REAL TIME DYNAMIC INDOOR POSITIONING SYSTEM USING MACHINE LEARNING TECHNIQUES

by

HASEEB BARI

Supervised By

DR. FAZLI SUBHAN

Submitted for partial fulfilment of the requirements of the degree of MSCS to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES,

ISLAMABAD

January 2020

REAL TIME DYNAMIC INDOOR POSITIONING SYSTEM USING MACHINE LEARNING TECHNIQUES

by

HASEEB BARI

Supervised By

DR. FAZLI SUBHAN

Submitted for partial fulfilment of the requirements of the degree of MSCS to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES,

ISLAMABAD

January 2020

NATIONAL UNIVERSITY OF MODERN LANGUAGES

FACULTY OF ENGINEERING AND COMPUTER SCIENCE

THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defence, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computer Sciences.

THESIS TITLE: Real Time Dynamic Indoor Positioning System using Machine Learning Techniques.

Submitted By: Haseeb Bari

Registration #: 01MS/MCS/S16

Master of Science
MSCS

Computer Science
Name of Discipline

Dr. Fazli Subhan
Name of Research Supervisor

Signature: ______________________

____
Name of Co-Supervisor (*If Any*)

Signature: ______________________

Col ® Jamil Ahmed Zia
Name of A/Dean (FE&CS)

Signature: ______________________

Brig. Muhammad Ibrahim
Name of Director General (NUML)

Signature: ______________________

16th January , 2020

CANDIDATE DECLARATION

I declare that this thesis entitled "*Real Time Dynamic Indoor Positioning System using Machine Learning Techniques*" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	______________________
Name	:	Haseeb Bari
Date	:	16th January, 2020

ABSTRACT

Position estimation is the process to find the actual location of an object with reference to some coordinate system or known landmark. This thesis focuses on position estimation of an object dynamically moving in an indoor environment. Previous studies focused more on static position estimation and used traditional position estimation techniques. In traditional position estimation techniques, RSSI measurements are used for distance estimation, which requires modelling of radio propagation to get distance estimates. Modelling of radio propagation in indoor environment is a challenging task due to multipath fading, reflection, refraction of light, temperature and presence of humans etc. All these parameters affecting the received signal and produces variations in RSSI. Due to variations in RSSI, distance and position estimation error occurs. To address the issue, this thesis presents fingerprinting based position estimation with the help of machine learning. Our proposed machine learning based indoor position estimation system consists of two steps. In step one, we perform real time experiments using Bluetooth Low Energy (BLE) Beacons and developed a radio fingerprints map. In second step, we investigated five different types of machine learning techniques. These techniques are Naive Bayes, K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Decision Tree in order to enhance position estimation accuracy especially for mobile objects.

Real time experiments are performed to evaluate performance of our proposed real time dynamic object tracking system, using five different trajectories in a 10 x 10 meters' indoor setup. These trajectories represent real time dynamic movement in different directions and speed. Experimental results show that LDA achieved highest mean accuracy of 79.34 % followed by SVM 78.38 %, while K-NN achieved 70.04 %.

Keywords: Position Estimation, Localization, Bluetooth, RSSI, Machine Learning.

DEDICATION

This thesis work is dedicated to my parents and my teachers throughout my education career who have not only loved me unconditionally but whose good examples have taught me to work hard for the things that I aspire to achieve.

ACKNOWLEDGEMENT

First of all, I wish to express my gratitude and deep appreciation to Almighty Allah, who made this study possible and successful. This study would not be accomplished unless the honest espousal that was extended from several sources for which I would like to express my sincere thankfulness and gratitude. Yet, there were significant contributors for my attained success and I cannot forget their input, especially my research supervisors, Dr. Fazli Subhan who did not leave any stone unturned to guide me during my research journey.

I shall also acknowledge the extended assistance from the administrations of Department of Computer Sciences who supported me all through my research experience and simplified the challenges I faced. For all whom I did not mention but I shall not neglect their significant contribution, thanks for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
RFID	Radio Frequency Identification
UWB	Ultra-Wideband
IR	Infrared
WLAN	Wireless Local Area Network
ToA	Time of Arrival
TDoA	Time Difference of Arrival
AoA	Angle of Arrival

CHAPTER 1

INTRODUCTION

1.1 Overview

Dynamic position estimation refers to object localization in real time where the object moves. Real time localization depends on accurate signal reception and conversion of received signals to distance estimates. These distance estimates are then used to localize object real time location [1, 2]. In most of the literature, the main focus is on static position estimation and using traditional lateration and fingerprinting based localization techniques. Machine learning approaches have been rarely used using small available data sets. This chapter briefly presents introduction to position estimation, its application in real environment, existing techniques, and possibility of machine learning approaches for position estimation. Moreover, we also summarized problem formulation, objectives, and research questions and finally summary of this chapter is presented in the end.

1.2 Introduction to Real Time Position Estimation

Real time or dynamic position estimation refers to object localization in real environment and on the go. Static position estimation refers to object localization while the object placed at one fixed position [3]. A wireless handheld device is attached to that physical object, or even human body which is going to be tracked. Sensors are installed which are also fixed in a specific indoor environment. Each fixe sensor node, collects RSSI samples from this fixed device and based on these received signal, the object position is estimated using traditional well known lateration, fingerprinting or combination of these two techniques, means hybrid techniques. The problem in these solutions is twofold. I.e offline database development in case of fingerprinting based solutions and signal to distance conversion if lateration approaches are used, which are environment specific and less accurate [4]. Machine learning approaches provides more accurate solutions for real time monitoring and object localization. The reason behind its accuracy is, we train machines with real time all possible scenarios, in our case, machines were trained with 100s of RSSI samples for each and every grid location. All possibilities of signal variations are considered and the solutions is less depended as compared to traditional localization techniques.

1.3 Motivation

Latest developments in smart phone industry, numerous tracking applications have been developed. These developments shows demands of the society, where tracking based solutions are required everywhere from gaming to guidance systems, from child tracking to industrial automation where robots needs accurate navigation, and from health care to defense based requirements. Real time localization or position estimation can be used in outdoor as well as in indoor environment as well. For outdoor environment, Global positioning systems is an idea solution exist which is been developed by the United states for military purpose. However, GPS is currently used for navigation purpose helping people to find the landmark using google map. GPS works with the help of satellites and requires direct line of sight technology, which is most suitable for outdoor application [5]. Due to line of sight issue, GPS signals are unavailable in indoor environment, or in other words we can say, the signal reception is week [6]. Therefore, researchers are working hard to develop an indoor solutions which provide an accurate real time indoor navigation, tracking. The word navigation and tracking or localization are used in different aspects. The word navigation means finding the location of known landmark, i.e in case of outdoor environment, where is Faisal Mosque, which is a popular land mark and geo tagged. In case of indoor environment, the scenario or requirement may be different i.e where is Lab-1, or where is designated office A. On the other side, the word tracking is used in different scenarios. For example tracking the location of a child inside environment, tracking the live movement of a worker working inside industry. So this thesis focuses on tracking, position estimation with more accurate solution and the use of machine learning based real time dynamic object or human localization. The next section presents possible applications of real time dynamic localization [7, 8].

1.4 Applications of Real Time Dynamic Localization

Real time tracking applications exist in daily lives ranging from child tracking to industrial solutions where dynamic location based solutions plays an important role [9, 10].

1.4.1 Real time Dynamic Indoor Navigation and Tracking

Environment can be categorized as two main types, i.e indoor and outdoor. We have already discussed outdoor navigation and GPS is a standard navigation system already developed and installed in most of handheld devices. In case of indoor navigation means developing such solutions which helps the visitors or unaware people locating a known landmark inside buildings like hospitals, airports, universities etc. On the other side, tracking means real time monitoring of someone location inside indoor environment such as employ live position tracking, patient tracking, child monitoring and tracking, and guiding blind people to reach their desired destination.

1.4.2 Localization in Industries

Localization or position estimation systems are installed in small and large industries in the shape of robots. Robots are intelligent machines which are installed where humans are unable to work. These robots have prerequisite information for example mines exploration, tunnels, medical diagnoses, latest surgical instruments where machines are installed through veins and guided from outside and also via computerized solutions to perform the required task.

1.4.3 Localization Systems for Disasters

Incidents happen, floods, earth quacks, etc where the physical infrastructure fall down or damage. Tracking or saving human lives in case of emergency situation is one of the application of indoor localization system. Humans can be tracking, inside the building, where such incidents happen. For this purpose, latest RFID tags, or sensor have already been installed before the incident happen. Device free based localization system can play an important role especially in case of disasters.

1.4.4 Virtual Reality based Localization Systems

The concept of virtual reality based indoor navigation systems can be used in games. Where tracking scenarios are created and the user is guided to find the target. Similarly these systems can help the tourist or indoor environments, which resembles the actual indoor scenarios, and helping the users to reach their destination.

1.4.5 Tourist Guiding Systems

One of the main application of all navigation systems are in tourism industry. Many tourist navigation systems have been developed for both outdoor and indoor scenarios. For example Mina Locater is one most recently and widely used application for Muslims pilgrims during Hajj. GPS is one of most widely used application of navigation system for outdoor environment.

1.4.6 Localization Systems in Hospital

In modern and large hospitals navigations system help the visitors and patients visiting hospitals to find their desired locations with the help of localization system specifically designed for hospital. These systems help both visitors and patients to locate the desired place in most convenient and easiest way.

1.4.7 Employs Tracking Solutions

Employ tracking system is also one of most widely application system offered by many cellular companies to big companies for monitoring and tracking their employs in real time environment. These system can work both in indoor and outdoor environment as well.

1.5 Real Time Dynamic Position Estimation Techniques

Real time dynamic position estimation techniques requires two things, i.e sensing technology, which can be a hardware device with embedded sensors to sense and transmit signal, and a complete system which can be a software to interact with the hardware, measure the signal, processing of signals to distance estimates and once the system get the distance estimates, these distance estimates can be used to predict and estimate object real time dynamic position with respect some coordinates [9]. Regarding sensing technologies, mostly indoor localization systems use Radio Frequency (RF) for static and dynamic position estimation. The advantage of RF is easy availability, and cost. Examples of RF is WIFI, Zigbee nodes, and Bluetooth. Among available technologies, we have used Bluetooth in our research work. Bluetooth is wireless technology embedded in almost every handheld device including smart phones. According to the latest Bluetooth standard released by Bluetooth Special Interest Group (SIP) version 5.1, its range can be extended to 40 to 400m as per latest specification. Bluetooth Beacons, BLE are specially designed for IoTs. Now a days BLE beacons is an ideal technology for indoor localization [11].

Indoor localization system, the second part i.e software to interact with hardware can be categorized as lateration based and fingerprinting based. These two methods are traditional localization techniques. Many researchers modified these techniques and used filtering to remove noise and signal variation together with position estimation for obtaining better accuracy. Examples of such techniques are Kalman Filter, Particle Filters. In lateration based approaches, the received signals are modeled using radio propagation to convert signals to distance estimates with optimized radio propagation constants specific to indoor environment. Once the distance estimates are available, then these distance estimates are used with the help mathematical models to estimate object location with respect to some coordinate systems. Examples of such techniques are Trilateration, Multilateration, MinMax, etc. The accuracy of lateration based techniques depends on accurate distance estimation and modeling of position estimation approaches in order to obtain better accuracy [4, 12]. Detail of these approaches and its working will be discussed in Chapter 2.

Fingerprinting approaches, consist of two steps. In step 1, the a radio offline map is generated by collecting signals of every grid location inside indoor environment, or where the system would be installed and operational. This is a challenging task. Building an offline map and collect fingerprints of every location totally depends on physical infrastructure, any sort of change in indoor setup will disturb the whole radio offline map. In step two, a pattern matching technique is used to match the signal or fingerprints when the object enters the locality, these fingerprints are then matched with the already stored database. The successful and most nearby matches are the estimated location. This approach is more accurate than lateration based approaches [4].

Researchers also introduced hybrid solutions, combinations of good features of lateration and fingerprinting based solution for achieving better accuracy. Examples of hybrid solutions can be found in [7, 13]. Researchers also used Machine learning approaches for static localization. Machine learning is a type of artificial intelligence, which provides systems to learn first and apply these learning to predict the current state with help of previous learned knowledge. Machine learning algorithms are of two types, i.e unsupervised and supervised machine learning techniques. In unsupervised machine learning algorithms, when the data or information is unavailable to train the system, clustering is used in unsupervised learning. In case of supervised machine learning algorithms, the date and learning patterns are available to train the system and predict the current state based on previous knowledge. The training phase train the system with correct and all possible values, which would produce more accurate results. The process of learning is used to improve the system performance [14, 15]. In case real time dynamic object tracking, we will use supervised machine learning techniques such as SVM, KNN, Decision Tree, Linear Discriminant Analysis (LDA) and Naive Bayes. These techniques shows better results in case of real time object tracking.

1.6 Problem Formulation

Real time dynamic localization of an object is a challenging task. Accurate position estimation or localization depends on multiple factors, such as accurate signal reception, use of specific hardware, use of technology, position estimation technique, and integration of all these parameters in one place collectively plays an important role in the design of an accurate real time dynamic indoor localization system. This research work is limited to consider two parameters only, i.e accuracy of signal reception, filtering and use of localization or position estimation approach. However, our main focus is on localization or position estimation technique to address its limitations and design an accurate solution which provide real time dynamic position estimation with optimal cost. Currently most of the available solution used traditional position estimation techniques for static localization i.e Lateration and Fingerprinting based localization techniques. The problem in lateration based position estimation is the conversion of RSSI samples to distance estimates and modeling of radio propagation according to indoor environment. Moreover, accuracy of lateration based position estimation techniques depends on distance estimates and distance estimation is extracted from RSSI samples. In case of Fingerprinting based solutions, generation of radio map is a challenging task and the problem in fingerprinting based approaches is its dependency on radio map which is environment specific. Any sort of change in indoor setup effect radio map and accuracy. To summarize it, following are the main challenges in existing solutions which will be addressed in this thesis [11, 15, 16].

a. Lateration based and fingerprinting based solutions are not accurate for real time dynamic position estimation for moving objects.

b. Environment in depended solutions are not accurate in case of movable objects in real time scenarios.

1.7 Research Objectives

The main objectives are as under.

i. To monitor and estimate real time object position which provides fast and accurate solution.

ii. To investigate Machine learning based solutions especially in case real time dynamic position estimation.

iii. To design an accurate indoor solution for dynamic and real time object localization

1.8 Research Questions

In this thesis we will answer the following questions.

i. Is there any machine learning based solution exist for real time dynamic object position estimation.

ii. Which machine algorithm performs better in case of indoor environment if there is a signal variation?

iii. Is machine learning approaches feasible for indoor environment?

1.9 Thesis Contribution

The main contributions and research findings are summarized as under.

i. To evaluate the performance of our proposed real time dynamic indoor position estimation using machine learning based techniques, requires real time RSSI samples. For this purpose, we performed practical experiments in real environments and collected RSSI samples. Statistical parameters are calculated and based on these samples, we further classified and extended these RSSI samples in to 1000.

ii. Based on real time RSSI samples, we analyzed these measurements in order to know, the variation in RSSI based on time and distance.

iii. We have developed fingerprinting based position estimation model, in which we developed radio map for training and testing of machine learning classifiers.

iv. We have simulated five different types of supervised and semi supervised machine learning techniques and combined these techniques with fingerprinting based position estimation model.

v. We have performed simulation studies of machine learning techniques for localization, and compared their performance with the help of five different trajectories. These trajectories resembles human monitoring in real time, dynamically.

vi. For comparative study, we measured their performance with help of accuracy, standard division and execution time.

1.10 Thesis Organization

This thesis is structured as follows. Chapter 2 discusses relevant literature, and existing techniques, chapter 3 discusses machine learning approaches, classifiers in detail, chapter 4 discusses experimental setup, data collection and classification of RSSI and proposed solution. Chapter 5 presents numerical results and detail discussion, and finally summary of this thesis is presented in chapter 6.

1.11 Summary

This chapter discussed an introduction to real time dynamic position estimation of an object or human body in indoor environment, existing solutions, its advantages and disadvantages, a brief introduction about machine learning approaches. Moreover, we also discussed motivation of our research work, applications of real time localization, objectives and contributions. In coming chapter 2, we will explain in detail existing possible techniques, solutions, available technologies for real time dynamic localization, and the need for adoption of machine learning techniques.

CHAPTER 2

POSITION ESTIMATION TECHNIQUES AND RELATED WORK

2.1 Overview

This chapter presents most relevant real time dynamic localization technologies and techniques for tracking movable object location. Furthermore, this chapter will also elaborate the research gap and challenges specific to machine learning techniques.

