qquad j=1...n \tag{16}$$

Electing K samples since the smallest value and calculate average coordinates as outputs in following equation:

$$(x,y)=\frac{1}{K}\sum_{i=1}^{K}(x_i,y_i) \tag{17}$$

where (x_i,y_i) is coordinates corresponding to ith sample.

K-NN is simple to implement and it provides reasonable accuracy. However, one drawback of the standard K-NN suffers from signal fluctuations because the RSSs detected in the same location can be varied from time to time. The fluctuations may cause errors in location estimation. This can be partially overcome by having multiple fingerprint sets for a given location, taken at different times, assuming that one or other finger print may cover that fluctuation.

4.4. Trajectory Estimation with a Kalman Filter

To improve trajectory accuracy, the Kalman Filter [Chan, Baciu et al.] is used to estimate the state $x \in \Re^2$ of a discrete-time tracking process that is governed by the linear stochastic difference equation. Our time update equations are

$$\begin{aligned} \hat{x}_k^- &= \hat{x}_{k-1} \\ P_k^- &= P_{k-1} + Q \end{aligned} \tag{18}$$

And our measurement update equations are

$$\begin{aligned} K_k &= P_k^-(P_k^- + R)^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \\ P_k &= (1 - K_k)P_k^- \end{aligned} \tag{19}$$

where Q is the process noise covariance and R *is* measurement noise covariance. P_k^- is the a priori estimate error covariance. P_k is the a posteriori estimate error covariance. $\hat{x}_k^-$ is the a priori state estimate at step *k*. $\hat{x}_k$ is the a posteriori state estimate at step *k*. K_k is the Kalman blending factor.

4.5. Performance Evaluation

This section describes experiments and results for the probability of obtaining a correct location fingerprint.

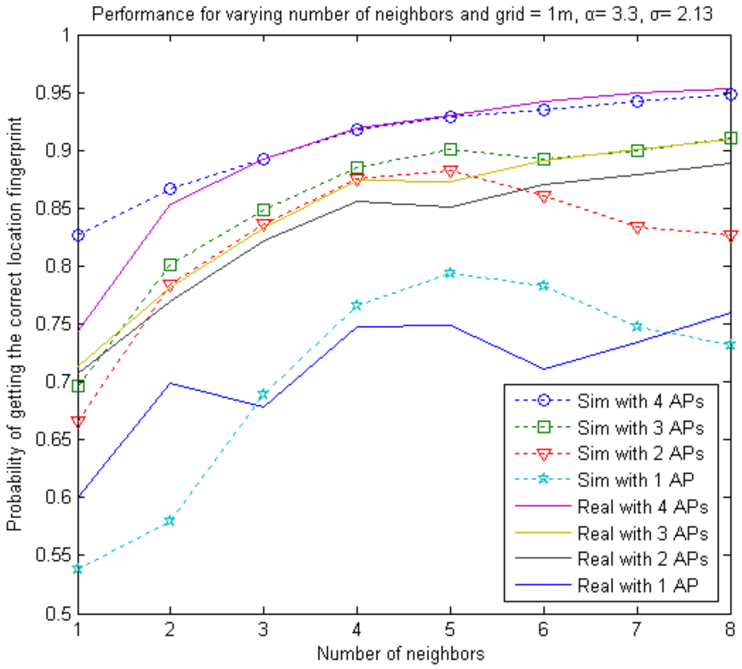


Figure 22 Influence of varying standard deviation σ

4.4.1. Results for the probability of getting the correct location fingerprint

Figure 22 shows the influence on the probability of getting the correct location fingerprint when varying the standard deviation, σ. Assuming the four nearest access points are used for calculation, when the standard deviation σ increases, the probability of getting correct location fingerprint decreases. Therefore, high localization accuracy is most likely achieved with standard deviation σ below 4dBm (In other reported measurements, σ was assumed to be 2.13dBm [Kaemarungsi and Krishnamurthy]) and more than three neighbors were required for estimation. However, the standard deviation will depend on environmental factors such as the fact that the human body absorbs some of the signal strength. So in a busy environment, the standard deviation will be larger.

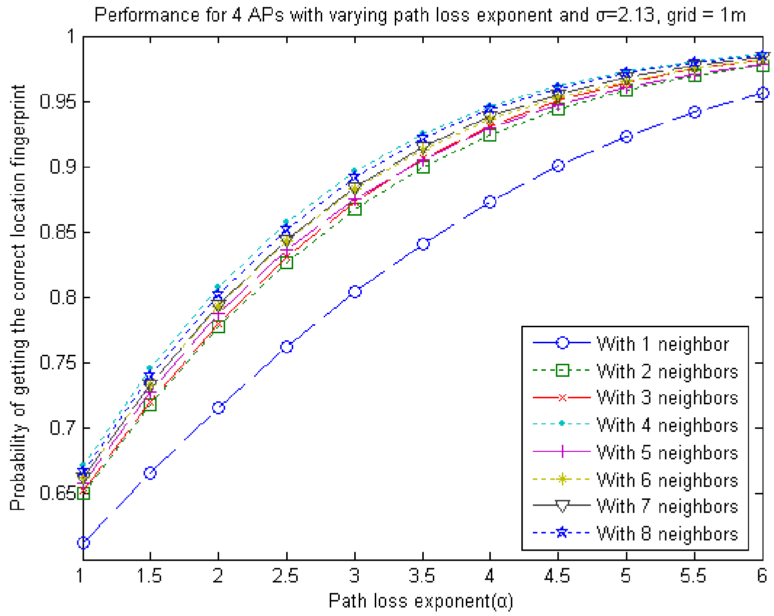


Figure 23 Influence of varying path loss exponent α

Figure 23 shows the influence on the probability of getting the correct location fingerprint when varying path loss exponentα. The path loss exponent α can be varied between 1 and 6. When the path loss exponent α increases, the accuracy also increases. Path loss exponent α represents the attenuation rate of the RSS. If the path loss exponent is large, it means that the RSS will be greatly affected by just a small shift in a distance. Therefore, a high path loss exponent makes it easier to recognize the location fingerprint and to increase the probability of getting the correct location fingerprint.

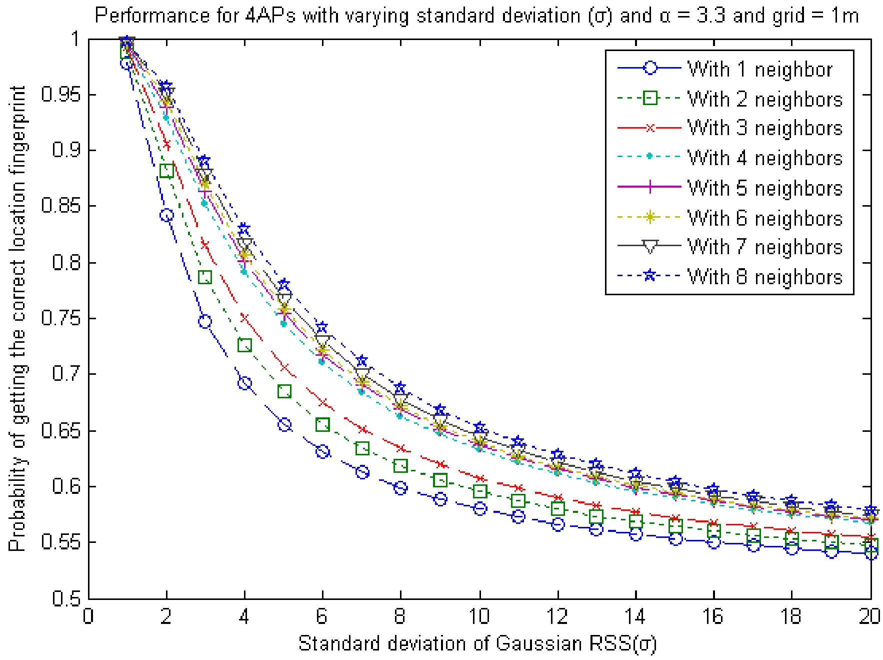


Figure 24 Influence of varying the number of neighbors

Figure 24 shows the influence of varying the number of neighbors on the probability of getting a correct location fingerprint. In our case, there are eight neighbors placed at eight directions. If more neighbors are used, the accuracy increases. This means that the more comparisons between neighbors and mobile sensors, the greater the probability of getting a correct location fingerprint. This probability depends not only on the number of neighbors but also on the number of access points. For example in our case, the probability in comparison with K= 8 neighbors increases approximately from 76% to 95%, when it increases the number of access points from 1 to 4. In conclusion, the more access points, the more likely we are to get a correct location fingerprint.

4.4.2. Results for adding the location tracking filter

The K-Nearest Neighbor (K-NN) algorithm is used to estimate the location of the mobile sensor. After the location is calculated, a Kalman filter is added to improve accuracy. The experiments in this section observe the trajectory of accuracy when a user with mobile sensor walks around the room. It is assumed that the Gaussian RSS noise measurement is 4.

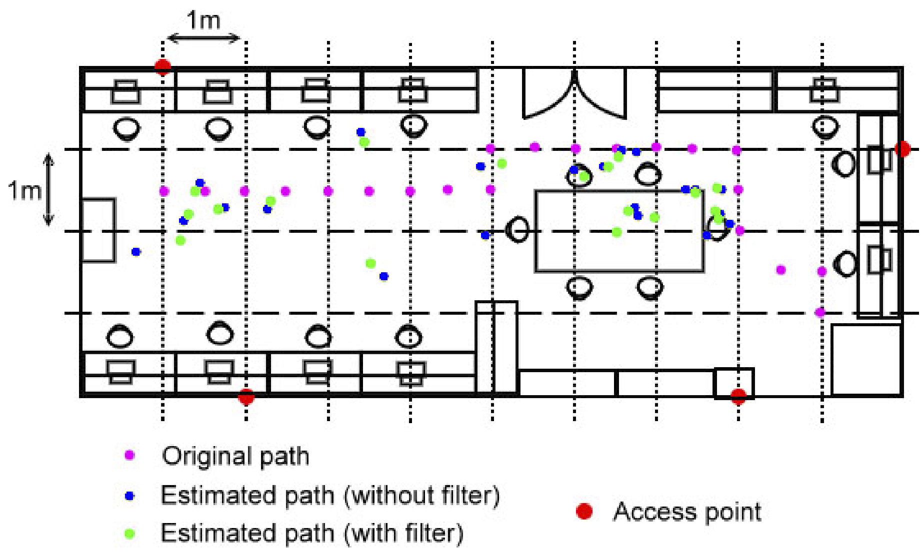


Figure 25 The actual and estimated (both with and without filter) paths of the user

Figure 25 shows the actual and estimated (both with and without filter) paths of the user. Figure 26 shows the reduction of error when using a Kalman Filter. The use of (15) and (16) slightly improves the trajectory accuracy, reducing the error from 1.16m to 1.10m. Figure 27 shows error on the X and Y coordinates. The error on the X-axis is greater than on the Y-axis (0.83m and 0.58m respectively). In the testing environment, the width of the room (X-axis) is longer than the length of the room (Y-axis). Therefore, the distance between mobile sensors and the access points in X-axis is longer than in Y-axis. Because on a longer distance, the probability of furniture blocking the signal is higher.

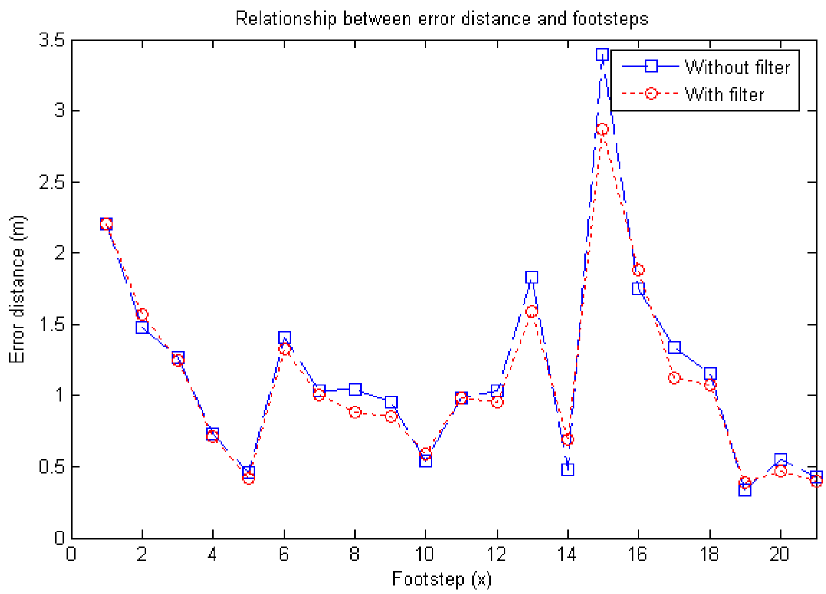


Figure 26 Reduction of error when using a Kalman Filter

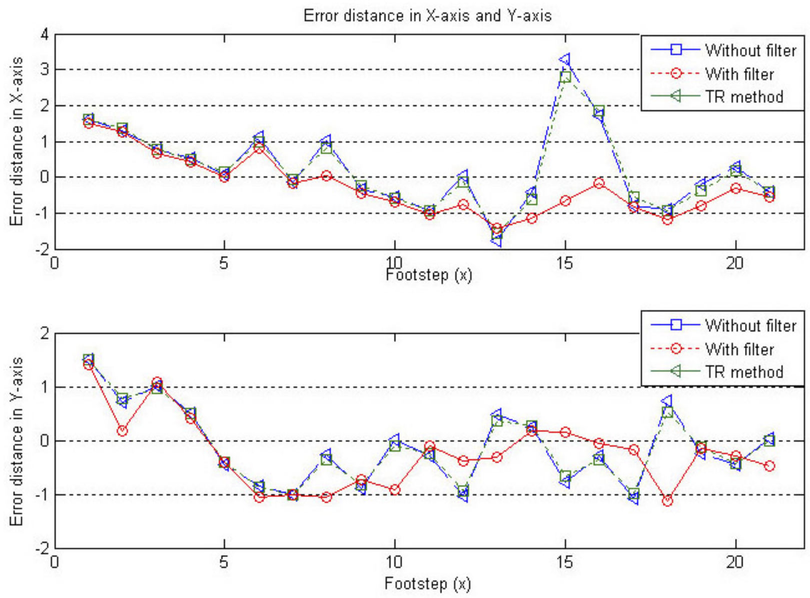


Figure 27 Error distance on the X and Y coordinates

Chapter 5: Using Newton Trust-Region Method to Localize in WLAN Environment

The Newton Trust-Region (TR) method is used to iteratively optimize the localization approach. Previous uses of the Newton Trust-Region method have been in areas other than in WLAN localization. This chapter proposes an accurate WLAN localization that uses the Newton Trust-Region method. The Newton Trust-Region method is used to optimize the Location Fingerprinting approach iteratively. The proposed method makes three main contributions. First, the WLAN received signal strength (RSS) can fluctuate a lot so the TR method makes use of the convergence of a trajectory to eliminate a noisy RSS. Second, the TR method is more effective than other iterative non-linear optimization methods. Third, the TR method requires 15% less access points and it is 10% more effective at precise localization than traditional Location Fingerprinting approaches. The experimental analysis shows that the proposed TR method produces a substantially more accurate and robust localization system.

The rest of this chapter is organized as follows: Section 5.1 introduces the background of how the Newton Trust-Region is used to localize in a WLAN environment. Section 5.2 presents the proposed Trust-Region algorithm. Section 5.3 describes the experimental setup. Finally, Section 5.4 discusses its performance in terms of location estimation accuracy.

5.1. Introduction

This chapter proposes a Newton Trust-Region (TR) method for localization based on the LF approach. The TR method [Liu and Chen 2004; Chang, Hong et al. 2007] has been used in applications such as solving magnetic field problems [Silvester and Chari 1970], tracking dim targets in infrared imagery [Wang, Zhang et al. 2006] and nonlinear programming problems [Ke 2007]. However, this method has not been used to optimize LF approaches. My previous work using the Kalman Filter [Chan, Baciu et al. 2009e] improved on the traditional LF by filtering noisy signals with a priori and a posteriori conditions. However, Kalman Filters do not make use of convergence factors of the trajectory. The proposed method makes use of the TR method and the LF-based approach in three phases. The first phase is to detect the IEEE 802.11b Wi-Fi signal strength and collect the LFs into a training database. In the second phase, the LFs (location fingerprints) are retrieved by the mobile WLAN-enabled device and used to estimate the location by applying the K-Nearest Neighbor method to the LF training database. The third phase is to acquire trajectory of a user (WLAN-enabled device). Normally, each step or estimated point in a trajectory falls within a region and has the same convergence. The TR method used a feedback process to track a trajectory. An experiment comparing the accuracy of the TR method with that of my previous research results using the

Kalman Filter [Chan, Baciu et al. 2009f] and traditional LF approaches was conducted on one floor of an on-campus building

The proposed approach offers a number of benefits. First, WLAN received signal strength (RSS) can fluctuate a lot but the Trust-Region method makes use of the convergence of a trajectory to eliminate noisy RSS. Second, the Trust-Region method is more effective than other iterative non-linear optimization methods. Third, the Trust-Region method addresses rotation of a WLAN-enabled object rather than working on some presumably possible discrete values of rotation angle. Finally, it is potentially suitable for every environment, indoor or outdoor, notwithstanding topography, the presence of man-made structures, or environmental conditions.

5.2. Trust-Region Algorithm

The TR method [Wang, Zhang et al. 2006; Ke 2007] is a nonlinear optimization algorithm that is used to optimize trajectory tracking. Consider a typical unconstrained minimization of a location error problem,

$$\min_{x \in \vec{V}} f(x) \tag{20}$$

where $\vec{V}$ is a vector space. $f(x)$ is derived from the propagation-based theorem in (1). At iteration k, with iterate x_k and TR radius $\triangle k$, the TR set is:

$$A_k = \left\{ x \middle| \in \vec{V} \middle| \|x - x_k\|_k \leq \Delta k \right\} \tag{21}$$

There are three essential elements to the TR method: 1) the TR subproblem, which involves approximating a location minimizer in the region, 2) TR fidelity, which involves evaluating the accuracy of the location and 3) TR radius, which is used to determine the size of a trust region.

5.2.1. TR subproblem

A quadratic model m_k is constructed to approximate f(x) within the TR. The inner product is denoted by <,>. A TR subproblem computes whether $x_k + s_k$ is in the region.

$$m_k(x_k + s) = m_k(x_k) + \langle g_k, s\rangle + \frac{\langle s, H_{x^S}\rangle}{2} \tag{22}$$

where $m_k(x_k) = f(x_k)$, g_k is the gradient or first derivative of $f(x)$ at x_k, and H_k is the Hessian of f or second derivative of $f(x)$ at x_k. When $H_K \neq 0$, m_k is said to be a second-order model. A TR subproblem compute an s_k, the TR region subproblem as follows:

$$s_k = \arg\min \psi_k(s) \stackrel{\Delta}{=} \langle g_k, s\rangle + \frac{\langle s, H_{x^S}\rangle}{2} \tag{23}$$

5.2.2. TR fidelity

After solving the TR subproblem, the trial point will be tested to see if it is a good candidate for the next iterate. This is evaluated by

$$p_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)} \tag{24}$$

Suppose an initial trust region is given and let η_1, η_2, γ_1 and γ_2 be constants satisfying $0<\eta_1<\eta_2<1$ and $0<\gamma_1<\gamma_2<1$.

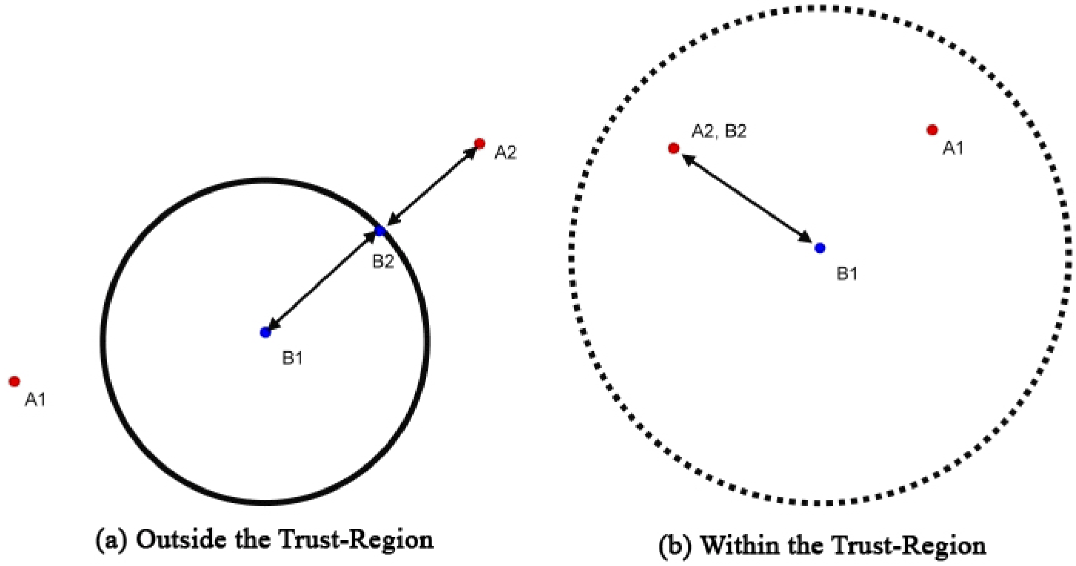
If $p_k > \eta_1$, then the trial point is accepted, i.e., $x_{k+1} = x_k+s_k$. Otherwise, $x_{k+1} = x_k$. When m_k approximates f well and yields a large p_k, the TR radius will be expanded for the next iteration. Otherwise, if $p_k < \eta_1$ or $p_k < 0$, m_k approximates f not well within the current region A_k. Therefore, the iteration remains unchanged and the TR radius will be shrunk to derive a more appropriate model and TR subproblem for the next iteration.

5.2.5. TR Radius

I can update the TR radius as follows:

$$\Delta_{k+1} = \begin{cases} \gamma_2 \Delta_k & if \quad p_k \geq \eta_2, \\ \Delta_k & if \ p_k \in [\eta_1, \eta_2] \\ \gamma_1 \Delta_k & if \quad p_k < \eta_1 \end{cases} \qquad (25)$$

where let η_1 = 0.3, η_2 = 0.95 and γ_1= 0.7, γ_2= 1.5. The iterative process for (22) will be repeated until the sequence of iterates x_k converges



.

Figure 28 Two cases of Trust-Region

Points may be either outside or within a TR, as in Figure 28. A_1 is the priori estimated location, B_1 is the priori amended location of A_1. A_2 is the current estimated location and B_2 is the current amended location of A_2.

A_2 was tested to see whether it was within the trust region. If it was within the trust region, then there was not need to amend the current location. Otherwise, it would be necessary to amend A_2 by using the TR radius to find the co-ordinate of B_2, as in Figure 28.

