Locating Senior Walking Frame Users in Crowded Indoor Environments

by

Kuranage Asok Aravinda Perera

A thesis submitted in partial fulfilment for the degree of Master of Engineering (Research)

at the
Centre for Autonomous Systems
Faculty of Engineering and Information Technology
University of Technology Sydney

October 2017

Declaration of Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed:

Date:

Abstract

This thesis presents a low cost indoor localization system, primarily intended for use by professional elder care supervisors for tracking elderly people in their excursions to a crowded shopping centre. The main requirement is that the system provides an approximate locations of multiple elderly people during an excursion to a crowded shopping centre. The residents are to use walking frames for locomotion, thus their motion is relatively slow and predictable. The resolution of localization is considered adequate if the care supervisor is able to locate a given person through visual contact relative to the estimated location. This thesis presents two novel localization methods that make use of these simplifying constraints and provides an industry strength implementation of one of these strategies.

The first method described is an image based place recognition technique that employs the Bag of Words model for generating image descriptors and a three layer feedforward neural network for producing location estimates. Shop fronts and their corresponding neighbourhood areas are used as classes for training the neural network. The performance of this approach that was evaluated in a real shopping centre environment is presented. Although the system developed performs well, it was found to require the user cooperation in crowded areas and was deemed to have potential privacy concerns.

An alternative solution, a Wi-Fi based indoor localization method is also presented. It estimates the current location of a subject using the Wi-Fi signal strengths received by a sensor module mounted on a walking frame. The environment is modelled as a collection of cells with sizes sufficiently small for locating a person through eye contact. A motion model, based on the knowledge of the floor plan of the environment is described. A probabilistic framework using the Bayes rule in combination with a Kernel Density method for estimating the probability density functions of received signal strengths at the cells is developed.

The Wi-Fi based indoor localization method was implemented on a unit that measures strengths of Wi-Fi signals received from the access points present in an environment,

computes the location and transmits it using the telephone network to a tablet held by a carer. An application on the tablet for visualizing the location of multiple walkers was also developed. The performance of this system was evaluated by conducting multiple trials, including a shopping centre excursion organized by a professional elder care provider named IRT.

From the localization accuracies obtained through the test trials, it could be concluded that the presented Wi-Fi based localization method is adequate to fulfil the requirement of IRT, which is locating elderly people in a crowded indoor environment. It is also found that the floor plan based motion model enables the localization algorithm to produce reliable location estimates, given the relatively slow motion of elderly people.

Acknowledgements

I would like to express my immense gratitude to the people who have supported me in many ways throughout my research.

Firstly, I thank my principal supervisor, Professor Gamini Dissanayake for his continuous support and supervision provided throughout this work, which eventually led to the completion of a quality thesis. His broad scope of knowledge and experience, as well as the patience and understanding nature helped me immensely to complete my research successfully. I'm also grateful to my co-supervisor, Doctor Ravindra Ranasinghe for the support and guidance provided to me. His experience and enthusiasm was invaluable for the completion of this work.

My thanks also go to the staff and residents in Illawarra Retirement Trust (IRT) care facility based in Sydney, for their collaboration in the designing, developing and evaluating processes of the indoor localization system proposed in this work.

I would also like to extend my gratitude to Doctor Lakshitha Danthanarayane, Doctor Michael Behrens and Janindu Arukgoda for their contributions to this work. Their assistance in various software and hardware related development tasks made this work a reality.

I also thank all my colleagues in Centre for Autonomous Systems (CAS), who made my stay in the lab wonderful and joyful. I was extremely fortunate to be a part of CAS circle.

Last, but not least, I would like to thank my loving family: Heartfelt thanks go to my wife Nilakshika, who stayed with me in all the good times and bad times. Thank you for your patience and understanding. I extend my sincere gratitudes to my parents, who made immense sacrifices to bring me to this state. Without your effort, nothing would have been possible.

Contents

List of Figures

Acronyms & Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
CAS	Centre for Autonomous Systems
UTS	University of Technology, Sydney
IRT	Illawarra Retirement Trust
GPU	Graphics Processing Units
RGB	Red Green Blue
RGB-D	Red Green Blue Depth
KDE	Kernel Density Estimation
LRF	Laser Range Finder
BoW	Bag of Words
RFID	Radio Frequency Identification
AMCL	Adaptive Monte-Carlo Localization
SLAM	Simultaneous Localization and Mapping
EKF	Extended Kalman Filter

MCL	Monte Carlo Localization
IMU	Inertial Measurement Unit
CNN	Convolutional Neural Network
SVM	Support Vector Machine
SURF	Speeded Up Robust Features
KLT	Kanade-Lucas-Tomasi
RANSAC	Random Sample Consensus
HOUP	Histogram of Oriented Uniform Patterns
UWB	Ultra-Wide Band
BLE	Bluetooth Low Energy
RSSI	Received Signal Strength Indication
ToF	Time of Flight
LoS	Line of Sight
FAF	Floor Attenuation Factor
WAF	Wall Attenuation Factor
AoA	Angle of Arrival
TDoA	Time Difference of Arrival
KNN	K-Nearest Neighbour
IAP	Important Access Points
MLP	Multi-Layer Perceptron
RBF	Radial Basis Function
GRNN	Generalized Regression Neural Network

cRBF Clustered Radial Basis Function

DAQ Data Acquisition

MAC Media Access Control

ROS Robot Operation System

Nomenclature

General Notations

$e^{\square}$	Exponential
α	Gain parameter in contrast/brightness function
β	Bias parameter in contrast/brightness function
$W_{\square,\square}$	Weight parameter in neural network
$B_{\square}$	Bias parameter in neural network
f	Frequency
X_g	Normal random variable which reflects the attenuation due to flat fading
γ	Path loss exponent
d	Distance
d_0	Reference distance
$\bar{d}$	estimated distance
FAF	Floor attenuation factor
WAF	Wall attenuation factor
n_w	Number of walls
C_w	Maximum number of walls upto which the attenuation factor makes an effect
$t_{\square,\square}$	Transition from one cells to another
N	Number of cells
M	Number of access points
(x, y)	Geometric x and y coordinates in a 2D plane

$s_{\square}$	Signal strength value of an access point
$\theta_{\square}$	Direct path angle of arrival of an access point
ρ	Case base parameter in *exponential* signal strength representation
ϕ	Case base parameter in *powed* signal strength representation
τ	Signal strength threshold
$F_{\square}$	A given value in floor weight vector
g	Decay constant for floor weight vector

Functions

$O(.,.)$	RGB pixel values at a given spatial coordinate of an output image
$I(.,.)$	RGB pixel values at a given spatial coordinate of an input image
$H(.,.)$	Cross entropy function of two distributions
$K(.)$	Kernel smoothing function
$p(.)$	Probability of a given variable
$P(.)$	Signal power at a given transmitter receiver distance

Operations

$\exp(.)$	Exponential
$\sum(.)$	Summation
$\prod(.)$	Product

Chapter 1

Introduction

1.1 Motivation and Background

Ageing is a natural biological process every human being undergoes which makes them less capable of performing their day-to-day tasks. The challenges elderly people experience might occur due to physical weaknesses and problems related to the psychological changes, such as normal age-related memory loss or more severe conditions like dementia, which heavily degrades their thinking skills. 'World Population Ageing 2015', a report by the United Nations has estimated that the number of people aged 60 years or over is going to grow by 56%, from 901 million to 1.4 billion between 2015 and 2030. By 2050 the global population of elderly people is estimated to more than double its size in 2015 [1]. Population growth of senior citizens in Australia demonstrates a similar trend, where 13% of people who are in between age 65-84 in 2014-15 is expected to increase up to 17.7% in 2054-55 [2].

This expected higher population growth of elderly people indicates the need for reliable solutions for the challenges they face, in order to continue their daily routines with the least inconvenience. These solutions should focus on enhancing their quality of life by addressing the challenges they face in daily activities due to age related physical and cognitive impairments. Recent improvements and innovations in technology have been successfully utilized in a number of ways for providing solutions to such challenges. In the context of elder care, technological solutions have been introduced and successfully applied in the domains of health monitoring, rehabilitation, therapy and mobility assistance for tackling the daily challenges experienced by elderly people [3–12].

Locomotion or moving from one place to another is a fundamental activity found in most of the daily routines followed by any person. Therefore, safe locomotion and navigation capabilities are crucial for elderly people as well, even with the presence of manageable cognitive and physical impairments to perform their daily routines and, subsequently live an independent life with improved quality. One of the most common routines in any person's social lifestyle where the locomotion and navigation is highly involved in is shopping in a crowded environment. Even though this could be seen as a simple task for a healthy and capable person, it may pose many challenges to an elderly person, such as getting lost in the crowd or forgetting the way to the exit. High crowd density and the complex floor structures in public indoor places like shopping centres may intensify the hindrance to a safe journey inside the building for elderly people and consequently lead to unexpected consequences. While these problems directly affect the safety of elderly people, they also constitute many difficulties to the people who are responsible for their safety such as professional elder care supervisors.

The practical solution presented in this thesis enables the professional elder care supervisors to retrieve the real time locations of elderly people who are navigating in a crowded environment. The locomotion of an elderly person with some physical impairments is usually supported by a walking frame, thus their motion inside the building is relatively slow and predictable. As a care supervisor usually confirm the safety of elderly people under their care through visual contact, it is adequate to provide an approximate position of a person (e.g. located subregion within the floor) instead of exact coordinates. The presented solution is a comprehensive indoor localization system which is developed in collaboration with a professional elder care provider, Illawarra Retirement Trust (IRT), for facilitating their care supervisors to retrieve locations (subregions) and find elderly people in frequent excursions to a nearby crowded multi-storey shopping centre. In these excursions, a group of elderly residents in IRT care facility is formed and accompanied by an IRT care supervisor to fulfil their shopping needs. As stated above, issues may emerge within this shopping routine such as elderly people losing their way back to the care supervisor due to their incapability of remembering the locations or travelled path. The IRT care supervisor who is responsible for this group should be able to cope with these situations in order to assure the safety of people under his/her care. Therefore in practical scenarios the care supervisor keeps the group together throughout their shopping tasks and stays with them to ensure their safety. This practice is highly inefficient and tiresome as the whole group has to travel together even for fulfilling a single person's shopping needs. The presented system in this thesis provides a method for a care supervisor to continuously track the group members under his/her care by observing their respective locations (subregions) within the shopping centre. As the main requirement of this specific application is to retrieve the current subregion (in contrast to the exact coordinates) of a person, employing a much simpler and computationally light localization technique is more suitable than a complex method which requires high processing power and expensive sensors.

The indoor localization system we present in this thesis consists of two primary components, i.e., self-localization unit and a hand-held tablet device for visualizing the real time locations of a group of people while they navigate in a shopping centre using walking frames. Self-localization units are mounted on wheeled walking frames while the carer holds the tablet. Communication between the walkers and the carer is through the telephone network, alleviating the need to log into the wireless infrastructure within the shopping centre. Our low cost and computationally efficient solution enables the care providers to easily accommodate multiple wheeled walker platforms, and to perform simultaneous localization.

1.2 Thesis Outline

This thesis is organized as follows.

Chapter 2 presents a computer vision based place recognition method for a crowded indoor environment. This method is developed in order to facilitate the professional elder care supervisors to track elderly people under their care, when navigating inside a shopping centre as mentioned in Section 1.1. This technique performs the place recognition in a shopping centre floor which is modelled as a collection of pre-determined cells, such a way that each cell contains a shop front. All the training and testing images of these cell based shop fronts are captured over multiple days inside a crowded shopping centre located in Sydney. The captured images include various dynamic properties such as inconsistent crowd density, changing advertisement boards and posters on a daily basis and illumination variations. The place recognition technique presented in this Chapter employs Bag of Words model and a feedforward neural network for producing location estimates for camera captured RGB images of shop fronts. We have also included results of various experiments conducted in order to evaluate the behaviour of the proposed feedforward neural network. As the conclusion, this Chapter presents the strengths in this approach as a place recognition technique and expected challenges when used in a practical localization system.

In practice, a localization system for tracking elderly people which employs the proposed image based place recognition method is required to be developed focusing on a wheeled walker platform. Such a system will rely on the image feed coming from the camera that is mounted on the wheeled walker. In such a setup, privacy concerns due to the images captured from the surrounding environment with the presence of crowd could be considered as a primary issue. Moreover, unexpected camera orientations, high occlusions due to crowd and extreme motion blur due to the locomotion pattern of the person could occur in such a scenario. In order to address these challenges the user is required to assist the system functionality by pointing the camera mounted walker platform at the closest shop front and wait until the crowd clears to enable the camera to capture a clear image. Even though this method is adequate to locate a person, this additional responsibility expected from the user, as well as the aforementioned privacy concerns should be best avoided if feasible. As a solution, Chapter 3 in this thesis presents a robust indoor localization method which relies on the existing Wi-Fi infrastructure inside a public indoor environment. While a Wi-Fi based localization method does not have a concern over privacy violation as in an image based place recognition system, the signal attributes could still be affected by the orientation of the receiver. However, such orientation dependent signal strength characteristics could be easily captured and included into an estimator in a training phase. The presented Wi-Fi based localization method adopts a Bayesian approach with Kernel Density Estimation (KDE), in order to produce probabilistic location estimations. As we do not require the exact coordinates of the localizing subject the locations are considered to be pre-determined cells on the floors. It is required to capture time, location and orientation dependent signal strength characteristics for each cell, when generating Wi-Fi access point based probability density functions of signal strengths using KDE. Moreover, this Chapter also presents a motion model we incorporated into the localization framework which exploits the information of floor transition events and prior knowledge of cell distribution.

Chapter 4 presents the complete process followed for developing a practical comprehensive Wi-Fi based indoor localization system based on the method presented in Chapter 3. This system consists of a small scale self-localization system which is capable of processing sensor data and producing location estimates, and a tablet application which can be used to visualize these locations. The presented self-localization system is designed to be included in a commonly used standard wheeled walker, as the aim of this system is to facilitate carers to track elderly people who use such walking aids. When in operation, the self localization system estimates its position and transmits this estimated information to the tablet application through standard text messages. The tablet application is capable of decoding the received message content and displaying the extracted location information on the floor map. This system is being evaluated in Roselands shopping centre located in Peakhurst, Australia, and the demonstrated results are presented in the Chapter.

As the final Chapter in this thesis, Chapter 5 presents the conclusions corresponding to the work performed in previous Chapters, their limitations and future work.

1.3 Contributions

- Image based place recognition technique for a crowded indoor environment
- Wi-Fi based indoor localization method for a crowded indoor environment
- Wi-Fi based comprehensive indoor localization system for tracking elderly people in a shopping centre

1.4 Publications

The following two publications by the author are related to the work presented in this thesis.

- Perera A., Ranasinghe R., Dissanayake G. A Neural Network Based Place Recognition Technique for a Crowded Indoor Environment. Conference paper presented in *The 12th IEEE Conference on Industrial Electronics and Applications (ICIEA 2017)*, Siem Reap, Cambodia, June 2017
- Perera A., Arukgoda J., Ranasinghe R., Dissanayake G. Localization System for Carers to Track Elderly People in Visits to a Crowded Shopping Mall. Conference paper presented in *8th International Conference on Indoor Positioning and Indoor Navigation (IPIN 2017)*, Sapporo, Japan, September 2017

1.4.1 Indoor Localization Competition - Track 1, IPIN 2017

The Wi-Fi based indoor localization method presented in Chapter 3 is accepted to be presented in indoor localization competition - track 1, organized by IPIN 2017 organization committee. The accuracies demonstrated in our method are competitive when placed against the performance of the teams during previous version of this competition organized by the same committee.

Chapter 2

Image Based Place Recognition in Crowded Indoor Environments

2.1 Introduction

Place recognition is an important requirement for applications such as location based context-aware applications, mobile robot navigation and people monitoring where the location of a person or equipment (commonly termed as the subject) is an important part of the task context. In this perspective, a robust place recognition technique that performs well in an indoor crowded environment has the potential to be used as an indoor localization system. Therefore, the primary aim of this Chapter is to present a computer vision based place recognition technique, which could be adopted by professional elder care providers to track elderly people, in their frequent excursions to a crowded shopping centre.

In addition to indoor localization methods depend on vision based place recognition, there exist many localization techniques which rely on various other sensors such as Odometry [13], RGB-D sensors [14], Laser Range Finder (LRF) [15–17] as well as wireless signal based methods which use, for example Zigbee modules [18] [19] or Bluetooth modules [20]. However, various shortcomings could also be identified in the majority of them, concerning our application of localization in an indoor crowded environment. While the concepts and

FIGURE 2.1: Sample images subjected to various factors which affect the place recognition task

the applicability of LRF, Odometry and RGB-D sensor based methods will be investigated in Section 2.2 in this Chapter, the wireless signal based localization techniques will be discussed in detail in Chapter 3.

In general, computer vision based place recognition techniques use a camera image to recognize the user location. Factors such as occlusion due to the presence of a large crowd, varying lighting conditions, dynamic shop-front setups and motion blur present in images captured from a moving camera increase the complexity of the vision based indoor place recognition task in a shopping centre. The technique presented in this Chapter aims to identify location from where the images were taken subjected to all of the aforementioned challenging environmental conditions (Fig. 2.1).

A subregion based place recognition approach, where the images are captured by covering the surrounding areas of multiple subregions (cells) in a larger indoor space is adequate for the target application of tracking elderly people in a shopping centre, as the exact coordinates of the people are not required. Based on this, a computationally light place recognition technique is developed, which is capable of processing images captured from a camera mounted on a walking frame and estimating the corresponding subregion. We have used the image data collected at Broadway shopping centre located in Sydney city, for conducting the experiments to evaluate this approach. As Broadway shopping centre is located within a busy area in Sydney, it resembles a highly crowded and dynamic indoor environment, thus an ideal setting for our experiments. Therefore, the images collected

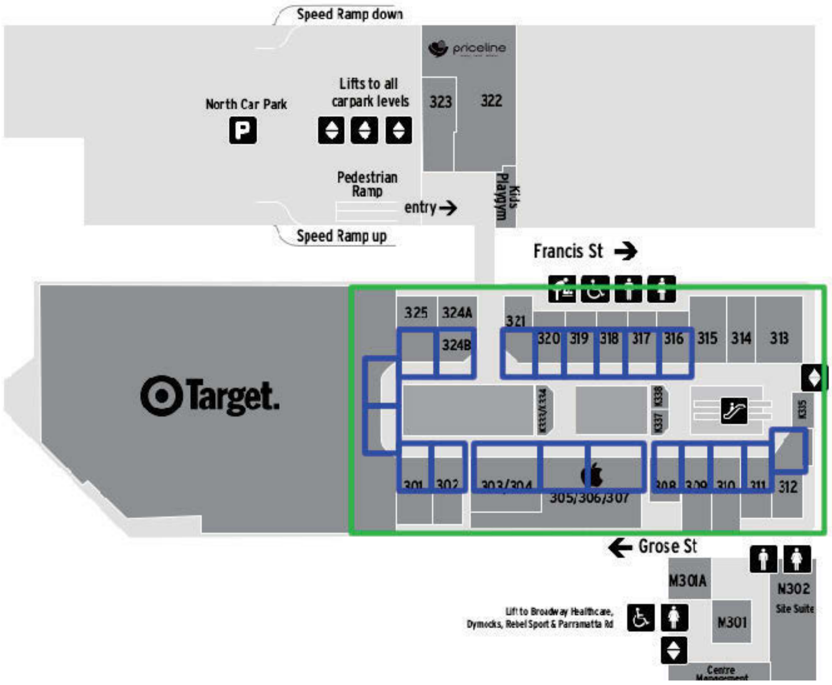

FIGURE 2.2: Top view of the third floor in Broadway shopping centre

inside this facility contained all the challenging elements such as occlusions, dynamic shop-front setups/posters. Plan view of the 3rd floor of the shopping centre building where the images were collected is illustrated in Fig. 2.2. The area inside the green square in Fig. 2.2 is the floor region used to collect image data. This area is divided into subregions as shown by blue squares, where each such region represents a class in the presented place recognition approach.

The approach we adopt for place recognition exploits the strengths of Bag of Words (BoW) technique and feedforward neural networks. BoW is a well known technique for producing image descriptors used in a range of computer vision based applications which are required to represent the content of images. A feedforward neural network with a single hidden layer is fed with BoW descriptors derived from the captured images. Its output layer produces a probability vector, where the value of each element represents the likelihoods of corresponding input image being captured from a particular subregion.

This Chapter is organized as follows: Section 2.2 describes the related works and in Section 2.3 we discuss the core concepts of our algorithm. Section 2.4 explains the setup and experimental results while Section 2.5 concludes the Chapter.

2.2 Literature Review

This Section investigates existing indoor localization methods based on Laser Range Finder (LRF), RGB-D sensor and Odometry, as well as the computer vision based place recognition techniques. Other localization techniques which rely on wireless signal based sensors such as Bluetooth, Zigbee, Radio Frequency Identification (RFID) and Wi-Fi will be comprehensively discussed in the next Chapter.

2.2.1 Laser Range Finder Based Indoor Localization

LRF based indoor localization methods are widely used in works related to recovering position and orientation of mobile robots in indoor environments. While most of the previous LRF localization methods are based on architectural floor blueprints, modern approaches adopt Simultaneous Localization and Mapping (SLAM) techniques for developing an environment map using an LRF sensor prior to the localization phase [21–24]. In general, LRF based localization techniques operate by processing the current laser scan observations for estimating position and orientation of a mobile robot within a given environment map. A commonly used approach here is Adaptive Monte-Carlo Localization (AMCL), which can produce probabilistic estimates for the robot position [15] [16]. In addition to that, Extended Kalman Filter (EKF) based approaches have also been applied in previous works [17], which usually requires an initial position of the robot to begin the localization task.