2.2 Overview of Real Time Dynamic Localization

Due to the recent advances in wireless technologies and mobile industry, use of social media, location and tracking based solutions is the demand of every one in recent years. In case of outdoor location based tracking or navigation GPS dominates and provides acceptable accuracy. However, in case of Indoor environment, GPS is not an optimal and accurate solution for navigation and tracking based solutions [5, 17]. The reason behind this, its line of sight and connectivity with satellites. Therefore developing indoor localization is a hot research area and researchers are discovering innovative solutions which provide sufficient and reliable accuracy for variety of applications [10]. Indoor localization requires sensing or technological support. Among available technologies, Bluetooth is an ideal low cost and easily available communication technology which is embedded in almost every smart phone, and many handheld devices. Other communication technologies do exist like Ultra-Wide Band (UWB), Radio Frequency Identification (RFID) tags, ZigBee, and WIFI etc. In this thesis we have selected latest Bluetooth as a communication and sensing technology for developing real time dynamic object localization only due to its low cost, easily availability, and extended range [8, 11].

Real time dynamic localization or position estimation refers to tracking object location in real time. Here the word real time means, the can be movable or static. We can track the actual location while the object is moving at slow or moving fast with in an indoor environment. The scope of this thesis is limited to dynamic object localization. Dynamic object localization system consist of two major steps i.e Sensing and localization. Accuracy is the main issue in real time dynamic localization, which depends on many factors [18]. These factors include, accurate signal reception, variation in the signal and its causes, use of localization technique. The term accuracy also depends on case to case, for example for a blind person navigation even 1 meter accuracy is still less and not ideal solution, while for other cases like child tracking inside the building even 10 meter accuracy is still acceptable. For a tourist to roam inside the building even 20 to 30 meter accuracy is still acceptable for find the required landmark. On the other side, for guiding robots in an industry, even 100 % accuracy is required. It means the word accuracy also depends on the deployment domain where the system will be installed. The same situation is also acceptable for outdoor tracking where GPS error of 10 meters is still acceptable. But in some cases moving on roads and finding actual track, even 2 to 5 meter error is also not acceptable and produces confusion for the users. Therefore, we must identify the area or domain where the system would work, and then design indoor localization system specific for that domain. Accuracy also depends on position estimation technique, its effectiveness and flexibility.

The main focus of this thesis is to address the key challenging factors in localization techniques and to address these issues. Considering alternative approaches, their feasibility, advantages and disadvantages of alternative localization techniques. The next subsection discusses localization or position estimation techniques in detail, its pros and cons.

2.3 Existing Localization Techniques

Figure 2.1 depicts a generalized view of localization system for real time object localization. In this figure, there are two main categories, i.e signal measurement, and position computation [10]. So here we will discuss first signal measuring methods, its pros and cons, installation cost and availability. In second phase we will explain existing localization techniques, its advantages, disadvantages, uses and accuracy as well. Following subsection discusses signal measurement methods, their uses, pros and cons.

2.3.1 Time of Arrival or Time of Flight (ToA) and Time Difference of Arrival (TDoA)

Time of Arrival is also referred as time of flight is the measurement of time, when the signal travel or transmitted from one antenna to another where the signal is received. This time is then translated to distance estimates using its velocity,

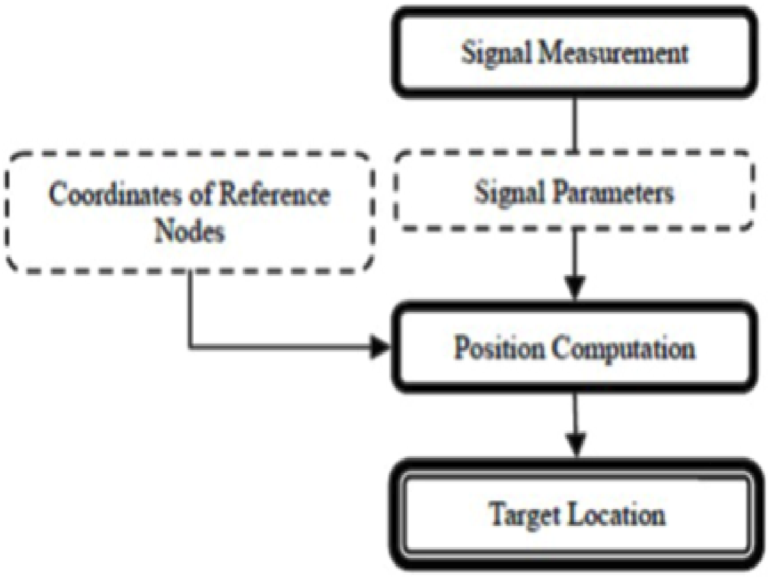


Figure 2.1: Generalized view of Localization System

which is already know when the signal is transmitted from base station to receiver side. On the other side, Time Difference of Arrival is another method for measuring the signal. This method is also used for navigation and tracking purpose where the time synchronization of different signals arrived are measured. In satellite based navigation i.e in GPS, TDoA is used where the moving object receive satellite signals at different coordinates. Earth coordinates are already know, and the satellites signals are transmitted and receiver receives these signals at different coordinates, location of the user or object with the help of TDoA is identified. Figure 2.2 depicts ToA and TDoA respectively [19].

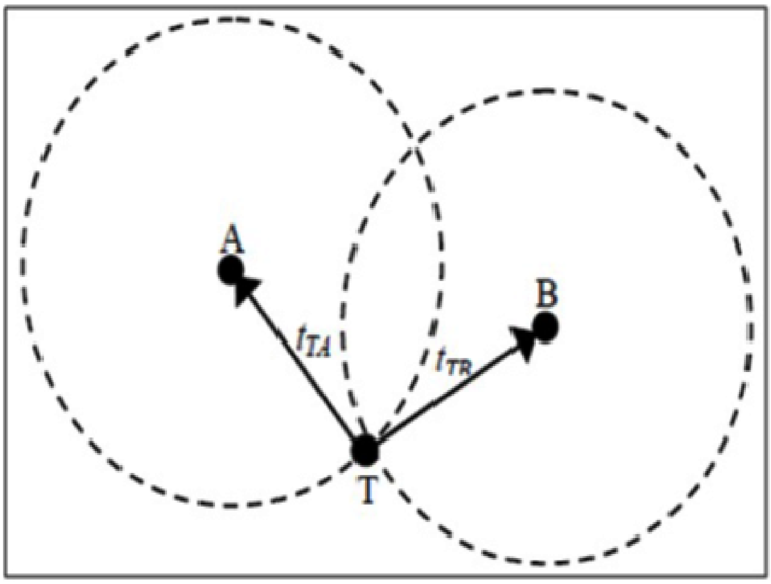


Figure 2.2: Time of Arrival (ToA) and Time Difference of Arrival (TDoA)

2.3.2 Angle of Arrival-AoA

This signal measuring method measures the direction of signal transmitted from antenna. AoA can be used together with TDoA. This method measures the direction and other measures the time of arrival, and the signals arrived after sometime. These methods are used in cellular networks for finding the locations of mobile phones. AoA requires expensive hardware to measure the direction of signal with accuracy. Time synchronization is also important along with installation and maintenance cost. Following Figure 2.3 depicts the concept of AoA. For indoor localization, these costly solutions are not feasible. Following figure showing two antennas which are named as AoA-1 and AoA-2. The central location is another hardware which receive the direction and time of incoming signal [20].

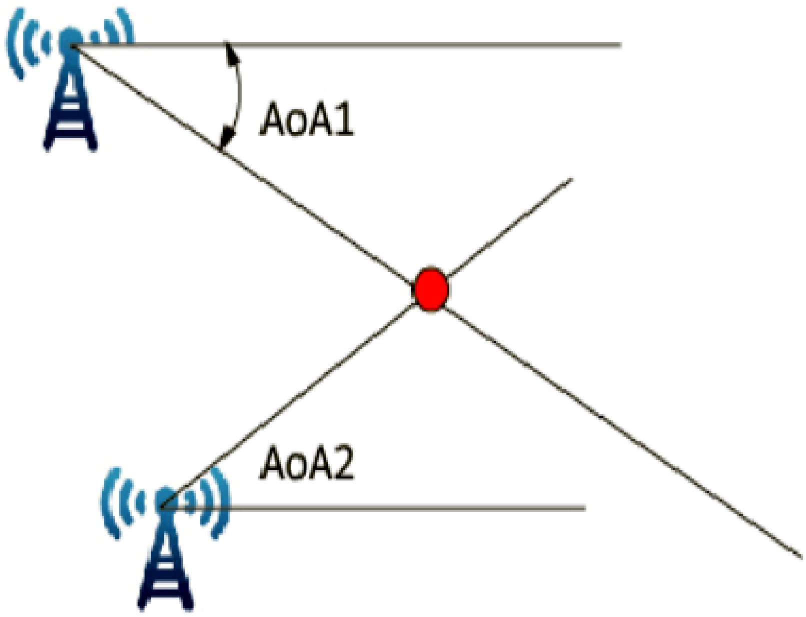


Figure 2.3: Angle of Arrival with Two Antennas

2.3.3 Triangulation

In Trigonometric based solution, position of the object is estimated with the help of triangles. In case of triangulation, at least two fixed nodes are required to measure the angle from these two fixed nodes. Trigonometric based solutions are of two types. i.e Triangulation's and Trilateration. In Triangulation, angles are measured to estimate the object real location, along with distance while in trilateration at least three fixed nodes are required and only distance estimates from atleast three fixed nodes are required to estimate object position [21].

A Lateration based Localization

Lateration is geometric based localization algorithm, which requires distance estimates from at least three or four fixed nodes. These algorithms are also called RSSI based localization techniques, the reason is distance estimates are obtained from RSSI. Figure 2.4 depicts typical Lateration based approach [22, 23]. When the fixed nodes are three, which is named as Trilateration and in case of four fixed nodes, which is called Multilateration. These fixed nodes, measure RSSI of the client node, or we can refer it as mobile node [24]. Mobile node is movable and dynamic while these anchor nodes are fixed. Anchor nodes measure RSSI of target node and these RSSI patterns are then converted to distance estimates using radio propagation model. Radio propagation modeling is itself a challenging task which requires modeling of environmental constants which minimize variation in RSSI. This variation in RSSI directly affect distance estimates. If there is an error in distance estimates, this will result position estimation error. Position estimation in case Lateration based algorithms can be seen, if the point of intersection is not unique. Figure 2.5, is a trilateration approach in which three fixed nodes are shown [3], i.e Beacons, and one target node, i.e unknown node. In case of no error, the circles will intersect at one point, if there is an error these circles will never intersect. These distances are basically extracted from RSSI patterns using radio propagation models [25]. Trilateration approach is a geometric based localization model which estimates object position if three distances are available, on the other side, Multilateration approach requires minimum four anchor nodes, i.e four distance estimates [26, 27].

2.3.4 Localization using Fingerprinting

Fingerprinting is a also considered to be an accurate localization algorithm, which is based on fingerprints of the area where the system would be installed. This localization system can also be referred as pattern matching technique or algorithm. There are mainly two phases involved in fingerprinting. In phase one, an offline fingerprints of the locality where the system would be installed are created for each grid location. Once the fingerprints are created, then these fingerprints are stored in an offline database with their respective coordinates [27]. These coordinates are defined before taking fingerprints. once the offline database is ready, the next step is online localization. Localization in fingerprinting approach means that, collecting real time fingerprints of the object or anything which is going to be tracked are measured by the fixed nodes installed within the premises. These fingerprints are then compared with the offline database. If these fingerprints matches with already stored fingerprints, then the estimated coordinates are displayed. The questions here arises, which algorithm in case of online phase is most suitable. The answer is any pattern matching approach. Following are few examples listed from recent literature [28].

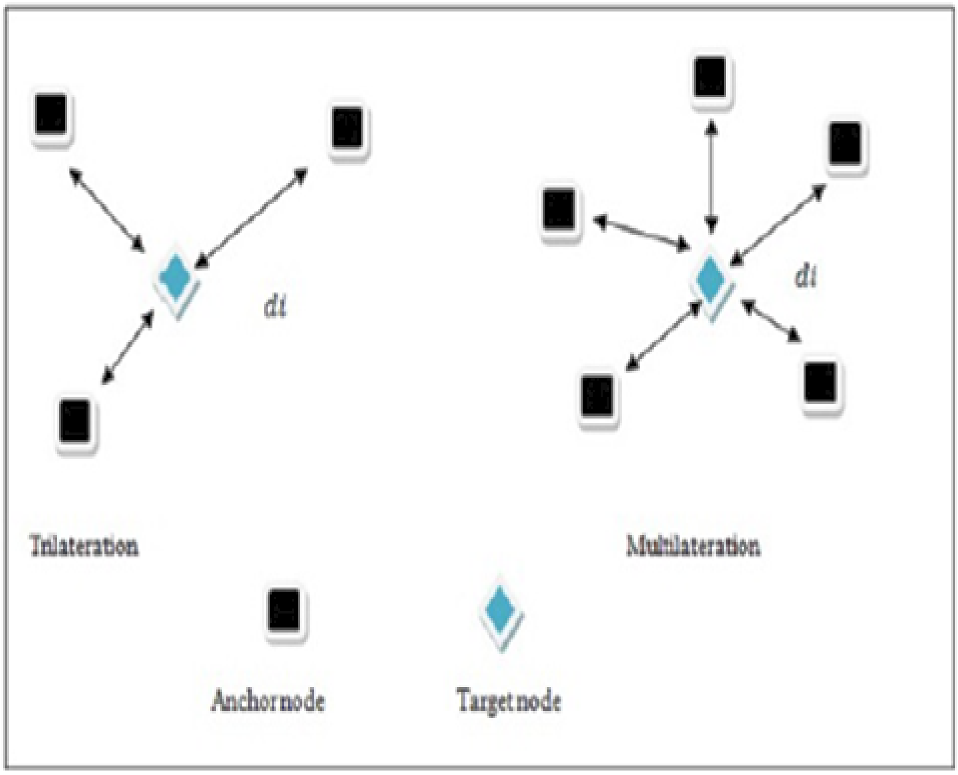


Figure 2.4: Lateration based localization

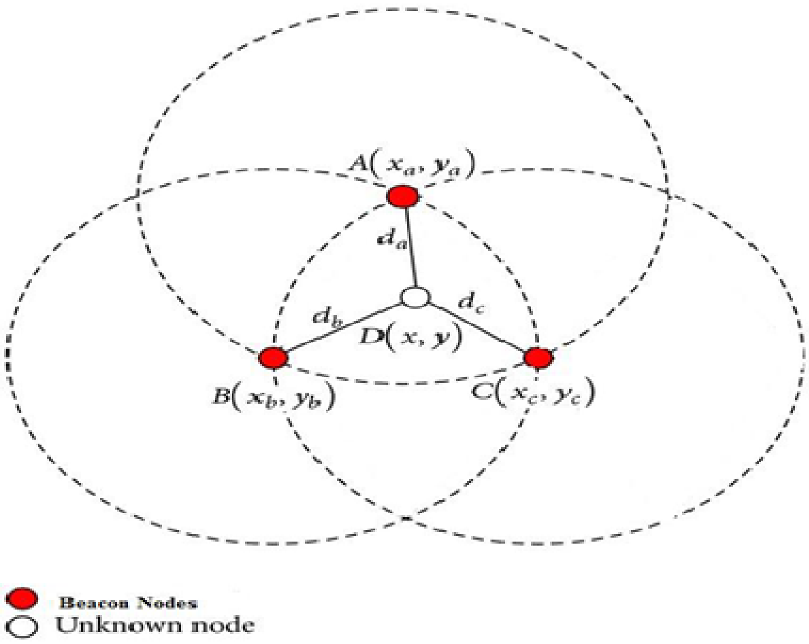


Figure 2.5: Lateration based localization (Trilateration)

a. Probabilistic method

b. K-NN

c. SVM

d. Neural Networks etc.

Any of the above technique can be used in online phase of fingerprinting based localization. Advantages of fingerprinting approaches is its accuracy, low complexity and low deployment cost. There is no extra infrastructure required for fingerprinting based localization. the available technologies such WiFi, Bluetooth can be used to design the system. The main disadvantage of this approach its time consuming offline map generation. Any sort of change in the existing system i.e furniture change, crowd, changing any physical setup inside the building or room where the system is being installed would affect localization accuracy. Figure 2.6 depicts typical system design of fingerprinting approach [23]. Two phases are shown in the figure, i.e offline and online. In offline, the area where the system need to in deployed needs to be divided in equal grids, these grid locations must be named as (x, y) numerically, and required number of fixed nodes i.e wireless beacons will collect RSSI patterns for each grid location. These patterns are then stored in offline database along with their respective coordinates. In online phase as shown in figure, once the object enters that locality, the fixed wireless beacons will detect its RSSI patterns, and these patterns will be given to localization or positioning algorithm for pattern matching and estimation of object real location with respect to a known coordinate.

2.3.5 Hybrid Localization Techniques

Hybrid localization techniques also exists which combine two different positioning algorithms or a combination of two different technologies. The main idea behind hybrid approaches is to improve performance in terms of minimizing mean error, computational complexity and more scalable solutions. There are many hybrid solutions do exist, and we can not summarize all, so few of the hybrid approaches are as under.