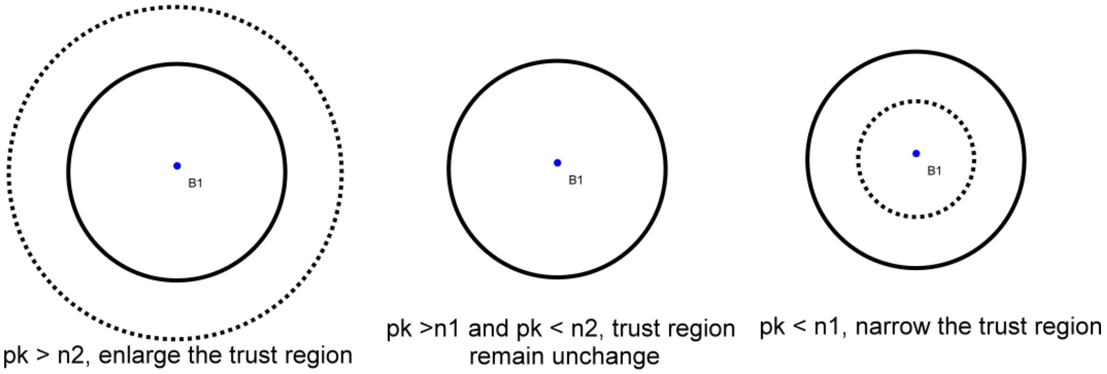


Figure 29 Three cases of radius adjustment

The radius of the TR method was then adjusted to bind the next estimated points with Trust-Region by (22). The distance ratio between $|A_1\text{-}A_2|$ and $|B_1\text{-}B_2|$ was calculated. If the ratio was greater then 0.95, then the radius of Trust Region was increased 1.5 times. If the ratio was smaller then 0.3, then the radius of trust region was reduced 0.7 times. Otherwise, the radius of the Trust Region remained unchanged. Figure 29 shows three cases of the radius adjustment.

5.3. Experiment Setup

In the following sections I describe a localization experiment to determine the localization accuracy of the traditional LF approach, the Kalman Filter, and my proposed TR method where the maximum number of APs is 15. (Use of the Kalman Filter is one of this author's previously proposed approaches [Chan, Baciu et al. 2009f])

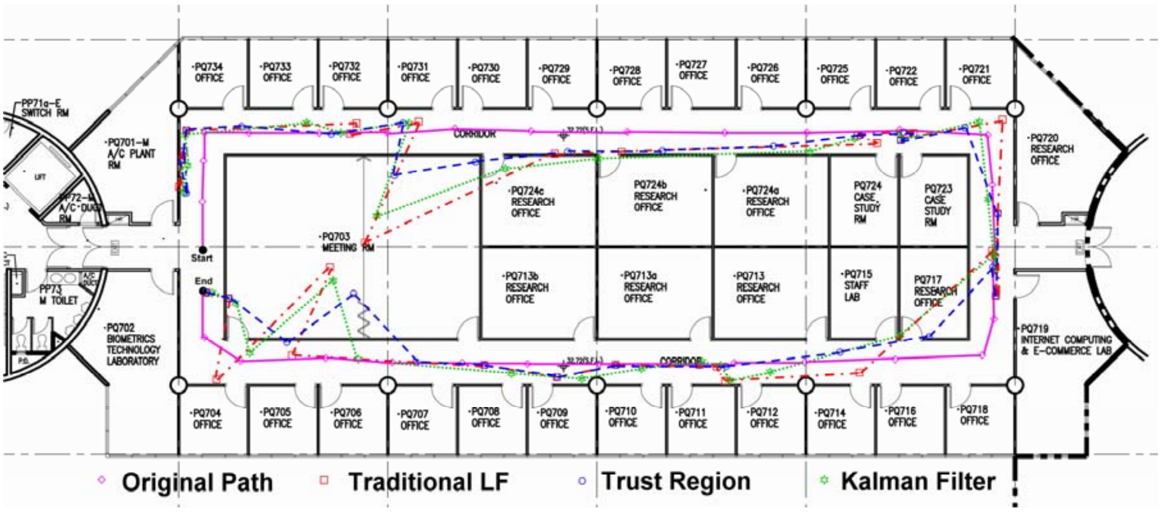


Figure 30 Walking trajectory and estimated trajectory on the 7th floor PQ building at The Hong Kong Polytechnic University

The experiment is conducted on the 7th floor of the PQ building, at The Hong Kong Polytechnic University. The experiment covers a total area of 50m x 20m. I walked through the hallway on the 7th Floor with a WLAN-enabled device, using (3) and (17-22) as the positioning algorithm to estimate the location of the device. The estimated and actual coordinates were calculated and collected at 30 locations in the hallway. At each location, each data set of samplings consisted of 25 RSS samplings collected at approximately four-second intervals. The signal strength ranged from -90dbm to -30dbm. Figure 30 shows the floor plan of the 7th floor of the PQ building at The Hong Kong Polytechnic University. Table 3 summarizes the setup.

Item	**Description**
Total area	50m x 20m

Sampling interval	4 seconds
Number of locations	30 locations
Maximum number of APs	15 APs
Range of signal strength	-90dbm to -30dbm

Table 3 SUMMARY OF EXPERIMENT SETTING

5.4. Performance Evaluation

The major performance metrics of interest for WLAN localization are the accuracy and the precision in estimating a position. This section investigates how the TR method influences the precision when varying the number of APs, resolution, the TR radius, and the TR boundary. Subsection 5.4.1 presents the results of trajectory estimation on the campus floor. Subsection 5.4.2 shows the relationship of the number of access points to accuracy under the TR, Kalman Filter and Traditional Location Fingerprinting methods. Subsection 5.4.3 also shows the relationship of resolution to accuracy. Subsection 5.4.4 shows that the TR method reduces the error distance of the positioning algorithm. Subsection 5.4.5 investigates the relationship of the TR radius and boundary to accuracy.

5.4.1. Result for Trajectory Estimation

Figure 30 shows the walking trajectory and estimated trajectory on the 7th floor of the PQ building at The Hong Kong Polytechnic University under the Trust-Region, Kalman

Filter and Traditional Location Fingerprinting approaches. The original path involved a walk around the hallways. As can be seen in Figure 4, due to signal fluctuation, the estimated path of the traditional LF and Kalman Filter bulged inside the room PQ703 and PQ717 sharply. The TR method partially eliminates inaccurate estimation due to signal fluctuation, maintains more stable estimation and locates more precisely.

5.4.2. Effect of number of APs on the localization accuracy

Figure 31 shows the relationship of the number of access points to accuracy under the three methods. The resolution is 3m. A higher number of APs improves the precision but the accuracy does not increase significantly with more than eight APs. TR method achieves from 50% to 80% accuracy if number of APs changes from three to seven. The superior performance of the TR method becomes progressively stronger as more APs are added and achieve 93% accuracy with only nine APs. Perhaps the most important point to take away here is that the traditional LF approach and Kalman Filter never at any point matches the accuracy of the TR method. In contrast, TR method requires 15% less APs in average to achieve effective localization. The results indicate that TR method is absolutely the best performer.

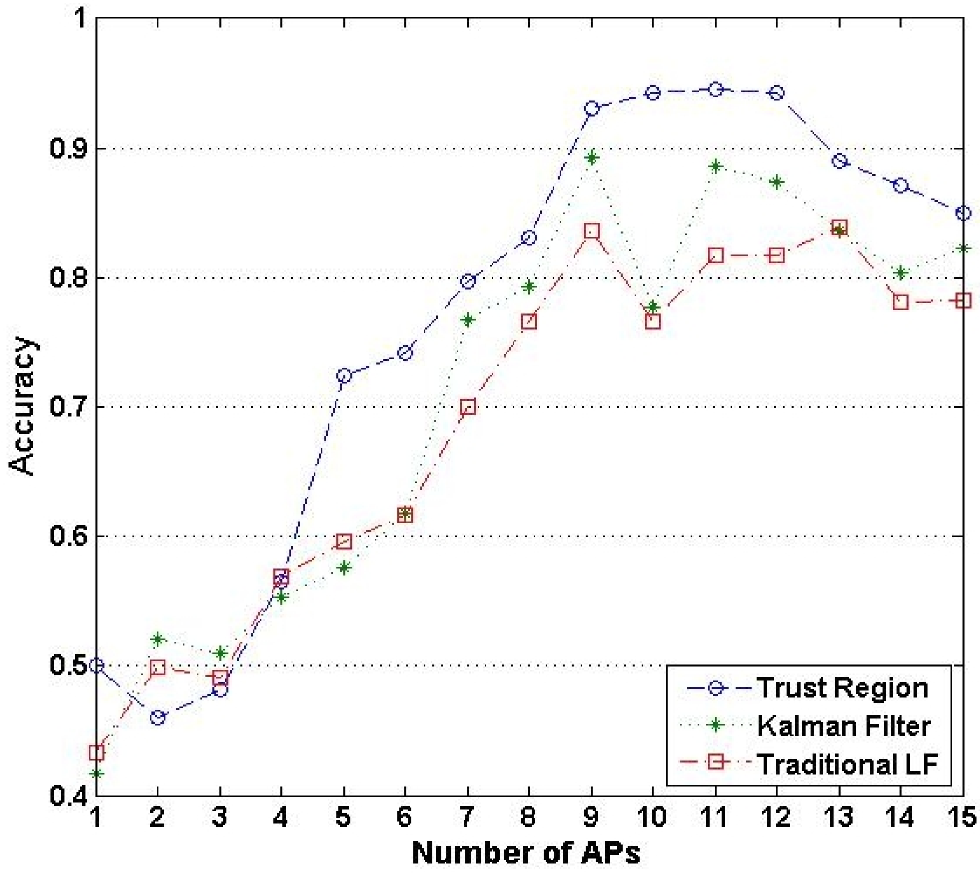


Figure 31 Relationship of number of access points to accuracy under Trust-Region, Kalman Filter and Traditional Location Fingerprinting approach

5.4.3. Effect of resolutions on the localization accuracy

Figure 32 shows the relationship of resolution to accuracy under the three methods. As shown in Figure 32, the higher resolution degrades the precision dramatically. The performance of the Kalman Filter and TR is similar in high resolution. The superior performance of TR becomes progressively stronger in high resolution, particularly, below 2.5 meters. Using the TR method produces over 71% and 92% accuracy when the resolution is 2 meters and 3.5 meters. TR will almost certainly achieve perfect accuracy when the resolution is 4.5m.

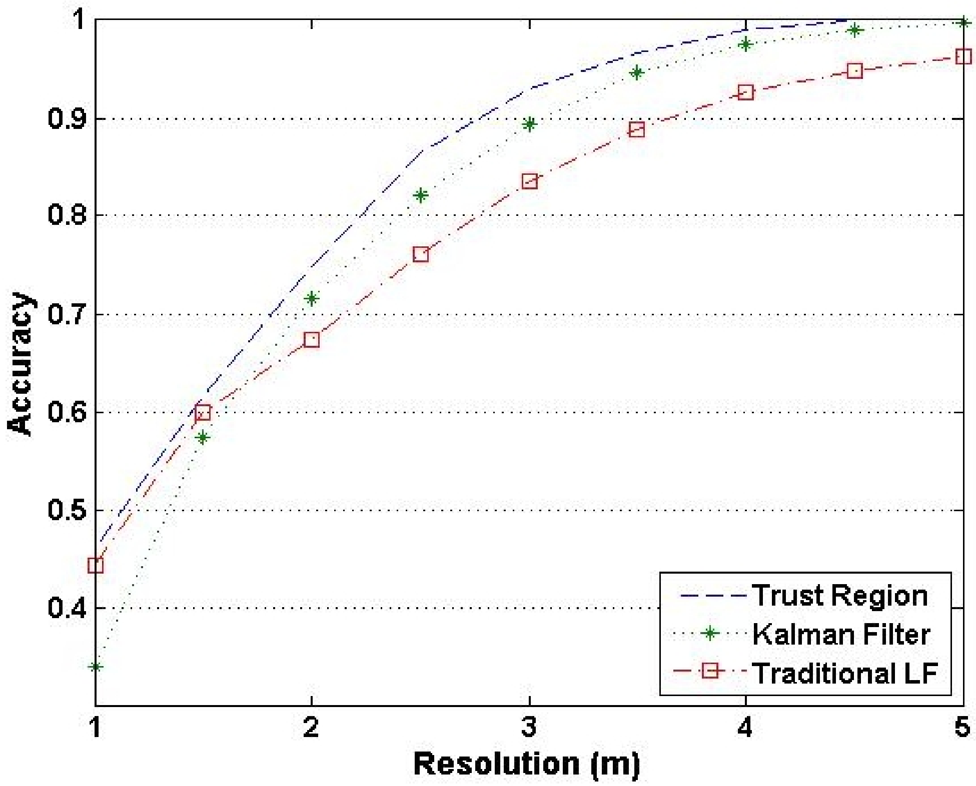


Figure 32 Relationship of resolution (m) to accuracy under Trust-Region, Kalman Filter and Traditional Location Fingerprinting approach

5.4.4. Results for the error distance in trajectories

This subsection considers the impact on the error distance under the TR, Kalman Filter and traditional LF approach. Figure 33 plots the error distance of position estimation. Figure 34 plots the error distance of estimation on the X and Y-axis. The TR method can reduce the average error distance from 2.49m (using a traditional LF approach) to 2.0m. Again, the TR method can reduce the error distance on the X-axis and Y-axis from 1.73m to 1.38m on X-axis and 1.21m to 1.11m on Y-axis. Interestingly, the error on the X-axis is greater than on the Y-axis (1.38m and 1.11m respectively). In the testing environment, the width of the floor (X-axis) is longer than the length of the floor (Y-axis). Because of a longer distance on X-axis, the probability of absorbing the energy of signal by air and furniture is higher. These results confirmed that the TR method reduces the error distance in position estimation significantly.

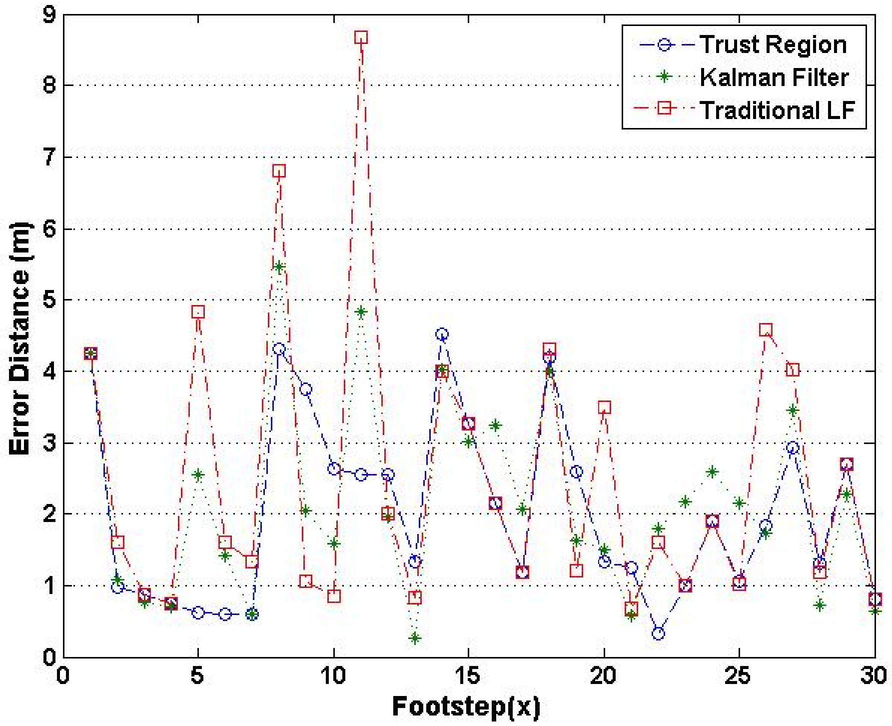


Figure 33 Error distance of position estimation under Trust-Region, Kalman Filter and Traditional Location Fingerprinting approach

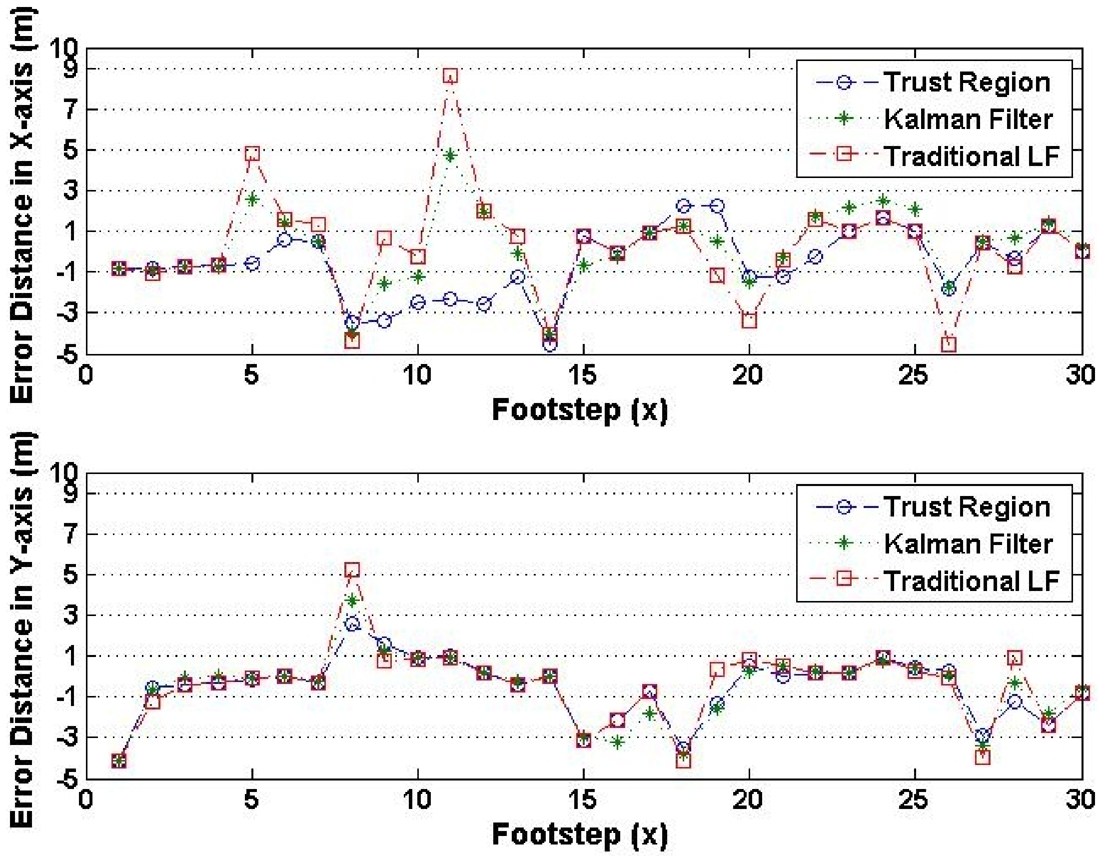


Figure 34 Error distance of position estimation on the X and Y coordinates under Trust-Region, Kalman Filter and Traditional Location Fingerprinting approach

5.4.5. Effect of TR radius on the localization accuracy

This subsection further investigates how the TR radius and the TR boundary contribute to achieving optimal precision. These factors are elaborated from (22). Figure 35 shows the relationship of the radius of the trust region to accuracy. When the radius increases from 4 to 5.5 meters, TR achieves from 82% to 93% accuracy. TR achieves the optimal point of 93% accuracy at around 6 meters but a radius higher than 6.5 meter degrades the precision dramatically. Interestingly, the results indicate that TR radius depends on the walking speed. In other words, TR achieves optimal precision when the TR radius is almost double the length of each stride.

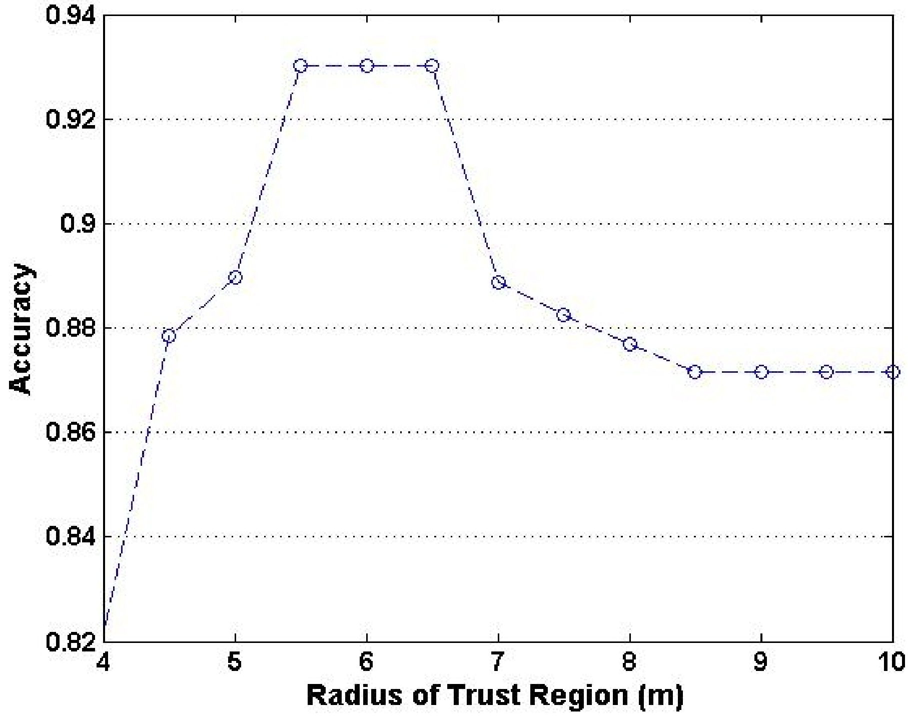


Figure 35 Relationship of TR radius to accuarcy

Figure 36 shows the relationship of the accuracy to upper bound, γ_1 and lower bound, γ_2. The optimum of upper bound is 1.1 and the lower boundary is 0.5 with 94% accuracy. The accuracy decreases rapidly when the lower bound is over 1.

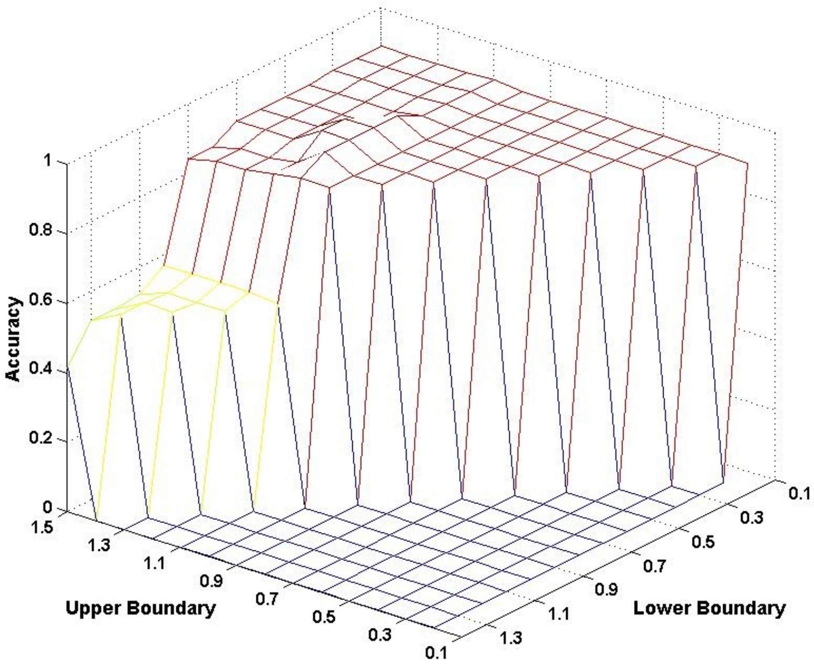


Figure 36 Relationship of TR boundary to accuracy

Chapter 6: Location Fingerprinting Based on Fourier Descriptors

The Location Fingerprinting (LF) method locates a device by accessing a pre-recorded database containing location fingerprints (i.e., the received signal strengths and coordinates). Most LF methods use the absolute received signal strength (RSS) to estimate the location. There are two drawbacks to this. First, the absolute RSS in a time interval may not represent the IEEE 802.11 signal, as the signal may fluctuate. Second, a manual error-pone calibration is needed across different mobile platforms. This chapter proposes using Fourier descriptors in LF. We first transform the IEEE 802.11b Wi-Fi signal into a Fourier domain. Then, the Fourier descriptors are used to estimate the location by applying the K-Nearest Neighbor algorithm. The experimental results show that the effectiveness of LF methods based on Fourier descriptors lead to substantially more accurate and robust localization.