Although LRF based indoor localization techniques demonstrate good accuracies, high performing and longer range LRF sensors required for estimating accurate position coordinates are expensive compared to other cheaper sensors like RGB cameras, Wi-Fi / Bluetooth / Zigbee modules. As the requirement of our target application could be satisfied by estimating the current subregion, instead of exact coordinates of the subject, highly accurate localization technique which relies on an expensive LRF sensor is not warranted.

2.2.2 RGB-D Based Indoor Localization

RGB-D sensors are capable of capturing RGB images together with corresponding depth measurements from the environment. The localization systems based on RGB-D sensors employ these readings together with a given map for estimating locations. The method proposed in [14] presents a dense depth sensor model which is capable of handling architectural floor plans as environment maps and incorporates Monte Carlo Localization (MCL) for estimating the position. Moreover, it presents a motion model based on visual-inertial odometry which makes use of camera feed as well as an Inertial Measurement Unit (IMU) sensor. A similar work presents *W-RGB-D* [25], an indoor global localization method based on Wi-Fi network and RGB-D camera. It incorporates Monte-Carlo localization and uses Wi-Fi signal strengths to estimate a coarse initial distribution of sensor position (as a particle initialization strategy in MCL).

When compared to a RGB camera, RGB-D sensor provides more information on its surroundings which could be used for improving the performance of the indoor localization task. However, an RGB-D sensor while relatively inexpensive, typically requires a computing platform with good processing capability.

2.2.3 Odometry Based Indoor Localization

Odometry based localization is another technique which uses readings from the motion sensors to estimate the position of a subject relative to a known starting position. Odometry based localization systems rely on dead reckoning, where it calculates the current position by advancing the previously determined position using estimated displacement and direction. Most commonly used motion sensors in odometry are wheel encoders, which are attached to a mobile robot's wheels. A well known issue in encoders could be identified as the accumulating error due to wheel slippage, floor roughness and discretized sampling of wheel increments [26]. The method presented in [13] aims to solve this issue, by employing an Inertial Navigation System (INS) that consists of an accelerometer and gyroscope. However, the drift due to cumulative errors in dead reckoning based localization systems are unavoidable. Additionally, the localization techniques that combine wheel encoders with IMU devices could enhance the accuracy of location estimate, but could be susceptible to magnetic influences from other electronic devices when a magnetometer is present in an IMU [27].

The prior knowledge of starting position is a prerequisite in an odometry based localization system. Despite the number of accuracy enhancement techniques available, an indoor location system based on odometry is not suitable for our application.

2.2.4 Computer Vision Based Place Recognition Techniques

Recognizing places or regions using a sequence of images from a camera could be regarded as a primary step in a computer vision based localization technique. Therefore, it is important to explore existing approaches related to vision based place recognition, and their suitability to our application of indoor localization in a crowded environment.

Based on the approach used, computer vision based place recognition techniques could be categorized into patch based techniques, feature and descriptor based techniques, logos based techniques or combined techniques.

2.2.4.1 Patch Based Techniques

Image patch is a subregion in an image which does not necessarily posses a semantic meaning. Patch based place recognition techniques are based on image subregions, that are consistent over various appearance changes in images corresponding to the same location. Therefore, identifying such image regions representative of the locations of interest, is a primary step in patch based place recognition methods.

An indoor and outdoor place recognition approach presented by Sunderhauf et al. [28], employs a reliable landmark proposal method together with strengths in Convolutional Neural Network (CNN). In order to extract the landmarks, it makes use of *Edge Boxes* approach proposed by Zitnick and Dollár [29], where the likelihood of an object existing inside a bounding box within an image is calculated based on the number of contours that are completely contained in this box area. Once this is performed, each extracted landmark is passed through a CNN to extract feature vectors which are then subjected to dimensionality reduction and compared for matching different images. The CNN used here is already being trained on *ImageNet* [30], a generic image database, hence no application specific off-line training is needed. As this work depends on landmark regions in an image for describing a scene in contrast to relying on the whole image, it demonstrates an improved robustness against view point changes or partial occlusions. However, the experiments performed in this work are based on more general outdoor and indoor environments such as roads, libraries, and rooms, hence a proper assessment on its performance inside a crowded environment is required. In addition, employing a CNN for a real time place recognition task usually requires significant amount of processing power. The main reason being the complex nature associated with the CNNs' network structure, such as the larger number of neurons and the number of network layers needed to achieve an acceptable performance. Highly efficient processing units such as Graphics Processing Units (GPU) are frequently used for implementing CNNs and for achieving a faster processing times.

McManus et al. [31] propose a method for learning place-dependent features for vision-based localization which is specifically targeted for outdoor place recognition in varying weather conditions. In this work, mid level patches representing distinctive visual elements are chosen within images in contrast to low level features, such that these patches are stable regardless of the appearance conditions of the environment. In the learning process, it uses a Support Vector Machine (SVM) classifier to find candidate scene signatures and then performs bundle adjustment to retrieve optimal landmark locations for each scene signature. Even though this system demonstrates a competitive accuracy level in terms of vision based localization, it has been particularly designed for the use in outdoor environments.

2.2.4.2 Features and Descriptors Based Techniques

An image feature represents a specific information content in a given location that is extracted based on predetermined image properties (such as intensity gradient, colour), while an image descriptor could be identified as an information content which describes the whole image, usually based on its detected features. In this context, image features and descriptors could be considered to be the basis in a large range of computer vision based techniques including place recognition.

A visual indoor localization method proposed by Piciarelli [32], detects Speeded Up Robust Features (SURF) [33] and stores them together with manually defined position information. While this is performed off-line, the localization is done on-line where visual features from each image are compared with stored data for identifying the best matching reference image. This algorithm employs a Kanade-Lucas-Tomasi (KLT) point tracker [34] to track feature positions in image frames and Random Sample Consensus (RANSAC) algorithm [35] is utilized to exclude outliers in the process of estimating projective transformation that best describes the displacement of features. However this work does not specifically assume a crowded indoor environment for localizing the subject where many occlusions and other feature inconsistencies may present for a given region.

In addition to feature based recognition, image descriptors have also been used in work related to place recognition. Work done by Sahdev and Tsotsos [36], employs Histogram of Oriented Uniform Patterns (HOUP) descriptors and utilizes an SVM classifier for place categorization. The performance of this system has been evaluated under varying illumination conditions within selected locations of a university building and a hotel. In addition, this has been further assessed using two public datasets: a dataset including images from a typical office environment and a dataset consists of images from random scene categories such as suburb, living room, forest. Another descriptor based place recognition approach is proposed by Sizikova et al. [37], where CNNs were used for generating descriptors. In order to derive depth descriptors and intensity descriptors, it assumes that the input images are RGB-D images. Once these two types of descriptors are generated, they are combined for deriving a joint descriptor which represents depth and intensity properties of the scene. These joint descriptors are then matched for estimating the most suitable region for a test RGB-D image. The performance evaluation of this approach has been done using public datasets which includes images from various locations such as domestic/office spaces, hotels, classrooms.

2.2.4.3 Logos Based Techniques

Another potential approach for vision based place recognition, especially in indoor environments such as shopping centres is to recognize logos/labels or similar symbols in shop fronts. There have been a number of works done for recognizing logos in the real world, such as methods proposed by Romberg et al. [38][39], *DeepLogo* by Iandola et al [40].

In work [38], a quantized representation of the regions in logos is derived based on the analysis of local features and the composition of basic spatial structures like edges and triangles. In order to evaluate this method, a new dataset has been built and published (FlickrLogos-32) which contains photos depicting logos which are downloaded from Flickr [41]. The images in this dataset includes logos in various setups such as vehicles, newspapers, outdoor/indoor environments, etc., and it has gained considerable recognition amongst the researchers in logo recognition, due to the large number of labelled logo images it provides. Romberg and Lienhart [39], have suggested another logo recognition technique using Bundle min-hashing approach. Min-hashing is a technique for estimating the similarity of two sets that is very commonly used in document matching applications such as web page comparison. In this work, the conventional min-hashing technique has been extended to a feature bundling technique, where local image features are aggregated with their neighbouring features, in order to create feature bundles. Then these feature bundles consisting of more information than single features, are employed for matching similarities in images, and subsequently for performing logo recognition task.

DeepLogo work proposed by Landola et al. [40], utilizes strengths of CNN for addressing the logo recognition problem. It presents three CNN architectures, where two of them are designed by introducing modifications to well known *GoogLeNet* CNN [42] while the remaining one closely follows the *GoogLeNet*. Even though this work attempts to address the logo recognition problem, it does not specifically consider logos in shop fronts, but rather focuses on recognizing logos from general setups, as it also uses FlickrLogos-32 data set for evaluation. In contrast to the general logo recognition, the image based indoor localization technique presented by Wang et al. [43], specifically focuses on text based logos found in shopping centre environments. The method proposed in this work employs a text detection technique in order to identify texts which appear in shop logos as well as in shop fronts, performs shop facade segmentation, and subsequently estimates the location in the shopping centre. However, it does not consider the logos which do not contain texts.

Even though logo recognition seems like a potential method for indoor localization applications, particularly in an environment like a shopping centre, observing logo only images for each localizing region is unlikely in practice. Therefore, such a method is required to be extended to recognize place dependent as well as consistent symbols and landmarks, that may not necessarily be logos.

2.2.4.4 Combined Techniques

Vision based place recognition methods have been designed by incorporating information from multiple techniques and fusing them together in order to enhance the reliability of the recognition task. A technique proposed by Xu et al. [44], performs feature fusion for shop front recognition, in order to localize a person in a shopping centre. This work incorporates style features and text features from an image, for identifying the user's position. In order to retrieve potential text features from the image, it employs a CNN based technique [45] and then executes a filtering method to reject false text candidates. Identifying style features is achieved by fine-tuning *AlexNet* CNN [46] using the collected data set. Once the style features and text features are retrieved, these two types of features are fused to produce the final result for place recognition. The feature fusion is done by combining these two features into a new feature vector and training a classifier on the joint vectors. This work uses shop front images collected from the Internet to evaluate its performance, where some of these images may represent more controlled environments compared to an active and crowded shopping centre. Even though performance improvements have been observed by this approach, results indicate the presence of scenarios where feature fusion fails.

Even though many image based place recognition approaches exist in the literature, most of them proposes computationally costly techniques such as CNN, while others target more general or controlled environmental setups. As our application require approximate location of a person in the shopping centre, it is adequate to recognize a nearby shop front for estimating his/her current subregion in the building. Therefore, we adopted a much simpler and computationally light approach for developing our place recognition technique.

In the following Section, we elaborate our three layer feedforward neural network based approach for place recognition in a crowded shopping centre.

2.3 Vision Based Place Recognition in a Crowded Environment

Place recognition in a crowded environment, especially in a place like a shopping centre, poses many challenges due to its dynamic nature. Extensive crowd and motion of people, varying lighting conditions, varying shop front setups and posters are a few examples for the aforementioned dynamic characteristics. In addition to these dynamic characteristics, it is possible for a camera to capture images that include some parts of nearby regions. This may occur due to the change in orientation of the camera or simply because the image is captured at a transitional stage between regions. Motion blur, due to the motion of the platform to which the camera is fixed as well as due to the fast moving crowd, might also increase the ambiguity of the estimated region. Therefore, it is important to adopt a probabilistic approach for the place recognition task, which can reflect the ambiguity of location estimates, especially in a highly dynamic environment like a shopping centre. In the proposed algorithm, a vector of probabilities is generated where each value represents the likelihood of the subject being in the corresponding region.

2.3.1 Image Preprocessing

The images collected for training and evaluating the proposed place recognition technique were captured from Broadway shopping centre in Sydney, Australia. In order to collect the images, we have divided an area on the 3$^{\text{rd}}$ floor within the shopping centre into 20 subregions based on the number of unique shop front setups as illustrated in Fig. 2.2. In addition, each subregion considered for data collection was set to contain an approximately equal geometric area on the floor. As the goal of this work is to recognize

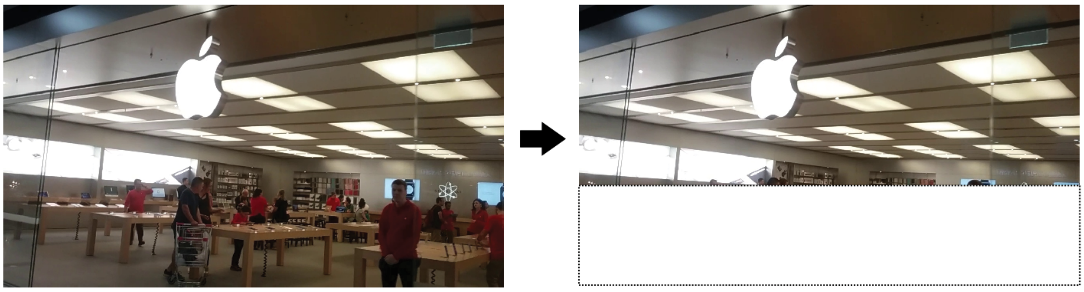

Figure 2.3: Bottom portion of images are cropped for removing crowd

places in the shopping centre, images are collected such that the camera is facing towards the corresponding shop front, while positioned in each subregion.

It could be observed that the majority of these images contains occlusions due to the appearance of crowd, primarily at the bottom portion of each image. As these occlusions has the potential to include spurious information into the data, one third portion from the bottom part of every image is cropped to remove the crowd appearances as illustrated in Fig. 2.3. The upper parts of the majority of images contained more persistent features coming from shop logos, fixed lights, roof patterns, etc. In parallel with the image cropping step, manual classification is performed for captured images by using an application we have developed, in order to label the training images.

2.3.2 Bag of Words Model

Bag of Words (BoW) technique is a popular approach used in text categorization [47] [48] which is also adopted by computer vision research community as an image descriptor, to represent the content of an image. Development of BoW descriptor follows three main steps; feature extraction, feature quantization and BoW histogram preparation. In the initial step, a set of features are detected and extracted from given images. Once this is done, the features are quantized by employing a suitable clustering method and each cluster is assigned with a word identifier. In this way, a vocabulary of words is created, where each word represents the corresponding cluster in feature space. In the last step, the image features are assigned with the relevant word identifier, and the frequency of each word is counted for images separately in order to prepare the word histograms. A given

BoW histogram prepared like this represents the list of word appearance frequencies for a given image and this histogram is considered to be the BoW descriptor.

For implementing the BoW model in this work, Speeded Up Robust Features (SURF) [33] are extracted from all training images and clustered using the K-Means algorithm. The purpose of this clustering step is to identify similar features in all images within the SURF space. Once the K-Means clustering is completed, the cluster centres are derived, such that each cluster centre represents the features in that particular cluster. The cluster centres are then assigned unique word IDs and a vocabulary is generated out of these words. The number of cluster centres decides the size of the vocabulary in the BoW model. This could be considered as a parameter which governs the quantization. In our BoW model, 1500 clusters have been created and consequently a vocabulary with 1500 different words is generated. Here, the number of clusters is determined experimentally by analysing the collected images in order to find out the feature quantization level they require to generate representative BoW descriptors.

Once the BoW vocabulary is created, all the detected features of training images are labelled as words from this vocabulary. This is the feature quantizing step in BoW model which reduces the scope of features by labelling them as representative words. In this way, the training images are now represented by a list of words instead of features. After this step, frequency of each word in each image is calculated and a list of word histograms is created for training images. A given word histogram from this list is considered to be the BoW descriptor for the corresponding training image. The process of preparing BoW descriptors for training images is illustrated in Fig. 2.4.

Once the BoW descriptors are prepared this way, the word frequencies in each histogram are normalized, and used for training the feedforward neural network. This process is described in the next Section.

2.3.3 Feedforward Neural Network

The feedforward neural network which is incorporated in our place recognition approach consists of an input layer, a single hidden layer and an output layer (Fig. 2.5). At

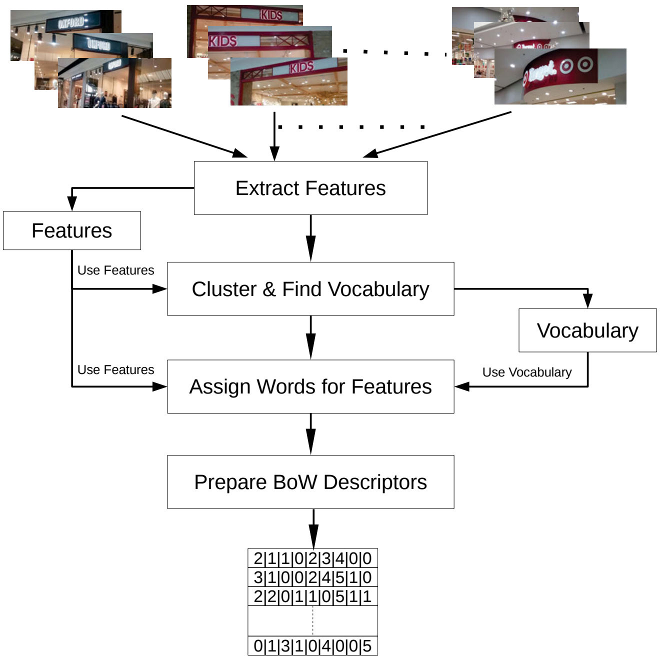

FIGURE 2.4: Bag of Words Extraction

the training phase of this network model, the input layer is fed with normalized BoW descriptors derived from the training images. Therefore, the size of the neural network input layer is set to be 1500, as the BoW implementation generates descriptors with the size of 1500 units. In other words, each input in the input layer corresponds to a word ID in the normalized BoW descriptor.

The proposed feedforward neural network is a fully connected network, where each input in the input layer is connected to all the neurons in the hidden layer, while each hidden layer neuron is connected to all outputs in the output layer. The neural network hidden layer consists of 150 neurons with rectifier as the activation function (2.1). It could be seen that selecting a smaller number of hidden layer neurons when compared to the 1500 inputs at the input layer has improved the generalization capability of the network by minimizing overfitting. The rectifier function used at the hidden layer in this implementation is a commonly used activation function in deep neural networks such as convolutional networks. Rectifying neurons are known to be better models of biological neurons compared to other functions such as hyperbolic tangent, and it has gained considerable attention due to its effectiveness on neural networks [49].

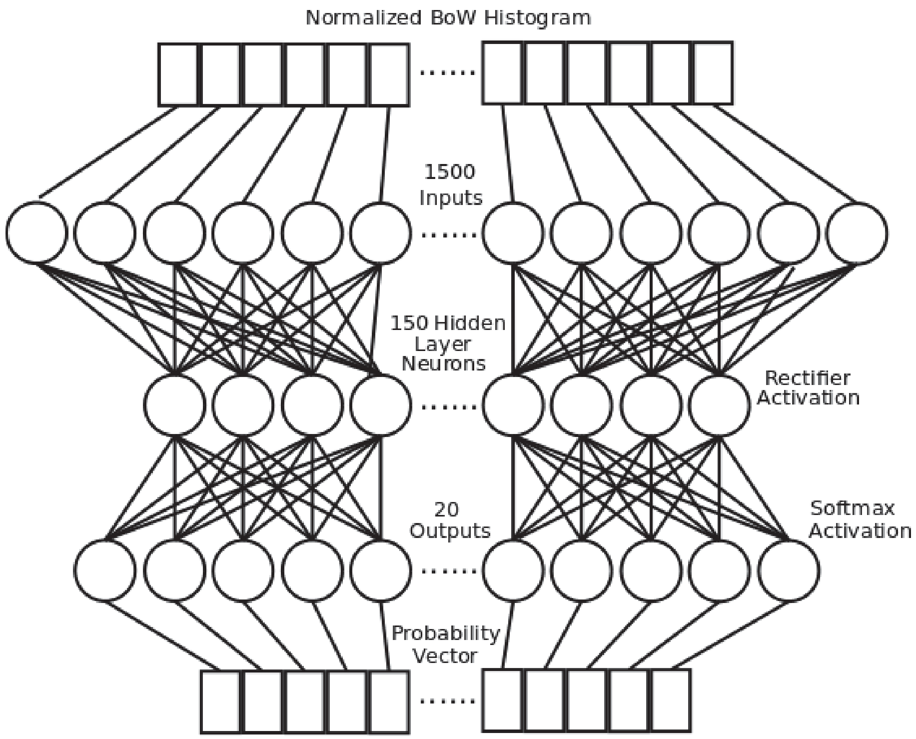

FIGURE 2.5: Neural Network Architecture

$$\text{rectifier}(x) = \max(0, x) \tag{2.1}$$

Here x denotes an input to the rectifier function. This input value in our implementation could be considered as the output generated from the first layer in neural network as shown in (2.2).

$$x_i = \sum_j W_{i,j} x'_j + B_i \tag{2.2}$$

Here x_i is the input to the rectifier function at ith hidden layer neuron and x'_j is the jth element in BoW descriptor in the input layer. $W_{i,j}$ represents the weight parameter corresponds to the ith input layer element and jth hidden layer neuron, and B_i denotes the bias value corresponds to the ith hidden neuron.

According to the rectifier function (2.1), the hidden layer neurons produce 0 for each negative input value while they output the input as it is for the positive values.

Since the goal of this place recognition technique is to produce probability estimates for each subregion in the shopping centre, a probabilistic approach is required to be adopted when designing the neural network model. Therefore, in order to produce a categorical distribution with each category being a subregion, the softmax activation function is employed at the output layer of the feedforward neural network (2.3).

$$\mathrm{softmax}(x)_i = \frac{e^{x_i}}{\sum_{j \in L} e^{x_j}} \tag{2.3}$$

Here x denotes the input to the softmax function while, L and i stand for the set of neurons and index in the output layer respectively. The softmax function at the output layer of neural network produces a 20 element vector, where each value in this vector is an approximation to the classification probability of the corresponding class [50].