In [29] the author designed a hybrid approach using Wireless Local Area Networks (WLAN) in an indoor environment. The author used fingerprinting based approach for modeling RSSI to distance estimates instead of radio propagation modeling. The reason behind this, is it is very difficult to model radio propagation constants for extracting ideal distance estimates. So for this purpose the author claim that, it is an option to find distance estimates from fingerprinting based approach. Here fingerprinting based approach means offline phase. The hybrid approach presented in [3A] consist of three steps, in step one a radio map of the distance to signal map, in second phase, they have calculated the distance between mobile device and access

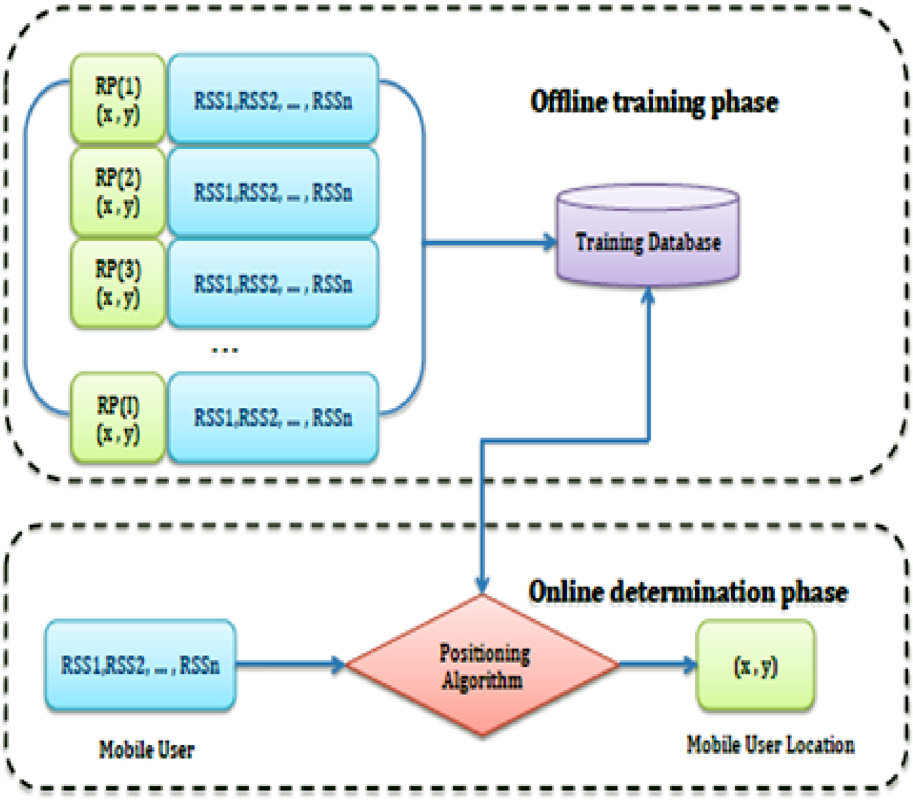


Figure 2.6: Localization using fingerprinting approach

points and in third phase they have used Trilateration approach for position estimation. They reported higher accuracy as compared to radio propagation based trilateration approach for position estimation.

In [30] this approach is also similar to [29]. In this approach the position estimation is carried out in two steps. First of all a fingerprinting based localization technique is used to estimate the mobile user location like in which room the mobile user is located. In the second phase, they also used Trilateration approach for final and exact position estimation. As per their claim, they have considered radio propagation modeling as a challenging task, which results in distance estimation error due to a tough and challenging task of modeling radio propagation constants in order to obtain accurate distance estimates. The accuracy reported by the authors is better than traditional Trilateration based solution based on radio propagation modeling but their accuracy is still lesser than KNN based position computation.

In [23] a modified hybrid approach based [3A, and 3B] is been proposed which combines fingerprinting and Trilateration approach with much higher accuracy. The reason behind this, study is, a gradient filter is used to refine RSSI measurements and to avoid communication holes in case of no signal been received. On the other side Kalman Filter is also used finally at position computation phase which further improved position estimation accuracy. In [] the authors combined two different technologies i.e Wireless Sensor Netxiong2013hybridworks with RFID tags. As per their observations they combine their good features and overcome their deficiencies and achieved reasonable accuracy. They also claim cost effectiveness solution where one of these technologies have already been installed. They concluded that, these two technologies can be combined successfully and obtained desired results. They used Extended Kalman Filter for improving position estimation accuracy together with heterogeneous technologies.

In most recent study [7], the researchers have suggested Bluetooth together with WIFI based indoor navigation systems. In this study they have performed an experiment on the table of size (0.9 x 1.8) meters square. Their reported accuracy in case of hybrid use of technology concludes with 97 % accurate values. They also suggested Machine learning based indoor positioning systems for getting better accuracy in case of dense indoor environment. In [31], the researchers combine Kalman filter based Trilateration with dead reckoning based position estimation. They performed real time experiments to evaluate the performance of their proposed hybrid indoor positioning system. They claim better accuracy in terms of mean square error compare to other techniques. They used Bluetooth Low Energy beacons for their hybrid solution.

To summarize the discussion, we have observed that, hybrid solutions also exist which can improve the position estimation accuracy. The next section discuses localization techniques using machine learning approaches. Machine learning approaches provides a more realistic and accurate solutions compared to traditional position estimation techniques.

2.3.6 Localization using Machine Learning

Machine learning approached have also been used by many researchers for static and dynamic real time localization. In this section we will discuss possibility of using machine learning based localization techniques for position estimation. The word machine learning is basically extracted from the field of artificial intelligence, in which the machines are trained with all possible scenarios, data patterns, possibilities in order to obtain best available results. Machine learning algorithms are not new. Many researchers have used machine learning for variety of applications ranging from image processing, language identification, speech recognition, object and feature identification also for human tracking and navigation purpose as well. In this thesis, our main focus is on machine learning based real time object localization, especially when the object is moving along a specific path anywhere in the premises of the system being deployed. Following are the main machine learning approaches in used typically for localization [11, 16].

A K-Nearest Neighbor (K-NN)

In Machine learning, KNN is one of best conventional supervised machine learning algorithm. This algorithm is widely used in fingerprinting based localization. This technique provides low cost and high accurate solution especially in case fingerprinting based indoor positioning or localization systems. The working mechanism of KNN is, when the target node enters the locality, whether it is static or movable, its RSSI patterns are measured and compared with already stored RSSI patterns. The alphabet K, means, K number of neighboring nodes are calculated. If we fix the value of K=2, Then two neighbors will be identified nearest to target node, if the value is “n” then “n” neighbors will be identified or calculated, then the point of intersection of these nodes, using Euclidean distance formula will be the target node real location. Here it is important to highlight that, KNN is a machine learning algorithm, but in case of fingerprinting approach, this algorithm is widely used in online phase of fingerprinting. The accuracy of this algorithm is better than conventional Lateration based techniques [32].

B Support Vector Machine (SVM)

Support Vector Machine (SVM) is also widely used machine learning algorithm, especially in statistical estimation. SVM is one of the most popular classification algorithm for wide variety of application such as image processing, health related projects, science, engineering, and also geo-location identification and matching. In case of localization this approach is also applicable in fingerprinting based localization systems [33].

C Decision Tree

Decision is also a machine learning and most popular hierarchical approach. The parent nodes, i.e non terminal, or inner nodes called decision nodes, while the non-parent, outer, or terminal nodes represents attributes, or features, classes etc. In simple words, Decision Tree algorithm can be used to estimate object position based. Decision Tree based indoor localization technique can be used in online phase of fingerprinting approach. Other than, Decision Tree, we can also use KNN, SVM or

Naive Bayes for position estimation during online phase of fingerprinting. This online phase of fingerprinting is also called localization phase of position estimation, where the patterns are matched with collected RSSI patterns by the fixed nodes known as anchor nodes [11, 33].

D Naive Bayes

The advantage of using Machine learning algorithms is its simplicity and accuracy. The word Naive Bayes is based on Bayes theorem which is popular for its easy and quick prediction of a class of data from where it belongs. This algorithm is most suitable for categorical input data instead of numerical variables. In case of Bluetooth, the working mechanism of Naive Bayes is the classification and training of RSSI patterns. For Indoor positioning system, it is recommended in case of using Naive Bayes to train the machine with more data in order to obtain better accuracy. Classification is fast and simple are the main advantages of Naive Bayes based machine learning technique for real time dynamic localization [15].

E Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical method for finding linear combination of features from two separate classes. In case of Indoor Localization, this technique is been used by few researchers and obtained good accuracy. This technique i.e LDA performs better if the pattern of data made on independent variables for each data and are continuous quantities. Also it can be used when the group of data is known to us. Although this technique is suitable for linear features, but still it is undiscovered whether it will work on nonlinear data pattern or not. The performance is yet to be discovered [34].

F Random Forest

Random Forest (RF) is also a machine learning algorithm and can be applied for localization of an object inside an indoor environment. This method can be used for classification and regression of data as well. This technique consist of many Decision Trees which are created randomly. There random trees are not linked or associated with each other. Once the RF is created, each individual decision decides on its own, which class of data this sample belongs to. This method is also been used for localization purpose, accuracy is satisfactory in some cases, but this method is not commonly used in case of indoor positioning systems developed so far [14].

The next section presents recently developed indoor localization systems based on machine learning approaches only.

2.4 Related Work

As discussed earlier, machine learning is the branch of artificial intelligence and artificial intelligence is the branch of computer science, in which machines are trained, learned to perform a specific task. In case of Indoor positioning systems, specially tracking object location, when the object is movable or dynamic not static, machine learning algorithm can be combined with fingerprinting based localization systems. Fingerprinting based solution is discussed earlier, which consist of two phases, offline and online. In offline phase, we require a radio map, fingerprints of each grid location. Fingerprints of each location means RSSI patterns of each grid point, coordinate. To use machine learning algorithms, fingerprinting based localization technique is been used by the researchers. So in this section the recent related work will be discussed which consist of fingerprinting based solutions.

RADAR was the first fingerprinting based indoor localization technique developed in 2000 using RSSI and used WLAN for navigation and tracking purpose [32]. In online phase, which is localization step used KNN approach for position estimation. Like in fingerprinting, the first step was radio map generation, i.e empirical measurements during offline phases. The accuracy reported in literature is about 2 to 3 meters using fingerprinting approach and 90 percentile about 5 to 6 meters approximately. This was the first innovation in indoor positioning system using fingerprinting based approach. In [35, 36] The authors used probabilistic approach and a joint clustering method for indoor localization, each candidate grid location was considered a class or we can say a category for minimization of distance estimation error. Real time experiments were performed, and the reported accuracy was 90 % with in the 2.1 meters limit. They also claim that, if the number samples are increased, this also improving the accuracy and mean standard deviation of Gaussian distribution. Also here it is important to mention that, they further enhanced the actual RADAR fingerprinting based solution proposed earlier.

Further more, in [37], the author presented and designed a grid based solution in a small geographical region inside indoor environment. They used Bayesian location monitoring system for tracking in less than 2 meters over 50 % of the proposed time. In [38] developed a localization system in indoor environment based on artificial neural networks. Both an online and offline phases were used, neural networks were trained by obtaining RSSI patterns, and in online phase, once the RSSI patterns are obtained, the trained neural networks estimates the object position. The documented position estimation error was 1.43. In [39], the author used decision tree for position estimation. The decision tree was built in offline phase in fingerprinting based approach and the target location was estimated in a decision tree. This method is considered more powerful as compared to neural networks and KNN. In [40] the author used decision for classification purpose and minimized computational complexity. They reported high accuracy of 2.1 meter compared to 1-NN and C4.5, bagging C4.5 techniques. The drawback in this technique is, they require RSSI patterns from each access point and also from the compass sensor. Due to this drawback, every sensor might not be able to extract RSSI patterns for training purpose.

In [33, 41] the author used various machine learning approaches for indoor position estimation. They performed real time experiments and used Wireless Local Area Networks (WLAN) as a technology for localization. They have concluded based on their comparative analysis that, Support Vector Machine based localization performs better than other statistical methods. They have also recommended SVM for localization. In [42] the researchers used kernel based learning approach and used Support Vector based localization in Wireless Adhoc Networks. In their proposed approach, it was assumed that, all the sensor nodes can get the RSSI from other neighboring nodes whose actual positions were fixed in advance. These fixed nodes are also called Access Points. They divided the sensor nodes in fixed categories and the data collected at the access points were used for training machines during offline phase of fingerprinting based position estimation technique. In the online phase the classification model of SVM is then used for object localization. Their reported accuracy was satisfactory.

In [43,44], the author claim an improved localization accuracy by dividing the actual space of the mobile node which is going to be tracked based on RSSI patterns or features. For each region and their respective RSSI patterns, a separate SVM model was trained in order to minimize the variation in RSSI. But the fact is, it is not confirm that, dividing the large region in to small one may minimize variation in RSSI. In [45], The researchers used KNN approach for position estimation based on Spearman distance for improving distance estimation and localization accuracy. They designed and developed this method in order to minimize the effect of noise due to multipath signal attenuation and other environmental effects. Their experimental results conclude mean error up to 2.7 meters. In [46], the researchers recommended KNN and SVM based machine learning approaches for indoor positioning systems. Recent studies are exploring machine learning techniques based on deep learning approach.

Based on the above literature studies, the researchers have recommended Machine learning based alternative approaches for real time dynamic object localization. These learning based statistical approaches are best suitable for dense indoor scenarios where traditional approaches are difficult to sustain due to frequent updating and changes in the indoor environment. The next section, discusses performance metrics to evaluate the design process.

2.4.1 Performance Metrics

To measure the performance of real time dynamic indoor localization system, only accuracy is not enough. Researchers have identified various performance metrics in order to bench mark standard localization systems such as accuracy, complexity, cost, precision, and scalability etc. The following subsection further elaborate each metric briefly [10].

A Accuracy

Accuracy or position estimation error is an important parameter considered to measure the performance a real time dynamic localization technique. Mean error is difference between actual coordinate and estimated coordinate with reference to some known location coordinates. The unit defined for measuring mean error is (meter). Another most important issue is the satisfaction level and deployment domain. i.e if the system is designed for industrial automation where the robots are installed and requires guidance to perform the desired work, the accuracy or mean error should be 100 %., even error in millimeters is not acceptable. On the other side, if guiding a visitor inside an indoor environment, then even mean error up to 10 meter is also acceptable for the visitor.

B Precision

Accuracy or we can say, mean error only consider distance estimation error, while precision also consider how consistent the system perform or estimate object location in iteration. Few researchers consider the performance metric precision as standard deviation which is another most important performance metric. Researchers also practiced cumulative probability function (CDF) of the distance estimation error as precision. In our research work, we used standard diviation as a performance metric together with mean error analysis and compare the effectiveness of localization techniques.

C Complexity

Complexity of an indoor localization system can be measured or attributed collectively using hardware solution or use of specific hardware, its processing time, its response time, etc, and software or algorithmic complexity and also other operational factors as well. In realistic scenario, if the complexity of the software is less, but the hardware response time or processing time is higher then, we cannot guarantee less complex design. On the, other side if the hardware processing is fast, response time is excellent but the algorithmic complexity is high still it is an issue. Similarly other factors relating to the device used to transmit the signal, here the device refers to the object or human body with whom the sensor node is attached for tracking purpose response slowly, then overall complexity may be affected. So to summarize it, here we only consider execution time of our algorithm for comparative analysis.

D Robustness

The system robustness is measurement, how the system will work if there is a blockage in the communication hardware, or even the available signal consist of outliers due to temporary issue, how well the system will work. Measuring robustness depends on several factors such as hardware failure, communication holes, means no signal at all, failure of the system, so a system should be robust in nature for best performance. In our thesis, it is out of our scope to measure it at this stage. This parameter can be measured after successful installation and deployment of the system.

E Scalability

Scalability is a parameter which measure how much can a system can grow both geographically and with density. By geographical extension of the system, whether the system will work, its distance or geographic region is extended or not. On the other side, scalability via density means, if the number of systems or tracking devices grow with in the same region, will the system perform would degrade or not. In case of indoor positioning system, we must ensure considering future scalability of the system in the designing phase. Considering extension in dimension will also affect overall performance, we must ensure extension in the design process so that the system work in two dimensional and three dimensional system.

F Cost

Cost is an important performance metric especially in case of real time dynamic localization system. Cost does not mean only hardware cost, it also mean, time, required space, and energy of the hardware used in the design. As discussed earlier, in this thesis, we are using Bluetooth for real time indoor localization system, which is one of most low cost solution among available options. So we have considered Accuracy, cost, precision, complexity and scalability in our design.

2.5 Summary

This chapter presented the basic idea of real time dynamic localization in indoor scenario. Also we have discussed existing localization techniques its pros and cons and improvement in the existing techniques have also been discussed. We also highlighted possibility machine learning approaches and recent research work carried out using machine learning for tracking real time dynamic localization of a movable object in indoor environment. In the next chapter we will discuss, our experimental setup, how the data is been collected and arranged.

CHAPTER 3

EXPERIMENTS AND DATA COLLECTION

Chapter 2 presented existing real time dynamic localization techniques, its advantages, disadvantages and possible improvements. We have also discussed machine learning techniques and presents recent related work focused on machine learning techniques. This chapter will discuss the steps to design a real time dynamic localization system. As discussed earlier, the system consist of two steps, i.e Signal measurement and localization. Following section discusses an introduction of the proposed system and will focus on sensing technology, experimental setup, data collection, and will discuss relation between distance and RSSI.