The rest of this chapter is organized as follows: Section 6.1 introduces the background of Fourier descriptors. Section 6.2 presents the LF methodology based on Fourier descriptors. Section 6.3 presents the experimental setup in a corridor environment. Section 6.4 discusses the performance evaluation of location estimation accuracy.

6.1. Introduction

The Fast Fourier transform (FFT) is an algorithm that is used to determine the power versus frequency graph for a signal. For example, the FFT algorithm shows the different amplitudes and waveforms that are combined to generate, for example, a complex sound, like the human voice. The Fourier curve is found by multiplying the periodic waveforms by the sum of the harmonically related sine waves. The Fourier transform aims to decompose a cycle of an arbitrary waveform into its fundamental periodic components. The Inverse Fourier transform converts a series of sine components into the resulting waveform. This chapter uses the FFT algorithm to convert the vector of Wi-Fi signals' amplitude of the access points and use the Fourier descriptor to recognize the pattern.

The Fourier descriptor is a method for measuring the shape of an object. It refers to the utilization of Fourier analysis, primarily the Fourier series as a curve fitting technique that can numerically describe the shape of irregular structures such as commonly found in living organisms. The Fourier descriptor is also invariant to rotation, expansion, contraction, and translation. If we keep only a low frequencies subset of descriptors, we get a curve that just approximates the outline of a shape. By increasing the number of components in the description, high frequencies are also rendered, and sharp curves or details can be generated. It is a well-studied algorithm for describing different wave signals and is applied in various fields (e.g., pattern recognition [Chen and Bui 1999], gait recognition [Mowbray and Nixon 2003] and image retrieval [Zhang and Lu 2002]) but could well be applied to solve previous addressed problems in LF methods.

In this chapter, the proposed Fourier descriptors approach uses LF-based techniques and sensors in three phases. The first phase detects the IEEE 802.11b Wi-Fi signal strength and uses a set of static LF sensors to collect the LFs into a training database in a time interval. In the second phase (location phase), the LFs (location fingerprints) will be transformed to Fourier domain and the Fourier descriptor is used to estimate the location by applying the K-Nearest Neighbor algorithm. Finally, a Kalman filter is used to track multiple points as a trajectory.

The proposed method offers a number of benefits. First, because it transforms the signal into a Fourier domain, the distortion of signal will be reduced. Second, it obviates the need for extensive manual calibration. Third, the series of Fourier descriptors of a signal at each access point (AP) is the same at different locations and it can be re-used. Finally, it is substantially more accurate.

6.2. Fourier Descriptors

In this paper, Fourier descriptors are used to estimate the location. The Fourier descriptors describe the Wi-Fi signal. The sample is taken and transformed into a frequency domain curve. The curve will be used in estimating a K-Nearest Neighbor location.

6.2.1. Definition and Properties

The modeling of Wi-Fi signals by the Fourier descriptors uses a pre-recorded location fingerprint database. The Wi-Fi signal in a location can be represented by a complex function of *v* defined by:

$$v(k) = x(k) + jy(k) \tag{26}$$

where k = 0, 1...,$N - 1$, N is the total number of period, x and y are the in-phase and quadratic components of the measurement Wi-Fi signal.

v can be transformed into the frequency domain by the Discrete Fourier Transformation (DFT) of $c(n)$.

$$c(n) = \sum_{k=0}^{N-1} u(k) e^{j2\pi nk/2}, n = 0,1,..., N-1 \tag{27}$$

By Euler's identity,

$$c(n) = \sum_{k=0}^{N-1} u(k)(\cos(2\pi nk / N) - j\sin(2\pi nk / N)) \tag{28}$$

The result can be transformed back into the spatial domain via the Inverse Discrete Fourier Transformation (IDFT) of $u(n)$.

$$u(n) = \frac{1}{N} \sum_{k=0}^{N-1} c(k) e^{-j2\pi nk/N}, n = 0,1,..., N-1 \tag{29}$$

6.2.2. Normalization

The complex coefficients $c(n)$ are called the Fourier descriptors (FDs) of the corresponding signal. The coefficients with low index contain information on the general form of the signal and the ones with high index contain information on the finer details of the signal. However, the first coefficient $c(0)$ depends only on distance from the AP of a Wi-Fi signal. Rotation invariance is obtained by ignoring the phase information. Scale invariance is obtained by dividing the magnitude values of all coefficients by the magnitude of the second coefficient $c(1)$. The new set of coefficients, f is given by the following equation:

$$f_k = \frac{|c(n)|}{|c(1)|}, n = 2,3,\ldots, N-1, k = 0,1,\ldots, N-3 \qquad (30)$$

These are sensitive to transformations of the signal such as translation, homothetic transformation and reverse description. Table 4 summaries the influence of these operations.

	Signal *u(n)*	Fourier Descriptors *c(n)*
Translation	$u(n)+u(0)$	$c(0)+u(0)c(n)$
Homothetic transformation	$\lambda u(n)$	$\lambda c(n)$
Reverse description	$u(N-n)$	$e^{-j2\pi n}c(-n)$
Rotation	$e^{j\varphi}u(n)$	$e^{j\varphi}c(n)$

Table 4 THE INFLUENCE OF FOURIER OPERATIONS

6.2.3. Key-frame Extraction

It is important to have enough samples and to distill key frames of the Wi-Fi signal. After samples are collected in a location, the difference between the samples and the mean is obtained and then the difference is transformed into the frequency domain. Using (2), the value is expressed in term of complex number. The DFT points are then used to generate Fourier descriptors using (27).

As an example, 20×1-point samples matrix *C* is [-56, -56, -56, -56, -57, -54, -55, -54, 55, -53, -52, -58, -52, -51, -52, -50, -51, -51, -54, -52]. *C* subtracts with the mean of sample. Figure 37 shows signal samples with 20 samples in a location. The new matrix *C* will be [-2.25, -2.25, -2.25, -2.25, -3.25, -0.25, -1.25, -0.25, -1.25, -0.75, 1.75, -4.25, 1.75, 2.75, 1.75, 3.75, 2.75, 2.75, -0.25, 1.75]. *C* will be transformed into *c(n)* by (2). Figure 38 shows the absolute value of each DFT point matches the original spectrum at the same location. It is the DFT magnitude spectrum of the signal. After the Fourier transform of the signal is obtained, it is converted into Fourier descriptors using (27). The normalized Fourier descriptor describes the signal in Figure 39.

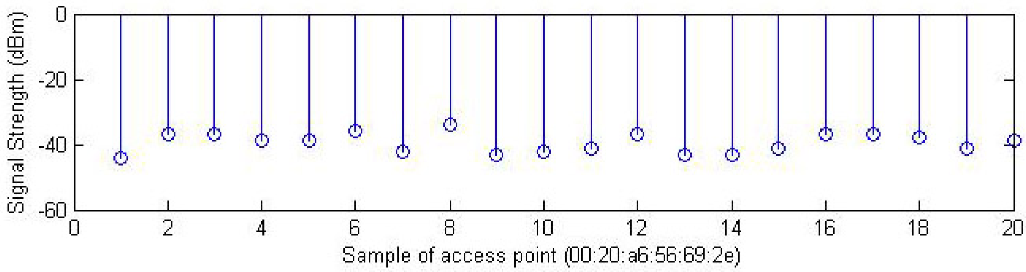


Figure 37 Signal Samples of Wi-Fi RSS

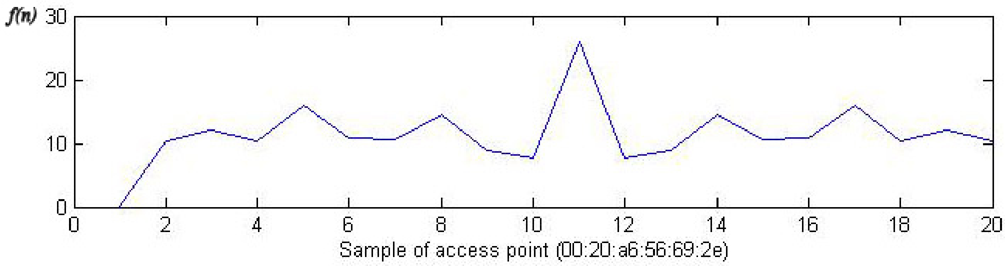


Figure 38 Discrete Fourier Transform (DFT) of Wi-Fi RSS

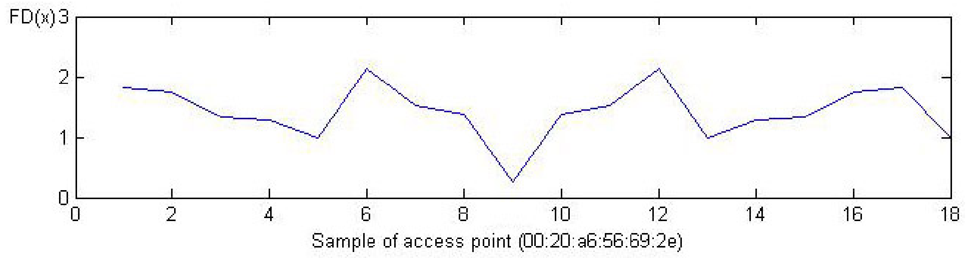


Figure 39 Fourier Descriptors (FD) of Wi-Fi RSS

6.3. Experiment Setup

This section describes the experiment setup in a corridor on the 8th floor of the PQ building, in Department of Computing, at The Hong Kong Polytechnic University. Actually, the testing environment on the 8th floor of the PQ building is quite similar to the 7th floor. For completeness, the system is implemented on both floors. In future work, these two results will be used to implement a 3D localization across different floors. This chapter discusses only the experimental settings and results on the 8th floor. To measure RSS, the same settings are used as in previous chapters. The layout of the laboratory is shown in Figure 40. The area of the floor is approximately 40m by 12m.

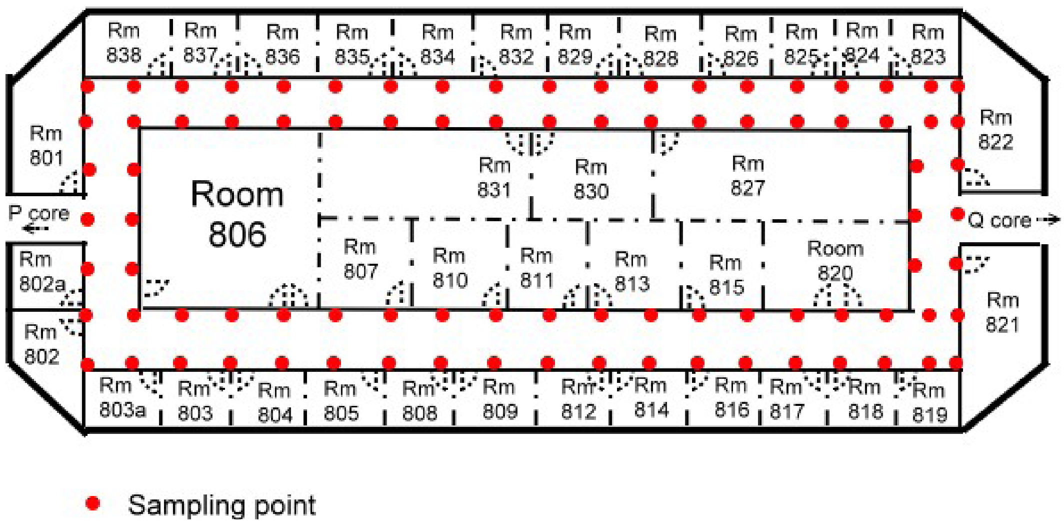


Figure 40 Laboratory experiment environment with 88 sampling locations

The surveying tool is a Wi-Fi enabled device for detecting RSS readings. A Place Lab is a piece of software produced by Intel research group at placelab.org [lab 2002]. Place Lab is a JAVA based radio beacons auditing tools, such as 802.11 access points, GSM cell phone towers, and fixed Bluetooth devices that already exist in large numbers around us in the environment. Its role is to detect wireless access points on declared and undeclared and provide information about these wireless networks. The software performs a periodic scan of local area detecting any wireless signals transmitted on the 802.11b standard. The user can manually set the time interval of these scans.

The received signal sensitivity also limits the range of the RSS to between -90 dBm and -30 dBm. Nevertheless, the highest typical value of the RSS is approximately -40 dBm at one meter from any AP. We collected samples at 88 locations as shown in Figure 5. Each data set of samplings consists of 20 samplings collected at approximately 0.25 second sampling intervals. We take the samplings every 2 meters. Figure 41 shows the frequency distributions for the P and Q core buildings.

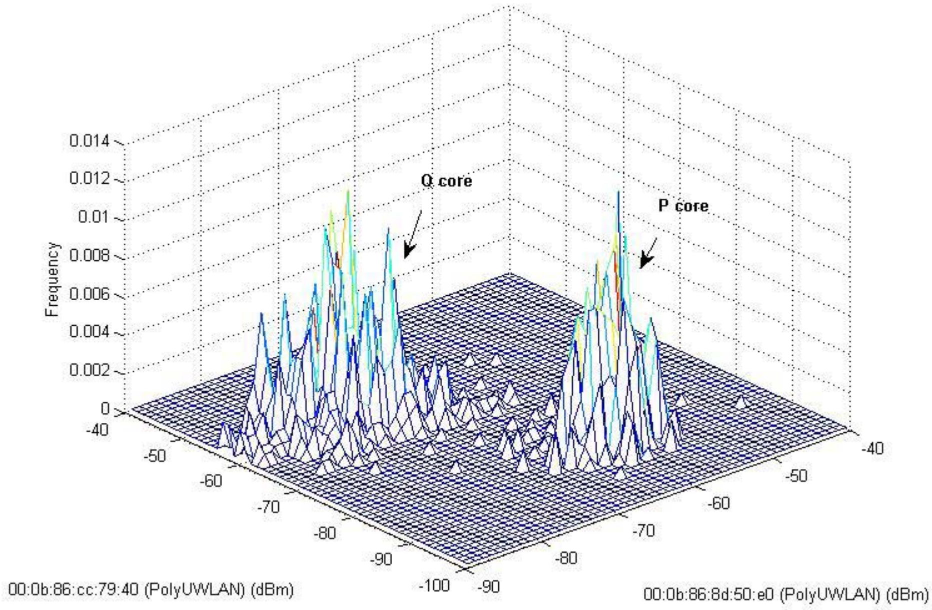


Figure 41 P and Q core buildings frequency distribution

6.4. Performance Evaluation

This section provides the experiment results for the performance of positioning accuracy evaluation.

6.4.1. Result for Varying the Number of Fourier Coefficients

Figure 42 shows the relationship between accuracy and the number of Fourier coefficients (descriptors). It increases from 87% in one Fourier coefficients to 95% in 18 Fourier coefficients. The accuracy is over 90% when there are more than three Fourier coefficients. The results show that if the number of Fourier coefficients increases; the positioning accuracy will increase significantly.

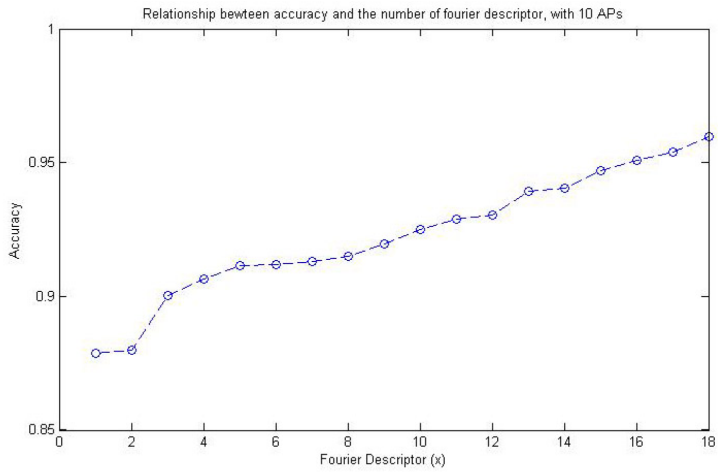


Figure 42 Influence of varying the number of Fourier coefficient on the positioning accuracy

6.4.2. Result for Varying the Number of Access Points

Figure 43 shows the relationship between accuracy and the number of access points (APs) when using the FFT method and the absolute value method. After using the FFT method, the accuracy increases from 78% to 90% when there is one access point. The highest accuracy rate occurs when ten access points are used to estimate location. The accuracy is 89% when using the absolute value and 96% when using the FFT method. The optimal number of APs should be around ten. The results show that the proposed FFT method is more accurate than the absolute value method.

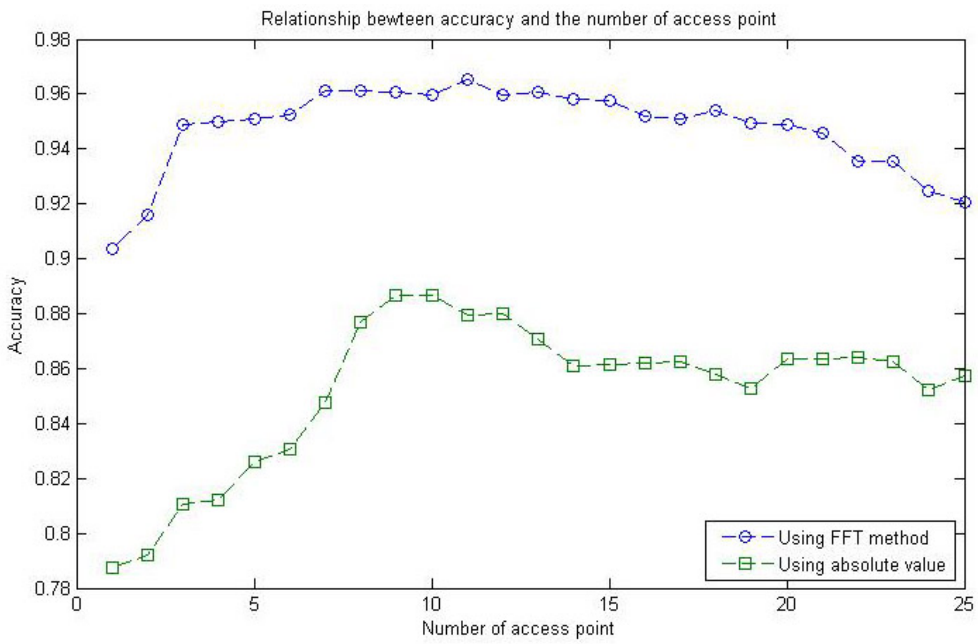


Figure 43 Influence of varying the number of access points on the positioning accuracy

6.4.3. Result for Adding the Location Tracking Filter

After the location is calculated using the K-NN method, a Kalman filter is used to track multiple points to characterize a trajectory. The following result observes the trajectory accuracy when a user with mobile sensor walks around the corridor. It is assumed that the Gaussian RSS noise measurement is 4. Figure 44 shows the original and estimated (both with absolute value and FFT) paths of the user.

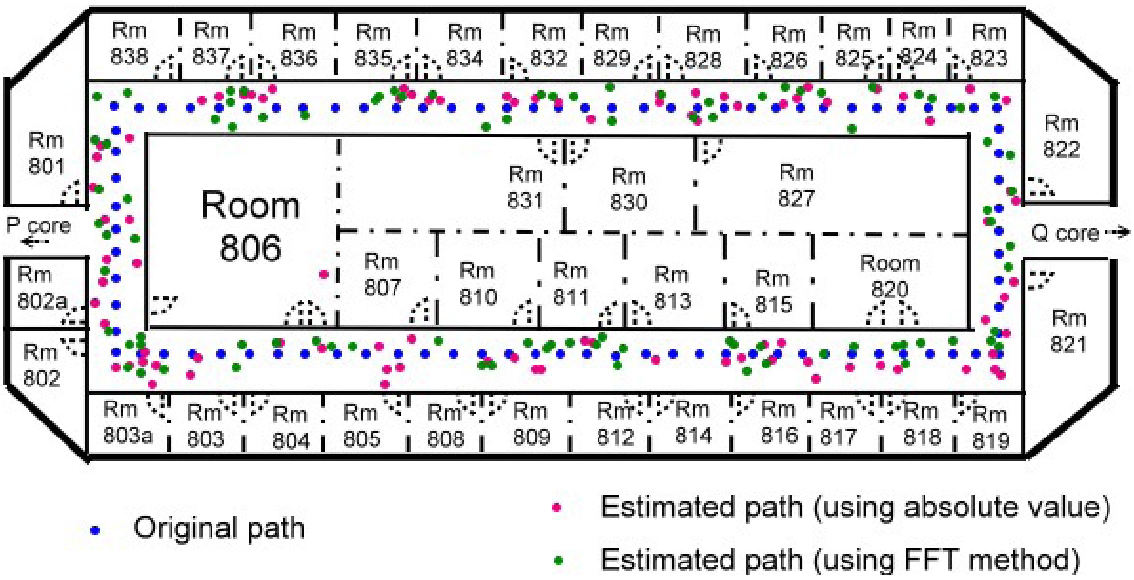


Figure 44 The comparison of both absolute value and FFT method to paths of the user

Figure 45 shows the reduction of error with absolute value and FFT method. It reduces the average error distance from 1.64m by using absolute value to 1.34m by using FFT method. FFT LF method reduces the 18% error in a general case.

Figure 46 shows error on the X and Y coordinates. The error on the X-axis and Y-axis improves after using FFT method, decreasing from 1.3m to 1.12m on the X-axis and 0.67m to 0.53m on the Y-axis. The error on the X-axis is greater than on the Y-axis (0.12m and 0.53m respectively). In the testing environment, the width of the corridor (X-axis) is longer than the length of the corridor (Y-axis). Therefore, the distance between sampling points and the access points on the X-axis is longer than on the Y-axis. Because of the longer distance, the probability of a signal being blocked by a wall is higher.