2.3.3.1 Training Phase

Number of epochs and training data batch size are important parameters to be considered in a neural network training phase. In the context of neural networks, an epoch corresponds to a single pass (forward and backward) of the complete training set, while a batch is defined as a subset of training data. Given these parameters, the training of a neural network occurs throughout the number epoch cycles (outer loop), where each epoch contains i number of forward/backward passes (inner loop) through the network, as given in 2.4,

$$i = \frac{\text{number of training samples}}{\text{batch size}} \tag{2.4}$$

In our network model, number of epochs is set to be 450 and batch size to be 15. In the data collection phase we collect 128 images for each subregion (class) in the floor. As the floor is being divided into 20 subregions, the size of the complete training data set is 2560. Therefore, the two loops that have been employed in training the given neural network include an outer loop with 450 steps for epochs and an inner loop with 170 steps as the size of complete training data set is 2560. At each iteration of the inner loop 15 random training data rows are selected from the loaded training data set to feed into the network. In order to receive consistent results we set a loop dependent seed value before calling the random selection function.

This network model adopts the well known cross entropy function to model the cost of our network output at training iterations (2.5).

$$H(l,p) = -\sum_{x} l(x) \log p(x) \tag{2.5}$$

Here, $H(l,p)$ denotes the cross entropy between probability distributions l and p. The distribution l is considered to be a binary vector which represents the correct label for a given training data row, while p is the output probability distribution from the neural network. The cross entropy provides a measure for the difference between correct answer (class label) and the output of neural network for a particular BoW descriptor which belongs to a training image.

In order to minimize the difference between true class label vector and estimated probability distribution, we use backpropagation in conjunction with Adam algorithm [51] at the training stage. Adam (derived from adaptive moment estimation) is a recently introduced algorithm which is capable of optimizing objective functions efficiently. This method performs stochastic optimization that only requires first-order gradients with little memory requirement. Adam employs individual adaptive learning rates which are computed for parameters from the estimates of first and second moments of the gradients.

This algorithm is empirically evaluated with different machine learning models, including logistic regression, multilayer fully connected neural networks as well as CNNs, where its efficiency has been demonstrated. Further information related to the steps of the Adam algorithm, it's advantages and the results demonstrated in the empirical studies performed for evaluating its performance could be found in [51].

2.3.3.2 Testing Phase

Once we complete the training of our neural network, we test it using test data collected on a different day to guarantee the inclusion of time based dynamic visual properties to the test images. Such properties could be identified as the illumination variations due to the changes in weather condition and lighting, changes to temporary posters/advertisement boards and crowd density based on the day of the week.

Once the test images are collected, BoW descriptors from these images were extracted by following the same procedure used for training images, except the feature clustering operation. Feature clustering is not required for test images, as we use the same vocabulary created for extracting BoW descriptors from training images. It is mandatory to use this same vocabulary for generating test descriptors in order to preserve word identities in BoW model. Once the test descriptors are generated, they are fed into the trained neural network for processing and subsequently the output layer of network generates a probability distribution per descriptor, representing the estimates for each subregion.

2.4 Experiments and Results

2.4.1 Setup

Data collection in the shopping centre for our experiments is performed by capturing video streams of 20 different shop fronts and similar regions, on different days and at different times. Reasons for capturing video streams are the capability of conveniently extracting as many images as possible for training and the presence of motion blur in some images which can be experienced in a practical implementation. An important factor to consider when selecting training images is the amount of information they provide for training the model. As the images are extracted from video streams, similar images (adjacent frames of video) were commonly observed. Such near duplicate images are excluded from the training set, as they do not introduce new information to the network model. Additionally, the images with extreme occlusions are also avoided as they do not contain valid information. We used Galaxy Tab A Android tablet to collect the video streams and made sure the data collection happened under varying lighting and shop front setups. In order to ensure these characteristics, we collected data before and after Christmas time, which introduced a considerable variety for the training images. We collected training video streams on three separate days and test video on a completely different day. Once the video streams were collected, they were separated into images and manually labelled based on the subregion in the floor.

We performed the manual classification of collected images by renaming their respective JPEG file names in order to utilize them at the training and testing stages. Once the manual classification is completed, corresponding BoW descriptors are generated using an application developed in C++. These BoW descriptors are normalized and loaded into the network model for training and testing. The presented three layer feedforward neural network is developed completely based on Python Tensorflow framework. Tensorflow is a well documented framework which facilitates rapid development of neural network models. It also provides the capability to modify an implemented network model conveniently in order to analyse its behaviour. We performed many experiments for understanding the neural network characteristics by adjusting parameters such as epoch count and batch size, which is elaborated in the next subsection.

In order to develop all software and algorithms in this work, we used C++ OpenCV 2.4.8 and tensorflow implementation in Python 2.7 on an Intel Core i5 machine with a memory of 8GB. In addition to that, MATLAB 8.6.0.267246 (R2015b) is used for various data analysis tasks in order to obtain a good understanding on the BoW descriptors related to training and testing images.

2.4.2 Characteristics of The Neural Network

Studying the neural network output accuracy by assigning different values to parameters such as epoch count, batch size, number of neurons in the hidden layer enables us to experimentally analyse the network characteristics. Applying the suitable activation function for network layers as well as using the most effective optimization algorithm are also key factors which effect the final result and efficiency of the network. In the conducted experiments, a clear trend in accuracy could be observed when the number of training epochs

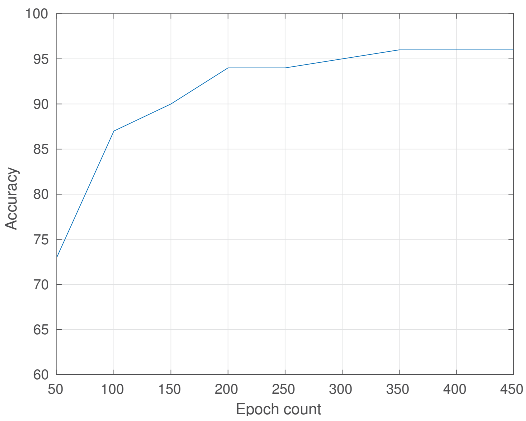

FIGURE 2.6: Accuracy vs epoch count

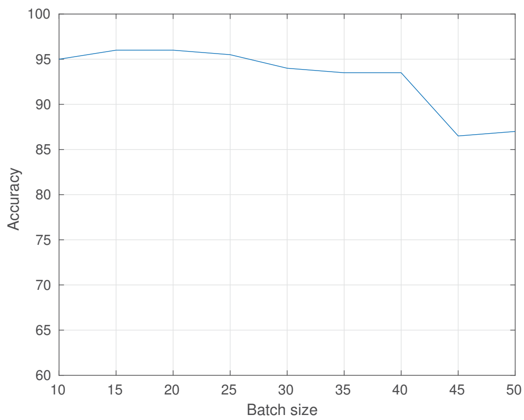

FIGURE 2.7: Accuracy vs batch size

is adjusted. As shown in Fig. 2.6, it is evident that the accuracy improves and converges to a stable value when the number of epochs is increased. Throughout this experiment the batch size is kept at a constant value of 15. The reason for this behaviour could be identified as the large number of training steps the neural network undergoes with the increase of epoch count. However, the accuracy converges to a stable value when the epoch count is continuously increased, as the number of training data is constant and there is no new information to learn further. We have also examined the behaviour of the proposed neural network by changing the batch size. The variation of neural network accuracy for different batch sizes are shown in Fig. 2.7. Here the epoch count is kept at a constant value of 450. As seen in this graph, the change in batch size does not result in a clear continuous trend in the accuracy of the presented neural network, as in the case with the number of epoch count. However, a sudden reduction in accuracy could be observed when the batch size is increased from 40 to 45, which might be due to a local minimum during the optimization process used for the training of the neural network.

The outer loop size and inner loop size in training phase, are directly influenced by epoch count and batch size parameters respectively. Therefore, increment in epoch count as well as reduction in batch size resulted in a higher training time as expected. However, the testing time is not affected by these changes in any way, as epochs and batch sizes are strictly related to the training phase. In the testing phase, our feedforward neural network took approximately 6.4 seconds to process 200 normalized BoW descriptors where each descriptor represents a single test image, hence takes around 32 milliseconds for one image.

We also could observe considerable accuracy variations as we changed the size of the hidden layers. According to our results, the neural network produced an accuracy of 96.0% for 150 hidden layer neurons, 95.0% for 300 neurons and 77.0% for 450 neurons, when epoch count and batch size are kept at 450 and 15 respectively. We could observe a growth in training time with the increase of hidden layer neuron count.

2.4.3 Results

The classification accuracy demonstrated by our place recognition technique is 96.0% for the test images captured from indoor shopping centre. The test data we have employed for this experiment consists of clear shop front images, as well as images with moderate to extensive motion blur. As the images were collected from a hand held tablet device and preprocessed for removing the bottom portion, they do not include the presence of a large crowd. However, when this method is employed in a practical system for tracking elderly people, an image feed from a camera which is mounted on a mobile wheeled walker could be expected as explained in Section 2.5. In such a scenario, the presence of crowd in captured images is possible.

However, based on the results obtained from the experiments, it could be seen that our technique is capable of correctly estimating the location using the clear images, as well as images with a moderate motion blur. Fig. 2.9 illustrates test images with moderate motion blur which were classified successfully. However, when an extensive motion blur is present in test images, the neural network tends to perform the classification incorrectly. The reason for this behaviour is the inability to extract correct features from the test images, due to the high level of distortion. Fig. 2.10 shows incorrectly classified test images with extensive motion blur. However, the presented place recognition technique demonstrates a competitive accuracy when estimating the cell based location using a reasonably clear image sequence captured near a shop front (Fig. 2.8).

For the comparison purpose, we tested the classification accuracy of Support Vector Machines (SVM) for the same test images we used in our approach. In order to train the SVM, training BoW image descriptors generated in our method were used. The testing was done similarly using the same test BoW descriptors, where it produced an accuracy of

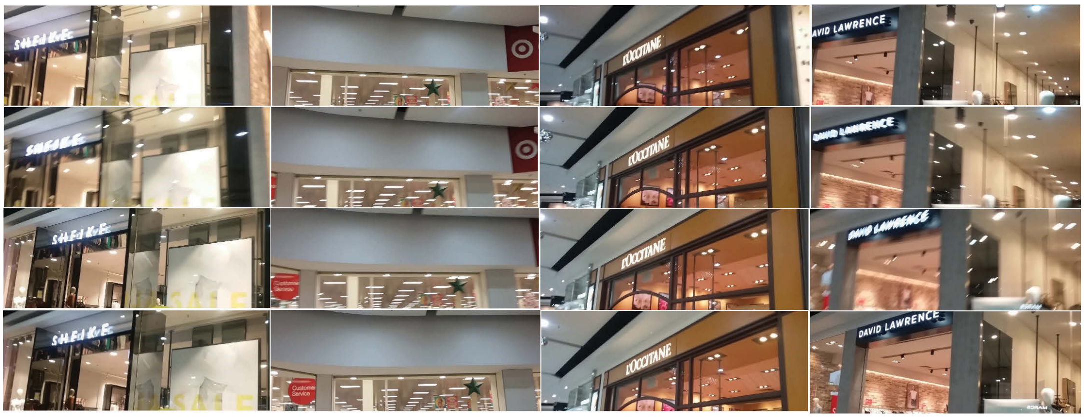

FIGURE 2.8: Sample image sequences near different shop fronts

FIGURE 2.9: Images with a moderate motion blur

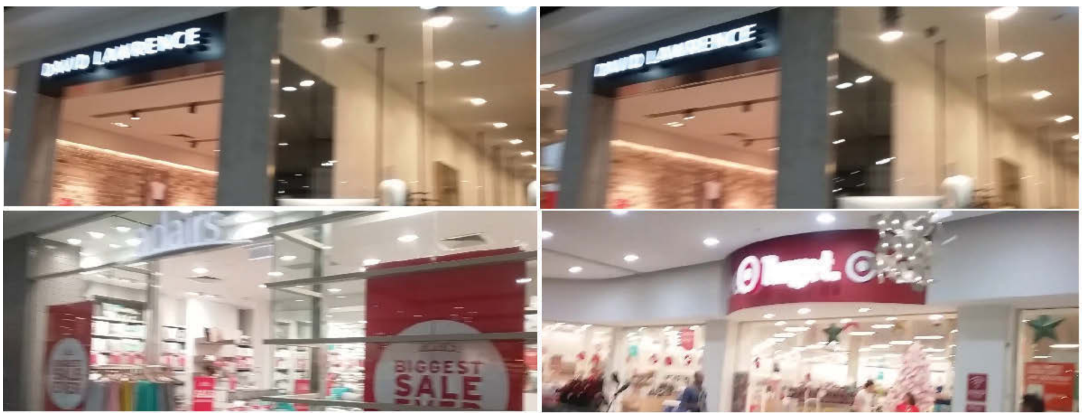

FIGURE 2.10: Images with an extensive motion blur

93.5% (Table 2.1). Therefore, it could be concluded that the approach presented here performs better than SVM in terms of place recognition task as it demonstrates an accuracy of 96.0%. These accuracy values are calculated based on (2.6).

$$\text{accuracy} = \frac{\text{number of correctly classified images}}{\text{total number of test images}} \text{x}100\% \tag{2.6}$$

TABLE 2.1: BoW-FFNN accuracy comparison with SVM

Method	**Accuracy**
Normalized BoW with FFNN	96.0%
Support Vector Machines	93.5%

2.4.3.1 Contrast and Brightness Variation Effect

The changes in illumination levels inside an indoor environment which occurs based on the lighting condition is an important dynamic property, which requires the attention of any vision based indoor place recognition technique. In order to simulate this property for evaluating its influence on the accuracy of the proposed neural network model, the collected test images are artificially modified by changing contrast and brightness levels. The contrast and brightness properties of the test images are modified by changing the α and β parameters according to (2.7),

$$O(i, j) = \alpha I(i, j) + \beta \tag{2.7}$$

where $O(i, j)$ represents a pixel in ith row and jth column in the output image and $I(i, j)$ is the pixel in input image. α and β parameters are referred to as gain and bias, which control the contrast and brightness of an image respectively. However, according to (2.7), $O(i, j)$ may be assigned values which are beyond the valid digital image pixel value range (0-255). In order to avoid this behaviour, such out of the range values are assigned the closest intensity extreme (0 or 255). In this experiment, 9 test image sets are generated by modifying the contrast and brightness levels of original test images. In this process, each

FIGURE 2.11: Test images with different illumination levels. Top row shows randomly selected three original test images, while second, third and fourth rows show same images modified with (α = 0.5, β = -50), (α = 1.5, β = 50) and (α = 2.0, β = 100) parameter values respectively.

test image set is manually assigned a unique combination of gain and bias values, and all images in each such set are subjected to the contrast and brightness changes, based on their corresponding gain and bias. Fig. 2.11 illustrates randomly selected original test images together with the corresponding modified images which are subjected to contrast and brightness changes.

Once different test image sets are generated by modifying original images, where each image set represents a unique value pair for gain and bias parameters, BoW descriptors are extracted using these modified test images by following the same procedure as illustrated in Fig. 2.4. These test BoW descriptors are normalized and subsequently tested with the network model that is trained using original training data. The table 2.2 demonstrates the classification accuracy for each test image set with different gain and bias values.

According to the neural network accuracies shown in table 2.2, it is evident that the accuracy is reduced when the contrast and brightness of test images are changed significantly from the original states. However, the presented approach still demonstrates reasonable accuracies for smaller variations of contrast and brightness levels in test images.

TABLE 2.2: Accuracies for different gain and bias values

Test set	Gain (α)	Bias (β)	Accuracy
1	0.5	0	89.5%
2	1.0	-50	94.0%
3	0.5	-50	78.0%
4	1.5	0	78.0%
5	1.0	50	93.0%
6	1.5	50	55.5%
7	2.0	0	45.5%
8	1.0	100	68.5%
9	2.0	100	13.5%

2.5 Conclusion

This Chapter presents an image based probabilistic place recognition method for a crowded indoor environment employing Bag of Words model and a three layer feedforward neural network. The image data are collected from a crowded shopping centre located in Sydney, Australia throughout multiple days. The Bag of Words approach is used for generating image descriptors and subsequently the generated training image descriptors are fed into the proposed feedforward neural network as inputs. Once the training is completed, the performance of the network is evaluated by passing test image descriptors through it and comparing the output with correct class labels. The output generated from the last layer of neural network is a vector of probabilities, where each element represents the likelihood of corresponding region being the correct region to which the test image belongs.

The place recognition technique presented here is required to fulfil the need of an indoor localization system, as the goal of this work is to develop a technique for locating elderly people in a crowded shopping centre. Therefore, a real time place recognition technique developed using this method is required to be deployed in a mobile hardware platform that remains with the elderly person all the time. In practice, the most feasible way to accomplish this task is to integrate such a mobile hardware system into the wheeled walker. Even though the accuracy of proposed vision based place recognition technique demonstrates competitive results for collected test images, there exist several challenges when such a practical scenario is considered.

The most fundamental requirement in any computer vision based localization system is capturing images from the indoor environment for processing and deriving position information. In a highly crowded environment such as a shopping centre, it is impossible to prevent such a system from capturing images which include human presence. As image capturing usually occurs without the consent of people in the surrounding area, the primary shortcoming in a practical computer vision based localization approach is the possibility of privacy invasion.

Assuming that the camera is fixed to the wheeled walker of an elderly person, it is unrealistic to anticipate the walker to be facing only in one direction when located in a given subregion. This inconsistent camera orientations produce completely different image streams, even when the person is inside the same area. In addition to that, high occlusions which could appear in images due to crowd is another challenge that needs to be addressed. If the camera is fixed in the walker platform, there exist a high possibility that camera will be completely covered by the crowd in the shopping centre. Additionally, the motion blur which appears in images is largely influenced by the way a person handles the wheeled walker in their motion. Sudden change of direction or speed variations of the walker platform due to the blockages of crowd or any similar behaviours will add an extensive motion blur to the captured images. As a consequence, the proposed method will be able to estimate the correct location only when the camera captures a moderately clear image of the shop front. In practice, clear shop front images will be captured intermittently when an elderly person is roaming inside the shopping centre, while pushing the camera mounted walker platform. However, such a behaviour will not be favourable for our application as retrieving the real time location of the elderly person is not feasible. In order to address this problem, the wheeled walker user will be required to assist the system by facilitating the camera to capture a clear view of the shop front. For an example, when a walker user is lost, he/she can point the camera mounted wheeled walker to the closest shop front and wait until the space is cleared. This will facilitate the camera to capture clear images of the shop front setups, subsequently enabling the place recognition system to estimate the position of a subject accurately.

In summary, the proposed image based place recognition system is capable of producing accurate location estimates, provided that the wheeled walker user is willing to assist the system functionality as mentioned above. However, expecting the walker user to be aware of these requirements and anticipating his/her assistance may add an extra burden on the user. This additional user responsibility, and the aforementioned privacy concerns are best avoided in a practical system. As a solution, the next Chapter presents an indoor localization method which is completely independent of the wheeled walker user, as well as protects the privacy of the surrounding crowd.

Chapter 3

Wi-Fi Based Localization: Theoretical Framework

3.1 Introduction

The primary shortcoming in the image based place recognition approach presented in Chapter 2 is the possible violation of privacy of people. In the imaging processing pipeline described, only the feature vectors associated with the images are important, therefore if the software is designed such that the raw images are never stored, it could be argued that there are no privacy issues associated with the deployment of such a system. However, in an image based place recognition system, perception that privacy is being violated is unavoidable, especially in a public place like a shopping centre as capturing of image feed from the surrounding area is a fundamental requirement. In addition to that, the presented place recognition method requires the user to point the camera to a nearby shop front and wait until the crowd clears to capture a clear image. This adds an additional burden on the user as he/she requires to be aware of these system requirements and act accordingly. Therefore, a robust indoor localization system which does not violate the privacy of the surrounding crowd as well as independent of the user's involvement in the process is more suitable in the context of the proposed application.

This Chapter presents a robust Wi-Fi based indoor localization method which is resilient to vision based dynamics, that could be directly implemented in a small scale hardware system to address the requirement of assisting care givers of aged care facilities to track elderly residents in an indoor crowded shopping centre. As the carer who is responsible for the group of elderly people confirms the presence of group members through visual contact, a coarse cell (subregion on the floor) based indoor localization technique is adequate for the carer to retrieve approximate locations of the people in his/her care. Therefore, the method proposed in this Chapter exploits the variation of Wi-Fi signal strengths based on the subregions on the floors, and determines the most probable cell based location of a subject by analysing the received signal strengths. Furthermore, as the users rely on wheeled walking frames, their motion is relatively slow and predictable, making it possible to use a motion model for enhancing the localization accuracy. Although many Wi-Fi based localizations systems have been proposed and evaluated in the literature, the above unique requirements makes it possible to design a localization system that is extremely robust and is computationally lightweight. The theoretical framework of the proposed Wi-Fi based indoor localization system is elaborated in Chapter 3. The next Chapter will focus on its practical implementation and experimental evaluation.