3.1 Overview

Real time dynamic indoor localization system track object location dynamically, when the object is moving inside the defined premises. Compare to outdoor environment, indoor localization system faces challenges from inside infrastructure, such as walls, buildings, furniture, light, room temperature, presence of human bodies, existing signals [18]. All these parameters reflects signals and causes signal attenuation, degradation which produces variation in signal. The signal transmitted from the antenna is not the same as the signal receive at the distance device. This causes distance estimation error and ultimately produces position estimation error. Indoor localization systems requires higher accuracy as compared to outdoor installations and navigation systems. There are some advantages over outdoor installation which facilitate indoor deployments. These plus points are less deployment area, relatively less dynamic because no frequent changes, existing buildings, further support less temperature, entries and minimum exits. Installation cost of the indoor navigation, tracking system is also comparatively lower than outdoor one. As discussed in chapter 2, among available technologies Bluetooth provides more low cost, less complicated, scalable, solution as compared to other available technologies [3]. Therefore, we have chosen Bluetooth for the design process.

Bluetooth is a wireless communication standard, and preferred technology for indoor localization. The parameter used in most of the existing localization system is RSSI, which is the power in the received signal. Its unit is dBm. Before 2010, and the emergence of BLE beacons, version 4.0, Blueooth based indoor localization system were not ideal and preferred. But after 2010, Bluetooth introduced low energy, long range solutions. Moreover, Android offers BLE support since Bluetooth 4.3, and also latest support in version 5.0 as well for discovering and extracting RSSI of nearby Bluetooth devices. These latest features makes Bluetooth an ideal solution for designing low cost indoor localization system [11]. The next section discusses the sensing technologies in more detail.

3.2 Localization Technologies

In order to investigate, available localization technologies and infrastructure support, we must consider performance metrics in mind before the design process. We have discussed performance parameters identified from the literature studies. i.e accuracy, complexity, cost, precision, and scalability etc. Selecting communication technology directly and indirectly affect accuracy, complexity, communication cost, precision and scalability. In Reference [9] the author identified building dependent and building independent. As this thesis focuses on indoor positioning, so we will consider building dependent available communication technologies and will discuss briefly considering the above mentioned performance metrics. Figure 3.1 depicts available communication technologies [9].

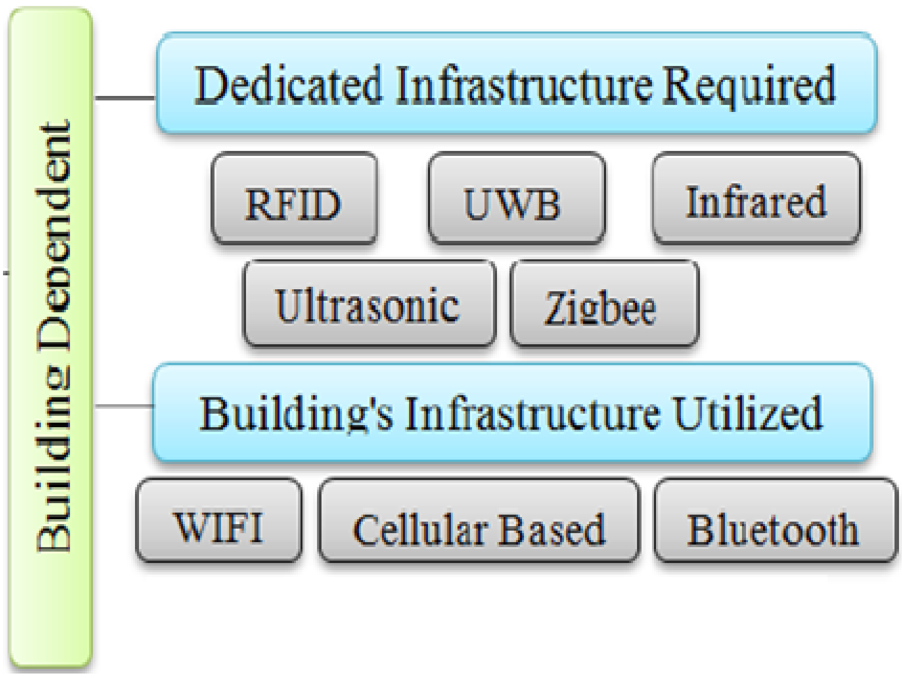


Figure 3.1: Indoor Localization Technologies

3.2.1 Radio Frequency Identification (RFID)

RFID tags is an available technology for variety of applications such as garments industry, sports, health care, defense and can also be used for location identification, tracking, navigation purpose as well. RFID tags uses electromagnetic signals to identify and tags attached to objects, cloths, sport items etc. These tags are of two types, i.e passive tags and active tags. Passive tags taken energy from nearby tags with capacity to read RFID embedded information using radio waves. On the other sides active tags have their battery and energy and there is no boundary limits. Both types of tags have their own coded information and can be attached to target node for location identification, tracking live information wherever the object move if the RFID tag is attached to the object. Many indoor localization techniques have been developed with the help of RFID. Now a days, even RSSI based ranging techniques can be combined with RFID based solutions for better accuracy and performance [24].

3.2.2 Ultra Wideband (UWB)

UWB is a radio technology designed for low energy transmission for short range communication. This technology is effect where the volume of data for transmission is large in small geographical range. Different researchers have used UWB for specific tracking solutions and attained desired results and accuracy. The signaling method used for distance estimation and localization is Time Difference of Arrival (TDOA). On the other side, the disadvantage of using UWB technology for indoor positioning systems is its range, this is not feasible for scalable solutions where geographical extension is required.

3.2.3 Infrared (IR)

Infrared is a short range wireless communication with the help of invisible light based spectrum. This technology was employed in old mobile phones for device to device short range data transfer. There are two main types of IR based communication, i.e called direct IR, which is for very short range direct point to point visible communication. The second type is diffuse IR range is much greater than direct IR, and not point to point. Its range is from 9 to 12 meters maximum. This technology also allows many devices to interact with the source device for communication. In case of indoor localization, Angle of Arrival (AOA) technique can be used with IR [47].

3.2.4 Ultrasonic

Ultrasonic is a sound wave above the human ability to hear. It is the same as other sound we are able to hear but its frequency is higher and humans are unable to hear it. Its physical properties are the same except higher frequency. The range of ultrasound waves are also short and it's not feasible solution for long range indoor localization systems. In case of indoor localization, Time of Arrival TOA, technique can be used in case ultrasound for localization. This is a costly solution in terms of expensive hardware installation and measuring synchronization of the received time difference of arrival signals. Some of the existing solutions have used ultrasound waves for indoor localization as well [48].

3.2.5 Zigbee

Zigbee is an IEEE 802.15.4 standard for short range communication. Zigbee is a standard adopted for short range wireless local networks, office automation, personal short range local networks. It is an ideal solution for short range indoor localization, due to its low energy consumption and operational cost. Zigbee is not commonly available communication standard, specific hardware are required for indoor location based solutions. Many researchers have used Zigbee for indoor positioning systems. The disadvantage of Zigbee based solution is its scalability and somehow cost. It is ideal for less geographical regions. In case of Indoor positioning systems, these devices are of two types full functional and reduced functional devices. Both types of devices can be used for localization. The parameter used in Zigbee nodes are the same as Bluetooth and WIFI i.e RSSI, which is used for ranging distance and positioning purpose. Accuracy of Zigbee nodes in case of indoor localization is also satisfactory and many researchers have used Zigbee for less geographical domains [2, 49].

3.2.6 Wireless Local Area Network (WLAN)

Wireless Local Area Networks and Bluetooth are the most widely communication technologies for navigation and indoor position estimation. Both technologies are wireless and free of cost. Free of cost means, we can use these technologies as a byproduct together with the main task. WLAN is an IEEE standard 802.11 (named as WIFI) available in almost every mobile phone and handheld device. Existing network for local connectivity is also based on WLAN and this technology can be used for indoor based localization, tracking and navigation purpose. WLAN can be generally classified in to two main categories i.e centralized and distributed. In centralized systems, the central device or machine will collect RSSI patterns of the local clients and then these clients location i.e target node location will be estimated from the centralized system. In distributed environment, the client that is the target node or user will locate himself by collecting RSSI samples from local access points. However this concept will lead further issues of self-computation power, energy requirements etc. on the other hand, the centralized system have two major issues i.e user privacy and network burden. As the user location is computed at the centralized location, so privacy is a major concern and also if many user at the same time, request for user location, the computational burden will increase. Another major issue of WLAN is its accuracy which is documented in literature is 3 to 30 meters [10].

3.2.7 Cellular Based

Cellular based indoor localization system also exist. Cellular based localization means Global System for Mobile Communication (GSM) and is been adopted in many countries. The advantage of GSM based indoor localization provides scalable solutions with extended geographic regions but with less accuracy as compared to WLAN and Bluetooth. Commercial based solutions also developed using GSM, but an expensive and less accurate as compared to other available solutions in the same cost. GSM based solutions also used RSS ranging technique for estimating object real and dynamic location [10].

3.2.8 Bluetooth

Bluetooth is an IEEE 802.15.1 standard for short range communication. Latest Bluetooth provides extended range wireless communication and its documented range in vesion 5.1 is 40 to 400 meters as per Bluetooth latest specification. Bluetooth is a low, cost highly energy efficient, easily available, and most widely used solution for managing indoor localization technology. Almost every mobile phone, latest digital watches, handheld devices, and almost every digital device installed in homes, industries, educational institutes, security devices etc. Moreover, Bluetooth Low Energy Beacons since version 4.0 providing energy efficient, solutions for Internet of Things (IoTs). Bluetooth is popular for its easily availability, cost, and energy efficiency. None of the other available technology provides all the features in one place. Even Bluetooth nodes with required functionality is also available and can be used for indoor navigation. Most popular solutions early developed are TOPAZ, Tadlyz provides Bluetooth based solutions for indoor navigation. Recently most of the researcher preferred Bluetooth over other technologies and obtained sufficient and acceptable accuracy. In Bluetooth RSSI is a parameter preferred for distance estimates and also indoor localization. RSSI is the power in the received signal, which can be obtained for measuring and estimating the distance between two Bluetooth enabled devices. To estimate the distance, Radio Propagation models can be used to estimate the distance, later on these distance estimates can be used in Geometric based indoor localization techniques for actual position estimation [19].

The next section presents an experimental setup and data collection.

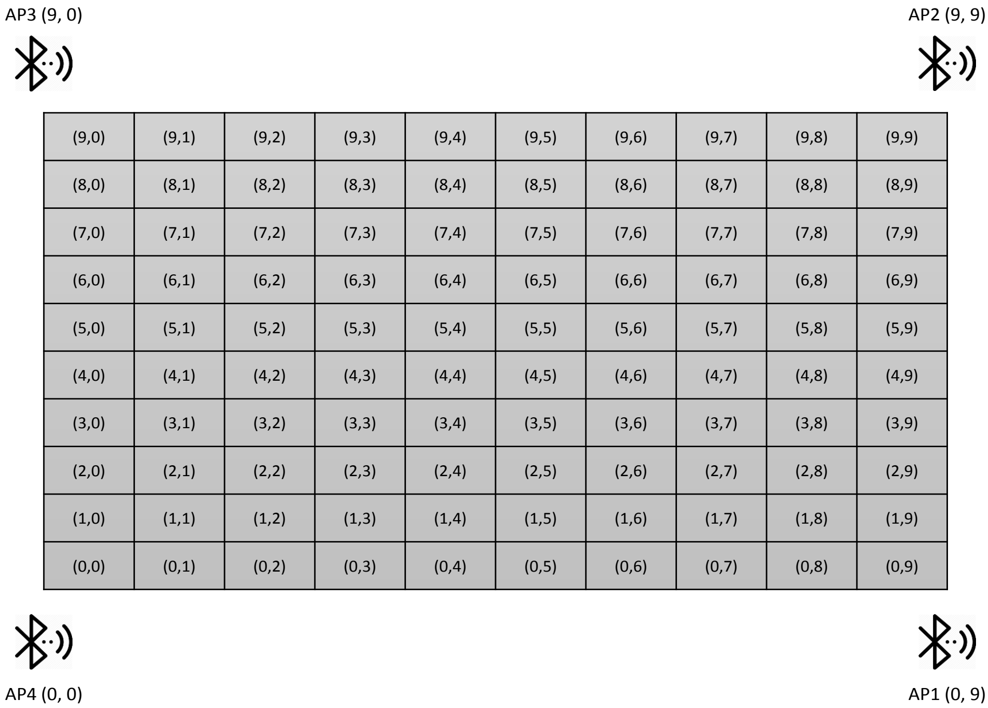


Figure 3.2: Experimental Setup

3.3 Experimental Setup

Real time position estimation and fingerprinting approach requires radio map generation which requires data extraction and mapping. For this purpose, real time experiments were conducted in an indoor setup. Figure 3.2 depicts graphical representation of experimental setup and data collection. This experimental setup is motivated from literature review, in which the experiments are conducted in an indoor fixed geographical region. In this thesis we have selected our computer lab, and inside our lab, we have selected (10 x 10) meters square shape region as depicted in Figure 3.2. Furthermore, we have placed four Bluetooth enabled smart phones with Bluetooth 4.0 and higher specification BLE modular support. We have divided this 10 x 10 region in 100 equal size grids. Each grid location is 1 meters square. Total number of grids where the data have been collected are 100. At each grid location we placed a mobile device with Bluetooth BLE support and took 10 RSSI samples. We have repeated this experiments for all grids and collected 1000 RSSI samples. Then the average of these 10 samples is calculated and stored against each grid location along with its respective (x, y) 2 dimensional coordinate location. Table 3.1 contains the RSSI values of mobile device from each AP at certain cell or position. Experimental setup consists of 100 cells and at each cell values from all four access points are taken. Table 3.1 only contains the value of first 50 cells along with their position w.r.t. X and Y. Similarly, Table 3.2 represents trajectory value of user along experimental setup.

Table 3.1: RSSI values obtained from AP

Cell	X	Y	AP1	AP2	AP3	AP3
1	0	0	-85	-91	-87	-16
2	0	1	-83	-87	-85	-64
3	0	2	-73	-88	-85	-69
4	0	3	-76	-89	-87	-74
5	0	4	-81	-86	-86	-75
6	0	5	-75	-86	-88	-79
7	0	6	-73	-85	-90	-76
8	0	7	-66	-87	-90	-73
9	0	8	-64	-85	-87	-83
10	0	9	-14	-84	-87	-85
11	1	0	-85	-87	-83	-64
12	1	1	-83	-88	-83	-64
13	1	2	-74	-87	-84	-68
14	1	3	-77	-87	-84	-75
15	1	4	-79	-85	-86	-75
16	1	5	-77	-87	-85	-79
17	1	6	-73	-82	-86	-75
18	1	7	-67	-83	-88	-73
19	1	8	-64	-83	-88	-84
20	1	9	-66	-82	-87	-85
21	2	0	-85	-88	-73	-67
22	2	1	-82	-86	-75	-69
23	2	2	-73	-86	-73	-74
24	2	3	-75	-84	-72	-76
25	2	4	-79	-83	-84	-75
26	2	5	-74	-82	-84	-79
27	2	6	-77	-73	-85	-76
28	2	7	-74	-72	-86	-73
29	2	8	-66	-73	-87	-84
30	2	9	-68	-74	-89	-86
31	3	0	-85	-88	-76	-74
32	3	1	-83	-86	-78	-75
33	3	2	-73	-86	-76	-75
34	3	3	-75	-83	-74	-76
35	3	4	-76	-85	-73	-81
36	3	5	-81	-74	-85	-76
37	3	6	-77	-77	-84	-77
38	3	7	-74	-78	-87	-73
39	3	8	-75	-77	-87	-84
40	3	9	-77	-78	-90	-85
41	4	0	-87	-87	-80	-76
42	4	1	-87	-85	-78	-74
43	4	2	-84	-84	-81	-77
44	4	3	-74	-84	-77	-80
45	4	4	-78	-73	-77	-79
46	4	5	-77	-77	-74	-76
47	4	6	-81	-80	-84	-76
48	4	7	-76	-80	-85	-85
49	4	8	-76	-81	-86	-87
50	4	9	-78	-80	-87	-87
...	...	...	...	...	...	...
...	...	...	...	...	...	...

The trajectory and position is defined using the cell distribution w.r.t. X and Y.