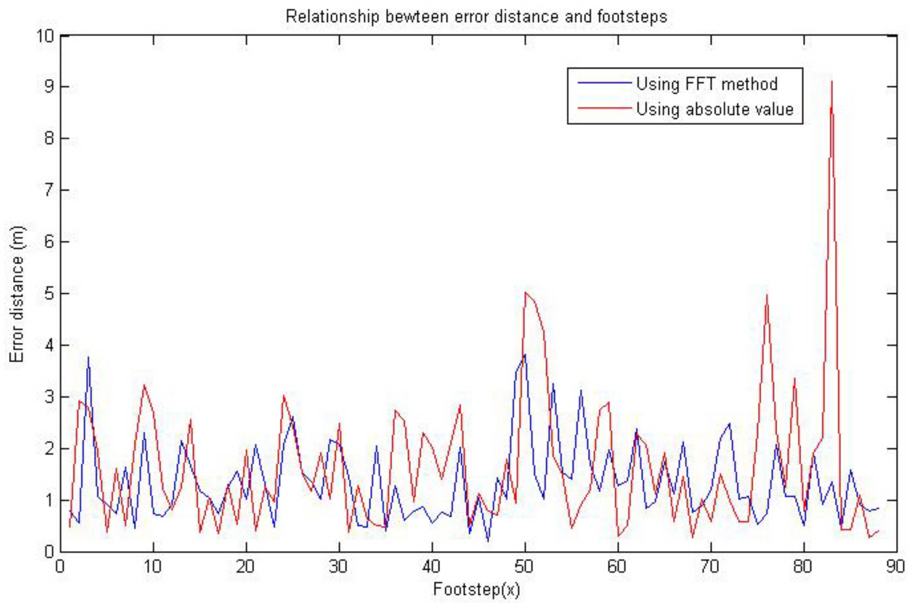


Figure 45 Reduction of error with absolute value and FFT method

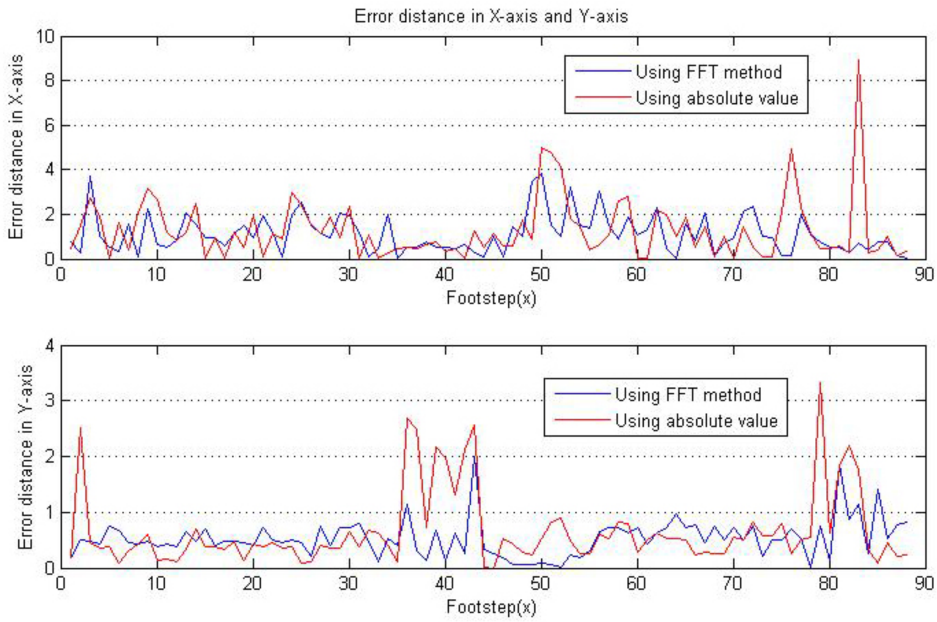


Figure 46 Error distance on the X and Y coordinates with absolute value and FFT method

6.4.4. Comparison of our approach against other 6 positioning approaches

We compare our approach against other 6 positioning approaches under 5 different criteria (accuracy, precision, complexity, scalability and cost). We briefly explain the scope of 5 different criteria. Accuracy considers the value of mean distance errors. Precision considers how consistently the system works in its performance over many trials. Complexity of a positioning system can be attributed to hardware, software, and operation factors. Measure of scalability is the dimensional space of the system.

Positioning Algorithm	Accuracy	Precision	Complexity	Scalability	Cost
AOA [Zhang and Wong, 2009]	5 m	90% within 4.2 m	Medium to High	2D	Medium
TOA [Mogi and Ohtsuki, 2008]	4 m	90% within 4.5 m	Medium to High	2D	Medium
TDOA [Yamasaki et al., 2005]	3 m	67% within 2.4m	Medium to High	2D	Medium
SMP [Gwon and Jain, 2004]	3 m	50% within 2.7 m	Medium	2D & 3D	Medium
NN [Fang and Lin, 2008]	3 m	63% within 2 m	Low	2D & 3D	Low
SVM [Qiu and Kennedy, 2007]	3 m	90% within 5.12 m	Low to Medium	2D	Low
Our approach [Chan, Baciu et al. 2010]	1.82 m	90% within 2.45 m	Low to Medium	2D	Low

Table 5 COMPARISON OF INDOOR POSITIONING TECHNIQUES

Table 5 summarizes the result of our recent publication in [Chan, Baciu et al. 2010] and other 6 positioning approaches. Our approach in accuracy and precision fields outperform other 6 approaches. Our approach is relatively easy to be implemented and the cost of implementation is low.

Chapter 7: Resource-effective and Accurate WLAN Infrastructure Design and Localization Using a Cell-structure Framework

A large-scale WLAN infrastructure requires the placement of many thousands of access points (APs). The current approach is to deploy these in an empirical and ad-hoc manner but these results in poor resource utilization and poor positioning due to signal overlaps and black spots. This chapter proposes and simulates three structured approaches to WLAN infrastructure deployment that would allow better positioning accuracy and optimal coverage. These three approaches make use of triangular, square, and hexagonal configurations which the later results show are all very much more effective in both 2-D and 3-D space than any of the current ad-hoc or empirical approaches to AP deployment. Overall, the hexagonal approach is the most cost effective and accurate and in fact allows perfect positioning with just half the number of APs normally used. As a further contribution, 3-D rendering of buildings and wireless signal coverage provide engineers with a very concrete visualization that helps them to see in advance where blind spots might occur how the signal will vary across multi-story buildings so that engineers can estimate the optimal number of APs and where they should be placed.

The rest of this chapter is organized as follows: Section 7.1 describes problems with existing WLAN infrastructure. Section 7.2 presents the design of the proposed cell-based

WLAN infrastructure. Section 7.3 presents the simulation setup. Finally, Section 7.4 offers results and discussion.

7.1. Introduction

Wireless Local Area Networks (WLANs) are deployed in a wide range of urban environments. Often these WLANS are made up of many access points (APs) or nodes deployed across extensive, topographically varied, heavily built up, and constantly changing environments that may carry heavy traffic. [Huang, Wang et al. 2005]. The basic requirement of an effective WLAN is that it should provide adequate coverage so that when users wish to access location-aware (e.g., pervasive computing-enabled) applications and services, the WLAN will permit user mobile devices to be accurately positioned. The network should also be deployed in a cost-effective and resource-efficient way, including in indoor environments. Current approaches to infrastructure design and deployment, however, apply an unstructured approach to WLAN infrastructure design that implies poor resource utilization, placing and positioning APs manually on the basis of empirical measurements of RSS (received signal strength) taken by engineers.[Budianu, Ben-David et al. 2006] Such an approach may use many APs but still leave blind spots or places where there are too many access points packed closely together, producing wasteful signal overlap and interference and reducing positioning accuracy.

Current research on the visual representation of WLAN signals [Kaemarungsi and Krishnamurthy 2004a; Swangmuang and Krishnamurthy 2008] is based on the accuracy of positioning systems and proximity graphs, such as the Voronoi diagram and clustering graphs. There are two drawbacks to these representations. First, none of them provide any way to visualize location uncertainty, for example, by showing on a 2-D or 3-D display where there might be blind spots. Second, they do not provide any way to estimate how many APs might be required for optimal coverage nor do they provide any guidance as to where they should be placed.

Location or positioning uncertainty occurs in areas of signal overlap. The simulations described later use the location fingerprinting (LF) approach [Kaemarungsi and Krishnamurthy 2004a; Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Li, Wang et al. 2005; Chan, Baciu et al. 2008]. In that case, positioning uncertainty occurs within the locus, a line (a series of points) having the same RSS measured within the overlap of two or more AP signals. This identity of RSSs makes it impossible to accurately localize points of uncertainty. Increasing the number of APs in an area reduces the zone of most common overlap and thereby shortens the locus (seen as a line). Given enough APs, positioning uncertainty can be eliminated and it becomes possible (in terms of a particular resolution) to achieve perfect positioning. But as long as an ad-hoc approach is used, it is not possible to estimate in advance the optimal number of APs. A more structured approach should produce economies and efficiencies as well as improved capabilities. In previous work [Chan, Baciu et al. 2009f; Chan, Baciu et al. 2009e] this author focused on the wireless tracking problem from the view of end-users. This work improves on the traditional positioning approach in that it uses 15% fewer access points and achieves a 10% more accurate positioning.

This chapter focuses on how to achieve accurate positioning from the point of view of wireless infrastructure deployment. It proposes a cell-based WLAN infrastructure deployment design that visualizes blind spots, predicts how many APs will be optimally required for positioning, and suggests where they should be placed. The proposed approach extends the previous work of this author and colleagues [Chan, Baciu et al. 2009d] from a two-dimensional to a three-dimensional approach which implements a three structured geometric configurations for use in WLAN infrastructure deployment: (1) triangular, (2) square, and (3) hexagonal.

The proposed approach both improves the speed and efficiency of large-scale WLAN infrastructure deployment and allows WLAN engineers and designers to choose an optimal deployment solution from a selection of different AP distributions. The simulations show that all of the proposed structured approaches are much more cost effective than the current unstructured AP deployment approach, all producing a regular and predictable disposition. The hexagonal approach is overall the most cost-effective way to achieve perfect accuracy, requiring only two-thirds as of the APs required by the unstructured approach. A further contribution offers a virtual, 3-D rendering of wireless signals in both small-scale laboratory and large-scale campus environment. It would be useful in any kind of built-up area but in particular where there are many multi-story buildings.

7.2. The Design of the Cell-based WLAN Infrastructure

A maximally accurate and resource-efficient cellular WLAN infrastructure will cover the biggest area with the minimum number of APs without allowing empty spaces and with the minimum of interference between cells. A tessellation is thus required and this can be achieved using regular polygons of three kinds: quadrangles, triangles, and hexagons. Each AP is placed at the center of the cell. A hexagonal plane tessellation will produce a honeycomb mesh. [Carle, Myoupo et al. 2001] Figure 47, 32 and 33 show a 2-D distribution of APs in triangular, square, and hexagonal tessellations. Figure 48, 33 and 35 show triangular, square, and hexagonal distributions of APs in 3-D space. The red points indicate the AP locations, the green regions are the cell area, and the yellow spheres represent the wireless coverage of the APs.

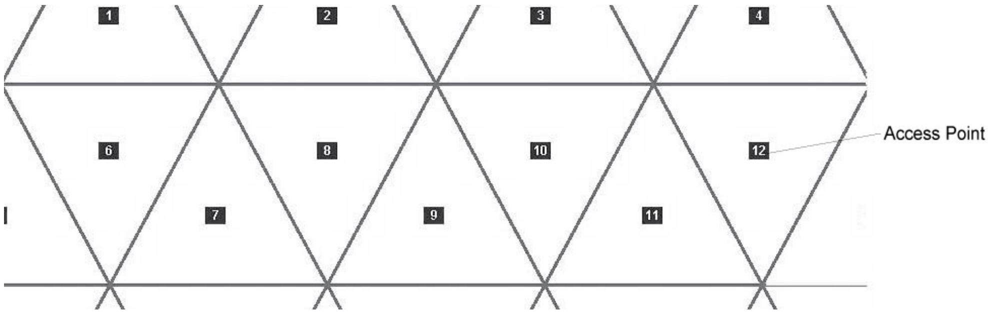


Figure 47 Triangular Tessellation

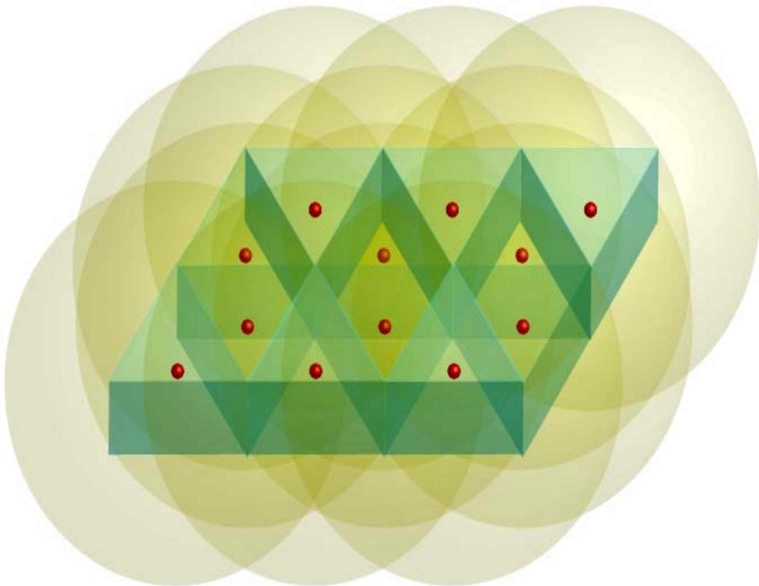


Figure 48 Triangular distribution

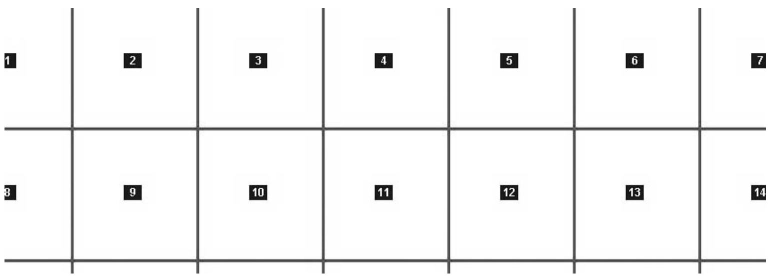


Figure 49 Square Tessellation

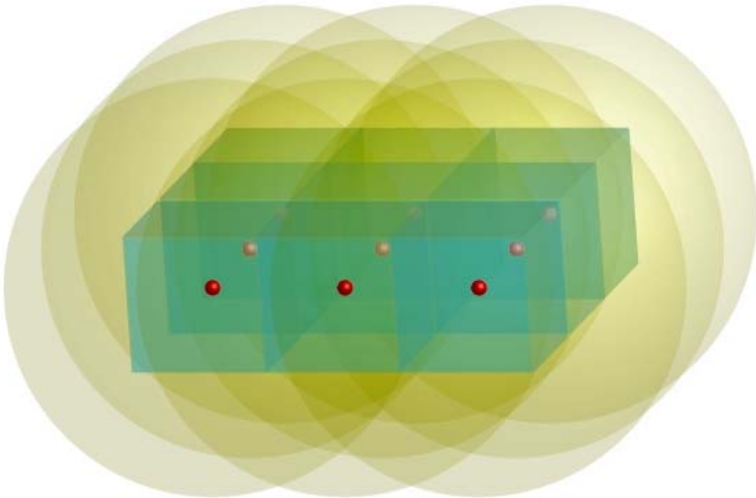


Figure 50 Square distribution

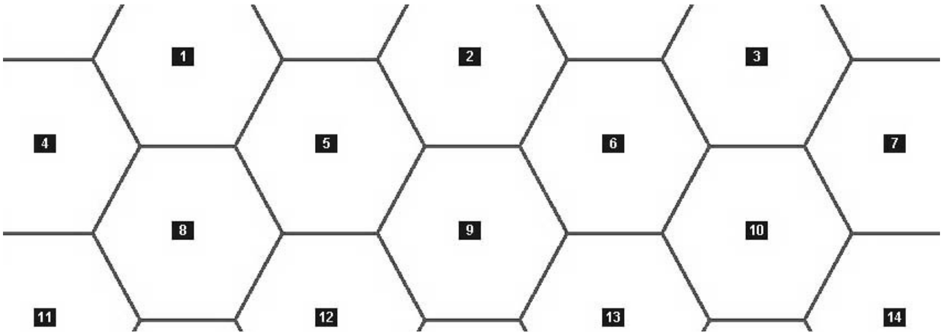


Figure 51 Hexagonal (or honeycomb mesh) Tessellation

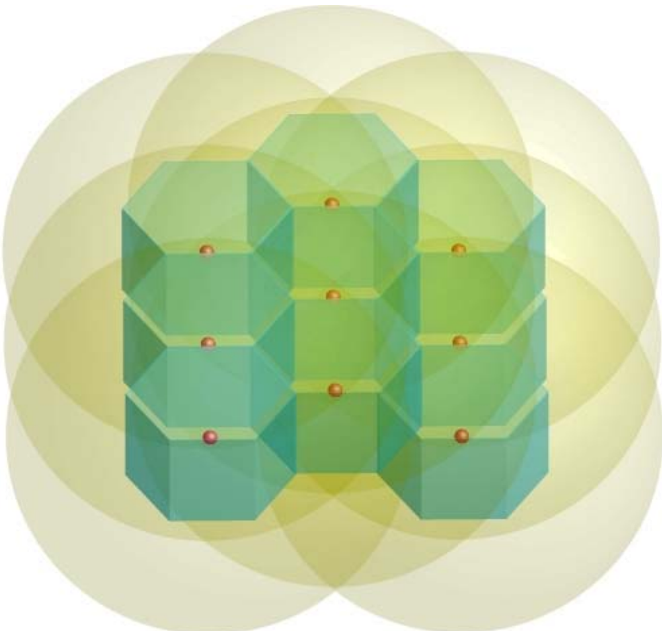


Figure 52 Hexagonal distribution

Consider a physical layout of the wireless network as shown in Figure 7. The hexagonal area is a cell. Each cell will contain one AP and, for simplicity, considers a circular Wi-Fi coverage. Thus there are areas of signal overlap. It may not be possible to localize a device in such an area due to the locus problem as described in the Introduction. In an LF-based approach, positioning uncertainty is dealt with by estimating the location using a pre-recorded LF database for each location. The positioning accuracy can be defined as the number of distinguishable location points. On this basis, then, APs would be evenly distributed, placed at the centre of hexagons as in Figure 53. In this configuration, each AP provides total coverage of seven cells, i.e., its own and six contiguous cells, and a user receives at least seven AP signals.

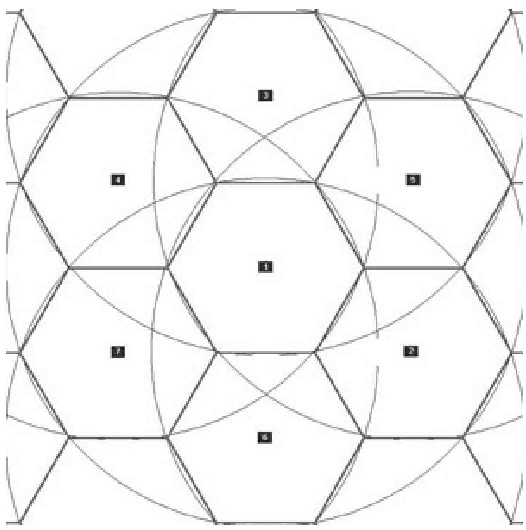


Figure 53 Hexagonal Cells satisfy with 1.5 cell radius coverage

7.3. Simulation Setup

The following sections describe three sets of experimental simulations. The first set of simulations (Set A) is designed to determine the positioning accuracy of my four basic distributions, triangular, square, hexagonal and ad-hoc where the number of APs is varied from the minimum required for positioning, two APs, until perfect positioning is achieved in all of the non-unstructured approaches. The second set of simulations (Set B) is designed to show how points of positioning uncertainty are distributed in 2-D space. The third set of simulations (Set C) is designed to show how points of location uncertainty are distributed in 3-D space and is made up of two simulations which seek to identify any variations, given equal cost (number of APs), in the distribution of positioning accuracy, on the assumption that this information may be of use in a structured WLAN planning process, particularly in real-world scenarios.

In all three sets of simulations, for reasons that have been mentioned, accuracy is measured using only the LF approach Accuracy is defined in a range from zero to one, with zero meaning not detected and one meaning perfectly accurate. The radius of coverage of each AP is 80m. The signal strength ranges between -85dbm and -30dbm in a testing area of 150m X 100m. The target in Set A is to achieve a positioning resolution of two meters. In Sets B and C, the target resolution is three meters. The accuracy of Sets A and B are considered only in 2-D space and Set C in 3-D space (150m x 100m x 10m). Table 6 summarizes the setup for Simulation A, Table 6 for Set B and Table 7 for Set C.

Item	**Description**
Total area	150m x 100m
Positioning resolution	2 meters
Wi-Fi coverage from each APs	80 meters
Range of signal strength	-85dbm to -30dbm

Table 6 SIMULATION SETTINGS (SET A)

Item	**Description**
Total area	150m x 100m
Number of Aps	9 Aps
Positioning resolution	3 meters
Wi-Fi coverage from each APs	80 meters
Range of signal strength	-85dbm to -30dbm

Table 7 SIMULATION SETTINGS (SET B)

Item	**Description**
Total area	150m x 100m

Number of Aps	9 Aps
Height	10 m
Positioning resolution	3 meters
Wi-Fi coverage from each APs	80 meters
Range of signal strength	-85dbm to -30dbm

Table 8 SIMULATION SETTINGS (SET C)

7.4. Results and Discussion

This section describes the results for simulation Sets A, B and C.

7.4.1. Result for Simulation Set A

Figure 54 and Table 9 show the relationship of the number of APs to positioning accuracy under the four different deployment scenarios. As can be seen in Figure 54, the performance of the four approaches is similar when few APs are used yet in every case and at every setting the structured approaches outperform unstructured approaches. The superior performance of the structured approaches becomes progressively stronger as more APs are added and ultimately they achieve perfect accuracy. Except at very low numbers of APs, when the triangular approach is the most accurate, the hexagonal distribution is the most accurate approach and is the first to achieve perfect accuracy, at 18 APs.

Table 9 compares the cost-efficiency of the structured and the ad-hoc or empirical approaches. The triangular approach is clearly the best performer, achieving perfect accuracy with just 18 APs, making it considerably more cost-effective than the next-most effective, square, and very much more effective than triangular, which requires more than 40% more APs to achieve the same accuracy. Perhaps the most important point to take away here is that the unstructured approach never at any point matches the accuracy of the structured approaches and never achieves perfect accuracy. In fact, the best performance of the unstructured approaches, just 96%, was achieved over 100 iterations and still required 23 APs, the number of APs required by the worst of the structured performances.