The reason for choosing Wi-Fi for our indoor localization method over many other possibilities such as Bluetooth [20], Zigbee [19], ultrasonic sensors [52], Odometry [13], RGB-D sensors [14] and LRF [15–17], is primarily due to the ubiquity of wireless access points (WAPs) in indoor environments such as shopping centres. While the shortcomings of LRF, Odometry and RGB-D are discussed in the previous Chapter, concepts and issues related to wireless signal based localization approaches such as Bluetooth, RFID, Ultra-Wide Band (UWB) and ZigBee are discussed in Section 3.2.

The core algorithm of the localization technique we present could be broadly categorized under 'Wi-Fi fingerprinting', a term used for characterizing methods which measure similarity between current Wi-Fi fingerprint (vector of received signal strengths for detected Wi-Fi access points at a particular location) and previously collected fingerprints employing a chosen distance function, and return locations corresponding to most similar fingerprints [53]. In contrast to this distance function based traditional fingerprinting approach, we selected Bayesian framework to fuse information from Wi-Fi signal strength readings to readings from the altimeter and prior knowledge of the floor plan and associated expected nominal motions. The probability density function of signal strength measurements in a given cell collected over multiple positions and orientations within the cell over multiple days, computed using Kernel Density Estimation (KDE), was found to adequately capture the likelihood of receiving a given signal strength reading. The proposed technique is, therefore, capable of producing the probability for a subject being in a particular cell in the floor making it easy for the carer to track down a resident even under significant disturbances to the Wi-Fi signal.

This Chapter is organized as follows: Section 3.2 describes the related works in the context of wireless signal based indoor localization methods, while Section 3.3 presents our localization algorithm which is based on Bayes rule and Kernel Density Estimation, Section 3.4 describes the integration of motion models with generated location estimates. Section 3.5 concludes the Chapter.

3.2 Literature Review

This Section explores fundamental concepts behind the localization methods which rely on wireless signals such as Bluetooth, Zigbee, RFID, Ultrasonic, UWB and Wi-Fi sensors. Moreover, the challenges related to the existing techniques when applied to our application related to tracking elderly people in a public environment will also be investigated.

3.2.1 Bluetooth Based Indoor Localization

Bluetooth sensors and Bluetooth Low Energy (BLE) technology have been used in indoor localization systems considering their characteristics such as direct signal measurement capability, low cost and low power consumption [20]. Most of the Bluetooth based localization systems use lateration technique, which derives the subject's (Bluetooth receiver) position using distances between receiver and each Bluetooth transmitter [20][54]. The distances between Bluetooth receiver and the transmitters are derived from the Received Signal Strength Indication (RSSI) at the receiver's end. In order to derive the position from calculated distances (lateration), the coordinates of each Bluetooth transmitter must be known prior to the localization. This could be identified as a drawback in such systems, as in practice, this is not suitable for use in the application scenario associated with this thesis, where the installation of explicit Bluetooth transmitters is unlikely to be authorized as well as cost prohibitive.

3.2.2 Zigbee Based Indoor Localization

Use of a network of Zigbee modules is another popular approach for indoor localization [18, 19, 55]. Similar to Bluetooth, these Zigbee based localization systems exploit the RSSI values for estimating the location of a Zigbee receiver. Fingerprinting approach is a popular technique in Zigbee based localization methods where a vector of current RSSI values are compared to a database of previously collected RSSI vectors (each corresponds to a known location) and then the most similar one/ones are retrieved for deriving the location, as explained in [19][55]. The *ZigBEACON* system proposed in [18] employs a method which calculates distances between mobile node and reference Zigbee nodes placed in the environment, and derives the position of mobile node using distance values to the k-nearest reference nodes.

However, a Zigbee based indoor localization system requires the installation of Zigbee modules within the environment prior to the real time localization. Similar to the Bluetooth based approaches retrofitting new Zigbee modules to the existing infrastructure is a cumbersome and expensive exercise.

3.2.3 RFID Based Indoor Localization

As another popular approach for indoor localization, Radio Frequency Identification (RFID) technology has been exploited in many previous works [56–58]. In contrast to passive RFID tags where an internal power source is not available, the majority of the RFID based localization systems use active tags which are capable of broadcasting signals. The RFID receiver module can be used to capture these signals and extract information such as tag identifier, signal strength, angle of arrival (AoA) and phase of arrival (PoA) for deriving the location information of the receiver.

While RFID based indoor localization techniques have received considerable attention due to the low cost of RFID tags, it is still required to install them in the target environment before the localization. This requirement limits the use of RFID techniques in our application, where the target environment is a public shopping centre.

3.2.4 Ultrasonic Based Indoor Localization

Ultrasonic sensors have been previously used in indoor localization techniques exploiting their distance measurement capabilities. In order to measure the distance between a sensor and an environment feature such as a wall or other obstacle, it follows the Time of Flight (ToF) principle assuming a constant speed of sound.

The extracted distances between the sensor and the surrounding features are then processed to derive the position of a subject in the environment. In order to achieve this, many works which use ultrasonic sensors follow perception-based localization where the clues that are detected in the environment are matched against a map of that environment [52][59]. However, extracting clues (observing distances) in the environment is not trivial as ultrasonic sensor readings contain a considerable amount of noise [52].

3.2.5 Ultra-Wide Band (UWB) Based Indoor Localization

Recent developments in indoor localization technologies include, UWB based radio systems which claim to produce location estimates with accuracy better than narrow band Wi-Fi based systems. Operating in both Line of Sight (LoS) and non-Line of Sight conditions, not interfering with narrow band transmission systems, low power consumption and low cost are the main advantages of UWB which makes it a preferable technology to use in indoor localization applications. The most commonly adopted principle behind UWB based localization is ToF which exploits the signal transmission time between UWB transmitter and receiver for deriving the position of a subject.

Previous works such as *EIGER* positioning system proposed by Kolakowski et al. [60] and indoor localization system proposed by Zwirello et al. [61] exploit the strengths in UWB based systems. However one notable shortcoming in similar systems is the requirement to install specialized hardware modules in target localization environment thus making the UWB technology unfavourable for our application.

3.2.6 Wi-Fi Signal Based Indoor Localization

Majority of the indoor localization methods that exploit Wi-Fi signal attributes can be primarily categorized into four groups (a) propagation model based localization methods; (b) angle of arrival based localization methods; (c) time of flight based localization methods; (d) fingerprinting based localization methods.

3.2.6.1 Propagation Model Based Localization Methods

A Wi-Fi propagation model is a mathematical formulation which characterizes the propagation of a signal as a function of distance, attenuation factors and other relevant parameters. The fundamental principal behind propagation model based localization is estimating the location by analysing signal characteristics that depend on these parameters under a chosen indoor Wi-Fi signal propagation model. In this localization technique, choosing a proper propagation model is crucial as it represents the Wi-Fi signal characteristics over the environment that are used to estimate the subject's position. Free Space Path Loss model (3.1) can be identified as the foundation for many other propagation models which have been proposed in the literature for indoor environments.

$$PL_{FS}(d) = 20log_{10}(d) + 20log_{10}(f) - 27.55 \tag{3.1}$$

where $PL_{FS}(d)$ denotes the free space path loss at transmitter-receiver distance d, and f is the signal frequency which is considered as a known parameter for a given wireless signal.

Even though free space path loss model is the fundamental model for deriving other propagation models, it is a theoretical model which may not be suitable for real life situations. Its incapability of producing reasonably accurate results when applied as a propagation model for indoor localization has been empirically presented in work by Shchekotov [62].

In contrast, Log Distance Path Loss model (3.2) can be identified as a practical loss model which considers the fact that the received power decreases logarithmically with distance.

$$PL_{LD}(d) = PL(d_0) + 10\gamma log_{10}\frac{d}{d_0} + X_g \tag{3.2}$$

where $PL_{LD}(d)$ denotes the log distance path loss at a given distance d between transmitter and receiver while $PL(d_0)$ represents path loss at the reference distance d_0. The constant γ is the path loss exponent which depends on the signal propagation environment and X_g represents a normal random variable reflecting the attenuation caused by flat fading. Flat fading is caused by partial cancellation of a signal by itself as the signal arrives at the receiver by two different paths. In case of no fading, X_g is set to 0. Readers are referred to [63][64] for further information on Free Space Path Loss Model and Log Distance Path Loss Model.

Even though Log Distance Path Loss model represents the true nature of signal propagation, more dependencies which are not captured in this model exist for the received signal strength at the receiver's end. Most prominent dependencies for received signal strength in indoor environments could be identified as the floor separations and walls in between Wi-Fi transmitter and receiver. As a solution, Floor Attenuation Factor (FAF) path loss model by Seidel and Rappaport [65], and Wall Attenuation Factor (WAF) model by Padmanabhan and Bahl [66], have been introduced for accommodating attenuation due to floor and wall separations (3.3) (3.4). The Wall Attenuation Factor model has adopted the original Floor Attenuation Factor model for formulating the effect of attenuation due to wall separations [66].

$$\bar{P}L(d) = PL(d_0) + 10\gamma log_{10}\frac{d}{d_0} + FAF \tag{3.3}$$

where $\bar{P}L(d)$ denotes mean path loss at transmitter receiver separation distance d, and FAF stands for the Floor Attenuation Factor that depends on the number of floors and building type. A similarly constructed signal path loss equation which relies on a given attenuation factor could be employed for deriving the signal power received at a receiver. The

WAF model derived this way for estimating signal power $P(d)$ received at a transmitter-receiver distance of d, could be formulated as (3.4).

$$P(d) = P(d_0) - 10\gamma log\frac{d}{d_0} - \begin{cases} n_w * WAF & n_w < C_w \\ C_w * WAF & n_w \geq C_w \end{cases} \tag{3.4}$$

where, C_w stands for the maximum number of walls up to which the attenuation factor makes an effect, n_w is the number of walls between transmitter and receiver and WAF is the Wall Attenuation Factor. Here WAF value depends on the construction material of the wall, hence it is derived empirically. Methods for empirically deriving these WAF and FAF values have been proposed in previous works and the effectiveness of these indoor models has been compared with Log Distance Path Loss model to demonstrate its adequacy for indoor environments [66][67]. These empirical methods involve collecting signal strength values under two scenarios; i.e., for FAF, a data set is collected at the same floor where the access point is established and the other data set is collected at a different floor. Similarly, for WAF, signal strength data are collected when the transmitter and receiver have line of sight and when obstacles (walls) are present between the transmitter and receiver. Then these collected data sets are compared for calculating the differences and deriving the FAF and WAF values. As WAF and FAF values are highly dependent on the material type of obstacles, such propagation models demonstrate high ambiguity in crowded environments where the people are moving continuously in an unpredictable manner. Further detail on WAF and FAF models are available in [65, 66].

A well known indoor localization system based on Wi-Fi Wall Attenuation Factor propagation model is *RADAR*, proposed by Padmanabhan and Bahl (Microsoft Research) [66]. *RADAR* employs empirical signal strength measurements as well as the theoretically generated signal strengths by a propagation model, in order to find the position of a subject in an indoor environment. In the propagation model based approach, it generates the signal strength values using the WAF model such that the data points corresponding to locations are spaced uniformly on the floor. After that, it estimates the user's position by matching the signal strengths measured in real-time to the theoretically computed signal strengths at these locations. In contrast to applying the same propagation model to every access point, the work proposed by Alonso et al. [68] introduces a method to improve the accuracy of localization by employing different propagation models for each of the access points. This is done by experimentally estimating the parameters of the propagation model for each access point and imposing these corresponding models when measuring the distance values from access points to the subject. The calculated relative distances are then used to extract the position of the subject by employing circular positioning algorithm (3.5).

$$\epsilon = \sum_{i=1}^{M} \left(\sqrt{(x_i - x)^2 + (y_i - y)^2} - \tilde{d}_i \right)^2 \tag{3.5}$$

where ϵ is the cost/error value that needs to be minimized, M is the total number of access points, (x_i, y_i) is the position of each access point, (x, y) is the position of subject and $\tilde{d}_i$ is the estimated distance between the subject and corresponding access point i. This non linear function can be minimized iteratively using an appropriate method such as gradient descent for estimating the optimal position of subject (x, y). In addition to the circular positioning algorithm, this work also considers the hyperbolic positioning algorithm (3.6) which converts the non-linear localization problem into a linear problem for calculating the position $\hat{\bar{x}}$, but concludes that circular positioning algorithm performs better based on the observed results. Further information on circular and hyperbolic positioning algorithms can be found in [68].

$$\hat{\bar{x}} = (Z^T Z)^{-1} Z^T \tilde{b} \tag{3.6}$$

where $\hat{\bar{x}} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}$, $Z = \begin{bmatrix} 2x_2 & 2y_2 \\ \vdots & \vdots \\ 2x_M & 2y_M \end{bmatrix}$ and $\tilde{b} = \begin{bmatrix} x_2^2 + y_2^2 - \tilde{d}_2^2 + \tilde{d}_1^2 \\ \vdots \\ x_M^2 + y_M^2 - \tilde{d}_M^2 + \tilde{d}_1^2 \end{bmatrix}$

Even though indoor localization methods based on Wi-Fi propagation models have been extensively studied in previous works, there are several inherent drawbacks in this technique. As stated previously the main issue with signal propagation models in localization application is their inability to capture the signal attenuation effect due to the presence of people in indoor environments like shopping centres. The complexity of this problem

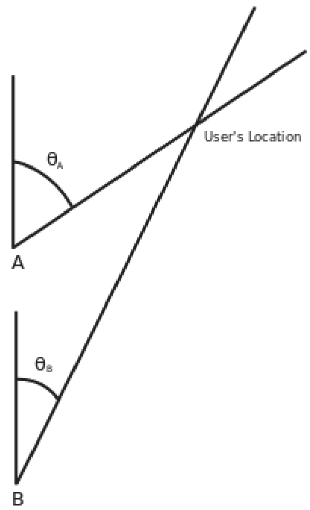

FIGURE 3.1: Angle of Arrival

becomes significant, given the fact that people in such crowded environments have an unpredictable motion throughout the building. Similarly, the positions of furniture and other objects in indoor environments also change over time and this also causes the same issue as it becomes challenging to establish a consistent propagation model which represents the correct attenuation. Another drawback in a propagation model based localization method is the requirement of prior knowledge of Wi-Fi access point locations.

3.2.6.2 Angle of Arrival Based Localization Methods

The Angle of Arrival method (also referred to as Direction of Arrival) is used for localizing a subject, by estimating the angle of incidence at which the Wi-Fi signals arrive at the receiver's end [69]. In the most fundamental form of Angle of Arrival (AoA) based localization methods, the position of a subject is derived by considering the intersection of two lines of bearing (which represent the incident signal on access points, A and B) as illustrated in Fig. 3.1. This method, which is known as direction based triangulation, could be used to derive a more accurate location estimation for the subject, by employing more than two access points (In 2D localization scenario).

In most of the practical systems, AoA is determined by sampling the received signal using multiple element antenna arrays or rotating antennas (for simulating a larger array). In order to derive AoA, the Time Difference of Arrival (TDoA) between elements of the antenna array is estimated. TDoA could be estimated by measuring the received signal phase difference at each array element. The complete procedure involves directly measuring the arrival time of signal at each antenna array element and calculating the TDoA between these elements and converting this information to an AoA measurement [69].

A major challenge in localization based on AoA technique is multi-path phenomena, where the emitting signal is reflected/refracted due to obstacles and arrives at the receiver through multiple paths. Multi-path propagation becomes significant in indoor environments due to the obstacle rich surrounding; thus it requires special measures for mitigating the effect on AoA estimation.

SpotFi [70], an Angle of Arrival based localization system proposed by Kotaru et al. claims to produce a decimetre level accuracy by incorporating novel filtering and estimation techniques for identifying AoA of direct path between the subject and AP. *SpotFi* does this by assigning likelihood values for each path (including the indirect paths and direct path) depending on how likely the particular path is the direct path. The calculations related to localization in *SpotFi* is done by exploiting the Wi-Fi signal information received at the Access Point antennas. A useful observation in indoor Wi-Fi signal propagation is the estimated Angle of Arrival and Time of Flight (ToF) values of direct path showing a much smaller variation over a few consecutive packets when compared to indirect paths [71]. *SpotFi* leverage this observation for estimating likelihoods for each path to be the direct path between the subject and access point. In order to perform the final localization *SpotFi* combines the direct path AoA estimates and their likelihoods corresponding to different access points. In addition to these values, it exploits the RSSI measurements for Wi-Fi signals received from each access point for the localization task. This system assumes the standard and widely used path loss model to relate RSSI to distance measurements (between transmitter and receiver) as described in [66]. All this information is then fused for localizing the subject. In contrast to the fundamental form of AoA based localization as illustrated in Fig. 3.1, this performs the localization by deriving the position that minimizes the following objective function (3.7).

$$\sum_{i=1}^{M} l_i[(\bar{s}_i - s_i)^2 + (\bar{\theta}_i - \theta_i)^2] \tag{3.7}$$

where M denotes the number of Wireless Access Points, θ_i is the direct path AoA of the ith access point, s_i is the observed RSSI by ith access point and $\bar{\theta}_i$ and $\bar{s}_i$ are the AoA and RSSI respectively that would have been observed at the ith access point if the subject was located at the position of interest. l_i is the likelihood value of most likely candidate for the direct path from ith access point. The goal of minimizing this objective function is to estimate the location that minimizes the deviation between AoA and RSSI values that would have been observed at each WAP if the subject was actually located at that position and corresponding values that are actually observed at each access point. Several drawbacks in *SpotFi* localization system concerning our application include the requirement of installing Access Points with the prerequisite of having at least three antennas and the prior knowledge of the positions of these Access Points. In addition to this, *SpotFi* also requires the Wi-Fi signal related information received at these AP antennas for estimating the location of the subject. In most public environments acquiring this information received at the already deployed access points is somewhat challenging.

Ubicarse [72], an AoA based localization system proposed by Kumar et al. uses standard triangulation for locating a subject, but instead of employing the subject as the transmitter and access points as the receivers, it treats the subject as the receiving end while access points are treated as transmitters. This system uses a SplitX2 hand-held tablet as the receiver and simulates the multiple antenna elements (as required by AoA based localization method), by twisting the tablet around its vertical axis. Then it generates the multi-path signal profiles (Relative observed signal power along each direction) for access points and isolates the corresponding direct paths for applying triangulation. In contrast to *SpotFi*, *Ubicarse* does not need to know the information received at the access points as the moving subject(hand-held tablet) is now acting as the receiving end. Even though the hand-held tablet is responsible for measuring the AoA in *Ubicarse*, the prior knowledge of access points' locations is still a prerequisite in this localization system. This requirement is not practical for our indoor localization system which operates inside a public environment where the prior knowledge of Wi-Fi access points is not available. Moreover, the

Ubicarse system expect the user to twist the hand-held device for simulating an antenna array for AoA estimation. This is impossible in our scenario as we are using a wheeled walker frame as the localization platform.

When considering the fundamental concept of AoA and implementations based on AoA localization method, it is evident that there are many challenges for such a system to be used in a public environment where the locations of access points are not known. Furthermore, the effect on Angle of Arrival in the presence of a larger crowd could be significant due to many moving obstructions. In such scenarios it could be nearly impossible to observe a direct path of signal propagation. Therefore, this will degrade the quality of estimated position as the direct path is crucial for most of the AoA based localization systems to perform.

3.2.6.3 Time of Flight Based Localization Methods

Time of Flight (ToF) method in Wi-Fi based localization technologies, exploits the signal propagation time between a Wi-Fi transmitter and a receiver in order to estimate the distance between these two nodes. After estimating distances between subject and each access point, it derives the position of subject using a technique such as trilateration (a method of estimating the location of a target employing relative distance measurements and geometry of circles).

The ToF value is estimated between two Wi-Fi nodes by considering the time interval between DATA and ACK frames. In Wi-Fi communication protocol, every DATA frame is acknowledged by the receiver's end with an ACK frame [73]. Since the internal interframe time between DATA and ACK frames (at the DATA frame receiver's end) is fixed under 802.11 standard, the time delay between DATA and ACK frames (at the DATA frame transmitter's end) can be used to infer the distance between two nodes.

The ToF measurements are considerably sensitive to noise since, at the speed of light, a measurement error of 1 μs result in a distance estimation error of approximately 300 meters [73]. These noises in ToF are caused by various sources such as time imprecision of off-the-shelf Wi-Fi devices and multi-path effect in indoor environments. Such a significant error in distance estimate will make positioning task impossible in a ToF based localization method. Techniques such as adaptive filters by Marcaletti et al. [73] and post processing using statistical estimators by Ciurana et al. [74]can be used to mitigate the effect of noise in order to produce a reasonable localization estimate. Even though the proposed methods indicate an enhancement of accuracy by reducing the noise the evaluation of those methods are carried out in controlled environments such as office rooms. The effect on ToF by a large crowd in public environments like shopping centres has not been captured in these evaluation. Moreover, even if the ToF methods are enhanced to generate a reasonably accurate result, the final localization is performed based on relative distance measurements between the subject and Access Points. This means the prior knowledge of Access Points' positions is a prerequisite in ToF based localization techniques as well. As stated previously, obtaining the precise locations of already deployed access points is not readily available in our localization application as we expect to perform the localization inside a public building.

3.2.6.4 Fingerprinting Based Localization Methods

Wi-Fi Fingerprinting method in localization technology is performed by comparing a current Wi-Fi fingerprint (A vector of Wi-Fi signal strengths from different Access Points at an instance) with a database of previously collected fingerprints based on the location, and then returning the locations which correspond to matching fingerprints [53]. If there is only a single fingerprint returned by the comparison method then the location corresponding to that fingerprint could be treated as the location of subject. In contrast, if multiple fingerprints are returned which are similar to the current fingerprint (obeying some similarity metric threshold), a final fingerprint location can be derived by combining the returned multiple fingerprints and corresponding locations.