Table 3.2: Sample trajectory obtained with user movement

Trajectory-3		**Trajectory-4**	
X	**Y**	**X**	**Y**
0	0	5	0
0	1	5	1
0	2	5	2
0	3	5	3
0	4	5	4
0	5	5	5
5	1	4	5
5	2	3	5
5	3	2	5
5	4	1	5
5	5	0	5
5	6	0	6
5	7	0	7
5	8	0	8
5	9	0	9

3.4 Variation in RSSI and its effect on Distance

RSSI is a parameter used in Bluetooth Low Energy modules is an estimate of the power in the received signal. Its unit is dBm. Variations occur in RSSI, as the signal transmitted is the not the same as the signal received. This variation in signal is due to multipath fading effects, noise, attenuation, temperature, presence of existing signals, humans and other physical objects in the indoor environment [50, 51]. As per our experimental observation, we have noticed almost 10 dBm variation in the signal. Figure 3.3 and Figure 3.4 clearly indicates variations in RSSI, when the two devices are at distance 1, and 2 meters and also 9 and 10 meters respectively. As discussed in experimental setup, the room size where we conducted experiments for data collection was 10 x 10 meters. So we took RSSI samples when the distance between two devices are 1, 2, and 9 and 10 meters in order to show, the minimum and maximum RSSI values at various distances. The maximum value of RSSI when the distance is 1 meter is -56 dBm it's the maximum value not the average and the minimum value is -90 dBm when the two devices are 10 meters apart from each other's.

3.5 Summary

This chapter presented the introduction to our proposed system for real time dynamic position estimation technique. In this regard, existing technologies and its advantages and disadvantages are summarized and then we presented our experimental

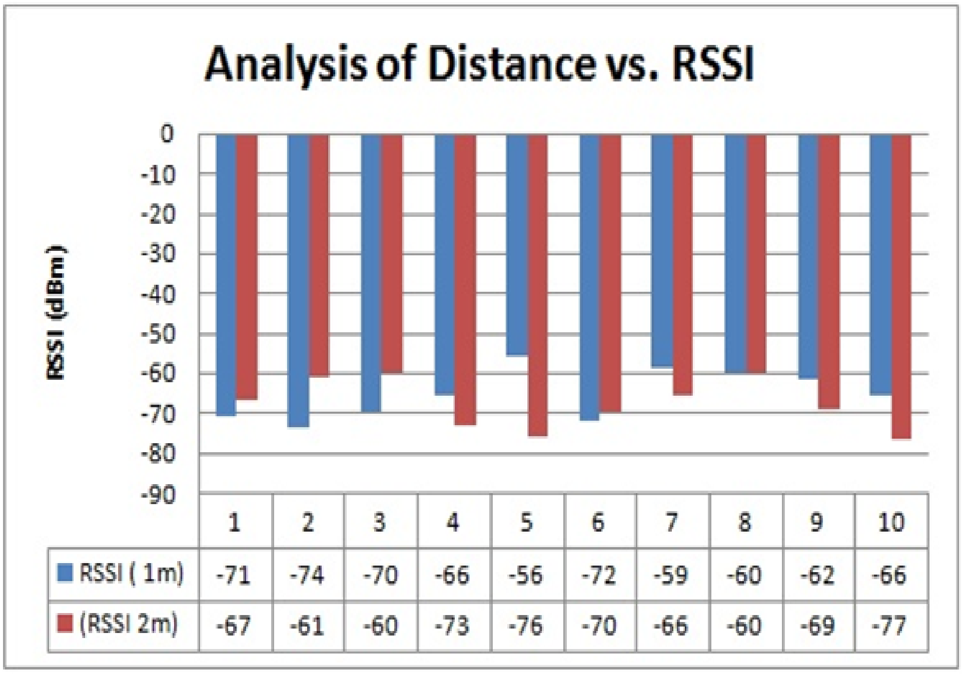


Figure 3.3: Variations in RSSI at distance 1 and 2 meters

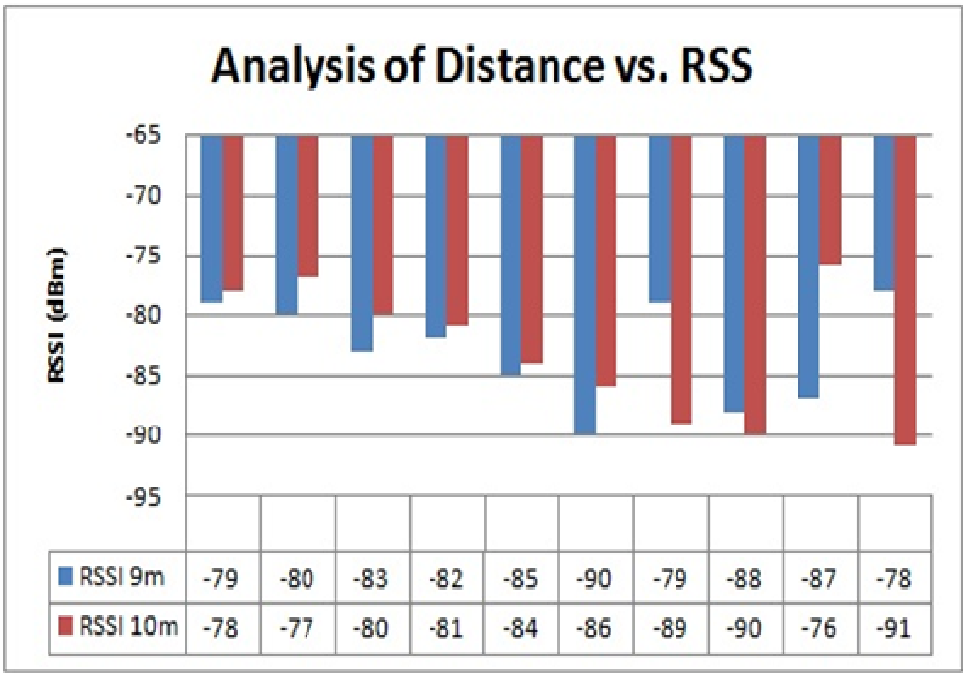


Figure 3.4: Variations in RSSI at distance 9 and 10 meters

setup, data collection, sampling of RSSI, testing scenarios depicting real time dynamic evaluation of our proposed system, then we discussed variations in RSSI and its relation with distance. The next chapter will discuss the design process of our proposed solution.

CHAPTER 4

PROPOSED REAL TIME DYNAMIC INDOOR POSITION ESTIMATION USING MACHINE LEARNING APPROACH

4.1 Overview

Chapter 3 discussed Data collection and experimental setup as well as variations in RSSI and its effect on distance estimation process. Distance estimation error produces position estimation error. In order to improve position estimation error, we are proposing fingerprinting and machine learning based position estimation techniques. In this chapter we will discuss first a fingerprinting approach and in phase two of fingerprinting based approach, we will discuss five different types of machine learning approaches for final position estimation specifically for real time dynamic position estimation using Bluetooth Low Energy modules.

4.2 Proposed System Design

Our proposed system architecture is basically an extension of fingerprinting based indoor position estimation. As discussed in detail, localization techniques are broadly classified as fingerprinting and lateration based techniques. In fingerprinting based localization techniques, the position estimation process is carried out in two steps. In step 1, an offline radio map is generated using real time experiments and collecting RSSI samples of the locality where the system would be deployed, and in second step, position estimation is carried out. In step, 2 there are different approaches, the most widely used localization process in step 2 is KNN. Researchers have investigated other techniques as well but mostly KNN is used in most of the literature studied carried out recently [11]. In this thesis, we have proposed Fingerprinting based real time dynamic position estimation using machine learning based approaches. For this purpose five different algorithms are identified and simulated. These techniques are Naïve Bayes, KNN, Linear Discrete Analysis, Support Vector machine and Decision Tree. System architecture of our proposed diagram is depicted in Figure 4.1. In this architecture, two stages are shown, offline and online. In offline, a radio map is visible where the actual location 10 x 10 is been divided in to 100 equal size grids.

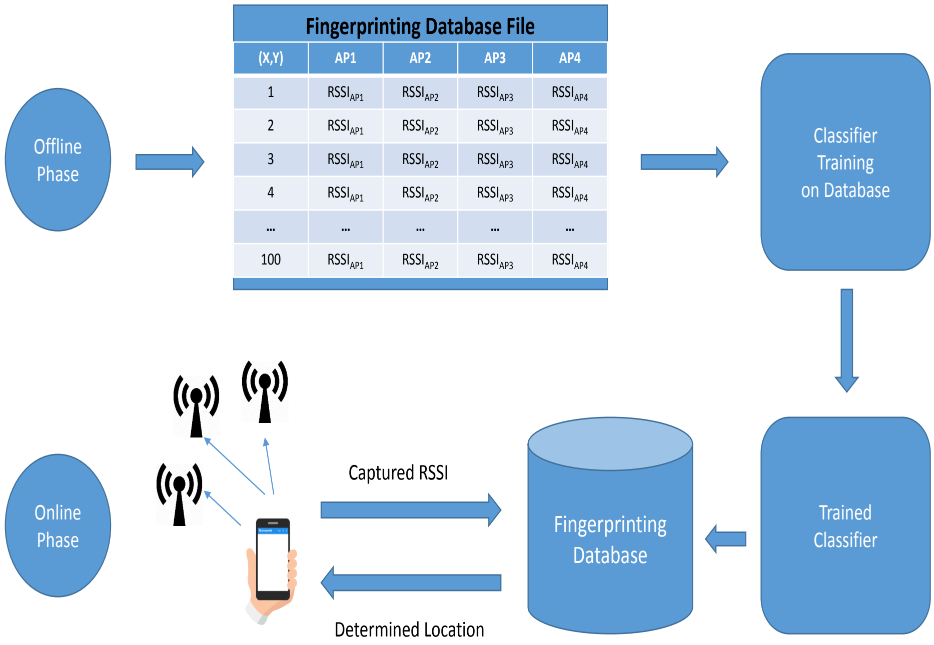


Figure 4.1: Proposed System Architecture

Their location coordinates are shown in (x,y) 2 dimensional coordinate system. The next coloum shows RSSI values of each access point. This is process is called offline mapping in which a radio fingerprints are generated and stored. Once a radio map or fingerprints are generated, then we used machine learning techniques. These RSSI values are then modeled and expended to 1000 RSSI values. This expansion is based on our standard deviation. In our experiments we observed the maximum 10 dBm variation in RSSI for each grid location. Further data was classified in two parts. 90 % of the RSSI patterns were used for training and 10 % of the data was used for test.

In step 2, when the user enters the locality, the sensors will detect user RSSI patterns and will be handed over to machine learning algorithms for matching. These classifiers will predict user real time dynamic location and will update user location, wherever the object move inside this region where the fingerprints were collected. Figure 4.2 further elaborate the complete process of our proposed real time dynamic position estimation using machine learning approach using step wise flow chart. For performance evaluation we have chosen accuracy and standard deviation. Standard deviation is somehow named as precision. As per literature survey, machine learning techniques performs better in terms of accuracy and precision. Machine learning also provides an easy and scalable solution. However, here the word scalability is applicable when the geographic region fingerprints are available. Extension of radio

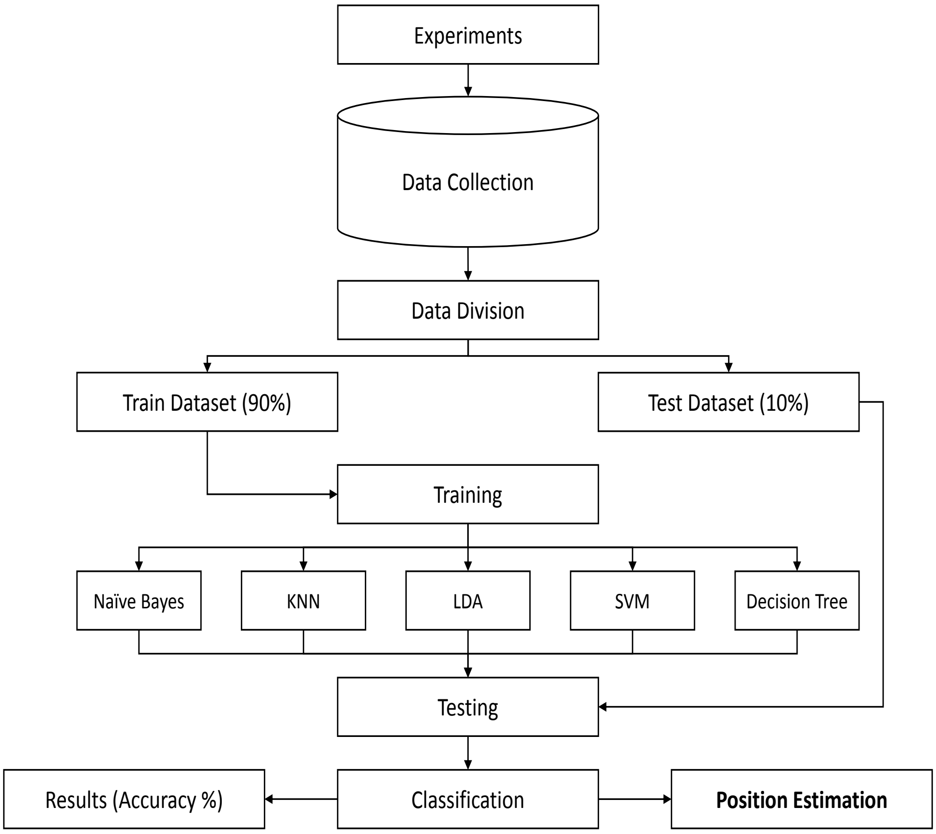


Figure 4.2: Flow Chart of The Proposed System Design

map would be required in order to extend the system. Following sections presents proposed machine learning techniques and its mathematical models in a more detail.

4.3 Naive Bayes

Naive Bayes classifier is a machine learning classifier based on application of Bayes theorem with assumption of independence between features [52]. It is a actually a probability based classifier and has been extensively studied and used for many classification problems of pattern recognition. This classifier is highly scalable In learning problem, Naive Bayes classifier requires number of linear parameters for variable such as features. By using Maximum Likelihood training the execution time can be reduced to linear time rather than iterative approximation which is an expensive with respect to time consumption [53].

Naive Bayes is a simple technique that assign a class label to a variable or set of features which represents a problem case. Some finite set can be used to draw the class labels. Several algorithm can be used to train these type of classifiers. The main aspect of this classifier is that it assumes each specific feature value is independent of any other feature value, given the class variable [54]. In supervised learning environment, naive Bayes classifiers can be effectively trained for probabilistic models. Parameter estimation for naive Bayes models can be achieved in many applications by using Maximum Likelihood. In many real-world cases where the problem is quite complicated, naive bayes classifiers performed reasonable better despite having simplified assumptions and apparently naive design.

As described earlier, naive Bayes is probabilistic model, therefore, if given a problem of a classification, it can be depicted using a vector $K = k_1, k_2, \ldots, k_n$ denoting n (independent) features or variables, assignment can be as follows:

$$p(C_A|k_1, k_2, \ldots, k_n) \tag{4.1}$$

for each A possible outcomes or classes C_A. Using Bayes Theorem, the above mentioned problem is reformulated, therefore, conditional probability can be dissolved in Equation 4.2 as:

$$p(C_A|K) = \frac{p(C_A)p(K|C_A)}{p(K)} \tag{4.2}$$

In simple English, the Equation 4.2 can be written as:

$$posterior = \frac{prior \times likelihood}{evidence} \tag{4.3}$$

If the naive conditional independence plays a part then, it is assumed that there exist mutually independent features in K. Because of this:

$$p(k_i|k_{i+1}, \ldots, k_n, C_A) = p(k_i|C_A) \tag{4.4}$$

Thus it can be represented by Equation 4.5 as:

$$p(C_A|K) = p(C_A)\Pi_{i=1}^{n} \quad p(k_i|C_A) \tag{4.5}$$

Under the assumption of independence, the conditional probabilistic model for class C can be written by Equation 4.6 as:

$$p(C_A|K) = \frac{1}{X}p(C_A) \quad \Pi_{i=1}^{n} \quad p(k_i|C_A) \tag{4.6}$$

where the evidence $X = p(K) = \sum_A \quad p(C_A) \quad p(K|C_A)$ is scaling factor. It is dependent only on $k_1, k_2, \ldots, k_n$ which remains constant if features are already known.

Equation 4.6 has derived a feature model which is independent and is known as naive Bayes probability model. By combining this model with decision rule, a naive Bayesian classifier can be computed. The most commonly used rule is called as *maximum a posteriori* or MAP decision rule which picks the most probable hypothesis for classification. So, a naive Bayes classifier, that assigns a class label $\hat{y}$ to class C_A for some A is given in Equation 4.7 as follows:

$$\hat{y} = argmax_{A \in 1,\dots,A} \quad p(C_A) \quad \Pi_{i=1}^{n} \quad p(k_i|C_A) \tag{4.7}$$

4.4 K-Nearest Neighbors

K-Nearest Neighbor (K-NN) is an algorithm which is used for regression and classification in different applications of pattern recognition [55]. In feature space there are k (particularly a positive integer) closest samples are available as input, whereas, class membership is output in k-nn classification. The classification of object or feature is carried with the help of its neighbor's plurality vote. It means the most common class among the k nearest neighbors of object is assigned to the object or feature. If $k = 1$, then the class of only nearest neighbor is assigned to the object. It is an instance based learning in which the function is locally approximated and computation are stayed until classification [56]. In K-NN classification, most commonly technique used is assigning of weights to the neighbors therefore, the nearest neighbor contribute more than the distant ones. A preferable weighting system is to grant a weight $1/d$, where d is distance of the neighbor.