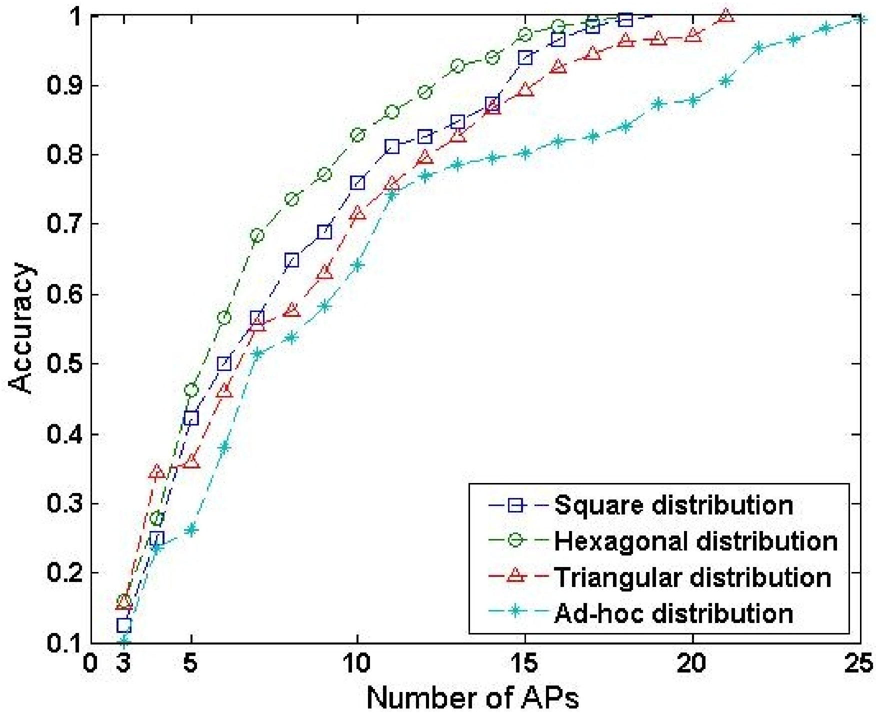


Figure 54 Relationship of number of APs to positioning accuracy under four different deployment scenarios

No. of APs	Accuracy (%)			
	Triangle	Square	Hexagon	Ad-hoc
2	6%	7%	8%	4%
5	36%	42%	46%	26%
10	72%	76%	82%	64%
15	89%	93%	97%	80%
16	92%	96%	98%	82%
18	96%	99%	100%	84%
20	97%	100%	100%	88%
23	100%	100%	100%	96%
>25	100%	100%	100%	99%

Table 9 POSITIONING ACCURACY WHEN VARYING NUMBER OF APS

7.4.2. Result for Simulation Set B

Figure 55, 56, and 57 represent WLAN infrastructure AP deployments under the simulation setup for my three proposed geometric deployment patterns. The testing area is 3x3 m 2-D grid. There are 1,734 grid points. The colored dots represent positioning uncertainty. Each pair of colored dots represent where positioning returned the same RSS reading for two coordinates. In other words, more colored dots means less accurate localization.

Figure 55, 56, and 57 show the distribution of location uncertainty in the three proposed types of WLAN infrastructure in 2D space. The colored dots in corners show places where accurate positioning is not possible. Figure 55 (triangular) shows 83 colored points and the overall positioning accuracy is 95.2%. Figure 56 (square) shows 52 colored points and a positioning accuracy of 97%. Figure 57 (hexagonal) shows 39 colored points and a positioning accuracy of 97.8%. The performances are similar to what was found in SET A. A hexagonal distribution of APs is the most accurate.

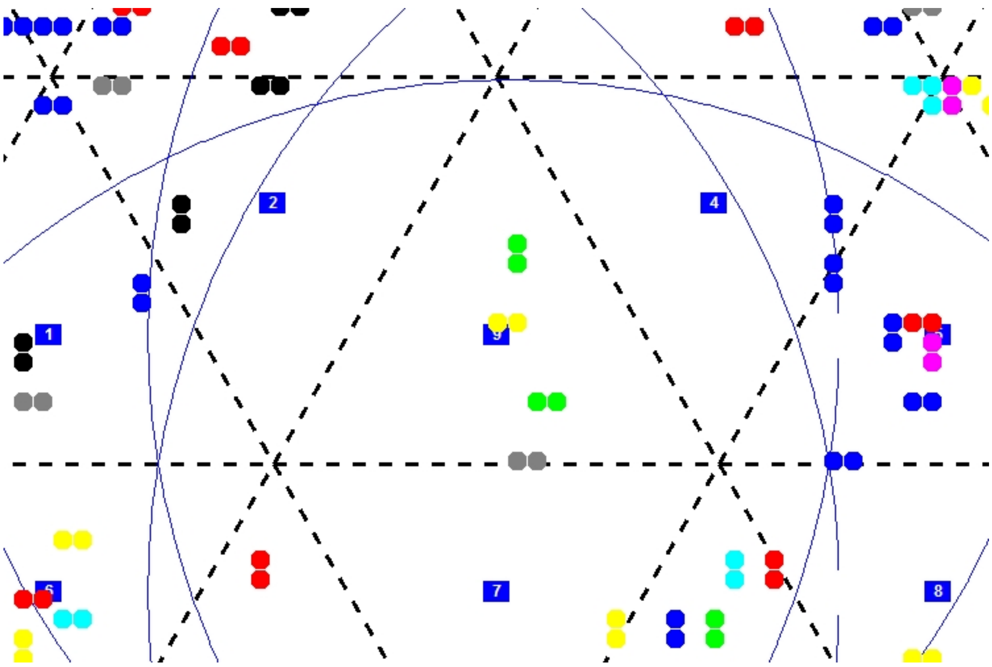


Figure 55 The distribution of location uncertainty in a triangular WLAN Infrastructure

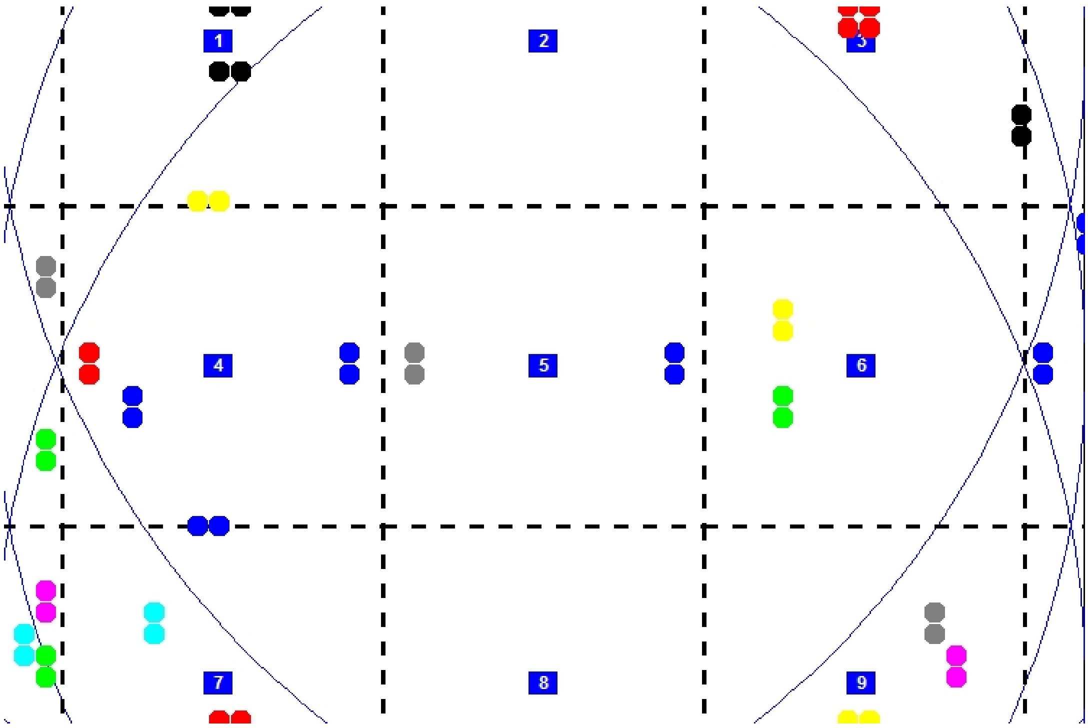


Figure 56 The distribution of location uncertainty in a square WLAN Infrastructure

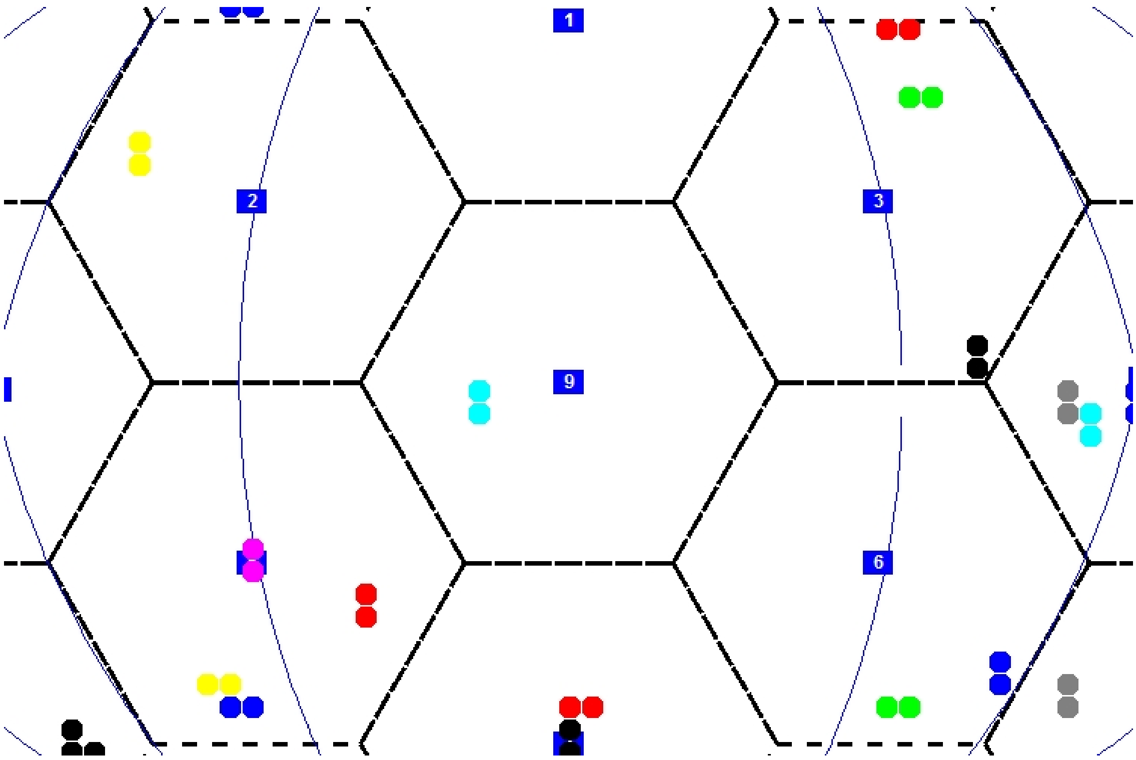


Figure 57 The distribution of location uncertainty in a hexagonal WLAN Infrastructure

7.4.3. Result for Simulation Set C

Set C is once again marked with colored dots but this time uses a 3x3x3 m 3-D grid with a total of 17,340 grid points. Figure 58, 42, and 43 shows the distribution of location uncertainty in the proposed three types of WLAN infrastructure in 3D space. The dots indicating inaccurate positioning in general correspond to corners, edges and walls where signals overlap and propagate weakly. Figure 58 shows 698 colored points in a triangular distribution in a 3-D space (150m x 100m x 10m). The positioning accuracy is approximately 96%. Figure 59 shows 514 colored points in a square distribution in a 3-D space. The positioning is approximately 97% accurate. Figure 60 shows 220 colored points in hexagonal distribution in a 3D space. The positioning is approximately 98.7% accurate.

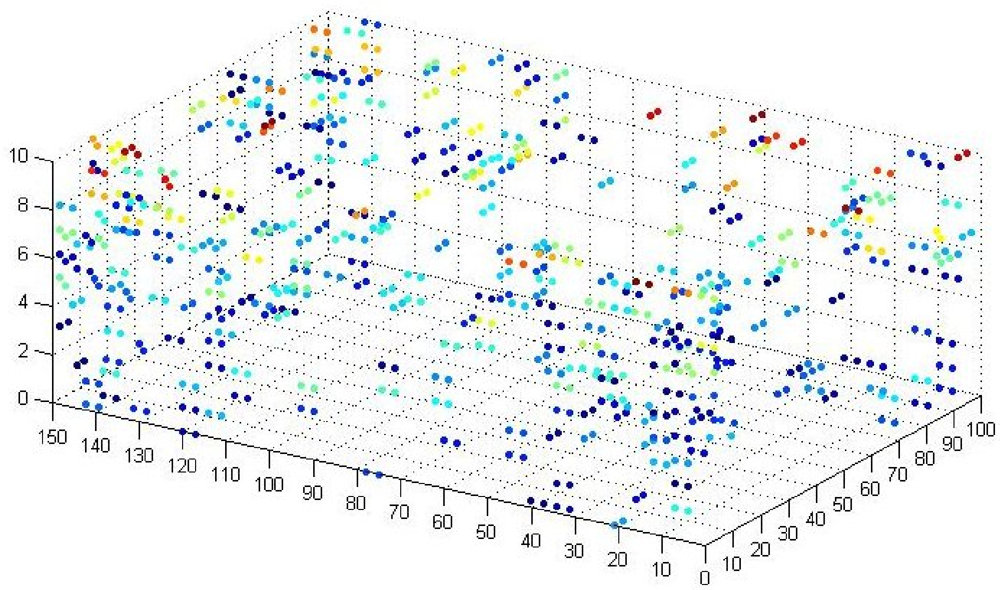


Figure 58 3-D Distribution of Location Uncertainty in triangular APs distribution

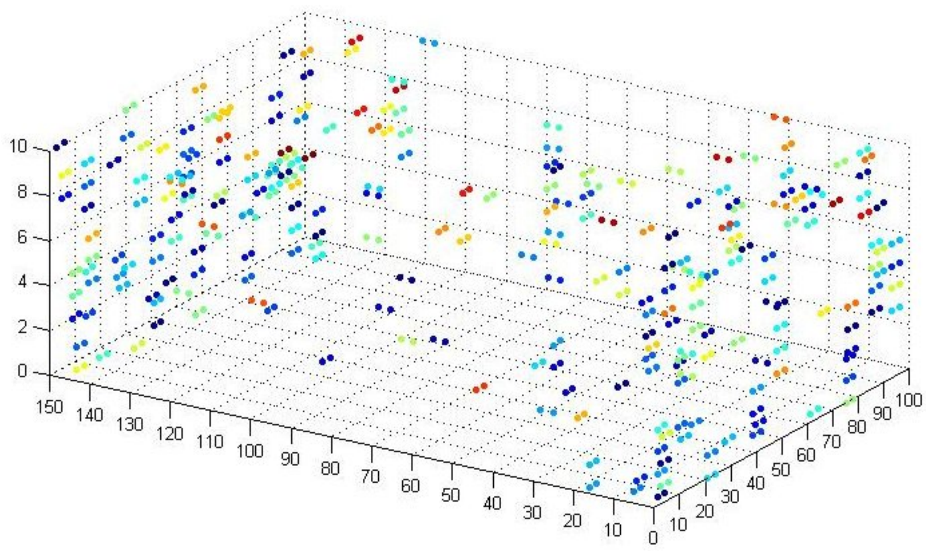


Figure 59 3-D Distribution of Location Uncertainty in square APs distribution

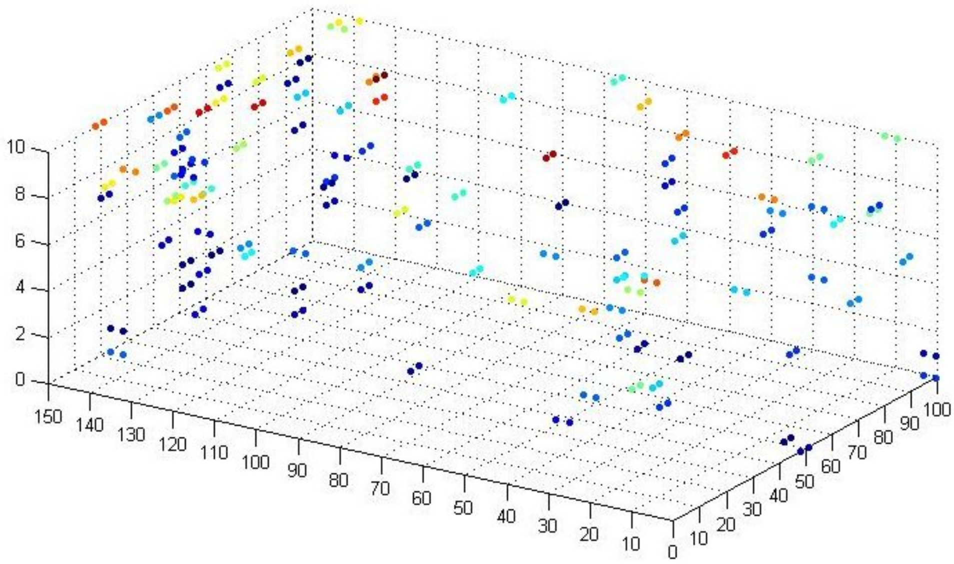


Figure 60 3-D Distribution of Location Uncertainty in hexagonal APs distribution

In both SET A and SET B, it is the hexagonal distribution of APs that is most accurate and efficient. Of particular interest is the fact that the hexagonal distribution outperforms other approaches even more strongly in 3D space even though it uses the same number of APs.

Chapter 8: 3-D Fuzzy Modeling in Wireless Signal Tracking Analysis

Fuzzy logic modeling can be applied to evaluate the behavior of Wireless Local Area Networks (WLAN) received signal strength (RSS), a pivotal part of WLAN tracking analysis. Previous analytical models have not effectively addressed how the WLAN infrastructure affects the accuracy of tracking. This chapter proposes a 3-D dimensional representation of WLAN infrastructure design to visualize the WLAN signal distribution and predict the localization blind spots. It also proposes a novel fuzzy spatio-temporal topographic model. The proposed model is implemented in a large (9.34 hectare), built-up university. More than 2,000 access points were surveyed and their WLAN received signal strength (RSS) collected. The Nelder-Mead (NM) method is applied to simplify this author's previous work on fuzzy color maps into a topographic (line-based) map. The new model can provide a detailed and quantitatively strong representation of WLAN RSS and serve as a quick reference and efficient analytical tool for improving the design of WLAN infrastructure.

The rest of this chapter is organized as follows: Section 8.1 provides some background to wireless signal analysis. Section 8.2 discusses the Fuzzy and Topographic Model Design. Section 8.3 presents the experimental setup in a university campus 9.34 of hectares. Finally, Section 8.4 presents a discussion of the results.

8.1. Introduction

A three-dimensional wireless local area network (WLAN) infrastructure design would greatly assist engineers and users to deploy and locate in an indoor environment. Currently, WLANs are made up of many access points (APs) or nodes. These access points (APs) are manually placed and positioned on the basis of measurements of received signal strength (RSS) taken by engineers empirically. An unstructured approach to WLAN infrastructure design implies poor resource utilization. [Budianu, Ben-David et al. 2006] The basic requirements of an effective WLAN are that it should provide adequate coverage where users wish to access location-aware (e.g., pervasive computing-enabled) applications and services and the WLAN infrastructure should support the accurate localization of mobile devices.

Current research on the visualization of WLAN signals [Kaemarungsi and Krishnamurthy 2004a; Swangmuang and Krishnamurthy 2008] are based on the accuracy of positioning systems and proximity graphs, such as Voronoi diagrams and clustering graphs. There are two drawbacks to these representations. First, none of them focus on the virtualization of location uncertainty. In other words, they do not display the prediction of localization blind spots in either two or three dimensions. Second, they cannot predict the optimal number of APs required for localization nor can they provide guidance as to where they should be placed.

This chapter proposes a three-dimensional representation of WLAN infrastructure deployment design that virtualizes the localization of blind spots and predicts in advance how many APs are optimally required for localization and where they should be placed. It makes use of the Nelder-Mead (NM) method to simplify this author's previous work on the multi-layer fuzzy color model [Chan, Baciu et al. 2009b] into a topographic (line-based) model. This topographic model will serve as an analytical tool for evaluating and visualizing where the RSS is denser and for clustering different RSS at different topographic levels. The proposed model offers two benefits. First, it is a speedy and efficient tool both for reference and analysis. Second, it can provide a detailed and quantitatively strong representation of WLAN RSS.

8.2. Fuzzy and Topographic Model Design

The basic idea of topographic model is to plot a curve connecting minimum points where the function has the same RSS value. The sets of APs are known as topographic line nodes. Topographic line nodes are the APs residing on the topographic lines around contour regions. This section introduces the major operations of the topographic model including the propagation-based algorithm and the fuzzy membership function introduced in this author's previous work [Chan, Baciu et al. 2009b], topographic line node measurement, the Nelder-Mead method, and topographic model generation.

8.2.1. Fuzzy Membership Function

This subsection will make use of the fuzzy membership function as proposed in my earlier work [Chan, Baciu et al. 2009b] which, for completeness, is described in the following.

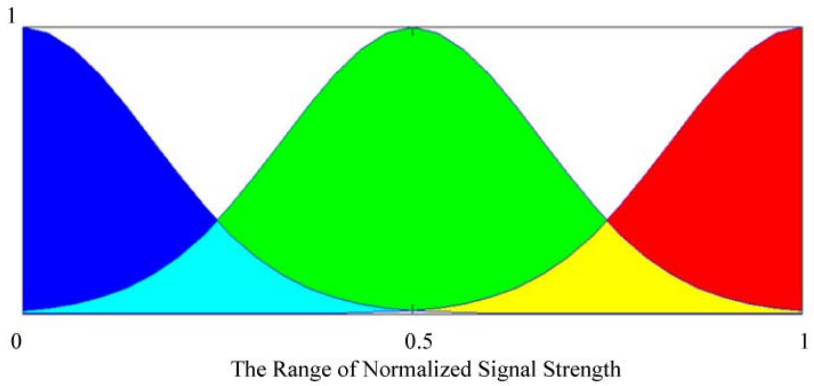


Figure 61 The RSS fuzzy membership graph

Using fuzzy logic, the proposed model offers an enhanced LF hyperbolic solution that maps the RSS from a 0 to 1 fuzzy membership function. This approach does not use a numeric value. Instead, it uses fuzzy logic to broadly categorize the RSS as “strong”, “normal”, or “weak”. The normalization distribution is used to represent the fuzzy membership functions.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(\alpha-\mu)^2}{2\sigma^2}} \tag{31}$$

where $p(x)$ is the probability function, x is the normalized RSS, σ is the standard deviation of normalized signal normalized strength in a region, μ is the mean of signal strength in a region. The WLAN network covers the entire campus.

The membership function of term set, μ(*RSSDensity*) is equal to the set of {*Red, Green,Blue*}. Red means the signal strength density is strong; green means the signal strength is normal; and blue means the signal strength density is weak. The fuzzy set interval of blue is [0, 0.5], [0, 1] is green and [0.5, 1] is red.

For the blue region, σ = 0.5, μ = 0.

$$\mu_{Blue}\left(0 < x < 0.5\right) = \frac{2}{\sqrt{2\pi}} e^{-2x^2} \quad (32)$$

For the green region, I substitute, σ = 0.5, μ = 0.5.

$$\mu_{Green}\left(0 < x < 1\right) = \frac{2}{\sqrt{2\pi}} e^{-2\left(x - \frac{1}{2}\right)^2} \quad (33)$$

For the red region, σ = 0.5, μ = 1.

$$\mu_{\mathrm{Re}d}\left(0.5 < x < 1\right) = \frac{2}{\sqrt{2\pi}} e^{-2\left(x-1\right)^2} \quad (34)$$

Figure 61 shows the fuzzy membership function. X-axis represents the normalized signal strength from 0 to 1 (from -93dBm to -15dBm). The width of membership function depends on the standard deviation of the RSS. The overlap area will be indicated by mixed colors.