The most fundamental form of fingerprinting could be characterized as a method of comparing the vector of received signal strengths with the stored fingerprints using a distance function (such as Euclidean distance and City Block distance) and then performing

K-Nearest Neighbour (KNN) algorithm to find the most similar fingerprints and corresponding locations. As the name states K-Nearest neighbour algorithm is used to find the most similar points to a given point in feature space (Received Signal Strengths could be treated as features regarding Wi-Fi based localization applications) in order to perform classification.

The choice of distance function and value for parameter K are crucial for KNN based fingerprinting methods. A comprehensive analysis of numerous distance functions and other parameters such as K value and thresholding for RSSI has been performed by Torres-Sospedra et al.[53]. This analysis demonstrates the suitability of alternative distance functions such as $S\phi renson$ distance and $NeymanX^2$ distance over traditional $Euclidean$ and $CityBlock$ distances for fingerprint comparison. Moreover, it concludes that applying a thresholding method to remove Wi-Fi signals with lower intensities is not necessary based on the calculated accuracy change in KNN based localization. Important contributions in this work are *Exponential* (3.8) and *Powed* (3.9) representations for signal strength data, in order to resemble the original nature of Wi-Fi signals and penalize fluctuations in strong signals (transmitted by closer access points).

$$\text{Exponential}_i(S) = \frac{\exp\left(\dfrac{\text{Positive}_i(S)}{\rho}\right)}{\exp\left(\dfrac{-s_{min}}{\rho}\right)} \tag{3.8}$$

$$\text{Powed}_i(S) = \frac{(\text{Positive}_i(S))^{\phi}}{(-s_{min})^{\phi}} \tag{3.9}$$

where,

$$\text{Positive}_i(S) = \begin{cases} (s_i - s_{min}) & \text{If access point}_i \text{ is present} \\ & \text{in the fingerprint } S \text{ and } s_i \geq \tau \\ 0 & \text{Otherwise} \end{cases} \tag{3.10}$$

where s_{min} denotes the lowest signal strength decreased by 1 dB considering all fingerprints, s_i is the signal strength of ith access point and S represents a received Wi-Fi fingerprint. The ρ and ϕ are case based parameters and τ is a threshold parameter.

Analysis performed in [53] demonstrates a reasonable accuracy improvement for localization in an indoor environment for certain distance functions as mentioned above. Even though the accuracy has been improved for these distance measures and novel data representations, approximately 7 meters of positioning error still exists according to the observed results. Furthermore, it evaluates these distance functions and data representations based on public 'UJIIndoorLoc' database which provides Wi-Fi signal strength data fingerprints with corresponding location coordinates. Therefore, it is arguable whether this would perform in the same way in a real life crowded environment such as a shopping centre.

The traditional KNN based fingerprint localization has been attempted to enhance with novel strategies such as Fuzzy K-Nearest Neighbour algorithm [75]. Even though the Fuzzy K-Nearest Neighbour algorithm used in work [75] has demonstrated a better performance over the basic KNN, this evaluation has been done in a simulated environment. Therefore its performance in a real life environment where a number of challenging dynamic obstructions are present with a more complex building structure is questionable. The suitability of KNN or derivations of KNN algorithm for localization in a crowded environment is uncertain due to the fact that KNN is a non-parametric algorithm. This means KNN does not make assumptions on the data distribution of interest(that is the RSSI distribution in Wi-Fi based localization scenario). This nature of KNN could be a disadvantage in scenarios where estimating an appropriate RSSI distribution for each access point per each location/region is possible.

In traditional fingerprinting methods, the computation time in Wi-Fi fingerprinting is a crucial factor since it compares the received Wi-Fi fingerprint with all other stored fingerprints. To address this, a solution has been proposed in [76] where it considers top ranked Access Points with highest RSSI as Important Access Points (IAP) in fingerprints when creating the database and in the phase of fingerprint comparison the stored fingerprints are considered for matching when it has the same IAP as the received fingerprint's IAP. Such strategies could be employed for reducing the computation and positioning time of the localization system.

In contrast to the traditional fingerprinting methods where the similarity between current Wi-Fi fingerprint and stored fingerprints are compared, employing a distance function, other approaches have been proposed based on techniques such as Artificial Neural networks[77][78], Random forest classifiers [79], Support vector machines[80] and Hidden Naive Bayes [81]. The basic concept behind these advanced approaches is to develop a representative model for RSSI distribution within the environment and utilize that model at the localization phase to classify a received Wi-Fi fingerprint. The model developing stage could be considered as an off-line training phase, which involves collecting Wi-Fi fingerprints with corresponding position labels and then feeding this information to train the model. For an example, the work done in [77] by Laoudias et al. evaluates different Artificial Neural Network models such as Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and Generalized Regression Neural Network (GRNN) for the implementation of localization method. Moreover, this work proposes Clustered Radial Basis Function (cRBF) model for the localization task where each centre of hidden layer neurons is set to the mean fingerprint (prototype fingerprint) value based on the location (class). As the RBF networks perform classification by computing the Euclidean distance between the input and hidden layer neuron centres (prototype fingerprints), cRBF reduces the number of hidden layer neurons dramatically when compared with a standard RBF networks where each reference fingerprint defines the centre of the corresponding neuron. Even though the proposed cRBF network demonstrates promising results, it has been evaluated under a typical office environment. Therefore its applicability as a localization method in a crowded indoor environment is still uncertain. In addition to RBF networks, Wi-Fi fingerprinting based localization has been addressed by employing a three layer feedforward neural network as proposed in the work [78]. Here, the Wi-Fi signal strength values in several directions are taken as the inputs. In order to verify the accuracy of this three layer neural network, it selects access point subsets from all available access points in the building, and performs the testing for each subset. According to the reported results, it is evident that the accuracy of the proposed network becomes highly sensitive to the selected access point subset. Similar to cRBF network based localization method, the three layer feedforward network proposed in [78], has also been verified in a controlled environment where a large crowd is not present.

Random forest classifier [79] and Support vector machines [80] have also been employed for Wi-Fi fingerprinting based localization and have demonstrated promising accuracy in controlled environments such as University buildings or simulated areas. Even though the accuracy is reasonable for such environments, the performance of these methods need to be verified in practical scenarios where a large crowd is present. A localization system based on Hidden Naive Bayes proposed by Song et al., employs a novel method to select best discriminating access points of each reference location which will be used as input features when the location is estimated [81]. The experiments related with this work have also been conducted in a university building hence its performance in a crowded environment requires further evaluation. However, model based Wi-Fi fingerprinting methods that incorporate techniques such as Artificial Neural Networks, Random forests, Support Vector Machines and Bayesian approaches are gaining preference over traditional KNN based fingerprinting due to their ability to develop a representative signal strength distribution model and produce a better location estimate using that model.

In summary, significant attention has been given for Wi-Fi fingerprinting as a localization method, over techniques based on propagation model, Angle of Arrival and Time of Flight as Fingerprinting does not require specialized hardware, line of sight to transmitter or prior knowledge of transmitters' locations [53]. Therefore we have chosen a Wi-Fi fingerprinting based approach as a suitable candidate for our indoor localization applications over other techniques.

3.3 Location estimation using the Bayes Rule

The signal propagation of a Wi-Fi access point in a crowded indoor environment like a shopping centre could be affected by a range of factors such as motion of people, repositioning of obstacles like wall partitions, advertisement boards, furniture and signal interferences from other wireless networks and electronic appliances. Primarily, the Wi-Fi access points that are located in a far distance will be highly susceptible for these effects as the signal strength becomes weaker with the distance due to attenuation. Consequently, the signal visibility of a distant Wi-Fi access point in a particular region could be practically uncertain. Therefore, identifying such access points and removing them from the data is important as those access points could introduce spurious information to the probability distributions generated in later steps.

Once the Wi-Fi signal strength data collection was completed, appearance frequency of each access point in a given data collection day was used to identify the prominent access points. For a particular access point to be considered as a prominent access point, it needs to produce a high number of valid signal strength readings for a predetermined number of days. As the initial step of determining prominent access points, we calculated ratios as shown in (3.11) for each access point a and day d.

$$r_{a,d} = \frac{\text{number of times a valid access point } a \text{ appeared in scans for day } d}{\text{number of total scans for day } d} \tag{3.11}$$

In (3.11), the number of time a valid access point appears in scans represents the number of times the received signal strength of a certain access point lies in between -90dB and -30dB. We generated $r_{a,d}$ for each access point for each day, and calculated the number of days the corresponding ratio exceeds a given threshold value. If this number was higher than another predetermined threshold, that particular access point was chosen to be a prominent access point. Changing these two thresholding parameters governs the level of appearance frequency required for an access point to be considered as a prominent access point. These two thresholds are determined experimentally based on the collected data.

Once the prominent access point list was derived, we selected signal strength values corresponding to those access points from the collected data in order to build the probability density functions of signal strengths.

These probability density functions can be used to obtain the likelihood of a given signal strength value being received from any given prominent access point at a given cell, using the Bayes rule. For an example, NxM number of distributions will be generated for N number of cells in the building and M number of prominent access points. We employed non parametric Kernel Density Estimation (KDE) method in order to build the probability density functions of signal strengths as this method generates more representative distributions of our finite data set. The formula for kernel density estimator is given in (3.12).

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big) \tag{3.12}$$

where x is the independent variable for kernel density estimator function, x_i represents the ith data sample and n denotes the number of data samples. Parameter h is called bandwidth which governs the smoothness of output distribution, and $K(.)$ denotes the kernel smoothing function. In this work we chose normal distribution as the smoothing function.

The discrete probability distributions generated using KDE method spans from -90 dB to -30 dB with a step size of 1 dB. Therefore we can estimate the probability of a signal strength being a given value within this range for a given access point and cell. We selected -90 dB to -30 dB signal strength range as most observable accurate values fall within this scope and lower intensities are highly susceptible to noise.

Once we generated the complete list of probability densities using KDE we stored it for later use in the localization algorithm. The graph in Fig. 3.2 shows the KDE based probability density of received signal strengths from four different Wi-Fi access points for a particular cell in the building.

The goal of our localization algorithm is to produce a vector of probabilities that represents the likelihood of the subject being positioned in any cell. In order to accomplish this, we employed the Bayes rule as shown in (3.13),

$$p(c_i|s) = \frac{p(c_i)\ p(s|c_i)}{p(s)} \tag{3.13}$$

where c_i denotes the ith cell while s stands for the received signal strength. As our location estimation task can use multiple signal strength readings from prominent access points, Bayes rule equation could be written as (3.14),

$$p_{post}(c_i|s_1, ..., s_M) = \frac{1}{p(s_1, ..., s_M)}\ p_{pri}(c_i)\ \prod_{j=1}^{M} p(s_j|c_i) \tag{3.14}$$

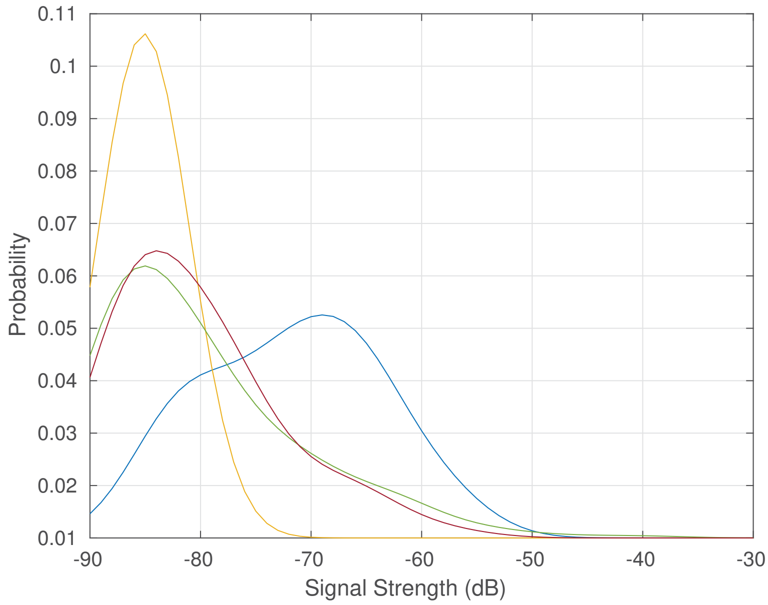

FIGURE 3.2: KDE based probability density functions of received signal strengths from four access points for a given cell

where c_i denotes the ith cell while s stands for the received signal strength, M denotes the number of prominent access points that produce signal strength values in the range of -90 dB to -30 dB, $p(s_j|c_i)$ is the likelihood of receiving a signal strength reading s_j while in cell c_i, and $p(s_1, ...s_M)$ is a normalizing constant. The likelihoods $p(s_j|c_i)$ are generated from the probability distributions generated during the training step and $p_{pri}(c_i)$ is the prior probability of being in cell c_i . At the beginning, the value of $p_{pri}(c_i)$ is assumed to be $1/N$ for all the cells where N denotes the number of cells in the building. $p_{post}(c_i|s_1, ..., s_M)$ is the posterior probability that the subject is in cell c_i given all the Wi-Fi signal strength readings.

As mentioned above we begin the estimation process by assuming equal prior probability $p_{pri}(c_i)$ for all the cells as initially we do not have any information regarding the location. These probabilities are then propagated to the next time instance using a motion model to obtain the probability of the location immediately prior to receiving signal strength measurements. Bayes rule is then used to incorporate the information available in the signal strength measurements to obtain the posterior probability of the location. This process is repeated every-time a signal strength measurement is received. The motion model used relies on floor transition and cell distribution information. The steps taken for the development of motion model and derivation of posterior probabilities will be discussed in the next Section.

3.4 Motion Model

The motion model which is incorporated into the localization framework exploits altitude information and prior knowledge of cell distribution. We adopted a hierarchical approach in order to perform this, where initially it employs the information on floor transition events, and subsequently exploits the prior knowledge of cell distribution.

Exploiting information on the current floor and floor transition events is crucial due to the fact that in some locations, particularly on the first floor, strong signals from Wi-Fi access points on the floor above are feasible. As floor transition events can be detected using the altimeter, a likelihood value of the subject being positioned on a particular floor can be achieved. As the initial step, a three element weight vector is established, with each element representing the likelihood of the subject located on the corresponding floor. Once this vector is initialized, the most probable floor for the subject to be currently located is found by examining the posterior probability vector generated in Bayes algorithm which exploits the received Wi-Fi fingerprints. This is achieved by retrieving the most probable cell in posterior probability vector and considering the floor corresponding to it. Once the floor is determined this way, the weight vector is updated by increasing the weight value corresponding to it, and this process continues throughout the runtime.

While the floor weight vector is built, floor transition events such as upward, downward or no transition are detected by the altimeter readings. These floor transition events are then used to modify the weight vector in such a way that if an upward floor transition is detected, the weights of lower and current floors are decayed significantly compared to the higher floors and vice versa, as shown in (3.15), (3.16) and (3.17).

When no floor transition is detected,

$$F_i' = \begin{cases} (1-g)F_i, & i = i' \\ gF_i, & i \neq i' \end{cases} \tag{3.15}$$

When an upward floor transition is detected,

$$F_i' = \begin{cases} (1-g)F_i, & i > i' \\ gF_i, & i \leq i' \end{cases} \tag{3.16}$$

When a downward floor transition is detected,

$$F_i' = \begin{cases} (1-g)F_i, & i < i' \\ gF_i, & i \geq i' \end{cases} \tag{3.17}$$

Here, F_i' denotes the new ith value of the floor weight vector, while F_i represents the previous ith value of floor weight vector. The parameter g is the decay constant which is set to be 0.4 in our experiments, and i' is the index of the currently estimated floor.

After a given time period, the floor weight vector is used to update the original posterior probabilities generated by Bayes theorem. In order to accomplish this, each posterior cell probability is multiplied by its corresponding normalized floor weight. The aim of this continuous procedure is to update posterior probabilities generated through Bayes theorem by employing true floor transition events detected by the altimeter, such that it minimizes spurious floor changes due to incorrect cell estimates. However, it should be noted that the floor weight vector is continuously updated, by exploiting the current floor estimation based on posterior probability vector generated by Bayes rule, as well as floor transition events detected through altimeter readings. The reason for not relying completely on altimeter readings for updating floor weight vector is the uncertainly of altitude readings. This uncertainty of altitude readings could be explained by the inconsistent pressure levels inside the building, which consequently affect the altimeter readings even when it is located at the same floor level. The procedure related to continuous update of floor weight vector is illustrated in Fig. 3.3.

Once the altitude information is integrated, we exploit the prior knowledge of cell distribution in floor plans to evolve the posterior probability vector, before processing the next set of Wi-Fi signal strength readings. This is achieved by incorporating a transition matrix which represents the likelihood of subject moving from a given cell to another cell.

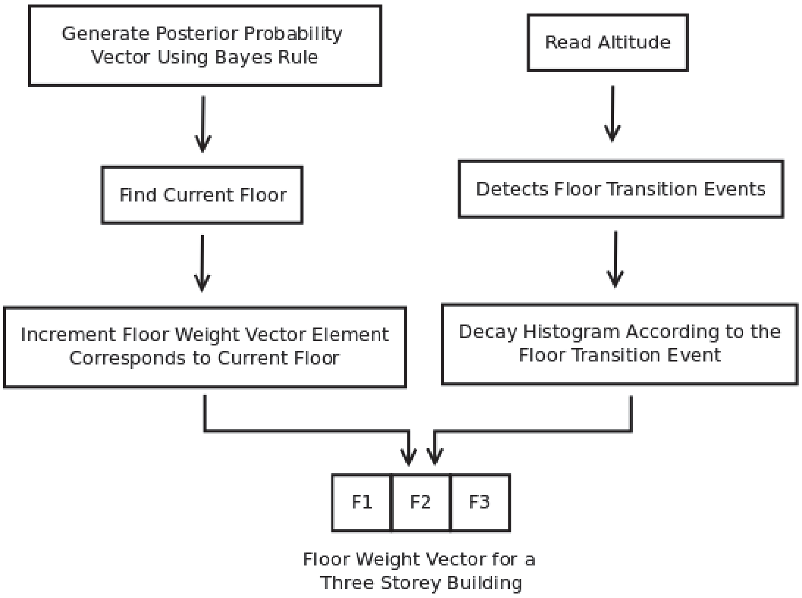

FIGURE 3.3: Process related to continuous update of floor weight vector

The transition matrix used here is generated in two different ways for the comparison of their effectiveness on the localization algorithm. In these two forms of transition matrices presented here, the current cell ID is represented by row number while the next cell ID is represented by column number. Therefore, the size of these transition matrices will be NxN, where N is the number of cells. The methods we employed to develop these forms of transition matrices will be discussed in the following sections.

3.4.1 Flat Transition Matrix

The flat transition matrix represents the simplest form of transition likelihoods between cells in a given floor. This matrix is prepared by assuming an equal probability for cell transitions between current cell and adjacent cells. Moreover it assumes the same probability for self cell transition which resembles the scenario where the subject stays in the same cell throughout consecutive iterations of localization algorithm. The distant cells which are not adjacent to the considered cell are assigned a smaller non-zero value as the transition probability. This is due to the fact that a transition to these distant cells without passing intermediate cells is an unlikely motion behaviour. This is especially true for an elderly person walking with a walker due to their relative slower speeds. A row in the transition matrix could be illustrated as in Fig. 3.4.

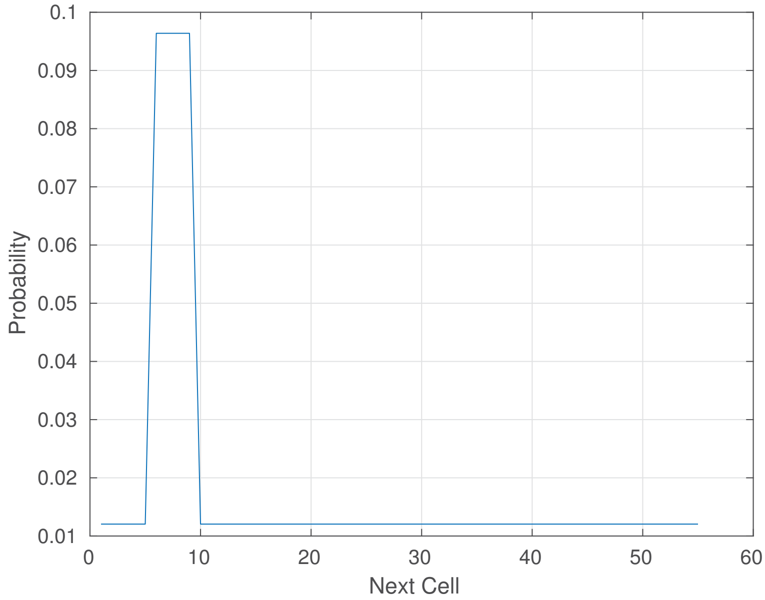

Figure 3.4: A single row in transition matrix representing the cell transition probabilities

3.4.2 Area Based Transition Matrix

In practice, it could be assumed that the likelihood of a moving walker's presence in a particular cell at a given instance is related to the geometric area of that cell. Similarly, it could be seen that the transition from a given cell to an adjacent cell depends on the boundary area where the transition occurs. In this context, an area based transition matrix exploits this information on geometry of cells for generating the transition probabilities.