The training examples, each having a class label are represented in multidimensional space with the help of vector. The training stage only consists of keeping record for training feature vectors and their class labels. During classification phase, k is constant and is defined by the user. An unlabeled query point or vector is classified with the help of its nearest training sample and the class label which is most frequent among them is assigned to the query point. As describe earlier, the weighting is based on distance between feature vector points, therefore, Euclidean distance is used extensively as distance metric in numerous applications where K-NN algorithm is used. In cases of text recognition or object localization from images, another metric named as Hamming distance is utilized. The accuracy of classification is improved if learning of distance metric is carried out with the help of specialized algorithms [57].

The selection of k, whether it is best or not, is typically depends on the data. Generally, if larger value of k is used, it helps in reducing the noise during classification but it makes less distinct boundaries between classes. A suitable value of k can be selected using different Heuristic approaches. The special case where the class of nearest training sample is assigned to query sample (i.e. $k = 1$) is known as nearest neighbor classification. In presence of unimportant or noisy features, the accuracy of k-nn classification is compromised and decreased. Nearest neighbor classifier is most intuitive type of nearest neighbor classification. The class of closest neighbor is assigned to point q in feature space. It can be represented by Equation 4.8 as:

$$C_n^{1nn}(q) = K_{(1)} \tag{4.8}$$

As the training dataset size is larger and reaches to infinity, the error rate of one nearest neighbor classifier increases to worse than twice the Bayes error rate.

4.5 Decision Tree

A decision tree that utilizes a model based on tree like structures consists of decisions and possible circumstances and consequences in order to make a decision or classify a feature vector point [58]. It is one way to representing a procedure containing only conditional control statements. It is specifically used in scenarios of decision analysis or in order to identify the correct strategy or direction to achieve the desired goal. it is also used frequently in machine learning as a popular classification tool [59].

Decision tree can be represented with the help of flow-chart like structure where each node depicts a ”query”, each branch shows the outcome of that query and class labels are shown by leaf nodes (decision formulated after all computations). Classification rules are the path between root node to leaf node. A decision tree is used as analytical and visual decision support utility in decision analysis in which the expected values of competing alternatives is computed [53].

Three types of nodes are used to represent a decision tree, which are as follows:

i. **Decision Nodes** - normally shown with squares.

ii. **Chance Nodes** - normally shown with circles.

iii. **End Nodes** - normally shown with triangles.

In situations of incomplete knowledge, and there requires a decision to be made then decision tree can be used as best choice model when paralleled with probabilistic model. Decision rules are defined for each decision tree. The contents of leaf node represents the outcome and using the if clause, the conditions along the path are described. Typically, a rule can be represented in a following way: if **condition 1** and **condition 2** or **condition 3** $then$ **outcome**

Generally, decision tree is drawn using symbols of normal flowchart because it is easy to understand and read for many people. Decision tree has several advantages which are as follows:

- They are easy to interpret and understand. A brief explanation is enough to describe decision tree model to people and they understand it easily.
- They can be easily combined with other approaches in order to obtain desired results.
- They generates valuable insight data based on expert description of scenario such as probabilities, alternatives or costs, also the preferable outcome.
- They are helpful in determination of expected, best and worst value for various situations.

Similarly, decision tree has some disadvantages as well which are as:

- They are not stable. A small alteration can lead up to huge changes in the decision tree structure.
- They are often incorrect. With same data, other predicting models performs better.
- Categorical variables having numerous levels in data leads decision tree to provide a biased information because of higher number of levels.
- Computation becomes complex when there are more uncertain values. Number of linked outcomes may also make complex calculations.

4.6 Linear Discriminant Analysis

Linear discriminant analysis (LDA) also known as Fisher's linear discriminant or discriminant function analysis, is a method which is used extensively in pattern recognition, machine learning and statistics in order to find a linear characterization between features or which distinct the two or more classes of objects. The result can be used for linear classifier. LDA is closely related to principal component analysis (PCA) and factor analysis. Both PCA and factor analysis also searches for a linear variable combinations that helps in better explanation of data [60]. LDA comprehensively tries to model the data classes differences. Whereas, PCA does not deals with any kind of differences in classes and factor analysis does not considers similarities instead it uses the differences to model feature combinations. In LDA, a distinction must be made between dependent and independent variables. Dependent variables are classes or groups whereas independent variables are object features [61].

Discriminant analysis is a technique that classifies an object or query in mutually exclusive classes on the bases of measurable set of object features. The main objective is to classify an object in one of two or more available classes based on the characteristics of objects that give their best explanation. Another main objective is to reduce the dimensionality of features without losing the information. This problem is known as ”Curse of Dimensionality” and it is occurs because of higher feature vector dimension. This results into undertrained classifiers and data sparsity. LDA is used in order to make an effort for optimizing the class separability [62]. Dependent variables are always be any category or class. Independent variable are any measurement scale which gives the best description about object is often known as features.

In order to use LDA, it is assumed that all the classes are linearly distinctive. There are two approaches used in LDA, class-dependent transformation and class-independent transformation.

- **Class-Dependent Transformation** - It maximizes class variance to within class variance ratio. It includes two optimizing criteria in order to independently transform the dataset.
- **Class-Independent Transformation** - It maximizes overall variance to within class variance ratio. It utilizes only one optimizing criteria in order to transform the dataset and therefore, all data points are transformed through this transformation irrespective of their class identity.

Considering two class problem having two data sets, LDA can be implemented through five steps. First step consists of computation of mean for every dataset separately and also for whole dataset. μ_1 denotes the mean for first set and μ_2 depicts the mean for second dataset. Mean of entire set is denoted by μ_3 and can be computed by:

$$\mu_3 = p_1 \times \mu_1 + p_2 \times \mu_2 \tag{4.9}$$

Step 2 includes the computation of between class scatter matrix S_B and within class scatter matrix S_W. Within class scatter matrix can be calculated with the help of

Equation 4.10 given as follows:

$$S_W = \sum_j p_j \times (cov_j) \tag{4.10}$$

where p_j is the probability of j^{th} class and $cov_j = (x_j - \mu_j)(x_j - \mu_j)^T$ is covariance matrix of j^{th} class. Similarly, between class scatter matrix can be calculated with the help of Equation 4.11 given as follows:

$$S_B = \sum_j (\mu_j - \mu_3) \times (\mu_j - \mu_3)^T \tag{4.11}$$

where μ_3 denotes the entire dataset mean and μ_j depicts j^{th} class mean. Step 3 of LDA involves the computation of eigenvectors. Eigenvectors are computed separately for class dependent transformation and class independent transformation. Firstly for class dependent transformation, it is given in Equation 4.12 as:

$$criteion_j = inv(cov_j) \times S_B \tag{4.12}$$

Similarly, for class dependent transformation, it is given in Equation 4.13 as:

$$criteion = inv(S_W) \times S_B \tag{4.13}$$

Step 4 is based on computation of transformed matrix. For class dependent approach of LDA, it is computed from Equation 4.14 as:

$$transformed\_set\_j = transform\_j^T \times set\_j \tag{4.14}$$

where $transform\_j$ is composed of eigenvectors form given in Equation 4.12. Similarly, For class dependent approach of LDA, it is computed from Equation 4.15 as:

$$transformed\_set = transform\_spec^T \times data\_set^T \tag{4.15}$$

where $transform\_spec$ is composed of eigenvectors form given in Equation 4.13. Step 5 is based on calculation of Euclidean distance which is given in Equation 4.16 as:

$$dist\_n = (transform\_n\_spec)^T \times x - \mu_{ntrans} \tag{4.16}$$

where μ_{ntrans} is transformed dataset mean, n is the class index and x is the test vector. For n classes, n Euclidean distances are obtained for each test points. Finally the smallest Euclidean distance among all n distances will be determined as classification result.

4.7 Support Vector Machine

Support Vector Machine (SVM) is one of most widely used machine learning technique in supervised learning category. This classifier is used for regression analysis and also for classification of data. SVM is used to find the margin between different classes of data. In localization, SVM is used for classification and prediction of user location based on training data samples [63]. Data Items, RSSI patterns are projected in n-dimensional space, then a hyper-plane is selected that discriminate two classes for classification. SVM uses different kernels, such as kernels that use polynomial and radial basis to improve the classification accuracy. This classification accuracy indicates minimizing position estimation error and increases accuracy for localization process. Important features of SVM are as under:

i. First of all, kernel uses vanish evolution for extraction of features space using SVM for minimizing complexity.

ii. In second step mean-squared error classifier is build using SVM to calculate precise measurements.

iii. Finally, a refined system is proposed compared to other available systems.

SVM solve two group classification problems using SVM. Here it is important to mention that, the input vectors are non-linearly mapped with a high dimension feature space. Feature space then predict decision space which is linear in nature [64]. The main idea is to map the training data error free. Support Vector Machine based model is presented using clear gap which further separate categories. This gap must be wide. The main reason of using machine learning approach is data or pattern classification by giving training data set of n points. Following equation represent the process:

$$(\overrightarrow{a}_1, b_1), \ldots\ldots\ldots, (\overrightarrow{a}_n, b_n) \tag{4.17}$$

Here b_i represents 1 or -1, which shows the class of $\overrightarrow{a}_i$. Each $\overrightarrow{a}_i$ is a t-dimensional real vector. Hyper-plane is found with a maximum margin. These margin is basically the division of two group points $\overrightarrow{a}_i$ for which $b_i = 1$ or $b_i = -1$. Any of the hyper-plane is considered as the set of grids or points $\overrightarrow{a}$ satisfying as in Equation 4.18:

$$\overrightarrow{d}.\overrightarrow{a} - c = 0, \tag{4.18}$$

where $\overrightarrow{d}$ is the normal vector of hyper-plane. $\overrightarrow{d}$ is not necessarily a unit vector. The offset of hyper-plane parameter $\frac{c}{(||\overrightarrow{d}||)}$ from the origin along the normal vector $\overrightarrow{d}$. Two parallel hyper-planes are selected that clearly separates the two data classes, so that the gap between them is as large as possible.

These box type or bounded region within two hyper-planes is ”margin”, and the hyper-plane with maximum margin is the one that is between them. With an ordered dataset, hyper-planes is expressed using Equation 4.19 and Equation 4.20 as under:

$$\overrightarrow{d}.\overrightarrow{a} - c = 1 \tag{4.19}$$

$$\overrightarrow{d}.\overrightarrow{a} - c = -1 \tag{4.20}$$

where anything on or above the boundary is of class 1 else class -1.

These Kernel techniques are considered as an algorithms class for pattern recognition in machine learning, and SVM is considered as its ideal member. Geometrically the distance between these two hyper-planes is measured by $\frac{2}{(||\overrightarrow{d}||)}$, therefore for maximizing the planes distance minimize$||\overrightarrow{d}||$. It is also prevented that data points fall upon the margin. This can be rewritten in Equation 4.21 as:

$$x_i(\overrightarrow{d}.\overrightarrow{y} - b) \geq 1, \quad \forall \quad 1 \leq i \leq n \tag{4.21}$$

Radial Basis Kernel function (RBF) are mostly used as a kernel in SVM. Using RBF kernel for two data samples a and $a^{'}$, represented as feature vectors in some input space, is defined in Equation 4.22 as:

$$F(a, a^{'}) = exp\left(-\frac{||a - a^{'}||^2}{2\sigma^2}\right) \tag{4.22}$$

$||a - a^{'}||^2$ is the Euclidean distance within the feature vectors and (σ) is a free parameter. An equivalent definition involves a $\gamma = \frac{1}{2\sigma^2}$. γ value is inserted in Equation 4.23:

$$F(a, a^{'}) = exp\left(-\gamma||a - a^{'}||^2\right) \tag{4.23}$$

The polynomial kernel is a kernel, which uses a polynomial function. This function is defined in Equation 4.24 as:

$$P(a, b) = (a^T b + c)^d \tag{4.24}$$

Here, a and b represents the input features, $c \geq 0$ is a unused or free parameter and d is a polynomial degree. Kernel is homogeneous when $c = 0$.

4.8 Summary

This chapter summarized our proposed fingerprinting based machine learning approach for real time position estimation in indoor environment. Fingerprinting based is basically a pattern matching base position estimation technique, which identify patterns and use a classification algorithm to estimate object real time location with reference to some coordinate system. As discussed in chapter 2, researchers have widely used KNN in fingerprinting approach. SVM, Naive Bayes, and decision tree based techniques also exist. But these techniques have been evaluated on small data sets. In this thesis we have evaluated five different types of machine learning techniques i.e Naive Bayes, KNN, LDA, SVM and Decision Tree on comparatively large data set. None of the previous work extensively studied these five techniques collectively. The next chapter, will evaluate the performance of these five machine learning based position estimation techniques. For depicting the real time dynamic scenario, we have created five scenarios and will test the performance.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Overview

Chapter 4 discussed proposed system design, fingerprinting and localization process using machine learning based approaches. This chapter presents simulations results of the proposed real time dynamic indoor position estimation using machine learning based techniques. Moreover, we have implemented five different types of machine learning techniques in phase 2 of localization process in fingerprinting approach. These five machine learning techniques are KNN, Navie Bayes, Decision Tree, Linear Discrete Analysis, and Support Vector Machine. In order to test the performance of our proposed system, we have simulated these techniques with the help of five different trajectories inside the room where the experiments were conducted.

5.2 Performance Evaluation

In order to evaluate the performance of our proposed real time dynamic indoor position estimation technique. We have combined machine learning based solution with fingerprinting based position estimation model. In fingerprinting based position estimation model, a radio map is generated of the locality where the system would be installed and work. Fingerprints of the system have already been collected and mapped in radio map, we call this phase as offline phase. In online phase, many researchers have used KNN, but here we are comparing different machine learning techniques in order to know their performance. The criteria for measuring their performance is accuracy, complexity, precision, cost, scalability and robustness. But researchers have mainly focused on accuracy, complexity and precision. In this thesis, we are measuring the efficiency of our proposed system using accuracy, standard deviation, and execution time. Standard deviation is an alternate of precision and also the word complexity is used as execution time. For testing purpose, we are using five different trajectories.

5.3 Trajectories

Trajectories refers to the actual movement of the object or human body inside the premises. This movement resembles real time dynamic tracking. These five trajectories are depicted in Fig 5.1.

5.4 Testing

For testing purpose, and validation of our proposed machine learning based indoor localization system. We perform further experiments and collected RSSI measurements for these five trajectories, which resembles real time movement of human body. These five trajectories are different from each others based on direction, speed, movement and time. As discussed in previously, researchers have investigated KNN and SVM, however, LDA is never been evaluated and embedded in fingerprinting based indoor position estimation model. However, in our thesis, we have investigate KNN, Naive Bayes, LDA, SVM and Decision Tree as well with extended and large data set. Comparative analysis of these five classifiers are presented using trajectories and their accuracy, standard deviation and execution time.