We can use different colored regions to represent the WLAN RSS distribution. Conceptually a spatio-temporal region is defined as follows: Assume that B is a finite set of RSS vectors belonging to a particular color region, where $B = \left\{b_1 \ldots b_n \mid b_i \in \Re^n\right\}$, i.e.,

$b_i \in S$, $\forall S \in R$, and $\forall S \in [l,u]$, where l is the lower bound of fuzzy interval and u is a upper bound of fuzzy interval. To analyze the distribution surfaces S, there always exists a spatio-temporal mapping, $q : B \to S$.

$$q(x) = \int_S h(x)b(S)dS \tag{35}$$

$$h(x) = \begin{cases} 1 & x \in S, \\ 0 & x \notin S \end{cases} \tag{36}$$

where $h(x)$ is the characteristic function of S, i.e., $b(S)$ is a weight function that specifies a prior on the distribution of surfaces S. I can explicitly define $b(S)$ by (8). By (28-33), the RSS distribution can be illustrated.

8.2.2. Topographic Node

Each topographic node consists of three components and can be expressed as < *l, d, g* >, in which *l* represents the topographic level, d represents the locations of WLAN received signals, and g represents the direction of the gradient of the RSS distribution. The spatial data value distribution mapped into the (*x, y, l*) space, where the co-ordinate (*x, y*) represents the location and *l = f(x, y)* describes a function mapping from (*x, y*) co-ordinates to level *l*. The gradient vector *g* denotes the direction of the RSS where to degrade in the space. The gradient vector can be calculated as follows:

$$g = -f'(x, y) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y} \right)^T \tag{37}$$

8.2.3. Nelder-Mead Method

The Nelder-Mead (NM) method is a commonly-used nonlinear optimization algorithm for finding a local minimum of a function of several variables. It was devised by Nelder and Mead [Mathews and Fink 1998]. It is a numerical method for minimizing an objective function in a many-dimensional space. Instead of using (1) and (2), this work estimates the location using the NM method.

The first step is to collect the location fingerprint *r* which is in an unknown location (*x,y*), using $f(n)=|n-r|$, where n is any location fingerprint. The next step is to select three location fingerprints (LFs) as three vertices of a triangle, initializing a triangle *BGW* and minimizing the function *f*. The vertices are *B*, *G*, and *W*, where *f(B)* is the smallest value (best vertex), *f(G)* is the medium value (good vertex), and *f(W)* is a largest value (worst vertex). There are four types of changes that may occur when using the NM method, re reflection, expansion, contraction and shrink. The NM method is applied recursively until the point is found which the local minimum is (nearest). That is, *B*, *G*, and *W* have the same value.

The midpoint of the good side is

$$M = \frac{B+G}{2} \tag{38}$$

Reflection Using the Point R

The function decreases as we move along the side of the triangle from *W* to *B*, and it decreases as we move along the side from *W* to *G*. Hence it is feasible that function *f* takes on smaller values at points that lie away farther from W on the opposite side of the line between *B* and *G*. A test point *R* is chosen by “reflecting” the triangle through the side *BG*. To determine *R*, it is necessary to first find the midpoint *M* of the side *BG* then draw the line segment from *W* to *M* and call its length *d*. This last segment is extended by a distance *d* through M to locate the point *R* (See Figure 62). The vector formula for *R* is

$$R = M + (M - W) = 2M - W \tag{39}$$

Expansion Using the Point E

If the function value at *R* is smaller than the function value at *W*, then I have moved in the correct direction toward the minimum. Perhaps the minimum is just a bit farther than the point *R*. In that case the line segment is extended through *M* and *R* to the point *E*. This forms an expanded triangle *BGE*. The point *E* is found by moving an additional distance d along the line joining *M* and *R* (Figure 63). If the function value at *E* is less than the function value at *R*, then a better vertex than *R* has been found. The vector formula for *E* is

$$E = R + (R - M) = 2R - M \tag{40}$$

Contraction Using the Point C

If the function values at R and W are the same, another point must be tested. Perhaps the function is smaller at M, but it is not possible to replace W with M because a triangle is required. Consider the two midpoints C_1 and C_2 of respectively the line segments WM and MR (see Figure 64). The point with the smaller function value is called C, and the new triangle is BGC. Note that while the choice between C_1 and C_2 might seem inappropriate for two dimensions, it is important where considering more dimensions.

$$C_1 = (M - W)/2 \tag{41}$$

$$C_2 = R - (M - W)/2 \tag{42}$$

Shrink toward B

If the function value at C is not less than the value at W, the points G and W must be shrunk toward B (see Figure 65). The point G is replaced with M, and W is replaced with S, which is the midpoint of the line segment joining B with W.

$$S = \quad (B - W)/2 \tag{43}$$

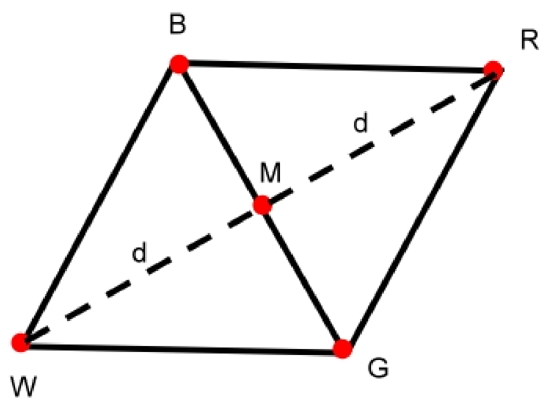


Figure 62 Reflection Using the Point R

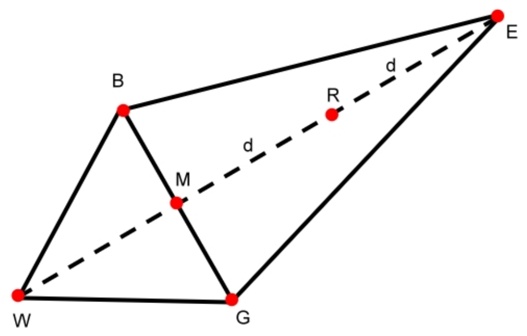


Figure 63 Expansion Using the Point E

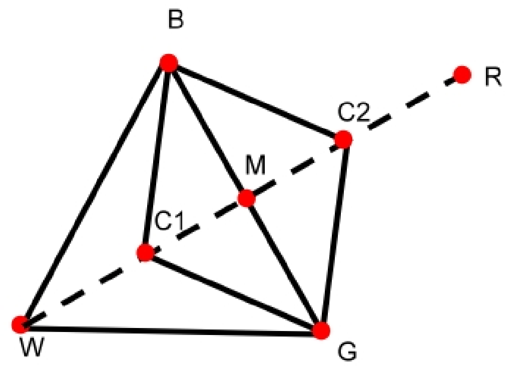


Figure 64 Contraction Using the Point C

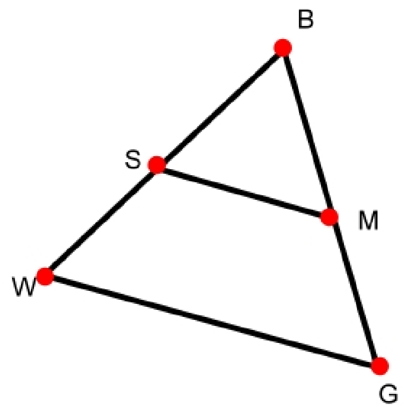


Figure 65 Shrink toward B

8.2.4. Topographic Model Generation

This section describes the generation of a topographic model based on this authors previous work [Chan, Baciu et al. 2009b] and the NM algorithm. Applying the NM method to a many-dimensional RSS distribution space problem allows the fuzzy color map to be simplified to a contour (line-based) map.

To do this, it is necessary to first select three LFs to be three vertices of a triangle: $\vec{B}$, $\vec{G}$, and $\vec{W}$, where $\vec{B}$ is a location with high RSS (best vertex), $\vec{G}$ is a location with medium RSS (good vertex), and $\vec{W}$ is a location with the low RSS (worst vertex). The location vector of RSS at x_k,y_k use in function, *N(x, y)*. In this case (1) is used to define *N(x, y).* There are four types of changes that may occur when using the NM method, reflection, expansion, contraction and shrink. The NM method is applied recursively until the point is found which is the local minimum in $\vec{B}$, $\vec{G}$, and $\vec{W}$ that they are the same value. Table 10 summarizes the procedure.

IF $f(R)<f(G)$, THEN Perform Case (i) {either reflect or extend}

ELSE Perform Case (ii) {either contract or shrink}

BEGIN {Case(i).}

IF $f(B)<f(R)$ THEN

replace *W* with *R*

ELSE

compute *E* and *f(E)*

IF $f(E)<f(B)$ THEN

replace *W* with *E*

ELSE

BEGIN {Case(ii).}

IF $f(R)<f(W)$ THEN

replace *W* with *R*

ENDIF

compute $C = (W + M)/2$

or $C = (M + R)/2$ and $f(C)$

IF $f(C)<f(W)$ THEN

replace *W* with *R*

ENDIF

ENDIF

END {Case (i)}

replace *W* with *C*

ELSE

compute *S* and *f(S)*

replace *W* with *S*

replace *G* with *M*

ENDIF

END {Case (ii)}

Table 10 NELDER-MEAD METHOD PROCEDURE

A contour function is then used to plot a curve connecting minimum points where the function has the same particular value. The minimum value is normalized between 0 and 1 and the contour line is 0.1 in each level.

8.3. Experiment Setup

This section describes the setup for a large-scale campus-based experiment. The same settings are used as in [Bahl, Padmanabhan et al. 2000; Jan and Lee 2003; Kaemarungsi and Krishnamurthy 2004a; Kaemarungsi and Krishnamurthy 2004b; Kwon, Dundar et al. 2004; Taheri, Singh et al. 2004; Wong, Ng et al. 2005]. The RSS site survey measurements are taken in The Hong Kong Polytechnic University (PolyU) campus. The approximate total area of the campus is 9.34 hectares. A standard laptop computer equipped with an Intel WLAN card and client manager software was used to measure samples of RSS from access points (APs) of PolyU campus. The WLAN card is a chipset inside the laptop.

In this arrangement, the campus is regarded as has 26 major buildings from Core A to Core Z as well as seven other large buildings, each with WLAN access. Each core building is covered by at least 13 APs. The radio frequency (RF) channels of IEEE 802.11b are in the 2.4 GHz band which is shared by other equipment in the industrial, scientific, and medical (ISM) band such as Bluetooth. There are three non-overlapping channels for 802.11b. The RSS value reported by the WLAN card is an average value over a sampling period and in increments of 1 dBm. The received signal sensitivity of the WLAN card limits the range of the RSS to between -93 dBm and -15 dBm. Nevertheless, the highest typical value of the RSS is approximately -30 dBm at one meter from any AP. The sampling schedule seeks to collect the RSS data every 5 seconds. The vector of the RSS data at each location forms the location fingerprint with around 20 RSS elements in the vector. The measurements were taken at a total of 27 locations on-campus.

Figure 66 and 67 show a satellite photo of the PolyU campus from Google Earth and a site plan for the same area showing the 27 buildings of interest. The radio channels used for each AP are channel 1, 6, and 11. The sampling will be taken in two periods: between 7:30am and 9:30am (not busy) and 4:30pm and 6:30pm (busy). As shown in [Chan, Baciu et al.], the presence or absence of people in a building significantly affects the RSS values.

The data were collected four times from four different directions, North, South, East and West. The duration of sampling was over two weeks on a total of 12 days (from Mon to Sat). The temperature, weather, sampling time and humidity were also recorded. The total number of RSS samples would be 12 days X 4 directions X 27 buildings X 20 APs X 2 = 51840.

Figure 66 Satellite photo of PolyU campus from Google Earth

Figure 67 Site plan of the PolyU Campus identifying the 27 buildings of interest

Figure 68 3-D rendering of PolyU Campus

Environmental readings were collected at 27 locations on the floor over a two-week period, At each test location a frequency was chosen, e.g. 2.4GHz and the average amplitude was calculated. Note that the RSS is the received signal from a beacon packet, while the spectrum energy is the ambient RF energy corresponding to a specific frequency range. Table 11 summaries the campus area measurement setup.

Item	Description
Total campus area	9.34 hectare 26 core + 7 extra buildings
Sampling period	2 weeks 7.30am - 9.30am 4.30pm - 6.30pm
RSS variation	Between -93 dBm and -15 dBm
No. of sample points	51,840 sample points
WLAN channel	1, 6, and 11
Facing direction	North, South, East, and West

Table 11 SUMMARY OF EXPERIMENT SETUP IN POLYU CAMPUS

8.4. Wireless Signal Behavior Study

This section discusses the effect of the presence of humans and the LOS factor in the proposed topographic model. There are three RSS features to be analyzed, LOS, the presence of humans, and RSS variation.

8.4.1. Effect of LOS on RSS

Figure 69 and Figure 70 show the effect of LOS in two major clusters of RSS. The two major centers of high intensity are located at F core and S core.

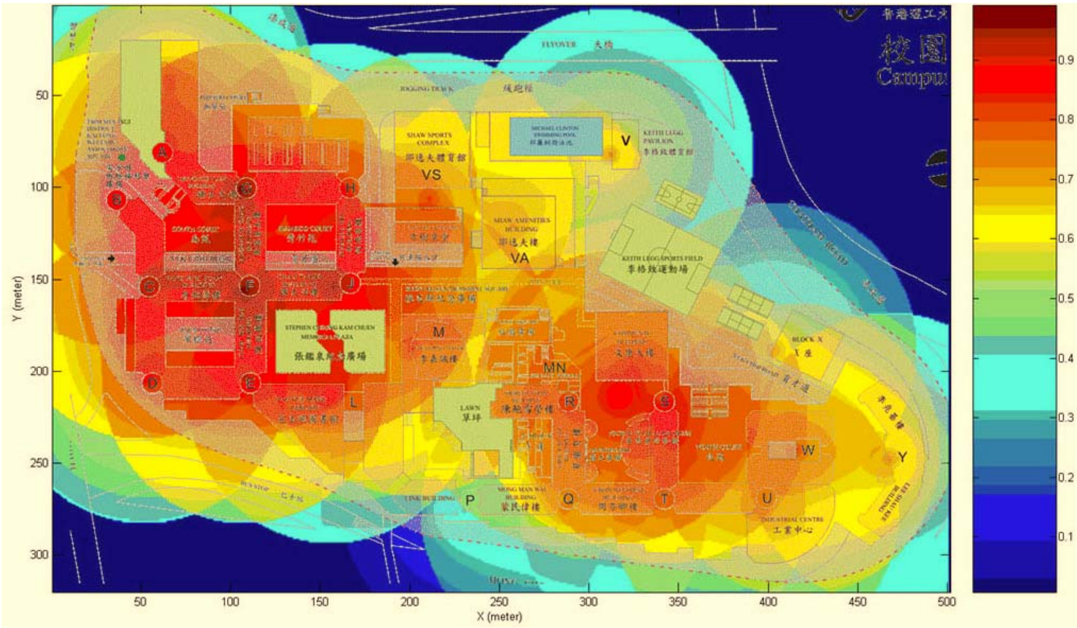


Figure 69 Fuzzy RSS Distribution with the campus floor plan

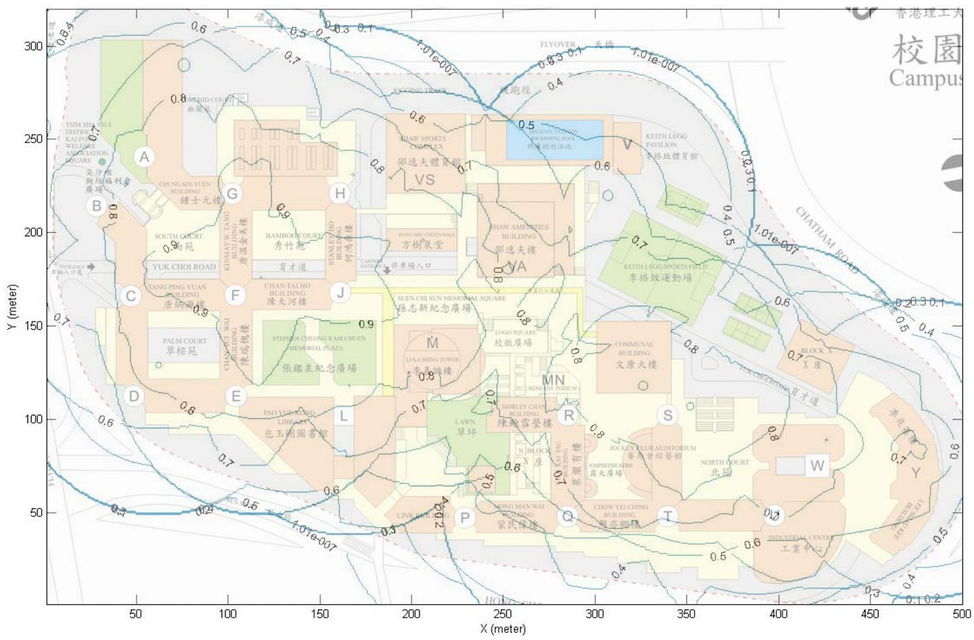


Figure 70 Topographic RSS Distribution with the campus floor plan

Figure 71 shows the 3-D Fuzzy WLAN signal distribution. The signal can be received to a height of around 75m and can cover almost every point in the PolyU campus.

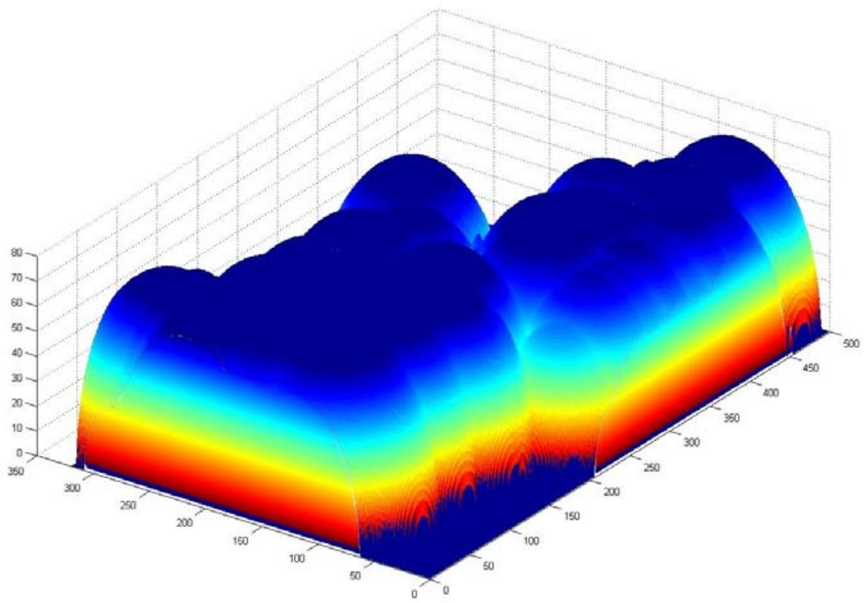
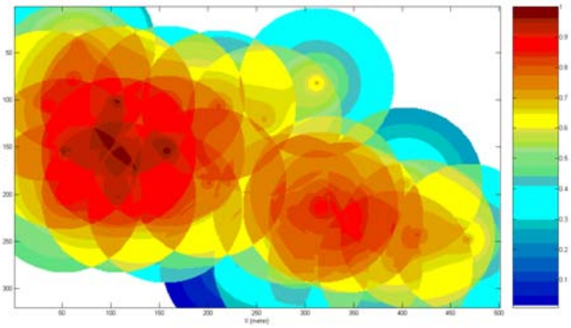
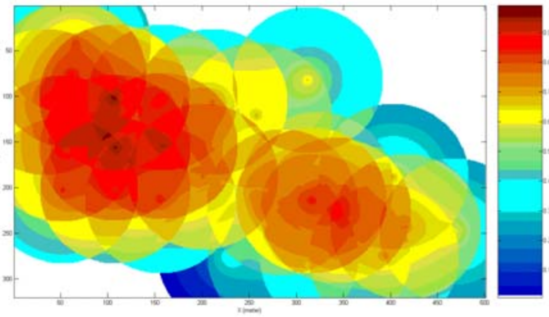


Figure 71 3-D Fuzzy WLAN Signal Distribution with the campus floor plan

Figure 69 shows the effect of LOS in two major clusters of RSS. The two major centers of high intensity are located at F core and S core. The distance from F core to S core is around 600m. In Table 1, the APs are evenly distributed. The signal should be distributed evenly. Between M core (Lee Ka Shing Tower) and R core (Shirley Chan Tower), the RSS distribution is relatively low. The heights of two buildings in M core and R core are around 80m and 70m respectively. The distance from M core to R core is around 200m. Figure 10 shows that from M core to R core there is low signal strength propagation. On the topographic map in Figure 70, the slope of the contour line from M core to R core is steep are the edge, which means that the RSS weakens quickly in the middle from M core to R core due to NLOS effects. The RSS transmission path between buildings can be LOS, partial LOS or shadow where NLOS propagation is possible. For LOS conditions, RSS should fit into lognormal distribution. A multi- story building in a campus area will experience lower signal strengths within tall buildings due to the absence of LOS propagation.

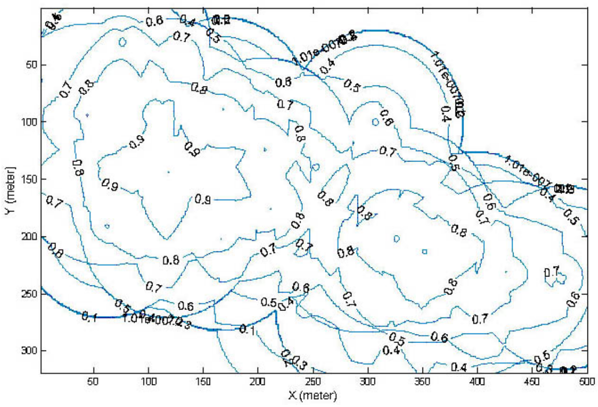


(a) In the leisure morning period

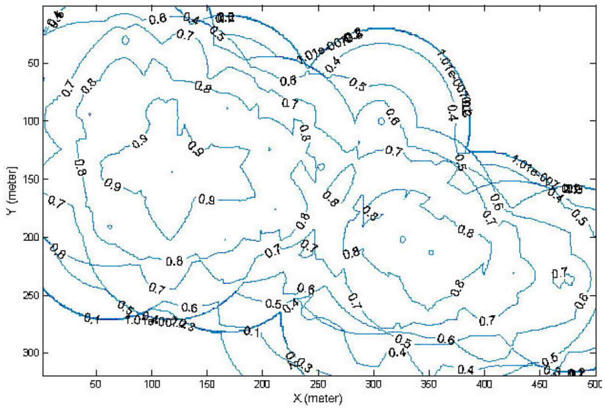


(b) In the busy evening period

Figure 72 RSS Distribution in Fuzzy Analytical Model



(a) In the leisure morning period



(b) In the busy evening period

Figure 73 RSS Distribution in Topographic Model

8.4.2. Behavior Study On the Human's presence

The human body is 70% water. This water can absorb some of the strength of signals more than the air [Kaemarungsi and Krishnamurthy] as can other clutter densities. A user's presence can thus affect the mean of the RSS value and the accuracy of location estimation in the system. To investigate this, the same data and collection methods are used as in the previous section. In this case, however, of especial interest are any differences between the measurements taken in the morning and in the afternoon.