Fig. 3.6 illustrates the transition boundary areas of a given cell (Cell A) to other cells in green colour and the area related to the self cell transition (remains in the same cell) in blue colour. The width of green area is approximately equal to the average distance a subject travels within the time period of two consecutive iterations of localization algorithm. Therefore, a subject positioned in green area is capable of travelling to another cell within this time period. In this perspective, probabilities related to inter cell transition and self transition are proportional to the areas related to transition boundaries and self transition. If a given boundary area corresponds to a single destination neighbour cell (Cell A to Cell C transition boundary), only half of that boundary area is considered for estimating that particular transition probability and the other half is added to the area related to self transition. This is due to the assumption that there is only a 50% chance of

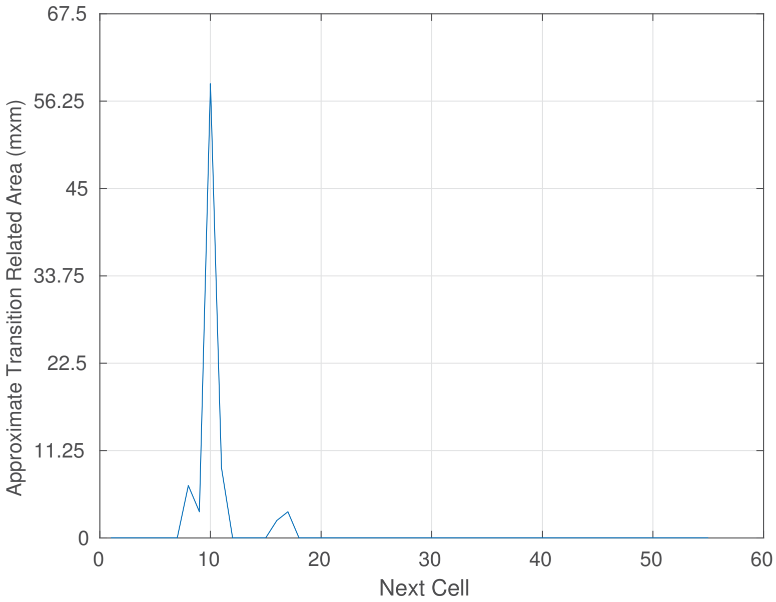

FIGURE 3.5: Transition areas related to a given cell. Similar area based information related to all the cells is used to develop the transition matrix

a walker located in this boundary area to move to the other cell. Boundary areas which are open to more than a single neighbour cell (e.g. around cell corners) are partitioned similarly based on the number of neighbour cells corresponding to that boundary. The transition probabilities for adjacent cells are determined in this way while the transition probabilities related to the distant cells are assigned a small non-zero value as in the flat transition matrix. This small non-zero value is experimentally determined to be 0.0225 m^2 before normalizing. However, the functionality of motion model could be affected by changing this value, as it governs the likelihood of motion from current cell to the distant cells. Transition areas related to a given cell is illustrated in Fig. 3.5.

Once the relevant transition matrix is selected, the posterior probabilities of cells are updated using the transition probabilities as in (3.18).

$$p'_{c_i} = p_{c_1}p_{t_{1,i}} + p_{c_2}p_{t_{2,i}} + p_{c_3}p_{t_{3,i}} + ... + p_{c_{55}}p_{t_{55,i}} \tag{3.18}$$

Equation (3.18) could be summarized to (3.19),

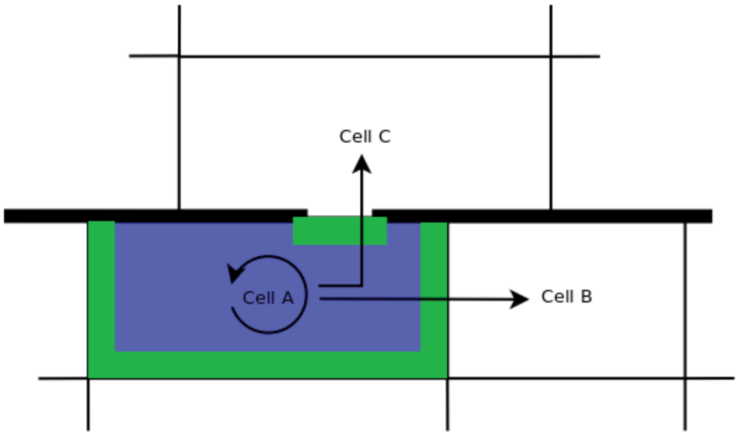

FIGURE 3.6: The two green areas represent transition boundary areas of Cell A to other cells (Cell A and Cell C are connected through a door, while other boundaries of Cell A do not have physical obstacles) and blue area represents the area corresponds to self transition

$$p'_{c_i} = \sum_{j=1}^{N} p_{c_j} p_{t_{c_j,c_i}} \tag{3.19}$$

where p'_{c_i} denotes the updated probability of ith cell, p_{c_j} is the posterior probability of jth cell and $p_{t_{c_j,c_i}}$ represents the transition probability from cell j to cell i. The variable N stands for the number of cells in the whole building. These updated probabilities are then used as the prior probabilities for the next iteration of Bayes rule, and also considered to be the final probability vector for representing the location estimates.

3.5 Conclusions

This Chapter presents the theoretical framework of a Wi-Fi based probabilistic localization method for indoor crowded environments, which could be implemented in a physical system and employed for previously specified real world application of tracking elderly people. This method is an alternative to the computer vision based place recognition method presented in Chapter 2, which has several shortcomings as mentioned.

The localization method presented in this work is based on the distribution of Wi-Fi signal strengths that are captured from an existing Wi-Fi network in a given environment. This method does not require the installation of specific equipment in the target environment as it relies on the existing network infrastructure. In addition, the proposed Wi-Fi fingerprinting based localization method does not rely on the positions of existing Wi-Fi access points in order to generate location estimates for the subject.

The presented method consists of an off-line phase and an on-line phase where in the off-line phase we divide the floor maps of shopping centre into cells and store the received Wi-Fi signal strength values for each cell in order to generate the probability distributions of signal strength values. In order to accomplish this, we employ KDE method to capture the probability density functions of signal intensities more accurately. These are then utilized at the on-line phase where we process real-time received Wi-Fi signal strength values using a Bayesian approach for estimating the location of a subject. A motion model which exploits floor transition events and the prior knowledge of cell distribution in the building is also used to enhance the performance of the technique.

The main aim of the work presented in this thesis is to implement the proposed Wi-Fi localization method in a comprehensive indoor localization system which is capable of providing assistance to elder care supervisors to track elderly people in a crowded shopping centre. The next Chapter elaborates the details of the development of this comprehensive indoor localization system in collaboration with a team of professionals and senior citizens from a Sydney based Illawarra Retirement Trust (IRT) elder care facility.

Chapter 4

Wi-Fi Based Localization: Practical Implementation

4.1 Introduction

This Chapter presents a comprehensive Wi-Fi based indoor localization system for tracking elderly people in a crowded shopping centre that was designed, developed and implemented in collaboration with Illawarra Retirement Trust (IRT), a leading aged care service provider in Australia. This system is based on the theoretical framework described in Chapter 3. As the main goal of this work was to develop a system for tracking elderly people under the care of an IRT care supervisor in an excursion to a crowded shopping centre, we first needed to comprehensively understand the problem domain in the context of providing professional elder care. This was primarily achieved by organizing a number of discussions and co-design workshops with professionals from IRT staff and residents. After these preliminary discussions we proposed the specifications of a localization system which could be efficiently used to address the localization problem. The proposed system consists of a self-localization package which is mounted inside a standard wheeled walker and a hand held tablet device with an application installed for visualizing the locations of these walkers. After that, several rounds of discussions were held to come to an agreement on the finer details of the final localization system. Initially, an experimental version of the system was developed by integrating a high performance computing platform and a wide range of sensors to the walker platform.

The standard wheeled walker has been retrofitted with a Wi-Fi dongle, an altimeter, a computer and a 3G dongle. An Inertial Measurement Unit (IMU), wheel encoders and Laser Range Finder (Fig. 4.2) were also included to capture the true location of the walking device (i.e. ground truth). Once this experimental version of the system was successfully evaluated in test trials, a mobile indoor localization package that is cost effective was designed and developed. This minimized version of the indoor localization system is designed with the aim of placing it inside any standard wheeled walker container bag without any additional electronics or wirings. In parallel with the development of the self-localization system an Android application for a tablet for visualizing the positions of the walker platforms was also built.

When this system is in operation the self localization system reads received signal strengths from wireless access points through the Wi-Fi module and processes them for producing location estimates according to the method presented in Chapter 3. Once the location estimates are generated, cells indicating high probability of possible user locations are selected and transmitted as location update messages to the hand held tablet device. An application running in the tablet decodes these update messages and displays the position of the walker in the shopping centre's floor map, allowing the care assistant to locate any group member under his/her supervision in real-time (Fig. 4.1).

4.2 Initial Discussions and Co-Design Workshops

Development of the comprehensive indoor localization system in order to fulfil the requirement of IRT required an in-depth understanding of the problem domain including the operations of IRT as an elder care provider and their general expectations, services provided by them to their residents including guided excursions to shopping centres, shortcomings of the current procedures related to these activities and their exact requirements. Therefore, the design and development of the localization system was commenced with a study of these important factors through a number of discussions and co-design workshops

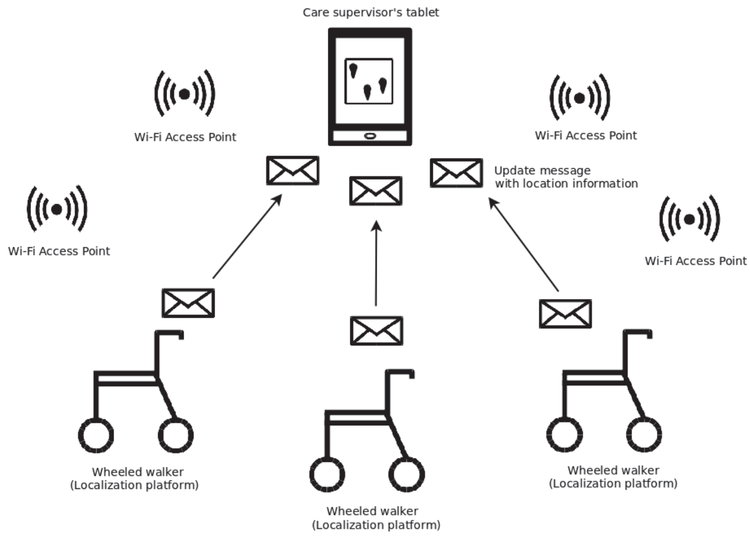

FIGURE 4.1: High level architecture of the localization system

held with a group of IRT staff members consisting of several care assistants, supervisors and the facility manager, and a few resident volunteers.

One of the primary services that IRT provides as a professional age care institute is assisting senior citizens to maintain a domestic lifestyle with necessary support, while ensuring their independence. While IRT offers basic care services for elderly residents who can reasonably manage their daily activities without an extensive involvement of a third party, it also provides personalized care services in IRT care centres for the residents who require specialized support. The basic care services IRT provides for their residents include assisting with household works, health and well-being support, as well as organizing social outings such as excursions to shopping centres.

In excursions to shopping centres, an IRT care supervisor is assigned to travel with a group of elderly residents in order to ensure the safety of the elderly people during the trip. Generally, this group of elderly residents typically includes people with varying degrees of cognitive capabilities and it is the responsibility of the care supervisor to ensure everyone's safety throughout the journey. To ensure that no-one separates from the group, the care supervisor always travels with the whole group. This practice is highly inefficient and tiresome as the whole group is required to stay together and travel to places just to fulfil a single person's shopping needs. As a solution, a comprehensive indoor localization system including a self-localization platform, and a hand held tablet for visualizing the positions which enables the group members to separate and roam throughout the building was proposed. This allows the IRT care supervisor to track the elderly people while staying in a different location.

Even though various indoor localization methods have been explored for implementing the self-localization system as discussed in previous Chapters, Wi-Fi fingerprinting based localization was chosen as the preferred method, primarily due to the ubiquity of wireless access points in public environments. Based on the outcomes of initial discussions and co-design workshops, it was then decided to develop a mobile self-localization package which implements the Wi-Fi based localization algorithm. It was integrated with a standard wheeled walker platform. Consequently, this design decision eliminated the requirement of carrying or wearing additional equipment by residents which could have potentially been uncomfortable to them.

Moreover, it was determined through consultations with IRT that accurate location to a region within which easy visual contact of the resident can be made is sufficient to satisfy the needs of the care supervisors. Therefore, based on the building structure and distribution of shops in the shopping centre where the routine excursions take place, we estimated the area which the care supervisors require to have visual contact with resident to be, approximately $300m^2$. The comprehensive indoor localization system presented in this Chapter accomplishes this localization task by locating residents in regions of $150m^2$ or less within the building.

4.3 Instrumented Walker Platform

Wheeled walker is a commonly used walking aid device of elderly people. The self-localization system that we developed is integrated into a walker as illustrated in Fig. 4.2, such that it does not interfere with the mobility of the user. The initial version of self-localization system was developed focusing on various experimental activities which require additional sensors, high performance computer and high battery life. After this experimental version was evaluated in detail it was minimized into a smaller and cost

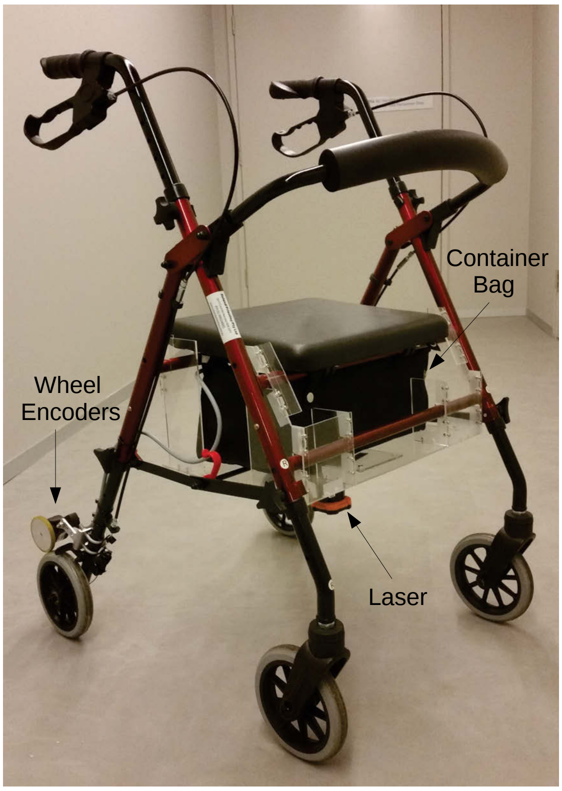

FIGURE 4.2: Instrumented wheeled walker

effective version which includes only the adequate sensory capabilities, computational and battery powers.

The experimental version of the wheeled walker platform is equipped with an AC600 USB Wi-Fi module (Fig. 4.4) which is capable of acquiring Wi-Fi access point information such as MAC address, signal strength values and frequency. In order to enhance the accuracy of location estimates our system exploits the floor transition events which occur while travelling inside the building. These events could be detected by identifying sudden altitude changes, hence we employed a Yocto Altimeter as the altitude measuring sensor for this task (Fig. 4.5). An Intel NUC-Core I3 computer is employed as the main computational unit for implementing software framework and algorithms. Furthermore, a 3G Dongle (Fig. 4.6) is used for sending text messages to the hand held tablet device for updating location information. In addition to this primary equipment, we integrated a Hokuyo laser range finder (Fig. 4.8), IMU (Fig. 4.9) and optical wheel encoders (Fig. 4.7) for generating ground truth position data.

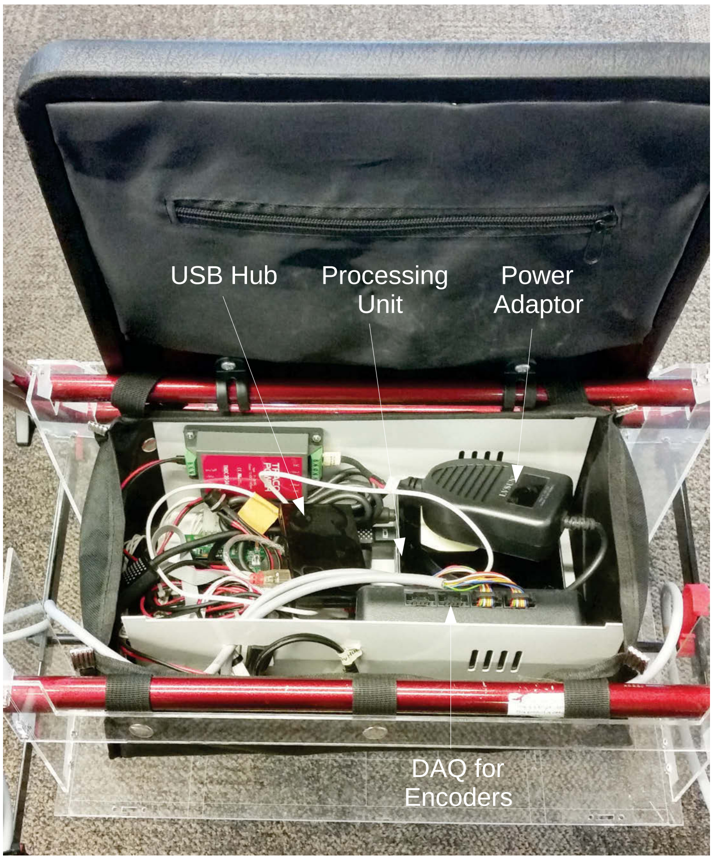

Figure 4.3: Experimental self-localization system placed inside walker container bag

Fig. 4.3 shows how the experimental self-localization system is placed inside the container bag which is located underneath the seat of the wheeled walker. Here, the Wi-Fi dongle and Data Acquisition (DAQ) equipment are directly connected to the NUC computer while altimeter, 3G dongle, laser range finder and IMU have been connected to the NUC through a powered USB hub. The pair of wheel encoders are mounted such that the encoder wheels touch and rotate with the walker rear wheels and encoder readings are directly passed to the DAQ device. The DAQ is capable of decoding the received encoder readings and sending them to the NUC for processing. Laser scanner is fixed to a Perspex Acrylic board that is placed underneath the walker container facing towards the environment in front of the walker platform. Other components including Wi-Fi dongle, altimeter, 3G dongle, IMU, NUC computer and batteries are placed inside the walker container bag. The wheeled walker equipped with these instruments is illustrated in Fig. 4.2.

FIGURE 4.4: Wi-Fi Adapter

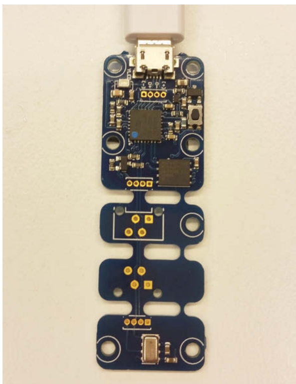

FIGURE 4.5: Altimeter

FIGURE 4.6: 3G Dongle

FIGURE 4.7: Wheel Encoders

FIGURE 4.8: Laser Range Finder

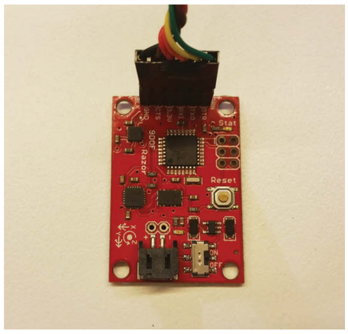

FIGURE 4.9: IMU

A smaller and more cost effective version of the localization system that could be packaged inside a smaller box is developed, by minimizing the capabilities of the experimental version. We employed the same off the shelf AC600 dual band Wi-Fi module and Yocto altimeter as the primary sensors, while the computational unit is changed to a less expensive Odroid-C1 computer. The same type of 3G dongle is used for transmitting text messages to the hand held tablet device, as in the experimental version. These sensors, Odroid-C1 unit and other necessary equipment are packed inside a small container box as illustrated in Fig. 4.10 and placed in the walker container bag.

4.4 Data Collection

Elderly residents of the IRT Peakhurst care facility frequently visit the nearby Roselands shopping centre during their shopping excursions. This shopping centre is a three storey

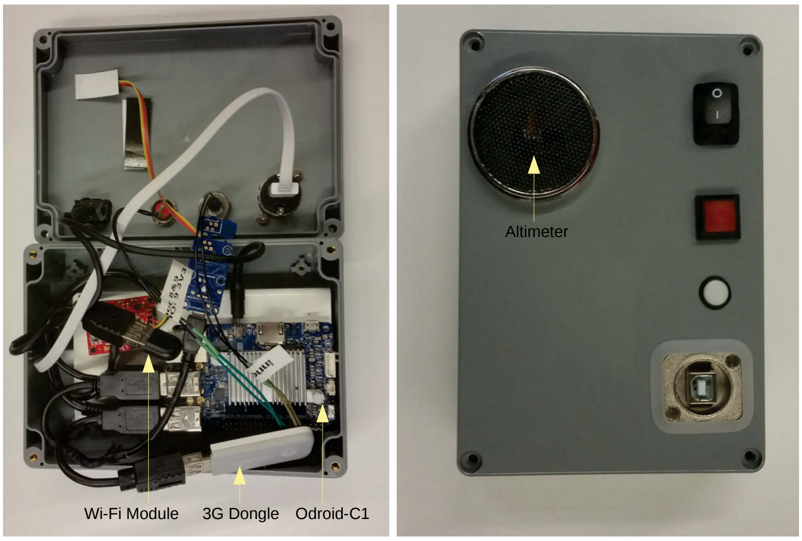

FIGURE 4.10: Self-localization system packed inside a container box

building with a large crowd being present at most of the time during the day. Given that the requirement is a system that determines the approximate location of a person and to find him/her through visual contact, the cell based localization approach described in Chapter 3 was selected. Maps from all three floors are separated into cells taking geometrical restrictions into consideration such as walls and partitions. While relatively smaller and enclosed spaces are treated as individual cells, larger open spaces like the food court situated on the third floor are partitioned into multiple cells such that a single partition does not exceed half of the full open area. We selected the maximum cell size to be approximately 150 m^2. In this way we partitioned the whole building into 55 individual regions and assigned ID numbers.