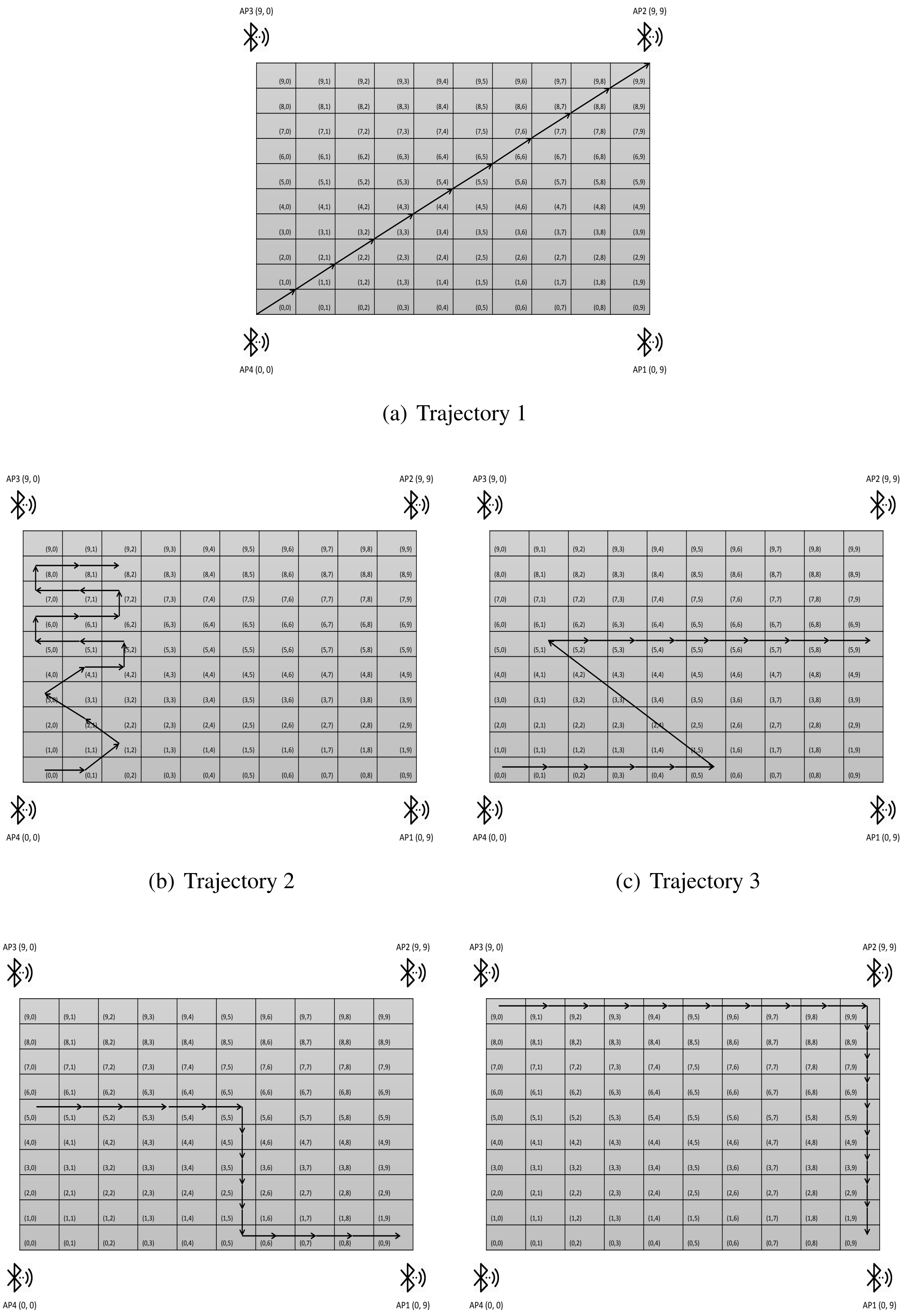
5.4.1 Comparison of Accuracy between Classifiers

Table 5.1 showing comparative analysis of Naive Bayes, KNN, LDA, SVM and Decision Tree in terms of accuracy for all the five trajectories. The accuracy of LDA is better than all the other classifiers. Figure 5.2 Further elaborates individual accuracy of each classifier in each trajectory as shown in Fig 5.1

Table 5.1: Comparison based on accuracy (%) between Naive Bayes, KNN, LDA, SVM and Decision Tree with five different trajectories

Classifiers	**Trajectory 1**	**Trajectory 2**	**Trajectory 3**	**Trajectory 4**	**Trajectory 5**
Naive Bayes	86.4	69.2	77.2	77.9	81.2
KNN, N=1	83.8	60.3	71.2	69.0	65.9
LDA	87.1	72.1	77.3	78.5	81.7
SVM	83.5	66.7	76.9	73.0	81.8
Decision Tree	82.9	64.1	73.2	67.0	71.5

Trajectory 1, is a linear movement, from one corner to another, therefore, the accuracy level compare to other movements are a bit higher. In case 1, i.e trajectory, Naive Bayes accuracy is also good i.e 86.4 % while KNN, SVM and Decision Tree accuracy is 83.8, 83.5, and 82.9 % respectively. Here we can see that, LDA performs better in tracking linear movement and its accuracy is 87. 1, which is better than all other classifiers. In case of KNN, we fixed the Nearest Neighbor node to 1, due to its better performance. Similar pattern is also observed in Trajectory 5, where



(a) Trajectory 1

(b) Trajectory 2

(c) Trajectory 3

(d) Trajectory 4

(e) Trajectory 5

Figure 5.1: Five different trajectories used in experimental setup

the movement is partially linear and horizontal, and performance of the classifiers showing 80 % accuracy. In this case also LDA performs better in terms of accuracy for trajectory 5. On the other side, in Trajectory 2, when the user moving slowly, in different directions, the accuracy level is not 80 %. In this scenario when the object movement is different, in a slow speed, and nonlinear, the performance of the classifiers are average which can be seen in Table 5.1. Numerical finding of Trajectory 3, and 4 also suggest LDA performs better as well. Each trajectory graphical result can visualized in Figure 5.2, 5.3, 5.4, 5.5 and 5.6 respectively.

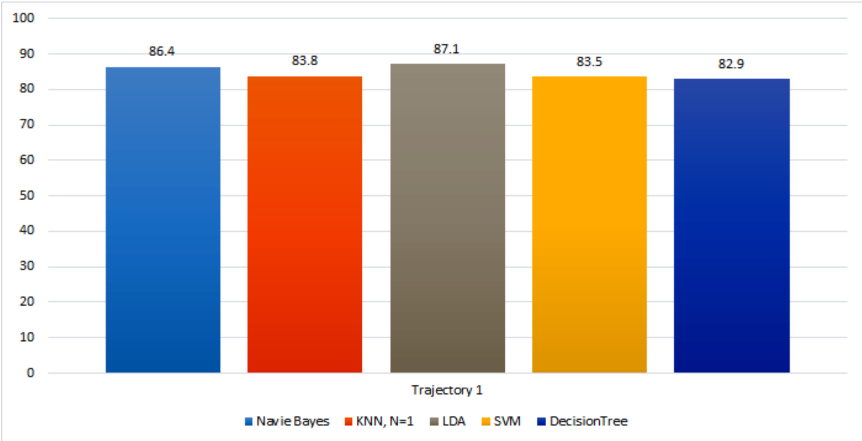


Figure 5.2: Classifiers accuracy for test trajectory 1

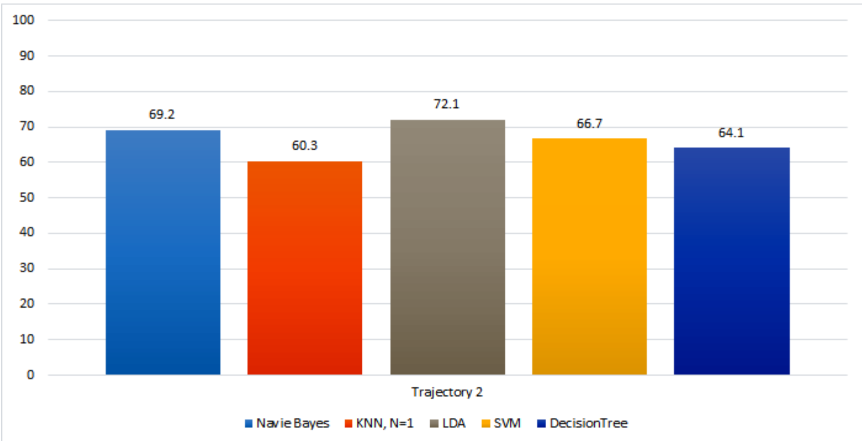


Figure 5.3: Classifiers accuracy for test trajectory 2

Other than, accuracy, we also calculated standard deviation to further investigate the difference among data pattern.

5.4.2 Comparison of Standard Deviation

Standard deviation reveals how much on average the values different from other. In comparative analysis, the lowest standard deviation results better results.

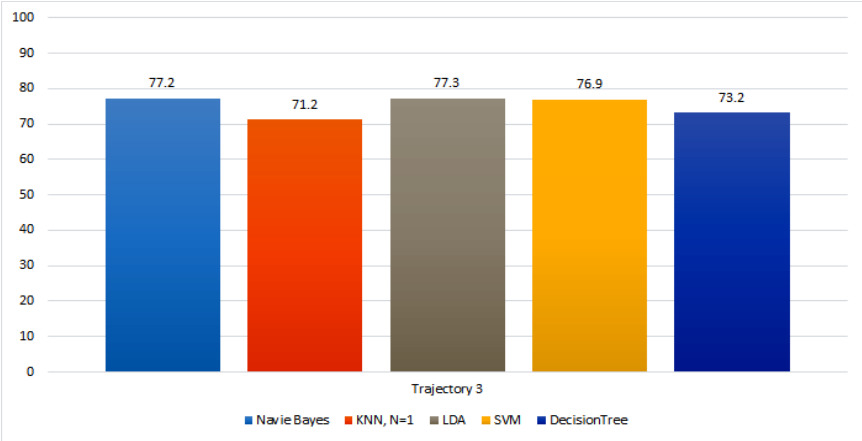


Figure 5.4: Classifiers accuracy for test trajectory 3

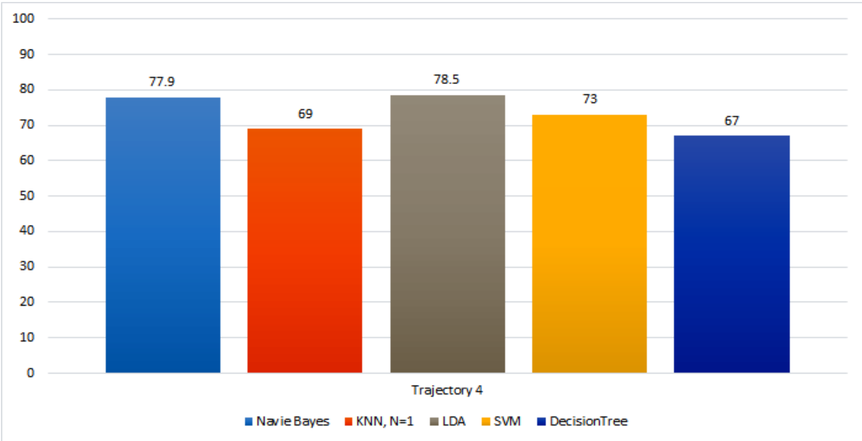


Figure 5.5: Classifiers accuracy for test trajectory 4

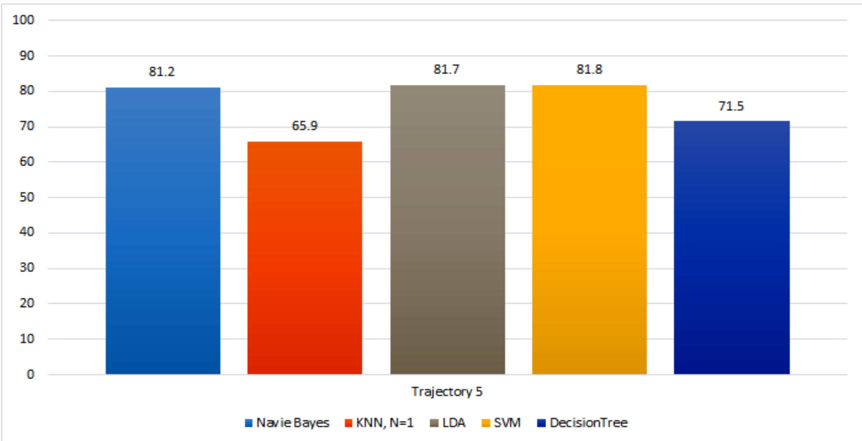


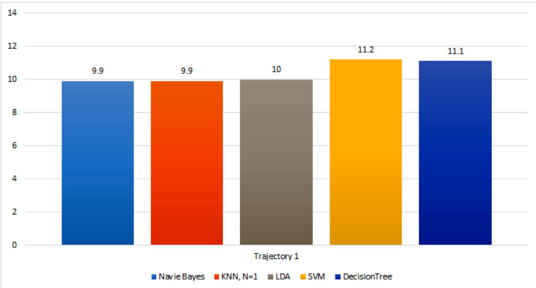
Figure 5.6: Classifiers accuracy for test trajectory 5

Table 5.2 shows numerical findings of each individual classifier and its standard deviation respectively. Based on statistical analysis, it can be seen that, the standard deviation of SVM is less as compared to other classifiers, however, its accuracy is less than LDA. Among all, classifiers the standard deviation of Decision tree is higher.

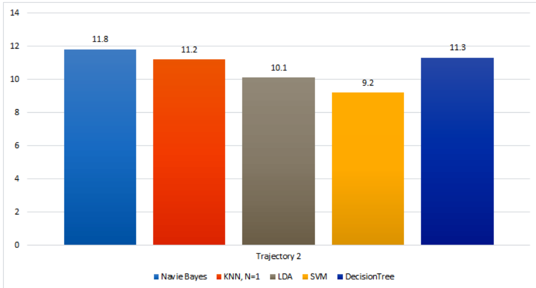
Table 5.2: Comparison based on standard deviation values between Naive Bayes, KNN, LDA, SVM and Decision Tree with five different trajectories

Classifiers	**Trajectory 1**	**Trajectory 2**	**Trajectory 3**	**Trajectory 4**	**Trajectory 5**
Naive Baye	9.9	11.8	8.7	10.2	8.6
KNN, N=1	9.9	11.2	9.9	9.2	10.1
LDA	10.0	10.1	9.9	11	8.6
SVM	11.2	9.2	8.4	9.0	8.0
Decision Tree	11.1	11.3	11	10.6	9.1

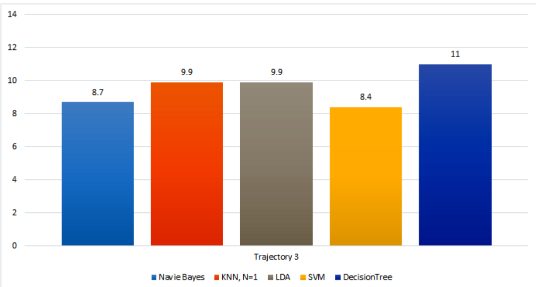
Figure 5.7 shows the visually depicts the standard deviation comparison between classifiers for test trajectories.



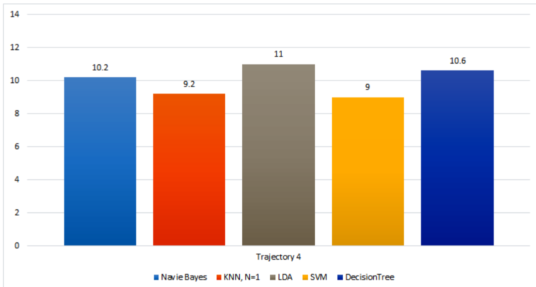
(a) Trajectory 1



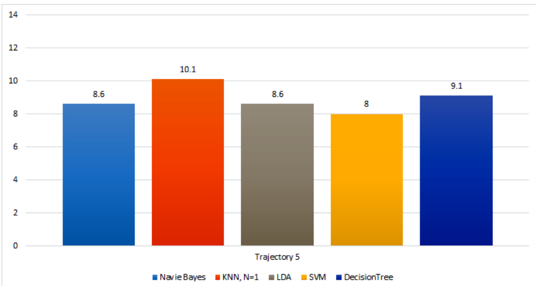
(b) Trajectory 2



(c) Trajectory 3



(d) Trajectory 4



(e) Trajectory 5

Figure 5.7: Average standard deviation value of classifiers for all test trajectories

5.4.3 Comparison of Execution Time

Execution time reveals, how much time, the classifier take during computation process. Table 5.3 shows comparative analysis of all classifiers based on time of execution. Here we can clearly observe that, the execution time of KNN with N=1, is less than the other competitors. While the worst execution time we observed is SVM. One of the reason for less execution time in case of KNN is finding neighboring nodes. Increasing number neighboring nodes also increases execution time. However, the accuracy and standard deviation of SVM is better than KNN.

Table 5.3: Comparison based on execution time required by Naive Bayes, KNN, LDA, SVM and Decision Tree with five different trajectories

Classifiers	Trajectory 1	Trajectory 2	Trajectory 3	Trajectory 4	Trajectory 5
Naive Bayes	0.463	0.477	0.514	0.472	0.46
KNN, N=1	0.009	0.008	0.007	0.009	0.007
LDA	4.092	4.193	4.231	4.116	3.903
SVM	8.227	8.248	8.482	8.957	7.979
Decision Tree	0.062	0.057	0.062	0.059	0.085

5.4.4 Mean Analysis

This subsection presents mean analysis of all the classifiers based on our proposed trajectories. These trajectories depicts real time dynamic movement of the physical object inside the indoor environment, where the system would be deployed. Table 5.4 presents mean analysis of accuracy, standard deviation, and execution time. It can be seen that LDA classifier has the highest mean accuracy i.e (79.34%) among all five classifiers followed by Naive Bayes i.e (78.38%). KNN has the lowest mean accuracy. In term of standard deviation value comparison, Decision Tree has the highest mean standard deviation value i.e (10.62) while SVM has lowest mean standard deviation i.e (9.16).

Table 5.4: Comparison based on mean accuracy % and standard deviation value for all classifiers

Classifiers	Mean Accuracy %	Mean Standard Deviation
Naive Bayes	78.38	9.84
KNN, N=1	70.04	10.06
LDA	79.34	9.92
SVM	76.38	9.16
Decision Tree	71.74	10.62

5.5 Summary

This chapter concludes all the results and findings of this thesis. Results have been obtained with five classifiers which are Naive Bayes, KNN (using K = 1), LDA, SVM and Decision Tree are compared using five different trajectories. These five trajectories, depicting real movement of the physical object inside the indoor environment. These numerical results are obtained and evaluated with respect to accuracy, execution time and standard deviation. These numerical findings shows that, Machine learning algorithms can also be used for monitoring, tracking of real time dynamic object inside indoor environment. Moreover, we have also concluded that, Linear Discriminant Analysis provides excellent accuracy with less computational cost.

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 Overview

Chapter 5 summarized the overview of the research work performed to address the issue of mean error, cost, complexity, scalability, and precision. This chapter will present contributions made in this thesis, limitations of our proposed design and future research directions to carry forward the process of an accurate, standard solution for indoor localization.