Figure 72(a) and Figure 72(b) show the different RSS patterns when the data was collected in the two different time slots. Figure 72(b) shows that the intensity of the red is lower during the busy period. Figure 73(a) and (b) show that in the less busy period the intensity of the red increases from 0.5 to 0.77. In Figure 73(a) the topographic region in 0.9 levels is larger than in Figure 8(b). The slope in Figure 73(b) is smaller than Figure 73(a). We can conclude with some confidence that the presence of humans' bodies will affect the mean of the RSS value. In other words, the RSS will be weaker in busy periods, reducing the accuracy of location estimation.

8.4.3. Effect of the RSS Variation on Accuracy

The accuracy of the tracking system is highly dependent on RSS variation. If the standard deviation of the RSS increases, the accuracy of the tracking system falls. In this it is assumed that to maintain high accuracy the suggested standard deviation of RSS should be under 4dBm. Other authors in other work have assumed a standard deviation of 2.13dBm. [Taheri, Singh et al. 2004]) But this work is in particular allowing for the exigencies of real environments, including dense human traffic, and so assumes a large standard deviation.

8.4.4. Effect of the RSS Path Loss Exponent on Accuracy

Accuracy of the tracking system is highly dependable on the RSS path loss exponent. The path loss exponent can be varied between 1 and 6 according to the RSS absorbing medium. Path loss exponent represents the attenuation rate of the RSS. When the path loss exponent increases, the accuracy of the tracking system becomes higher. Therefore, high path loss exponent is an easy way to track the target.

Chapter 9: Location-Aware Information Retrieval by Agent-based Semantic Matching

Agents operating in both wired and wireless networks find and retrieve location-aware information. Agents in the proposed system must have the full range of abilities, including perception, use of natural language, learning and the ability to understand user queries. The speed and accuracy of retrieval and the usefulness of the retrieved data depends on a number of factors including constant or frequent changes in its content or status, the effects of environmental factors such as the weather and traffic, and the techniques that are used to categorize the relevance of the retrieved data.

This chapter proposes semantic TFIDF, an agent-based system for retrieving location-aware information that makes use of semantic information in the data to develop smaller training sets, thereby improving the speed of retrieval while maintaining or even improving accuracy. This proposed method first assigns intelligent agents to gathering location-aware data, which they then classify, match, and organize to find a best match for a user query. This is done using semantic graphs in the WordNet English dictionary. Experiments will compare the proposed system with three other commonly used systems and show that it is significantly faster and more accurate.

The rest of this chapter is organized as follows: Section 9.1 presents the background of location-aware retrieval systems. Section 792 describes semantics TFIDF technique in the proposed system. Section 9.3 presents the system implementation and architecture of the proposed agent-based semantics retrieval system. Section 9.4 describes the experimental design and setup. Finally, Section 9.5 discusses the performance of the proposed system in terms of precision and speed.

9.1. Introduction

One of the most challenging problems in retrieving the location-aware information is to understand the behavior of users and how it suits the current location. Wireless tracking applications are popular ways to help user navigate that may make use of current location-aware information. However, very often, the information retrieved in common search engines is both excessive and unstructured from the user's point of view. The basic requirement of an effective location-aware retrieval system is that it should match user queries and provides accurate information where users access location-aware (e.g., pervasive computing-enabled) applications and services. The retrieval system should do this in an organized and efficient way.

In recent years, researchers have focused on how to provide higher accuracy and faster retrieval by making use of keywords and textual semantics. However, the results so far have been unsatisfactory. First, the truth-conditional semantics that are often applied provide only a very limited account of meaning. Second, information is derived from few sources and the information from those sources is not structured. In these circumstances, the value of the information that can be extracted is very limited. Third, information collection is an expensive, time-consuming process that is often carried out manually. This makes it very difficult to build, maintain and grow comprehensive databases. Fourth, some approaches, such as Naïve-Bayes [Danesh, Moshiri et al. 2007] and K-Nearest Neighbors (K-NN) classifiers [Weiss, Kasif et al. 1996] machine learning approaches ignore the semantic meaning in text classification. This leads to inadequate search results. Finally, location-aware information is distributed in different locations and some information, traffic information for example, they will quickly become out-of-date so location-aware information must be updated frequently. Figure 74 summarizes problems of current information retrieval systems.

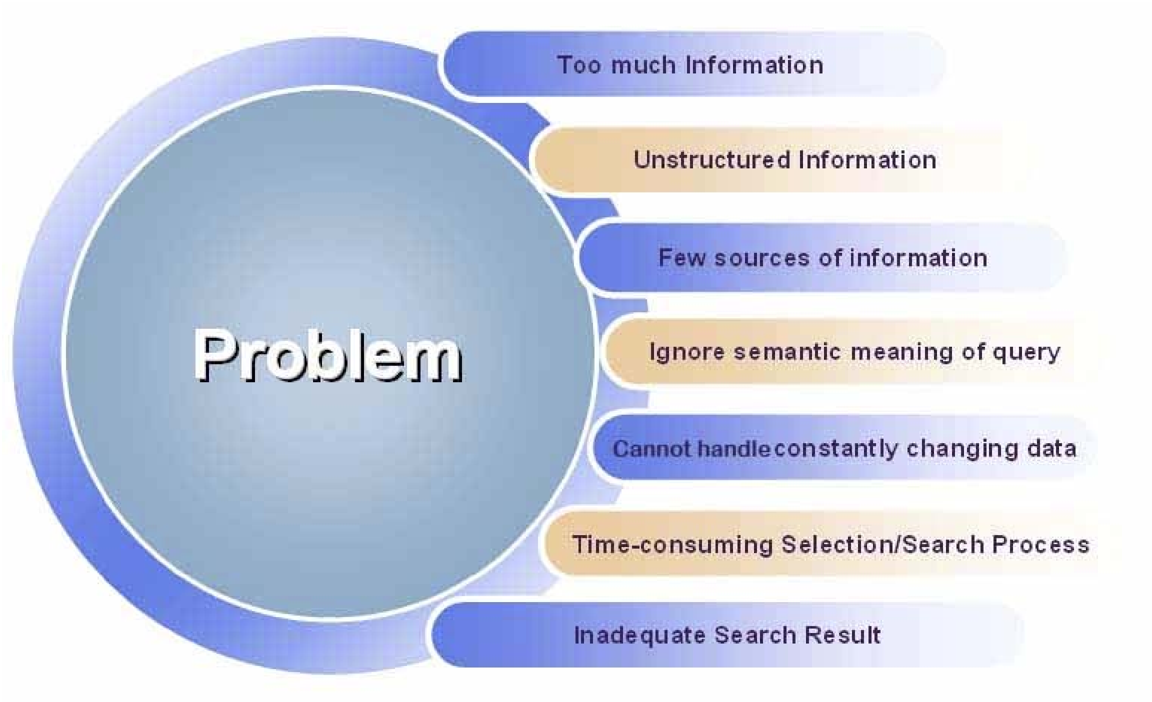


Figure 74 Problems of current information retrieval systems

This chapter proposes an agent-based semantics retrieval system for location-aware information. The proposed system uses the WordNet [WordNet 2004] dictionary to construct the semantics graph structure of a location. This graph structure sets the weight values between edges and nodes (words) of the semantics and the classic information retrieval, term-frequency-time-inverse-document-frequency (TFIDF) technique is modified with semantics weight values. As the location-aware information is distributed in different locations, the approach also implement agents using the IBMAglets Agent Software Development Kit [IBM 2004]. These mobile network agents are programs that can be dispatched from one computer and transported to a remote computer for execution. Arriving at the remote computer, they present their credentials and obtain access to local services and data. The remote computer may also serve as a broker by bringing together agents with similar interests and compatible goals, thus providing a meeting place at which agents can interact. The proposed system uses four types of agents: one each to gather, classify, match, and organize information.

The proposed system offers a number of benefits. First, the proposed semantics graph structure provides a hierarchical structure for the location-aware information. Second, the proposed agent system obviates the need for extensive manual information grasping. Third, the modified TFIDF with semantics weights improves the accuracy of matching keywords and provides a more meaningful result from the point of view of semantics. Forth, the agent can update information directly by communicating with its neighbor agents. Finally, it is fast and cost-effective.

9.2. Semantics TFIDF Technique

Normally, a location-aware application receives a keyword query from users. This section begins by describing how to use the keyword feature selection technique. It then introduces the TFIDF technique for forming a space vector model. The WordNet [WordNet 2004] English dictionary is then used to form a graph of the semantic structure of a term/word and to set weight values to each edge of the graph. Instead of using Euclidean distance directly, the approach calculates the distance including the weight value of each vector. Finally, the information is clustered using K-NN algorithm.

9.2.1. Feature Selection Technique

Feature selection is a common information retrieval technique. Usually, there is three feature selection criteria: 1) pruning of infrequent features, 2) pruning of high frequency features and 3) Choosing features, which have high mutual information with the target concept.

The first step is to prune the infrequent words from my training sets of data. This removes most spelling errors and speeds up the later stages of feature selection. The nest step is to prune the most frequent words. This technique should eliminate non-content words like "the", "and", or "for". Finally, the remaining words are ranked using the

TFIDF technique according to their mutual information with reference to the target concept/category.

9.2.2. The TFIDF Technique

TFIDF [W.Wen, Liu et al. 2001; Lo, Cheng et al. 2008] is an information retrieval technique commonly used in query searching. This technique formulates the significance of a term/word according to its frequency in a document or a collection of documents. This work uses TFIDF to determine the weights that are assigned to individual terms. If a term t occurs in document d,

$$w_{di} = tf_{di} \times \log\left(N / idf_{di}\right) \tag{44}$$

where t_i is a word (or a term) in the document collection ,w_{di} is the weight of t_i, tf_{di} is the term frequency (term count of each word in a document) of t_i, N is the total number of documents in the collection and idf_{di} is the number of document in which t_i appears. The TFIDF technique learns a class model by combining document (vectors) into a vector space model.

9.2.3. Semantics Graph Structure

The WordNet English dictionary provides a hierarchical representation of English words organized according to the strength of their associations in a number of defined domains, which I will refer to here as a semantics graph. WordNet defines a weight value for each relationship in a graph.

Figure 75 shows a graph for the word "taste". The words "tart" and "unpleasant" are at the same level of the hierarchy and have relation with each other so they are very similar. Their distance in the tree structure can measure the similarity of two words. The ontology-based term frequency can be obtained by comparing the meanings of two terms,

$$otf_1 = tf_1 \text{ x } (1+(1/D(t_1,t_2))^{tf2} \qquad (45)$$

where t_1, t_2 are different terms; otf_1 is the ontology-based term frequency of t_1; tf_1, tf_2 are the term frequencies of t_1 and t_2; and $D(t_1,t_2)$ is the depth between t_1 and t_2. $D(t_1,t_2)$ can be calculated.

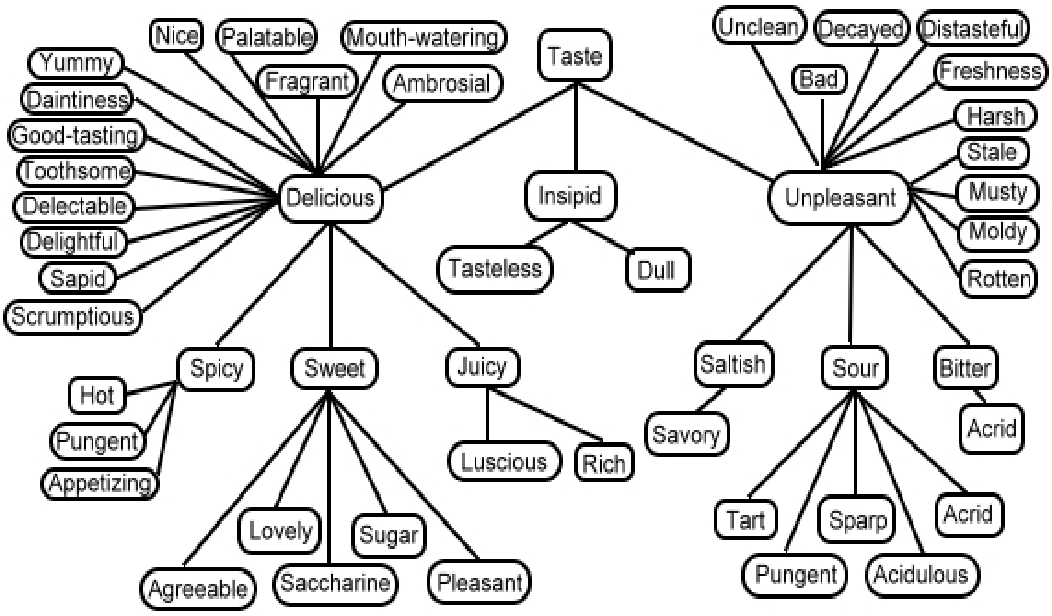


Figure 75 A word graph example for the word "taste"

For example, assume the term frequencies of "tart" and "unpleasant" are 3 and 2 respectively and the depth from "tart" to "unpleasant" is 3. The ontology-based term frequency of "tart" will be otf_1 = 3 x $(1+(1/3))^2$ = 5.33. The ontology-based term frequency of “unpleasant” will be otf_2 = 2 x $(1+(1/3))^3$ = 4.67. After adjustment of each term frequency, their term frequency value may increase. The two terms could thus become more significant after computing TFIDF.

9.2.4. Term Document Matrix

A term document matrix [Sudarsun, Kalaivendhan et al.; Jaber, Amira et al.] is widely used in natural language processing and information retrieval as a way to reduce space requirements and speed up searching. A term document matrix is a large grid representing every document and content word in a collection. A term document matrix is generated to store my data. (41) is used to get a new (ontology-based) TFIDF value for each term and then a document vector denoted by $A = \{a_1, a_2 ..., a_n\}$ where n is the total number of documents. A document vector contains three elements, a term, an ontology-based TFIDF, and a term frequency. These elements allow each document vector to be formed as a matrix in which each row stands for a term and each column stands for a document. Each cell will contain the ontology-based TFIDF weight of a particular document. If a given term appears in a given document, it can be seen at the intersection of the appropriate row and column. Finally, each document matrix is normalized and scaled down to prevent large documents with many keywords from overwhelming smaller documents. Smaller documents are considered to be more important and larger document are penalized, making every document equally significant.

Once a normalized term document matrix has been created, singular value decomposition is run to allow faster operations on the matrix.

9.2.5. K-NN Algorithm

After developing a set of term document martices, the K-NN algorithm is applied to a document w and a set of document clusters $C = \{c_1, c_2 \ldots c_i\}$. D = $\{d_1, d_2 \ldots d_i\}$ is a set of distances between w and c_i. It is possible to estimate the similarity between a document and a set of document clusters by calculating the Euclidean distance $|w- c_i|$. In a hierarchical clustering, two clusters c_i and c_j are randomly chosen and their similarity $sim(c_i ,c_j)$ is calculated. They are merged if the similarity value is greater than the threshold value. Otherwise, this step repeats until reaching the termination condition with K clusters.

$$d = \frac{\sum_{i=1}^{n} \frac{d_i}{|w - c_i|}}{\sum_{i=1}^{n} \frac{1}{|w - c_i|}} \tag{46}$$

9.3. System Implementation

This section introduces the proposed agent-based approach to implementing the system. It uses four different types of agent, which handle information grasping, classification, matching, and organizing. Figure 76 shows the system infrastructure of the agent-based semantics retrieval system while the following subsections describe each of the four types of agents.

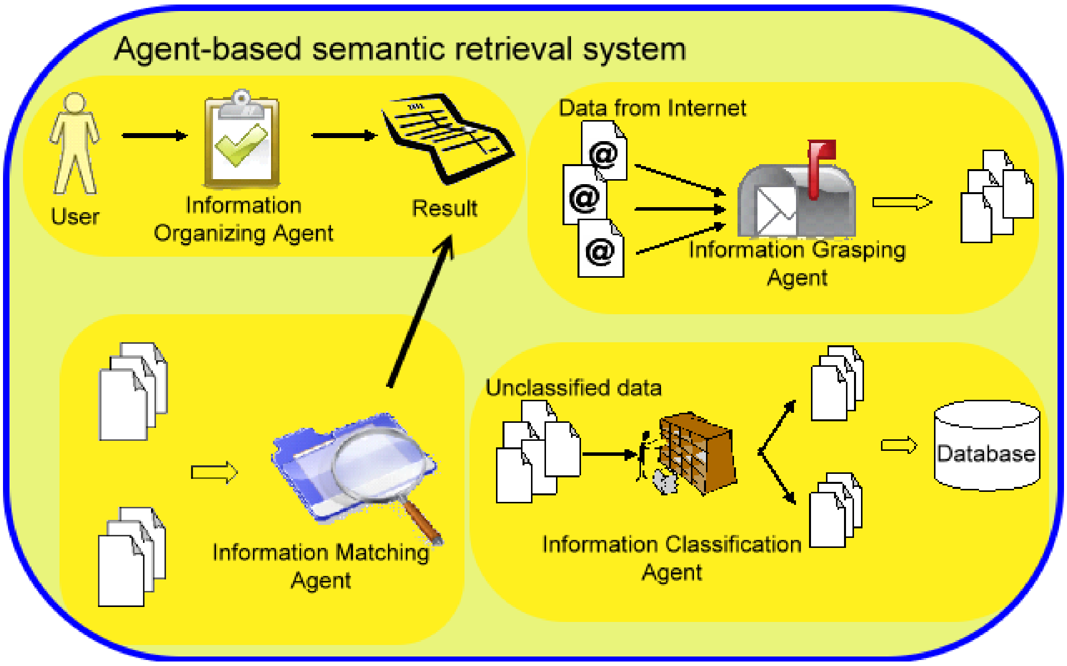


Figure 76 System Architecture of Agent-based semantics retrieval system in the proposed system

9.3.1. Information grasping agent

The information-grasping agent consists of two sets of agents, one static and the other dynamic. The static information-grasping agent is installed in a server to collect location-aware information from URL websites, such as weather forecast, transport, and tourism websites. Static agents are always online and maintain a 24-hour record of RSS (rich site summary) information from popular websites.

Dynamic information grasping agents collect information in a mobile environment. In most of cases, they collect location-aware information through information exchanges between users or sometimes collect data invoked by other real-time systems such as surveillance systems or supply chain RFID tracking systems. Increasing the number of dynamic information agents will help keep information more up-to-date but on the other hand puts an added burden on network traffic. A user-driven process will allow only the dynamic information-grasping agent to be triggered. When the agent is invoked at a user's request, it gathers different sources of information automatically and then updates the information with neighbors or other systems. It thus contributes to and receives the latest information through the system.

9.3.2. Information classification agent

The information classification agent classifies output information into suitable categories and calculates the Euclidean distance between the words to form a word graph, which expresses relationships between words (Figure 1). It is a core part in the system because word classification directly affects the accuracy of searching.

9.3.3. Information matching agent

The information-matching agent matches user queries and categorizes information. According to the users' preferences, the agent analyzes and searches for information in the categorized information. The best match for the user's requirements is found using (41) and (42).

9.3.4. Information organizing agent

The information-organizing agent organizes the search result. It provides a user-friendly interface, receiving user preferences and changing the display and organization of the content.

9.4. Experiment Design and Setup

This section makes use of the Chinese Learner English Corpus (CLEC) [PolyU 2004] and in particular 500 location-related articles subdivided into six categories. There are 83 articles each for the domains Hotel, Restaurant, Traffic and Shopping mall and 84 for the domains of College and Clinic. Figure 3 illustrates the number of articles in each category.

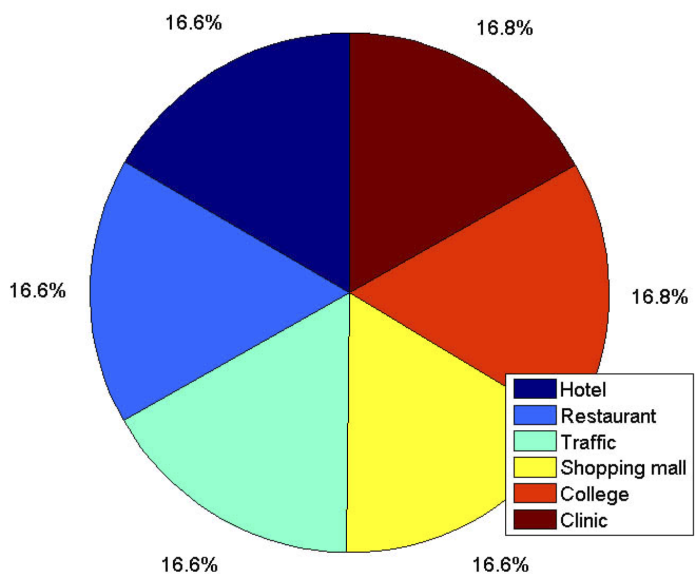


Figure 77 The number of articles in each category

9.4.1. Experiment Settings

Two sets of experiment are conducted to compare three types of techniques, the three-layer neural-network (NN), TFIDF with hierarchical clustering, and the proposed semantics TFIDF techniques. The first experiment tests the precision of categorization/clustering. The second experiment measures the computation time required to cluster using these three techniques. An additional three test sets of 100 user queries is input to find suitable information from the system. Each user query consists of at most ten words. All the experiments are conducted on a common desktop computer with a CPU of Intel Pentium 4.3 GHz, 2GB DDR2 SDRAM and physical storage of 200GB with 7200 rpm.

9.4.2. Determining the effectiveness of categorization

Determining categorization effectiveness is a binary decision problem which is either positive or negative. The decision made by the classifier can be represented in a structure known as a confusion matrix (or contingency table). The confusion matrix has four categories: True positives (TP) are examples correctly labeled as positives. False positives (FP) refer to negative examples incorrectly labeled as positive. True negatives (TN) refer to negatives correctly labeled as negative. Finally, false negatives (FN) refer to positive examples incorrectly labeled as negative. Categorization effectiveness is usually measured by the precision and recall rates [Davis and Goadrich 2006]:

$$\text{Precision} \quad \frac{TP}{TP+FP} \tag{47}$$

$$\text{Recall:} \quad \frac{TP}{TP+FN} \tag{48}$$

9.5. Performance Evaluation

This section evaluates the effectiveness of categorization in terms of precision and speed.

9.5.1. Precision rate

Table I and Figure 4 show the result for precision when using the three different information retrieval techniques. The average precision across different categories of semantics TFIDF is 95.68%, which is the highest of the three techniques.