During the data collection stage, we used the wheeled walker equipped with the Wi-Fi dongle to record received signal strengths and the Media Access Control (MAC) address for each Wi-Fi access point along with the corresponding cell ID. These signal attributes were collected at each cell covering the complete cell space including cell boundary and the centre area. We also ensured the Wi-Fi signal strength data collection under different orientations of wheeled walker within the cell. As we needed the ground truth location information in order to derive the associated cells corresponding to the collected signal data, we performed LRF based localization using Adaptive Monte Carlo Localization (AMCL) [15] approach using Robot Operation System (ROS) amcl package during the

Wi-Fi data collection stage.

As LRF based localization implementations such as ROS amcl package requires accurate 2D maps (Fig. 4.11,4.12,4.13), we adopted chamfer distance based scan-to-map matching method (CD-Mapping) presented in [82] to generate these maps. Capability to generate highly accurate laser maps which leads to perform accurate LRF localization is a significant advantage in CD-Mapping over other map creation methods like gmapping SLAM (Simultaneous Localization and Mapping) implemented in ROS. However, this laser based map creation technique was not successful in certain areas of the building due to the presence of glass partitions which could not be detected by the laser. In these areas, we manually labelled the collected signal data with relevant cell identifications.

We conducted the Wi-Fi data collection for four different days in order to include possible time based signal variation characteristics into the data. In addition to the time based signal variations, a data set that spans over multiple different days assisted us to capture Wi-Fi signal strength variations due to the change in human motion and population density inside the building based on the day. Even though the exact number of Wi-Fi scans collected per cell varies based on the cell size and the time we spent on collecting data inside the cell area, on average 400 Wi-Fi scans are recorded for a single cell.

4.5 Software Architecture of System Implementation

The complete software framework of our Wi-Fi based indoor localization system consists of two main parts, i.e., the self-localization software stack and an application running on an Android based platform for real time visualization of walker platform positions. This Section elaborates on the development and operation aspects of these two components.

4.5.1 Self-Localization System Implementation

The software stack of our self-localization system was developed based on the well known Robot Operating System (ROS) under a Linux environment (Ubuntu operating system), primarily due to its inbuilt features such as low level device control capability, convenient

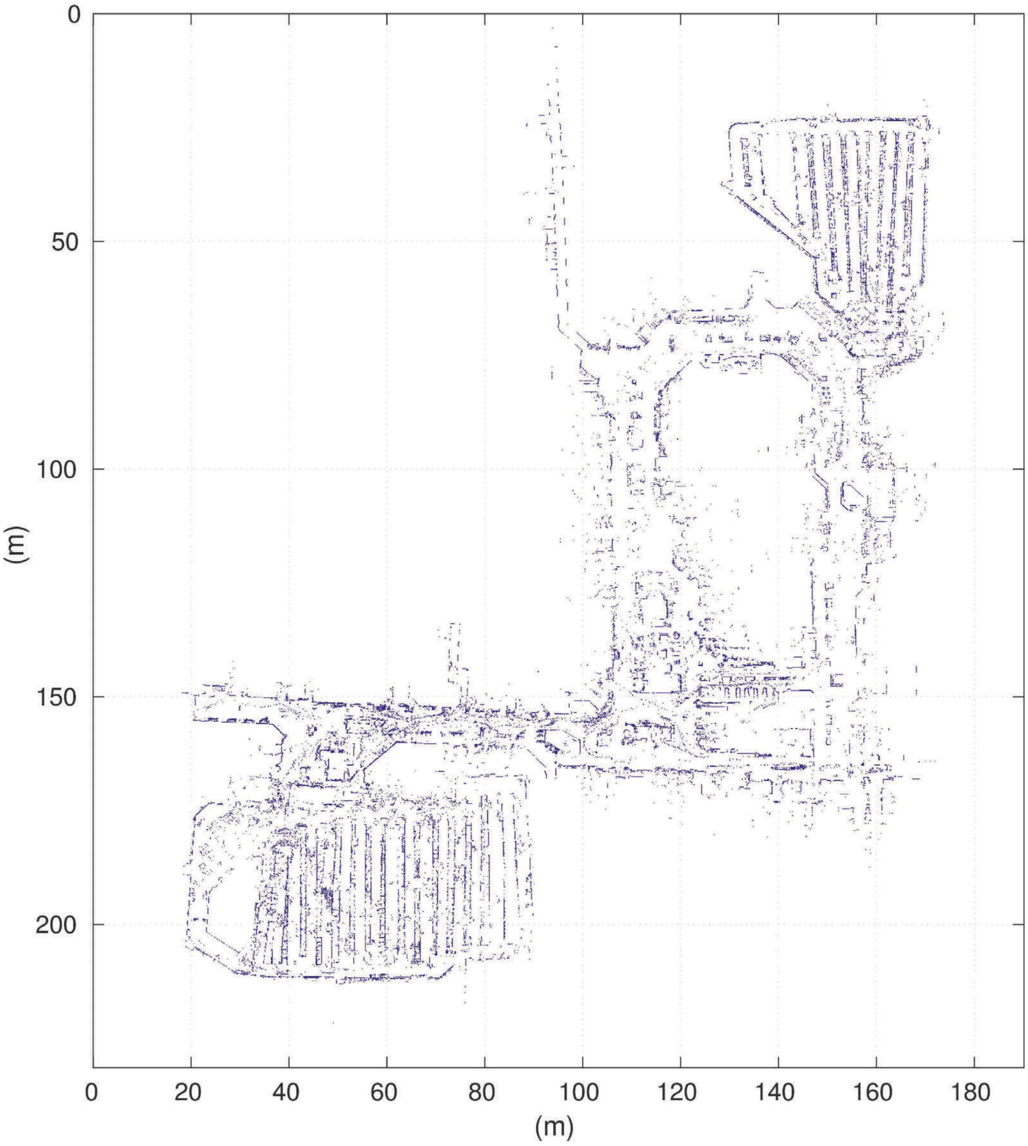

FIGURE 4.11: Laser based map for floor 1

modular structure and inter-module messaging facility. We employed ROS *Indigo Igloo* distribution as the base framework for our system while Python and C++ programming languages were used for developing its software modules (also known as ROS nodes).

A software module or a ROS node in the ROS framework represents a process which can perform tasks such as computations, reading sensor measurements and actuating devices, while the number of ROS nodes and their intended functionalities could be determined by the developers. A properly designed software architecture is composed of ROS nodes

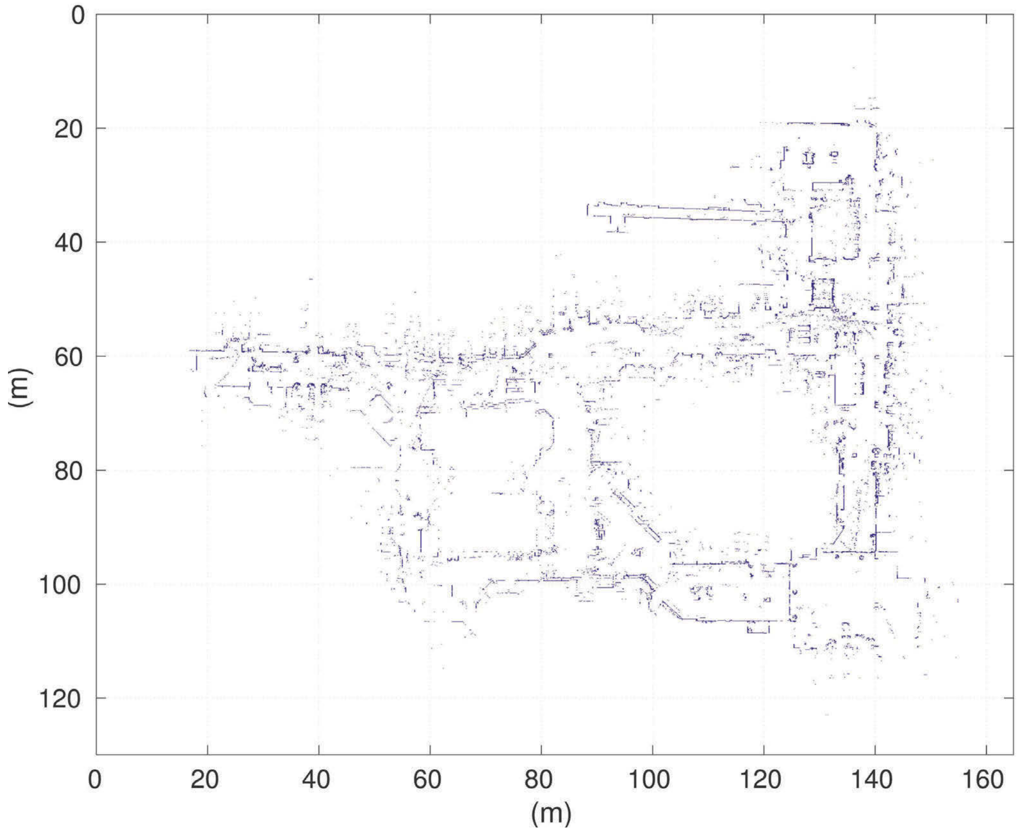

FIGURE 4.12: Laser based map for floor 2

with unique functionalities guaranteeing their loosely coupled modular nature. This type of an architecture facilitates the system developers to conveniently integrate any future modifications or enhancements without affecting the overall system architecture. This best practice is properly followed when designing our self-localization system architecture based on ROS framework.

Our self-localization software stack consists of five ROS nodes each of them having a unique functionality within the process of probabilistically estimating the location of a walking device. The following list describes these function of the five ROS nodes.

- **Wi-Fi scanner node**

  This ROS node reads Wi-Fi signal strength related data such as Wi-Fi signal strength, MAC address, signal frequency through the Wi-Fi hardware module. Once these data are retrieved from the Wi-Fi module, necessary attributes are filtered and re-structured for publishing into the ROS space making them accessible to other nodes.

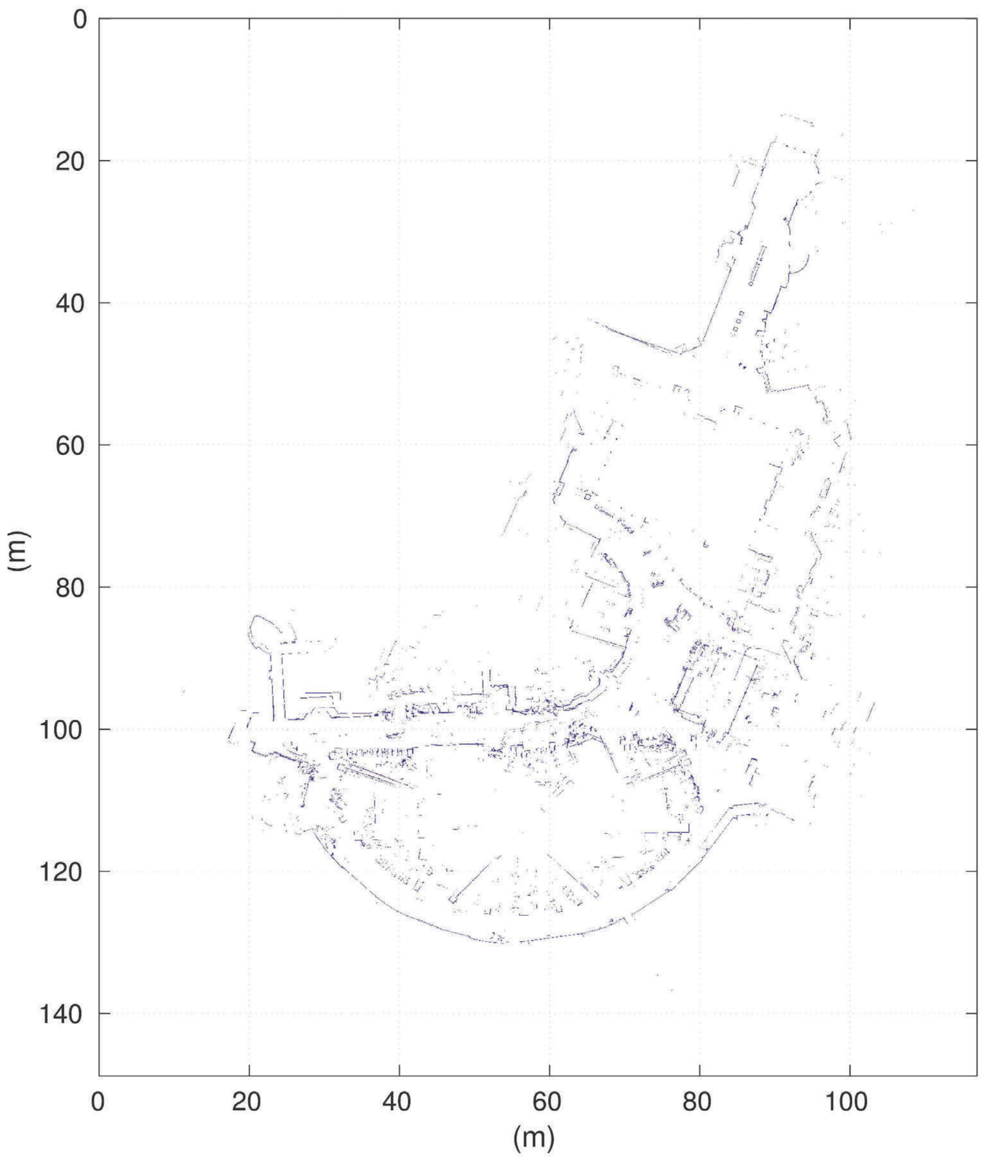

FIGURE 4.13: Laser based map for floor 3

- **Location estimator node**

  Location estimator node reads the filtered scan messages published by Wi-Fi scanner node and processes them as described in Section 3.3. The output of this node is a vector of probabilities corresponding to the cell based location estimates. Once these location estimates are computed, they are published as ROS messages. This ROS node could be identified as the core module in our self-localization system.

- **Altitude reader node**

  The Altitude reader node is capable of accessing the Yocto Altimeter device and reading the altitude. These retrieved altitudes are then restructured and published into the ROS space.

- **Motion model node**

  This node accesses the published altitude data as well as cell distribution information of the building to implement the motion model explained in Section 3.4. We have used a cell area based transition matrix for this implementation due to its better accuracy as demonstrated in Section 4.6. Once the probabilities are propagated using the motion model, they are sent back to the Location estimator node to update prior probabilities for the use in the next iteration of Bayes rule.

  Moreover, this node performs a selection of most probable cells based on two criteria: (a) selected cells should belong to the same floor which includes most likely cell; (b) their localization probabilities should exceed a given threshold, which was experimentally found to be 0.1. Once the best cells are selected and subjected to aforementioned criteria they are published into the ROS space.

- **Text message node**

  Text message node controls the 3G dongle device attached to the computer, which can send standard text messages. Functions of this node include retrieving the final cell information published by Motion model node, encoding this into a predetermined format and sending this encoded content via a text message to the hand held tablet.

The communication between our self-localization platform and the tablet device occurs through standard text messages. Once the relevant cell information is included in a text message, it is sent to the tablet and an application installed in the device reads and displays this information on a map. This complete process is illustrated in Fig. 4.14.

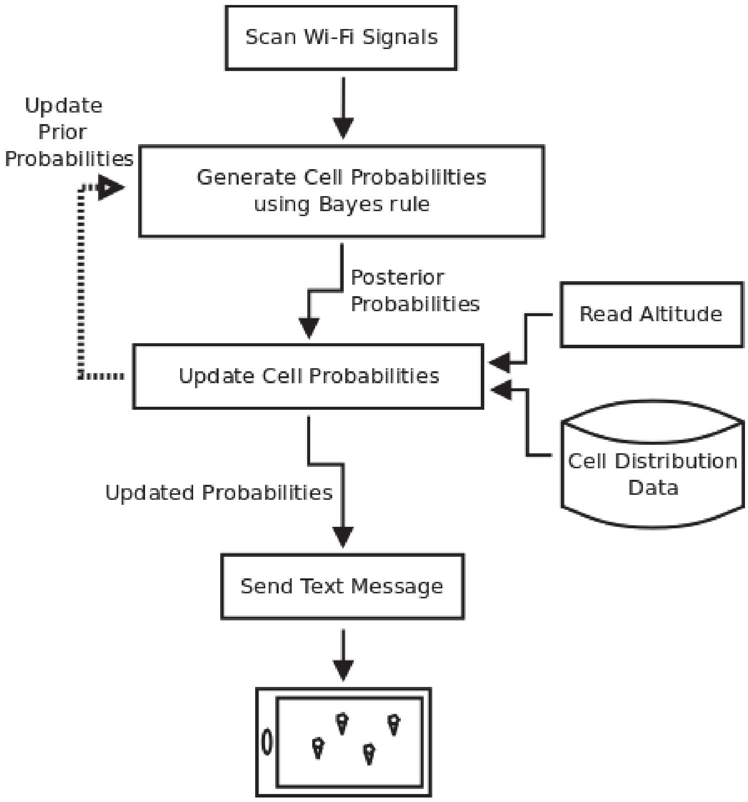

Figure 4.14: Software architecture for Localization framework

4.5.2 Tablet Application for Visualization

We developed an Android application and installed it in a Galaxy 'Tab A' device for the care supervisor to monitor wheeled walkers inside the building. This tablet application is capable of reading the text messages transmitted from the self-localization platform and extracting the location information in order to display on the corresponding floor map. The application is able to support multiple localization platforms (i.e. wheeled walkers) by uniquely categorising the incoming text messages based on the MSISDN (Mobile Station International Subscriber Directory Number) of message sender. Once the text message is received the application identifies the corresponding wheeled walker and decodes the message and displays its cell location in the correct floor map.

This Android application can display estimated locations based on two criteria: (a) wheeled walker based localization; (b) floor based localization (Fig. 4.16). In the wheeled walker based localization scenario, the user selects a preferred walker for tracking and this allows the application to process the continuous messages being received from this selected

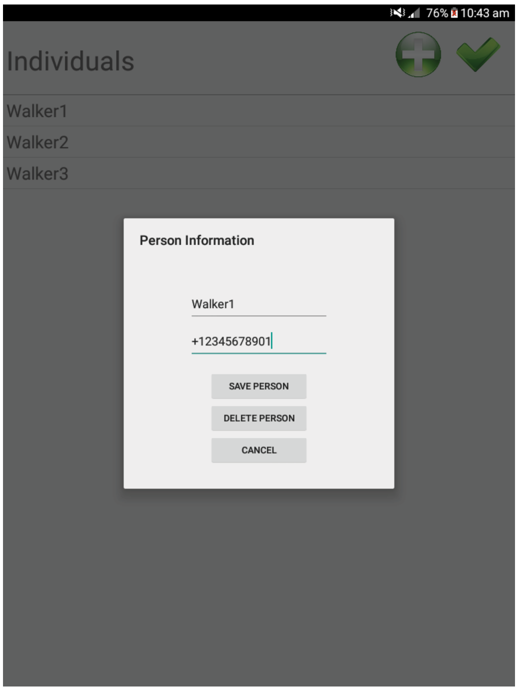

FIGURE 4.15: Registering/Modifying walker information

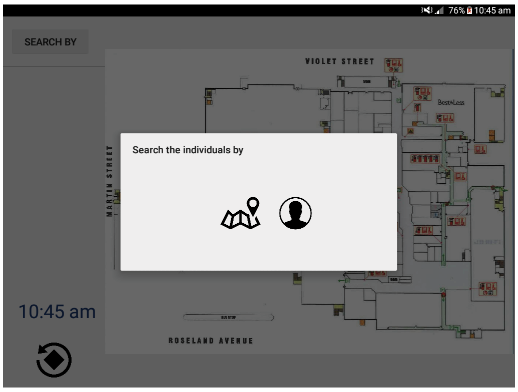

FIGURE 4.16: Selecting the localization criteria

wheeled walker for displaying its location. In the floor based localization, user is given the option to select the preferred floor map in order to process messages coming from wheeled walkers in that particular floor and to display them on the map.

Additionally, our application allows the tablet user to perform fundamental operations such as registering, removing or modifying walkers as illustrated in Fig. 4.15.

The application also includes a test mode where the tablet user is allowed to observe cell locations together with the corresponding localization probabilities. This feature enables the user to understand the behaviour of the self-localization platform in real-time test trials, especially on occasions where multiple cells exist with similar probabilities. The localization screen of our Android application is illustrated in Fig. 4.17, where the coloured circles show located wheeled walkers.