6.2 Summary of the Research Challenges and our Proposed Solution

Localization or position estimation is the process to find an actual real time location of the mobile node with respect some known coordinate or landmark. The process of localization can be broadly classified according in two main categories based on its deployment and geographic location. i.e Outdoor and Indoor environment. In case of outdoor geographic location. GPS is the standard developed and fully functional available navigation system originally launched in 1978 by the United States for guiding their military. GPS is based on satellites signals, which requires direct line of sight, and also it does not requires the user to transmit the data, The user device must be equipped with GPS receiver to get the satellite signals. The techniques used to locate an object or navigate a landmark are TDOA, and Geometric based position estimation techniques. The accuracy in GPS depends on user acceptance level, for example, if finding the location of a known landmark such as Faisal Mosque in Islamabad, then acceptable accuracy of more than 10 meters error is also acceptable. On the other side, GPS is not suitable for indoor tracking, due to its attenuation issue. GPS signals can't penetrate the physical infrastructure such as buildings, etc. The reason is its line of sight technology, therefore indoor localization is a key challenge for researchers.

Currently none of the standard indoor localization system so for developed which competes GPS and acceptable for almost every domain. The challenges in indoor environment is the accuracy, precision, cost and many other factors. Among all the key challenging issue is the position estimation error, which is due to many factors such as signal attenuation, multipath fading effects, noise, travel distance, and physical surroundings, means physical infrastructures. In this thesis, the research work is conducted to design a real time dynamic indoor localization technique, which estimate physical objects with higher accuracy and precision. Provides an optimal solution in terms of cost, scalability, and less complex solution. To address these issues, we have thoroughly investigated, existing technologies, existing traditional and conventional indoor localization techniques, and its advantages and disadvantages in detail. Based on our literature review, we have noticed that, artificial intelligence can also play a vital role in this field. The word artificial intelligence means training and learning a system based on all available scenarios, cases, assumptions and then decide what to do. In this regard, we have studied various machine learning techniques such as KNN, Naive Bayes, Decision Tree, Linear Discrete Analysis and Support Vector Machine. Among all these techniques, KNN is the most widely indoor localization techniques which provides excellent accuracy. KNN is as a part of fingerprinting based indoor localization technique. On the other side, the researchers have also investigated these machine learning techniques, but Linear Discrete Analysis is rarely used by the researchers for indoor localization. In this these we have compared five different types of machine learning techniques and found that, Linear Discrete Analysis (LDA) is one of the best machine learning technique among the available techniques.

6.3 Research Findings and Contributions

Chapter 2, discussed various parameters for evaluating the performance of indoor localization systems. These performance metrics are Accuracy, cost, precision, complexity, scalability and Robustness. These parameters are related to the use of technology, localization techniques and implementations. In this thesis we have evaluated our proposed real time dynamic indoor localization technique using Accuracy, and precision. Also the parameters cost and scalability are related to technology which we have already selected the lowest cost wise available solution, which provides a less complex and scalable solution. So only two parameters have been considered. These two parameters have been used by many researchers as well. The main contribution of this thesis are summarized as under.

i. We performed real time experiments and collect RSSI samples in dense indoor environment. At each grid location we took 10 RSSI samples, their mean, and standard deviation are observed.

ii. Based on standard deviation, we extended our data set from 10 RSSI samples to 1000 RSSI similar patterns, while keeping the standard deviation we observed for each grid location, and access point respectively.

iii. We have performed extensive analysis of RSSI, to know the variation with respect to time and distance.

iv. We have studied and reviewed machine learning based solutions proposed by the various researchers for static and dynamic based position estimation inside the indoor environment.

v. After complete understanding and knowing its use, we have implemented Five different types of machine learning techniques, i.e Navie Bayes, KNN, Linear Discrete Analysis, Support Vector Machine and Decision Tree. All these techniques are simulated with a 1000 samples for each grid. In training phase, we use 90 % of our RSSI data in training and 10 % data in testing phase.

vi. Comparative analysis of machine learning techniques using five trajectories, in terms of accuracy, execution time and standard deviation.

6.4 Limitations

Every architecture and design have some limitations. These limitations can be technological, or algorithmic. In our case, we have chosen machine learning for indoor localization. Machine learning requires large amount of data for more accurate results. Also more data and features can extended its performance. In our proposed Real Time Dynamic Indoor Positioning System using Machine Learning, the limitations as under.

a. Real time dynamic indoor localization systems have two main parts, i.e signal reception, and modeling of user localization. In this thesis we focused more on user localization process. However, variation in signal still exist and a challenging task which affect distance and positioning accuracy. This is one of the limitation of this work, that signal variation can be minimized with filtering modeling of the received signal for getting more ideal and better results.

b. Machine learning would perform better if the data set and features are extended. In our thesis, we used RSSI patterns features from four access points and used 1000 samples of data for each actual grid location. One of the limitation of this thesis is limited data set. The results would improve if the data set was large.

6.5 Future Research Direction

More research work is required to investigate Deep learning process with extended data set and feature selection. Also filtering of the received signal is a challenging research domain. To design more accurate, scalable and less complex solutions, it is recommended to investigate Deep learning with filtering RSSI signals. Following are the future research directions.

a. Filtering of RSSI is required for accurate distance and position estimation.

b. large data set is required along with extended geographic area would be more interesting to investigate.

c. Deep learning with extended data set is recommended with filtered RSSI measurements.

d. In our research, we used Bluetooth as communication technology, however, the concept of Internet of Things (IoTs) based devices can be used for Indoor localization systems.

REFERENCES

[1] Pandey, S. and Agrawal, P. A Survey on Localization Techniques for Wireless Networks. *Journal of the Chinese Institute of Engineers*, 2007. 29(7): 1125–1148.

[2] Goldoni, E., Savioli, A., Risi, M. and Gamba, P. Experimental Analysis of RSSI-based Indoor Localization with IEEE 802.15.4. *European Wireless Conference*. 2010. 71 –77.

[3] Jayakody, J., Lokuliyana, S., Chathurangi, D. and Vithana, D. Indoor positioning: Novel approach for Bluetooth networks using RSSI smoothing. 2016.

[4] Luo, X., Brien, W. and Julien, C. Comparative Evaluation of Received Signal-Strength Indicator (RSSI)-based Indoor Localization Techniques for Construction Jobsites. *Advanced Engineering Informatics*, 2011. 25(2): 355 – 363.

[5] Hofmann-Wellenhof, B., Lichtenegger, H. and Collins, J. *Global Positioning System: Theory and Practice*. 4th Edition, Springer Verlag. 1997.

[6] Dedes, G. and Dempster, A. Indoor GPS Positioning - Challenges and Opportunities. *Global Positioning system, Technical Report*, 2005: 1–10.

[7] Nuradha, T., Gnanarathne, I., Perera, L., Denipitiyage, D. and Dias, D. Beacon Placement Algorithm for Hybrid Indoor Positioning with Wi-Fi and Bluetooth Low Energy. *2019 Moratuwa Engineering Research Conference (MERCon)*. IEEE. 2019. 135–140.

[8] Rodríguez Damián, M., Vila, X. A. and Rodríguez Liñares, L. Accuracy of Bluetooth based Indoor Positioning using different Pattern Recognition Techniques. *Journal of Computer Science & Technology*, 2019. 19.

[9] Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M. and Al-Khalifa, H. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors*, 2016. 16(5): 707.

[10] Liu, H., Darabi, H., Banerjee, P. and Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man,*

and Cybernetics, Part C, 2007. 37: 1067–1080.

[11] Lovón-Melgarejo, J., Castillo-Cara, M., Huarcaya-Canal, O., Orozco-Barbosa, L. and García-Varea, I. Comparative Study of Supervised Learning and Metaheuristic Algorithms for the Development of Bluetooth-Based Indoor Localization Mechanisms. *IEEE Access*, 2019. 7: 26123–26135.

[12] Kaemarungsi, K. *Design of Indoor Positioning Systems based on Location Fingerprinting Technique*. Ph.D. Thesis. University of Pittsburgh, United States. 2005.

[13] Xiong, Z., Song, Z., Scalera, A., Ferrera, E., Sottile, F., Brizzi, P., Tomasi, R. and Spirito, M. A. Hybrid WSN and RFID indoor positioning and tracking system. *EURASIP Journal on Embedded Systems*, 2013. 2013(1): 6.

[14] Dai, P., Yang, Y., Wang, M. and Yan, R. Combination of DNN and Improved KNN for Indoor Location Fingerprinting. *Wireless Communications and Mobile Computing*, 2019. 2019.

[15] Farahiyah, D., Romadhoni, R. M. and Pratomo, S. W. Navie Bayes Classifier for Indoor Positioning using Bluetooth Low Energy. *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*. ACM. 2018. 181–185.

[16] Li, Z., Xu, K., Wang, H., Zhao, Y., Wang, X. and Shen, M. Machine-Learning-Based Positioning: A Survey and Future Directions. *IEEE Network*, 2019. 33(3): 96–101.

[17] Enge, P. and Misra, P. The Global Positioning System (GPS). *Proceedings of the IEEE*, 1999. 87(1): 3–172.

[18] Curran, G. D. K. and Condell, J. *Evaluation of Smoothing Algorithms for a RSSI-based Device-Free Passive Localisation*, Springer. 2010, 469–476.

[19] Saad, B. and Yiming, J. A Survey and Comparison on Localisation Algorithms for Wireless ad hoc Networks. *Internation Journal of Mobile Communication*, 2005. 3: 374–410.

[20] Rong, R. and Sichitiu, M. L. Angle of Arrival Localization for Wireless Sensor Networks. *IEEE Sensor and Ad Hoc Communications and Networks, Reston*. 2006. 374–382.

[21] William, S. and Murphy, J. *Determination of a Position using Approximate Distances and Trilateration*. Master's Thesis. Trustees of the Colorado School of Mines, Colorado, USA. 2007.

[22] Ting, W. and Scott, B. Indoor Localization Method Comparison :

Fingerprinting and Trilateration Algorithm. *International Journal of Control*, 2006: 1–4.

[23] Subhan, F., Hasbullah, H. and Ashraf, K. Kalman filter-based hybrid indoor position estimation technique in bluetooth networks. *International Journal of Navigation and Observation*, 2013. 2013.

[24] Vorst, P., Sommer, J., Hoene, C., Schneider, P., C. Weiss, T. S., andA. Zell, W. R. and Carle, G. Indoor Positioning via Three Different RF Technologies. *Fourth European Workshop on RFID Systems and Technologies*. 2008. 1–6.

[25] Buratti, C., Conti, A., Davide, D. and Verdone, R. An Overview on Wireless Sensor Networks Technology and Evolution. *Sensors*, 2006. 9(9).

[26] Guoqiang, M., Bar, F. and Brian, A. Wireless Sensor Network Localization Techniques. *Computer Networks*, 2007. 51: 2529–2553.

[27] Hakan, K. and Shuang, Y. A Survey of Indoor Positioning and Object Locating Systems. *Journal of Computer Science*, 2010. 10(5): 121–128.

[28] Kriz, P., Maly, F. and Kozel, T. Improving indoor localization using bluetooth low energy beacons. *Mobile Information Systems*, 2016. 2016.

[29] Hongpeng, W. and Jia, F. A hybrid modeling for WLAN positioning system. *2007 International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE. 2007. 2152–2155.

[30] Li, B., Dempster, A., Rizos, C. and Barnes, J. Hybrid method for localization using WLAN. *Spatial Sciences Conference*. Citeseer. 2005. 341–350.

[31] Huang, K., He, K. and Du, X. A Hybrid Method to Improve the BLE-Based Indoor Positioning in a Dense Bluetooth Environment. *Sensors*, 2019. 19(2): 424.

[32] Bahl, P., Padmanabhan, V. N., Bahl, V. and Padmanabhan, V. RADAR: An in-building RF-based user location and tracking system. 2000.

[33] Brunato, M. and Battiti, R. Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks*, 2005. 47(6): 825–845.

[34] Mo, Y., Zhang, Z., Lu, Y., Meng, W. and Agha, G. Random forest based coarse locating and kpca feature extraction for indoor positioning system. *Mathematical Problems in Engineering*, 2014. 2014.

[35] Youssef, M. A., Agrawala, A. and Shankar, A. U. WLAN location determination via clustering and probability distributions. *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE. 2003. 143–150.

[36] Youssef, M. and Agrawala, A. Handling samples correlation in the horus system. *IEEE infocom 2004*. IEEE. 2004, vol. 2. 1023–1031.

[37] Roos, T., Myllymäki, P., Tirri, H., Misikangas, P. and Sievänen, J. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*, 2002. 9(3): 155–164.

[38] Mehmood, H., Tripathi, N. K. and Tipdecho, T. Indoor positioning system using artificial neural network. *Journal of Computer science*, 2010. 6(10): 1219.

[39] Yim, J. Introducing a decision tree-based indoor positioning technique. *Expert Systems with Applications*, 2008. 34(2): 1296–1302.

[40] Sánchez-Rodríguez, D., Hernández-Morera, P., Quinteiro, J. and Alonso-González, I. A low complexity system based on multiple weighted decision trees for indoor localization. *Sensors*, 2015. 15(6): 14809–14829.

[41] Battiti, R., Brunato, M. and Villani, A. *Statistical learning theory for location fingerprinting in wireless LANs*. Technical report. University of Trento. 2002.

[42] Nguyen, X., Jordan, M. I. and Sinopoli, B. A kernel-based learning approach to ad hoc sensor network localization. *ACM Transactions on Sensor Networks (TOSN)*, 2005. 1(1): 134–152.

[43] Deng, Z. A., Xu, Y. B. and Wu, D. Indoor positioning in WLAN environment based on support vector regression and space partitioning. *Advanced Materials Research*. Trans Tech Publ. 2011, vol. 204. 1599–1602.

[44] Vahidnia, M. H., Malek, M. R., Mohammadi, N. and Alesheikh, A. A. A hierarchical signal-space partitioning technique for indoor positioning with WLAN to support location-awareness in mobile map services. *Wireless personal communications*, 2013. 69(2): 689–719.

[45] Xie, Y., Wang, Y., Nallanathan, A. and Wang, L. An improved K-nearest-neighbor indoor localization method based on spearman distance. *IEEE signal processing letters*, 2016. 23(3): 351–355.

[46] Wang, X., Gao, L., Mao, S. and Pandey, S. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 2016. 66(1): 763–776.

[47] Aitenbichler, E. and Muhlhauser, M. An IR Local Positioning System for Smart Items and Devices. 23^{rd} *International Conference and Distributed Computing Systems Workshops*. 2003. 334 – 339.

[48] Yick, J., Mukher-jeeand, B. and Ghosal, D. Wireless Sensor Network Survey.

Computer Networks, 2008. 52(12): 2292–2330.

[49] Rong, R. and Sichitiu, M. L. Indoor Localization System using RSSI Measurement of Wireless Sensor Network based on ZigBee Standard. *IASTED International Conference on Wireless Sensor Networks, Alberta, Canada*. 2006. 1–6.

[50] Pu, C. and Chung, W. An Integrated Approach for Position Estimation using RSSI in Wireless Sensor Network. *Journal of Ubiquitous Convergence Technology*, 2008. 2(2): 121–128.

[51] Peneda, L., Azenha, A. and A.Carvalho. *Indoors Localization Using Mobile Communications Radio Signal Strength.* Advanced Trends in Wireless Communications. 2011.

[52] John, G. H. and Langley, P. Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 1995. 338–345.

[53] Yotsawat, W. and Srivihok, A. Inbound tourists segmentation with combined algorithms using K-Means and decision tree. *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE. 2013. 189–194.

[54] Rish, I. *et al.* An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*. 2001, vol. 3. 41–46.

[55] Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 1992. 46(3): 175–185.

[56] Aha, D. W., Kibler, D. and Albert, M. K. Instance-based learning algorithms. *Machine learning*, 1991. 6(1): 37–66.

[57] Shin, B., Lee, J. H., Lee, T. and Kim, H. S. Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems. *2012 8th International Conference on Computing Technology and Information Management (NCM and ICNIT)*. IEEE. 2012, vol. 2. 574–577.

[58] Quinlan, J. R. Induction of decision trees. *Machine learning*, 1986. 1(1): 81–106.

[59] Prasasti, N. and Ohwada, H. Applicability of machine-learning techniques in predicting customer Defection. *2014 International Symposium on Technology Management and Emerging Technologies*. IEEE. 2014. 157–162.

[60] Martínez, A. M. and Kak, A. C. PCA versus LDA. *IEEE transactions on pattern analysis and machine intelligence*, 2001. 23(2): 228–233.

[61] Mika, S., Ratsch, G., Weston, J., Scholkopf, B. and Mullers, K.-R. Fisher discriminant analysis with kernels. *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*. IEEE. 1999. 41–48.

[62] Xu, C., Firner, B., Zhang, Y., Howard, R., Li, J. and Lin, X. Improving RF-based device-free passive localization in cluttered indoor environments through probabilistic classification methods. *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE. 2012. 209–220.

[63] Eriksson, A. Aural/acoustic vs. automatic methods in forensic phonetic case work. In: *Forensic Speaker Recognition*. Springer. 41–69. 2012.

[64] Cortes, C. and Vapnik, V. Support-Vector Networks. *Machine Learning*, 1995. 20(3): 273–297.