Category	**Method used**		
	NN (3-Layer)	***TFIDF with hierarchical clustering***	***Semantics TFIDF***
Hotel	68.40%	90.30%	95.13%
Restaurant	58.45%	93.37%	96.72%
Traffic	73.43%	93.88%	95.14%
Shopping mall	67.00%	94.33%	97.44%
College	70.22%	93.78%	94.38%

Category	**Method used**		
	NN (3-Layer)	***TFIDF with hierarchical clustering***	***Semantics TFIDF***
Clinic	50.23%	91.67%	95.24%
Average	64.62%	93.72%	95.68%

Table 12 COMPARISON OF THE PRECISION RATE (%)

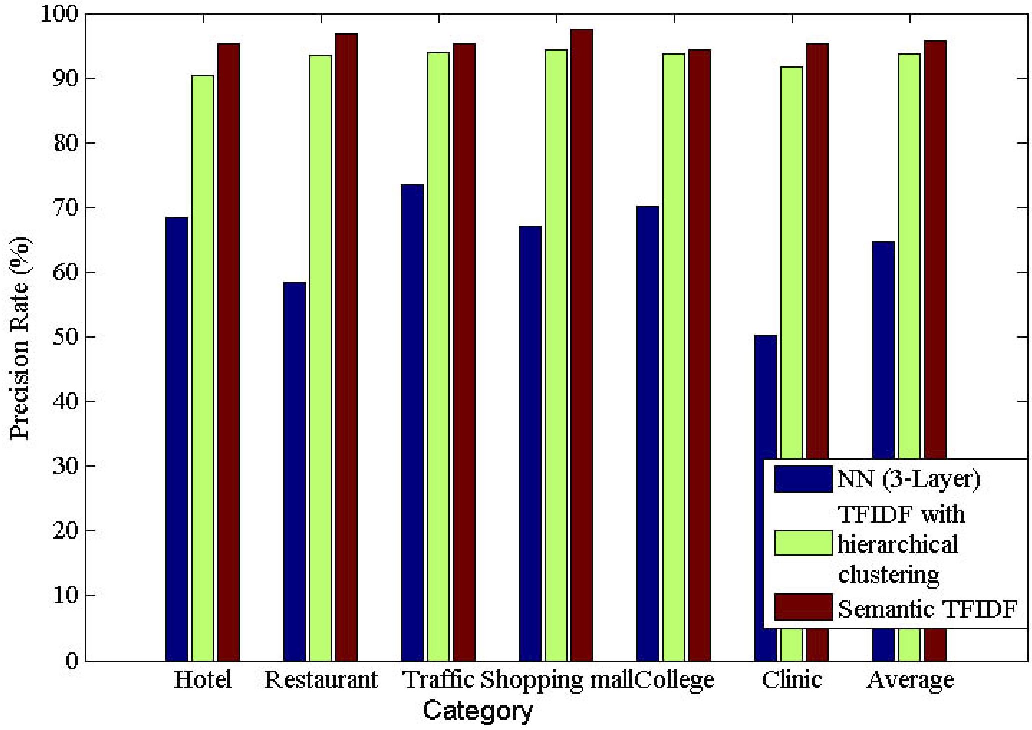


Figure 78 The precision rate for the classification with 3 different information retrieval techniques

9.5.2. Comparison of time taken for clustering

Figure 5 shows the computational time required to find suitable information when using three different information retrieval techniques. The NN technique takes 5295 seconds for clustering, which is approximately five times longer than the other two methods. The shortest time required finding suitable information is TFIDF with hierarchical clustering technique, at only 1138 seconds to finish clustering. The semantics TFIDF technique requires 1185 seconds.

Category	**Method used**		
	NN (3-Layer)	***TFIDF with hierarchical clustering***	***Semantics TFIDF***
Set 1	5277 seconds	1114 seconds	1175 seconds
Set 2	5432 seconds	1123 seconds	1186 seconds
Set 3	5175 seconds	1176 seconds	1195 seconds
Average	5295 seconds	1138 seconds	1185 seconds

Table 13 THE COMPUTATIONAL TIME TAKEN FOR FINDING SUITABLE INFORMATION WITH 3 DIFFERENT INFORMATION RETRIEVAL TECHNIQUES

Chapter 10: Conclusion & Future Work

This chapter concludes the thesis and is organized in four sections. Section 10.1 summarizes the research work. Section 10.2 summarizes the major contributions of the thesis. Section 10.3 discusses the limitations of the proposed system. Finally, Section 10.4 suggests some possible further work on the proposed system.

10.1. Summary of Research Framework

This research has presented an accurate and robust indoor localization framework. There are five major goals to be achieved by this framework, to provide

(1) a stable WLAN signal transmission,

(2) an accurate positioning algorithm,

(3) a structural WLAN infrastructure that could support intensive tracking,

(4) a model that could visualize the WLAN signal distribution to prevent signal black spots and interferences to be happened and finally

(5) a perception retrieval system that could provide location-aware information according to users' requirements.

Satisfying these five goals in turn produced the following five proposed methods:

(1) reducing channel interferences between APs,

(2) using the Newton Trust-Region Method to improve positioning

(3) establishing a cell-based plan for the WLAN infrastructure,

(4) visualizing the WLAN signal distribution, and

(5) retrieving location-aware information according to the users' need.

The approach to reducing channel interferences between APs process used location fingerprinting to investigate the influence of channel interference properties and using orthogonal channel assignment to maintain the stability of wireless transmission. The study of channel interference is essential for accurate indoor positioning system. The localization experiments and simulations were conducted on the IEEE 802.11 test-bed and investigated the channel assignment of APs, the distribution of received signal strength (RSS) values, and variations in coverage and distance between APs. The resulting analysis has provided insights into how to assign channels, how to separate each AP at an appropriate distance so as to reduce interference, and how many access points are required to uniquely identify a location at a given accuracy and precision. These results would be of interest to engineers designing a WLAN channel assignment.

The proposed accurate WLAN localization using a Newton Trust-Region (TR) method implements the TR method and LF-based approach in three phases. The first phase was to detect the IEEE 802.11b Wi-Fi signal strength and collect the LFs into a training database. In the second phase, the LFs (location fingerprints) were retrieved by the mobile WLAN-enabled device and used to estimate the location by applying the K-Nearest Neighbor method to the LF training database. The third phase was to have a trajectory of a user (WLAN-enabled device). Normally, each step or estimated point in a trajectory falls within a region and has the same convergence. The TR method tracked a trajectory by using a feedback process.

In establishing the cell-based plan WLAN infrastructure process, a proposed cell-based WLAN infrastructure deployment design was used to visualize blind spots, predicted how many APs would be optimally required for positioning, and suggests where they should be placed. The proposed approach extends my previous work [Chan, Baciu et al. 2009d] from a two-dimensional to a three-dimensional approach and uses three structured geometric configurations for use in WLAN infrastructure deployment: (1) triangular, (2) square, and (3) hexagonal.

To visualize the WLAN signal distribution process, I proposed a three-dimensional representation of a WLAN infrastructure deployment design that virtualizes the localization blind spots and predicts in advance how many APs are optimally required for localization and where they should be placed. Use was also made of the Nelder-Mead (NM) method, simplifying my previous work on the multi-layer fuzzy color model [Chan, Baciu et al. 2009b] to topographic (line-based) model. I develop a topographic model as an analytical tool for evaluating and visualizing where the RSS is denser and clustering different RSS at different topographic levels.

This work has proposed the retrieval of location-aware information using an agent-based semantics retrieval system. The proposed system is a preliminary attempt to use the WordNet [WordNet 2004] dictionary to construct a semantics graph structure for a location. This graph structure makes use of semantics weight values between edges to nodes (words). The classic information retrieval TFIDF technique has been modified with semantics weight values and as the location-aware information is distributed at different locations, the system also implements the agent with the IBMAglets Agent Software Development Kit [IBM 2004]. These mobile network agents are programs that can be dispatched from one computer and transported to a remote computer for execution. Arriving at the remote computer, they present their credentials and obtain access to local services and data. The remote computer may also serve as a broker by bringing together agents with similar interests and compatible goals, thus providing a meeting place at which agents can interact. The proposed system implements four types of agents, which gather, classify, match, and organize information.

10.2. Summary of Contributions

The proposed accurate and robust indoor localization framework provides

(1) a stable WLAN signal transmission, (2) an accurate positioning algorithm, (3) a robust and structural WLAN infrastructure that can support intensive tracking, (4) a comprehensive model that can visualize the WLAN signal distribution to prevent signal black spots and interference and (5) an accurate perception retrieval system that provides location-aware information according to user behaviors. The major contributions of this research are summarized as follows:

1. Stable WLAN signal transmission: a comprehensive analysis is provided of the channel interference to positioning accuracy based on location fingerprinting. Beginning from the position that channel interference worsens positioning accuracy and that more channel interference implies less accurate positioning, the notion is refuted that interference could help to achieve more accurate positioning. The experiments clearly show that channel interference between APs using the same frequency channel has a significant worsening impact on the positioning system. The experimental analysis also shows that the channel interference usually obeys a right-skewed distribution. These findings are an important step towards the accurate virtual modeling of the environment of channel interference.
2. The effect of channel interference on positioning accuracy depends on the number of APs, the SINR and the channel assignment scheme. A higher number of APs gives more accurate positioning and as the number of APs increases, the channel

interferences between APs increases, decreasing the rate at which positioning accuracy increases. It is also noted that a higher value of SINR is associated with more accurate positioning. This chapter emphasizes the fact that channel assignment is a critical issue in a positioning system. The 2-D simulation result of location uncertainty shows that orthogonal channel allocation should be the first consideration in improving a positioning system. Choosing orthogonal channel allocation could reduce by 15% the number of APs required while improving positioning accuracy by an average of 10%.

3. Accurate positioning algorithm: A Newton Trust-Region method of WLAN localization is proposed and it is tested and its accuracy is compared with that of the Kalman Filter and traditional LF approaches. The experimental analysis shows the effectiveness of proposed Trust-Region method leads to substantially more accurate and robust localization system. The proposed method requires 15% fewer APs and in localization is 10% more effective than the traditional LF approach.

4. Robust and structural WLAN infrastructure: three geometric structured approaches to WLAN infrastructure deployment are deployed and simulated to compare their accuracy with that of random approaches. A particular contribution in this chapter is to show that the current random approach to AP deployment is very much less cost effective than any of the structured approaches no matter how

many APs are involved or what the desired level of accuracy, even before the real-world costs related to empirical or manual (that is to say, essentially random) deployment are factored in. Interestingly, in certain circumstances, that is, where few APs are to be deployed, the triangular approach outperforms the hexagonal approach. But overall the hexagonal approach is by far the most cost-effective and infrastructure designers and administrators may wish to take account of this in their infrastructure designs and deployments. In contrast, the random approach requires 50% more APs than the hexagonal structured approaches to achieve perfect localization.

5. Comprehensive WLAN signal visualization model: The proposed visualization model of the wireless signal distribution can operate in either a color-based fuzzy format or a line-based topographical format. The proposed framework offers two benefits. First, it serves as a quicker reference and efficient analytical tool. Second, it can provide a three-dimensional detail and quantitatively strong representation of WLAN RSS.

6. Accurate location-aware information retrieval system: The proposed system makes use of a semantics TFIDF technique and implements intelligent agents for information grasping, classification, and matching. Experimental evaluations of the system show that it produces more accurate and faster search result.

10.3. Limitations of the Framework

There are several limitations of my framework, as follows:

1. Insufficient IEEE 802.11b/g channels: Although the experimental result clearly shows that orthogonal channel allocation scheme gives a significant improvement of positioning, it is very expensive to use this scheme. Orthogonal channels are limited and IEEE 802.11b only offer three non-overlapping channels, leaving the remaining channels idle for the purposes of accurate positioning.
2. Physical limitations to install wireless infrastructure strictly according to cell-based plan: Although the proposed approach gives a significant improvement in positioning and clearly shows that a structural deployment of AP is far better than an empirical approach, the proposed approach suffers from physical limitations such as the depth of walls and floors and their composition (e.g. concrete). These would affect issues of AP placement and visualization of wireless signal. In real world scenarios, engineers need to go through every floor of every building in a construction site to detect whether a region can be covered by wireless APs. After they collect wireless signal samples, they must decide where APs should be added to increase signal coverage or be taken away to reduce signal interference. As mentioned, a strict implementation of the proposed approach is an ideal because the architecture of a building may not allow for an

entirely strictly structured deployment of APS.

3. It is difficult to match location-aware multimedia data according to user's preference. The proposed approach in Chapter 9 is a preliminary attempt that uses "text" to semantically match user's preference. However, it is difficult to retrieve useful information from non-text or multimedia content.

10.4. Future Work

In my future work, I will continue to investigate in two directions. The first one is the real-time WLAN signal visualization system. One possible solution is to build a wireless sensor network that could retrieve WLAN signal distributions automatically.

The second one is to provide a mathematical analysis to verify the result of cell-based plan approach. Recently, there is a paper [Hossain and Soh 2010] published in 2010 which provides some insight in mathematical verification. This paper aims to find the CRLB for location fingerprinting method. The covariance of a location estimator θ is bounded by the inverse of Fisher information J(θ). The bound of inverse of Fisher information, $[I(\theta)]^{-1}$ is called CRLB. They find out that the lower bound for a square configuration of the APs (K = 4) would be smaller than that of an equilateral triangle configuration (K = 3). I may further extend this result to have a complete mathematical analysis of structured deployment of AP.

Bibliography

Bahl, P. and V. Padmanabhan (2000). RADAR: an in-building RF-based user location and tracking system, The 19th IEEE International Conference on Computer Communications, INFOCOM, vol.2.

Bahl, P., V. Padmanabhan, et al. (2000). A Software System for Locating Mobile Users: Design, Evaluation, and Lessons, Microsoft Online Research Document.

Budianu, C., S. Ben-David, et al. (2006). Estimation of the number of operating sensors in large-scale sensor networks with mobile access, Signal Processing, IEEE Transactions, vol.54, pp.1703-1715.

Carle, J., J. Myoupo, et al. (2001). Higher Dimensional Honeycomb Networks, Journal of Interconnection Networks, vol.2, pp.391-420.

Chan, C. L., G. Baciu, et al. (2008). Wireless Tracking Analysis in Location Fingerprint, IEEE Wireless and Mobile Computing, Networking and Communications, pp.214-220.

Chan, C. L., G. Baciu, et al. (2009a). Cognitive Location-Aware Information Retrieval by Agent-based Semantic Matching, The 8th IEEE International Conference on Cognitive Informatics, pp.435-440.

Chan, C. L., G. Baciu, et al. (2009b). Fuzzy Topographic Modeling in WLAN Tracking Analysis, International Conference on Fuzzy Computation, pp.17-24. Madeira, Portugal.

Chan, C. L., G. Baciu, et al. (2009c). Properties of Channel Interference for WLAN Location Fingerprinting, submitted to IEEE Transactions on Wireless Computing.

Chan, C. L., G. Baciu, et al. (2009d). Resource-effective and Accurate WLAN Infrastructure Design and Localization Using a Cell-structure Framework, The 5th International Conference on Wireless Communications, Network and Mobile Computing, vol.6, pp.9-15.

Chan, C. L., G. Baciu, et al. (2009e). Using the Newton Trust-Region Method to Localize in WLAN Environment, The 5th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications., pp.363-369. Marrakech, Morocco.

Chan, C. L., G. Baciu, et al. (2009f). Using Wi-Fi Signal Strength to Localize in Wireless Sensor Networks, IEEE International Conference on Communications and Mobile Computing, vol.1, pp.538-542.

Chan, C. L., G. Baciu, et al. (2010). Orientation-based Wi-Fi Positioning on the Google Nexus One, accepted by the 6th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. Niagara Falls, Canada. .

Chang, K., L. Hong, et al. (2007). Stochastic trust region gradient-free method (strong): a new response-surface-based algorithm in simulation optimization, The 39th

Conference on Winter simulation, pp.346-354.

Chen, G. and T. Bui (1999). Invariant Fourier-wavelet descriptor for pattern recognition, Pattern Recognition Journal, vol.32, pp.1083-1088.

Chen, W.-P., J. C. Hou, et al. (2004). Dynamic Clustering for Acoustic Target Tracking in Wireless Sensor Networks, IEEE TRANSACTIONS ON MOBILE COMPUTING, pp.258-271, IEEE Computer Society.

Chiu, H. S., K. L. Yeung, et al. (2006). J-CAR: an Efficient Channel Assignment and Routing Protocol for Multi-channel Multi-interface Mobile Ad Hoc Networks, IEEE Global Telecommunications Conference, pp.1-5.

Danesh, A., B. Moshiri, et al. (2007). Improve text classification accuracy based on classifier fusion methods, The 10th International Conference on Information Fusion, pp.1-6.

Davis, J. and M. Goadrich (2006). The Relationship Between Precision-Recall and ROC Curves, ACM International Conference Proceeding Series, vol.148, pp.233-240, Proceedings of the 23rd international conference on Machine learning.

Dil, B., S. Dulman, et al. (2006). Rangebased localization in mobile sensor networks, European Conference on Wireless Sensor Networks, pp.164-179.

Dutta, P., S. Jaiswal, et al. (2007). Routing and Channel Allocation in Rural Wireless Mesh Networks, The 26th IEEE International Conference on Computer

Communications., pp.598-606.

Fang, S. H., T. N. Lin, et al. (2008). Location fingerprinting in a Decorrelated Space, IEEE Transactions on Knowledge and Data Engineering, vol.20, pp.685-691.

Goor, M. v. d. (2009). Indoor Localization in Wireless Sensor Neworks. Computer Science, Radboud University. **PhD Thesis**.

Haidar, M., R. Ghimire, et al. (2008). Channel assignment in an IEEE 802.11 WLAN based on Signal-To-Interference Ratio, Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference, pp.001169-001174.

Hightower, J. and G. Borriello (2001). Location Sensing Techniques, IEEE Computer magazine on Location Systems for Ubiquitous Computing, pp.57-66.

Hossain, A. K. M. M. and W.-S. Soh (2010). Cram´er-Rao Bound Analysis of Localization Using Signal Strength Difference as Location Fingerprint, The 29th IEEE International Conference on Computer Communications, INFOCOM, pp.1-9.

Hu, L. and D. Evans (2004). Localization for mobile sensor networks, Proceedings of the 10th annual international conference on Mobile computing and networking, pp.45-57.

Huang, J.-H., L.-C. Wang, et al. (2005). Deployment strategies of access points for outdoor wireless local area networks, The 61st IEEE Vehicular Technology

Conference, VTC, vol.5, pp.2949-2953.

IBM. (2004). "IBM Aglets." Retrieved 2004-05-20, 2004, from http://www.trl.ibm.com/aglets/.

Jaber, T., A. Amira, et al. (2007). Empirical study of a novel approach to LSI for text categorisation, The 9th International Symposium on Signal Processing and Its Applications, pp.1-4.

Jan, R. H. and Y. R. Lee (2003). An indoor geolocation system for wireless LANs, International Conference on Parallel Processing Workshops, pp.29-34.

Kaemarungsi, K. and P. Krishnamurthy (2004a). Modeling of indoor positioning systems based on location fingerprinting, The 23th IEEE International Conference on Computer Communications, INFOCOM, vol.2, pp.1012-1022.

Kaemarungsi, K. and P. Krishnamurthy (2004b). Properties of indoor received signal strength for WLAN location fingerprinting, The 1st International Conference on Mobile and Ubiquitous Systems: Networking and Services, MOBIQUITOUS, pp.14-23.

Ke, S. (2007). A Modified Filter Trust Region Method for Nonlinear Programming, Chinese Control Conference, pp.287-290.

Kj rgaard, M. B. and C. V. Munk (2008). Hyperbolic location fingerprinting: a calibration-free solution for handling differences in signal strength, The 6th

Annual IEEE International Conference on Pervasive Computing and Communications, pp.110-116.

Kwon, J., B. Dundar, et al. (2004). Hybrid algorithm for indoor positioning using wireless LAN, The IEEE 60th Vehicular Technology Conference, vol.7.

lab, I. R. S. (2002). "PlaceLab." Retrieved 2005-06-30, 2005, from http://www.placelab.org.

Li, B., Y. Wang, et al. (2005). Method for yielding a database of location fingerprints in WLAN, IEE Proceedings on Communications, vol.152, pp.580-586.

Liu and H. Chen (2004). Real-Time Tracking Using Trust-Region Methods, IEEE Transactions on Pattern Analysis on Machine Intelligence, pp.397-402.

Lo, C.-C., D.-Y. Cheng, et al. (2008). Design and Implementation of Situation-Aware Medical Tourism Service Search System, The 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp.1-5.

Mathews, J. and K. Fink (1998). Numerical Methods Using MATLAB, Simon & Schuster.

Metsälä, M. (2003). Positioning Framework for Symbian OS, University of Turku. **Master Thesis**.

Mowbray, S. and M. Nixon (2003). Automatic Gait Recognition via Fourier Descriptors of Deformable Objects, Lecture Notes in Computer Science, pp.566-573.

PolyU. (2004). "Chinese Learner English Corpus (CLEC)." Retrieved 2004-05-20, 2004, from http://langbank.engl.polyu.edu.hk/corpus/clec.html.

Prasithsangaree, P., P. Krishnamurthy, et al. (2002). On indoor position location with wireless LANs, The 13th IEEE International Symposium, Personal, Indoor and Mobile Radio Communications, vol.2, pp.720-724.

Rubin, I. and C. W. Choi (1997). Impact of the location area structure on the performance of signaling channels in wireless cellular networks, IEEE Comumunications Magazine, vol.35, pp.108-115.

Silvester, P. and M. Chari (1970). Finite Element Solution of Saturable Magnetic Field Problems, IEEE Transactions on Power Apparatus and Systems, pp.1642-1651.

Subramanian, A. P., H. Gupta, et al. (2008). Minimum interference channel assignment in multiradio wireless mesh networks, IEEE Transactions on Mobile Computing, vol.7, pp.1459-1473, IEEE Computer Society.

Sudarsun, S., D. Kalaivendhan, et al. (2006). Unsupervised Contextual Keyword Relevance Learning and Measurement using PLSA, Annual India Conference, pp.1-6.

Sun, W., Z. Qin, et al. (2008). Research on Interference-Based Channel Assignment Methods in 802.11-Based Wireless Mesh Network, The 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp.1-4.

Swangmuang, N. and P. Krishnamurthy (2008). Location Fingerprint Analyses Toward Efficient Indoor Positioning, The 6th IEEE International Conference on Pervasive Computing and Communications, pp.100-109.

Taheri, A., A. Singh, et al. (2004). Location fingerprinting on infrastructure 802.11 wireless local area networks (WLANs) using locus, The 29th IEEE International Conference on Local Computer Networks, pp.676-683.

Tiemann, C. O., S. W. Martin, et al. (2000). Aerial and Acoustic Marine Mammal Detectionand Localization on Navy Ranges, U.S. Department of Commerce and Department of the Navy, Joint Interim Report: Bahamas Marine Mammal Stranding Event.

W.Wen, C., H. Liu, et al. (2001). A Distributed Hierarchical Clustering System for Web Mining, The 1st International Conference on Web-Age Information Management, pp.103-113, Springer-Verlag Berlin Heidelberg.

Wang, J., H. Zha, et al. (2006). Coarse-to-Fine Vision-Based Localization by Indexing Scale-Invariant Features, IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol.36, pp.413.

Wang, Z., K. Zhang, et al. (2006). Tracking Dim Target in Infrared Imagery Using the Trust Region Embedded Particle Filter, The 8th International Conference on

Signal Processing, vol.2.

Weiss, S., S. Kasif, et al. (1996). Text Classification in USENET Newsgroup: A Progress Report, AAAI Spring Symposium on Machine Learning in Information Access Technical Papers. Palo Alto.

Wong, W., J. Ng, et al. (2005). WLAN positioning with mobile devices in a library environment., The 25th IEEE International Conference on Distributed Computing Systems Workshops, pp.633-636.

WordNet. (2004). "WordNet." Retrieved 2004-05-20, 2004, from http://www.wordnet.com.

Zeng, G., B. Wang, et al. (2007). Multicast algorithms for multi-channel wireless mesh networks, IEEE International Conference on Network Protocols, pp.1-10.

Zhang, D. and G. Lu (2002). Generic Fourier Descriptors for Shape-based Image Retrieval, IEEE International Conference on Multimedia and Expo, vol.2, pp.425-428.