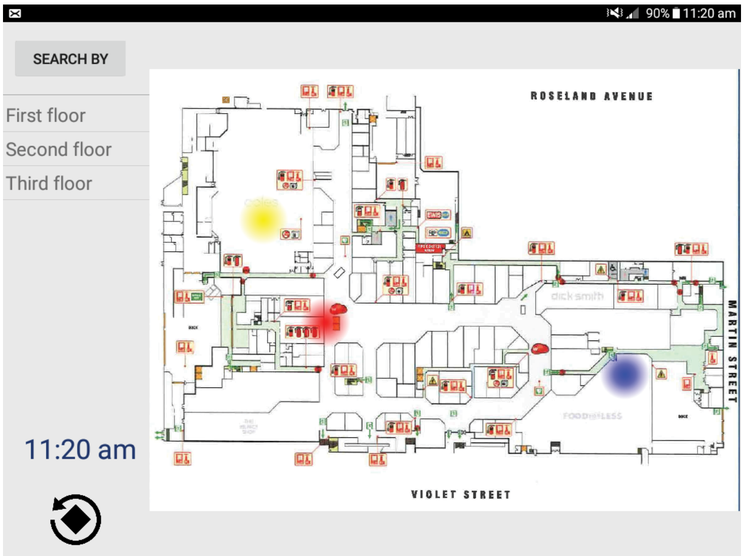

FIGURE 4.17: Localization screen of Android application which shows the positions of the wheeled walker users in first floor of the building. The three different colour circles represent different wheel walkers

4.6 Experiments and Results

All the experiments and test trials related to the localization system were conducted in Roselands Shopping centre in Peakhurst, Australia. This shopping centre is typically a highly crowded environment and has dynamic and unpredictable variation to the Wi-Fi signal strengths at a given position in the building. Therefore, this shopping centre provides a challenging environment for evaluating our localization framework. In the test trials, we roamed inside the building with a wheeled walker and recorded location estimates (cell IDs) generated through the system together with corresponding ground truth data. The ground truth data were generated using the Adaptive Monte Carlo Localization algorithm with the recorded laser scans and wheel encoder readings. Then the location estimates from our framework were compared with the ground truth data to compute the accuracy.

Two different types of accuracies are calculated for the generated location estimates, i.e., primary accuracy and secondary accuracy. Primary accuracy is the percentage of location estimates that matched with the correct cell ID while the secondary accuracy is percentage of location estimates that matched with either the correct cell or any adjacent cell. Therefore, the secondary accuracy is always equal or higher than the primary accuracy.

Floor	Flat Transition		Area Based Transition	
	Primary	Secondary	Primary	Secondary
Floor 1	36.66%	85.41%	66.11%	89.57%
Floor 2	51.09%	99.39%	87.94%	100.0%
Floor 3	51.22%	98.44%	77.68%	100.0%

TABLE 4.1: Floor accuracies of transition matrix models

These accuracies for our localization framework are listed in Table 4.1. The accuracies are based on the two transition models i.e., the flat transition model and the cell area based transition model that are comprehensively discussed in Section 3.4.

According to the observations, in certain instances our localization algorithm detects an adjacent cell as the most probable cell specially when the walker is close to the cell boundaries. This is due to the negligible change demonstrated in Wi-Fi signal strength values at opposite sides of a cell boundary where each side represents a different cell ID. This characteristic could be identified as the reason for the secondary accuracy being higher than the primary accuracy in each transition model. The distinguishable difference between primary accuracy and secondary accuracy in the flat transition matrix model indicates that this mismatching happens more often as the transition probabilities corresponding to adjacent cells and current cell are equal in this case. However, a higher secondary accuracy satisfies the requirement of locating the elderly person in a public environment, as estimated location being either correct cell or adjacent cell is adequate for finding him/her through visual contact.

It could also be observed that, while secondary accuracy of floor 2 and floor 3 are 100.0%, secondary accuracies of floor 1 in both transition matrix forms are relatively smaller (85.41% and 89.57%). One possible reason for this reduced accuracy is that in certain instances localization algorithm estimates a non-adjacent but relatively nearby region as the most probable cell. Another possible reason is that the first floor is connected to the upper floor through an open area, thus the localization algorithm may estimate the most probable cell as the corresponding cell in the upper floor close to the open area. This could be explained by the similarity of Wi-Fi signal characteristics due to the absence of obstacles/partitions between cells close to open areas within adjacent floors. Even though this has been addressed by the floor weight vector to a certain degree complete elimination of such behaviours by adjusting floor weight parameters based on altimeter readings adversely impacts the correct functionality of the localization algorithm in other areas. The reason for such an outcome is that the altimeter readings are affected by pressure level variations inside the building even when the sensor is located at the same floor. Therefore, completely relying on the altimeter for detecting floor information could affect the overall localization accuracy.

4.7 Discussion and Conclusions

The focus of this Chapter is the development and implementation process of the proposed Wi-Fi based indoor localization system capable of tracking elderly people in a crowded shopping centre. This system is developed in collaboration with Illawarra Retirement Trust (IRT), a leading aged care service provider in Australia. The goal of our scalable and comprehensive localization system is to improve the safety of elderly residents during shopping excursions by enabling their IRT care supervisor to track them within a crowded shopping mall.

The presented indoor localization system consists of a portable self-localization system which could be conveniently placed inside a commonly used standard wheeled walker and an application installed in a hand-held tablet which is used to visualize the real-time positions of these walker platforms. The Wi-Fi based position estimation method presented in Chapter 3 is employed as the core algorithm in the self-localization system. Moreover, the cell area based motion model is incorporated with the localization system due to its higher accuracy as shown in Section 4.6.

The performance of this Wi-Fi based localization system is evaluated in Roseland shopping centre located in Peakhurst, Australia through a number of test trials. We have demonstrated highly satisfactory accuracy rates in terms of facilitating a care supervisor to find residents through visual contact. Specifically, the higher secondary accuracy rates of location estimates as described in Section 4.6, satisfies the requirement of locating an elderly person in a public environment and subsequently finding him/her through visual contact by a care supervisor. Once the successful functionality of our Wi-Fi based indoor localization system is confirmed in test trials, it was also evaluated with a group of IRT residents and a care supervisor during their real shopping centre excursions.

Chapter 5

Conclusions

This thesis presented a comprehensive study, design and development process of a low cost indoor localization system for tracking elderly people in crowded indoor environments such as shopping centres. This work includes an introduction of a computer vision based place recognition technique and a discussion related to its applicability, introduction of a robust Wi-Fi based localization method, and finally the development process of a practical indoor localization system based on the proposed Wi-Fi based localization method.

The presented indoor localization system has been designed and developed in collaboration with Illawarra Retirement Trust (IRT), one of the leading professional elder care provider in Australia, for tracking their elderly residents in frequent excursions to a nearby multi-storey shopping centre. The developed system has been assessed in a number of test trials and finally it was evaluated with IRT care providers during their excursions to the shopping centre with elderly residents. Localization system performed well during these trials and feedback as to its behaviour was excellent.

The next Section summarizes the contributions related to our work presented in this thesis, followed by the limitations and future works.

5.1 Contributions

5.1.1 Image Based Place Recognition Technique for a Crowded Indoor Environment

Image based place recognition is a potential technique that could be incorporated in the process of developing an indoor localization system. Therefore, as our first contribution we developed a probabilistic place recognition technique for a crowded indoor environment based on Bag of Words (BoW) descriptors and a feedforward neural network.

The proposed neural network in this work consists of an input layer, a single hidden layer and an output layer. We have used a rectifier as the activation function in the hidden layer while softmax activation function is used in the output layer. This allowed us to probabilistically estimate the class of an input image. The training of the neural network is performed by minimizing the cross entropy cost of the network output by employing Adam algorithm for backpropagation. The performance of our place recognition technique has been successfully verified using a dataset of shop front images captured from a real crowded shopping centre located in Sydney, Australia.

However, a localization system that employs an image based place recognition technique designed to be operated in a crowded shopping centre may violate the privacy of people in the shopping centre due to unintentional exposure to the camera. Although this can be avoided by designing the system such that images are never stored and are not accessible to anyone, the presence of a camera leads to the perception that this system is not appropriate for use in a populated environment. Additionally, to perform well this method requires the user's involvement to capture clear images such as pointing the camera to a shop front and waiting for the crowd to clear from the surroundings. Even though this method performs well provided that the user interaction is available, it could add an extra burden on the walker user, as he/she needs to be aware of these requirements and act accordingly.

Therefore, as a practical solution we developed a Wi-Fi based indoor localization method which is independent of user interaction as well as free from the privacy concerns.

5.1.2 Wi-Fi based Indoor Localization Method for a Crowded Indoor Environment

The Wi-Fi based indoor localization method presented in this thesis exploits the variation of signal strength values based on the position of a Wi-Fi receiver for indoor localization. As the expected application of this indoor localization method does not require exact coordinates of the Wi-Fi receiver, a cell based localization is performed in a Bayesian framework.

Initially, the Wi-Fi fingerprints are required to be collected for each pre-determined 2D cell in the target building to generate the probability distributions of the Wi-Fi signal strength values. KDE is employed for this task due to its ability to produce more representative signal strength distributions. These signal strength distributions are then used with the Bayes rule for determining the posterior probability vector which represents the location of the subject in a given cell. The localization framework is further enhanced by employing a motion model which incorporates floor transition events and prior knowledge of the cell distributions in the floor. Here, we exploit the motion of the subject within the target building, in terms of changing the floor and moving from one cell to another.

5.1.3 Wi-Fi Based Comprehensive Indoor Localization System for Tracking Elderly People in a Shopping Centre

The final goal of this work is to develop a comprehensive indoor localization system which facilitates professional elder care supervisors to track elderly people in their excursions to crowded shopping centres. We accomplished this by developing a portable self-localization package which employs the Wi-Fi based indoor localization method proposed in this thesis and an application that is installed in a hand held tablet device for visualizing the location estimates. Our small scale portable self-localization package could be easily placed inside the container bag found in any commonly used standard wheeled walker without any additional wiring.

Our comprehensive indoor localization system has been tested by IRT care supervisors during their excursion to a crowded shopping centre located in Peakhurst, Australia. The system demonstrates a high accuracy in terms of facilitating IRT supervisors to find elderly people through visual contact using the location estimates displayed in the hand held tablet device. Based on the good accuracy levels demonstrated by the system, we received positive feedback from the participants from IRT and the system has been acknowledged as a possible solution for their requirement of tracking elderly people in a crowded shopping centre.

5.2 Limitations and Future Works

The prominent access point selection and the generation of probability distributions of Wi-Fi signal strength values using KDE in the proposed system were based on the existing Wi-Fi infrastructure in a shopping centre and were done during an initial survey of the environment and were then stored in a database to be used during localization. Changes to the existing Wi-Fi network infrastructure due to changing network requirements such as addition of new Wi-Fi routers/transmitters and removal of malfunctioning ones will change the behaviour of the Wi-Fi network. The Wi-Fi signal strength at a particular location depends on factors such as attenuation due to obstacles which appear in the signal propagation path and the distance between transmitter and receiver. Therefore, building renovations such as adding and removing wall partitions, changing the positions of large advertisement boards and furniture, moving Wi-Fi routers/transmitters to new locations also cause similar changes. In order to mitigate this the probability distributions are required to be updated from time to time by collecting new signal strength data. Data collection is a labour-intensive task and requires a localization system that relies on other sensors such as a laser range finder. A map maintenance strategy that can automatically detect changes to the Wi-Fi network as well as changes to the received signal strength due to changes in the building structure is likely to help avoid such manual intervention leading to a more user friendly system.

Implementing the localization system developed in this thesis in a mobile phone application is feasible as a majority of the mobile phones facilitates inbuilt Wi-Fi scanning capabilities. This will make it possible to use this system with a wider group of elderly residents who do not use a wheeled walker. However, this will require modifying the motion models as the behaviour of such residents will be significantly different to those studied during this work. A more extensive signal strength database will be necessary as the location of where the phone is held will also change from time to time. It may be necessary to use additional sensors available in the mobile phone, such as the magnetometer and the tilt sensors, in order to achieve acceptable performance in such situations.

Bibliography

[1] World Population Ageing. Technical report, Department of Economic and Social Affairs Population Division, United Nations, New York, 2015.

[2] Intergenerational Report 2015. Technical report, Australian Treasury, Commonwealth of Australia, 2015.

[3] R Gnanavel, P Anjana, K S Nappinnai, and N P Sahari. Smart home system using a Wireless Sensor Network for elderly care. In *2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM)*, pages 51–55, 3 2016. doi: 10.1109/ICONSTEM.2016.7560922.

[4] M N A Hossain, A Pal, and S K A Hossain. A wearable sensor based elderly home care system in a smart environment. In *2015 18th International Conference on Computer and Information Technology (ICCIT)*, pages 329–334, 12 2015. doi: 10.1109/ICCITechn.2015.7488091.

[5] B J Driessen, H G Evers, and J A van Woerden. MANUS–a wheelchair-mounted rehabilitation robot. *Proceedings of the Institution of Mechanical Engineers. Part H, Journal of engineering in medicine*, 215(3):285–290, 1 2001. ISSN 0954-4119.

[6] L Tonin, T Carlson, R Leeb, and J del R. Millán. Brain-controlled telepresence robot by motor-disabled people. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4227–4230, 8 2011. doi: 10.1109/IEMBS.2011.6091049.

[7] N Pavón-Pulido, J A López-Riquelme, J J Pinuaga-Cascales, J Ferruz-Melero, and R M d. Santos. Cybi: A Smart Companion Robot for Elderly People: Improving Teleoperation and Telepresence Skills by Combining Cloud Computing Technologies and Fuzzy Logic. In *Autonomous Robot Systems and Competitions (ICARSC), 2015 IEEE International Conference on*, pages 198–203, 4 2015. doi: 10.1109/ICARSC.2015.40.

[8] C Park, S Kang, J Kim, and J Oh. A study on service robot system for elder care. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2012 9th International Conference on*, pages 546–547, 11 2012. doi: 10.1109/URAI.2012.6463070.

[9] S Ohta, H Nakamoto, Y Shinagawa, and T Tanikawa. A health monitoring system for elderly people living alone. *Journal of telemedicine and telecare*, 8(3):151–156, 1 2002. ISSN 1357-633X.

[10] M White, M Vining Radomski, M Finkelstein, D A S Nilsson, and L I Eugen Oddsson. Assistive/Socially assistive robotic platform for therapy and recovery: patient perspectives. *International journal of telemedicine and applications*, 2013:948087, 1 2013. ISSN 1687-6415. doi: 10.1155/2013/948087.

[11] M Topping and J Smith. The Development of Handy 1, A Robotic System to Assist the Severely Disabled. Technical report.

[12] K T Song and S Y Jiang. Force-cooperative guidance design of an omni-directional walking assistive robot. In *2011 IEEE International Conference on Mechatronics and Automation*, pages 1258–1263, 8 2011. doi: 10.1109/ICMA.2011.5985842.

[13] B S Cho, W S Moon, W J Seo, and K R Baek. A dead reckoning localization system for mobile robots using inertial sensors and wheel revolution encoding. *Journal of Mechanical Science and Technology*, 25(11):2907–2917, 11 2011. ISSN 1738-494X. doi: 10.1007/s12206-011-0805-1.

[14] W Winterhalter, F Fleckenstein, B Steder, L Spinello, and W Burgard. Accurate indoor localization for RGB-D smartphones and tablets given 2D floor plans. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3138–3143. IEEE, 9 2015. ISBN 978-1-4799-9994-1. doi: 10.1109/IROS.2015.7353811.

[15] S Thrun, W Burgard, and D Fox. *Probabilistic robotics*. MIT Press, 2005. ISBN 9780262201629. URL `https://mitpress.mit.edu/books/probabilistic-robotics`.

[16] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo localization for mobile robots. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, volume 2, pages 1322–1328. IEEE. ISBN 0-7803-5180-0. doi: 10.1109/ROBOT.1999.772544.

[17] L Dantanarayana, G Dissanayake, R Ranasinghe, and T Furukawa. An extended Kalman filter for localisation in occupancy grid maps. In *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*, pages 419–424. IEEE, 12 2015. ISBN 978-1-5090-1741-6. doi: 10.1109/ICIINFS.2015.7399048.

[18] C N Huang and C T Chan. ZigBee-based indoor location system by k-nearest neighbor algorithm with weighted RSSI. *Procedia Computer Science*, 5:58–65, 2011. ISSN 18770509. doi: 10.1016/j.procs.2011.07.010.

[19] J V Marti, J Sales, R Marin, and E Jimenez-Ruiz. Localization of Mobile Sensors and Actuators for Intervention in Low-Visibility Conditions: The ZigBee Fingerprinting Approach. *International Journal of Distributed Sensor Networks*, 2012(8):1–10, 8 2012. ISSN 1550-1329. doi: 10.1155/2012/951213.

[20] D Schwarz, M Schwarz, J Stuckler, and S Behnke. Cosero, find my Keys! Object Localization and Retrieval using Bluetooth Low Energy Tags. In *18th RoboCup International Symposium, Joao Pessoa, Brasil*, 2014.

[21] J. Yin, L. Carlone, S Rosa, M L Anjum, and B Bona. Scan Matching for Graph SLAM in Indoor Dynamic Scenarios. 2014.

[22] G. Weiß, G. Weiß, and E. V. Puttkamer. A Mapbased On Laserscans Without Geometric Interpretation. 1995.

[23] K Ryu, T Furukawa, S Antol, and G Dissanayake. Grid Based Scan-to-Map Matching for Accurate Simultaneous Localization and Mapping: Theory and Preliminary Numerical Study. In *ASME Mechanisms and robotics conference*, Portland, 2014.

[24] K Ryu, L Dantanarayana, T Furukawa, and G Dissanayake. Grid-based scan-to-map matching for accurate 2D map building. *Advanced Robotics*, 30(7):431–448, 4 2016. ISSN 0169-1864. doi: 10.1080/01691864.2015.1124025.

[25] S Ito, F Endres, M Kuderer, G Diego Tipaldi, C Stachniss, and W Burgard. W-RGB-D: Floor-plan-based indoor global localization using a depth camera and WiFi. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 417–422. IEEE, 5 2014. ISBN 978-1-4799-3685-4. doi: 10.1109/ICRA.2014.6906890.

[26] K S Chong and L Kleeman. Accurate odometry and error modelling for a mobile robot. In *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, volume 4, pages 2783–2788, 4 1997. doi: 10.1109/ROBOT.1997.606708.

[27] O J Woodman. An introduction to inertial navigation. Technical report, University of Cambridge, ISSN 1476-2986, 2007. URL `http://www.cl.cam.ac.uk/techreports/`.

[28] N Suenderhauf, S Shirazi, A Jacobson, F Dayoub, E Pepperell, B Upcroft, and M Milford. Place Recognition with ConvNet Landmarks: Viewpoint-Robust, Condition-Robust, Training-Free. In *Robotics: Science and Systems XI*. Robotics: Science and Systems Foundation, 7 2015. ISBN 9780992374716. doi: 10.15607/RSS.2015.XI.022.

[29] C L Zitnick and P Dollár. Edge Boxes: Locating Object Proposals from Edges. pages 391–405. Springer, Cham, 2014. doi: 10.1007/978-3-319-10602-1_26.

[30] J Deng, J Deng, W Dong, R Socher, L Li, K Li, and L Fei-fei. Imagenet: A large-scale hierarchical image database. *IN CVPR*, 2009.

[31] C McManus, B Upcroft, and P Newman. Learning place-dependant features for long-term vision-based localisation. *Autonomous Robots*, 39(3):363–387, 7 2015. ISSN 0929-5593. doi: 10.1007/s10514-015-9463-y.

[32] C Piciarelli. Visual Indoor Localization in Known Environments. *IEEE Signal Processing Letters*, 23(10):1330–1334, 10 2016. ISSN 1070-9908. doi: 10.1109/LSP.2016.2593958.

[33] H Bay, A Ess, T Tuytelaars, and L Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 6 2008. ISSN 10773142. doi: 10.1016/j.cviu.2007.09.014.

[34] S Jianbo and Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, pages 593–600. IEEE Comput. Soc. Press, 1994. ISBN 0-8186-5825-8. doi: 10.1109/CVPR.1994.323794.

[35] M A Fischler and R C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 6 1981. ISSN 00010782. doi: 10.1145/358669.358692.

[36] R Sahdev and J K Tsotsos. Indoor Place Recognition System for Localization of Mobile Robots. In *2016 13th Conference on Computer and Robot Vision (CRV)*, pages 53–60. IEEE, 6 2016. ISBN 978-1-5090-2491-9. doi: 10.1109/CRV.2016.38.

[37] E Sizikova, V K Singh, B Georgescu, M Halber, K Ma, and T Chen. Enhancing Place Recognition Using Joint Intensity - Depth Analysis and Synthetic Data. In Gang Hua and Herv Jégou, editors, *ECCV Workshops (3)*, volume 9915 of *Lecture Notes in Computer Science*, pages 901–908, 2016. ISBN 978-3-319-49408-1.

[38] S Romberg, L G Pueyo, R Lienhart, and R van Zwol. Scalable logo recognition in real-world images. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval - ICMR '11*, pages 1–8, New York, New York, USA, 4 2011. ACM Press. ISBN 9781450303361. doi: 10.1145/1991996.1992021.

[39] S Romberg and R Lienhart. Bundle min-hashing for logo recognition. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval - ICMR '13*, page 113, New York, New York, USA, 4 2013. ACM Press. ISBN 9781450320337. doi: 10.1145/2461466.2461486.

[40] F N Iandola, A Shen, P Gao, and K Keutzer. DeepLogo: Hitting Logo Recognition with the Deep Neural Network Hammer. 10 2015.

[41] Multimedia Computing and Augsburg University Computer Vision Lab. FlickrLogos-32. URL `http://www.multimedia-computing.de/flickrlogos/`.

[42] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich. Going Deeper with Convolutions. 2015.

[43] S Wang, S Fidler, and R Urtasun. Lost Shopping! Monocular Localization in Large Indoor Spaces. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2695–2703. IEEE, 12 2015. ISBN 978-1-4673-8391-2. doi: 10.1109/ICCV.2015.309.

[44] Z Xu, M Pang, G Zhou, and L Fang. Feature Fusion for Storefront Recognition and Indoor Navigation. 10 2016.

[45] M Jaderberg, A Vedaldi, and A Zisserman. Deep Features for Text Spotting. pages 512–528. Springer International Publishing, 2014. doi: 10.1007/978-3-319-10593-2_34.

[46] A Krizhevsky, I Sutskever, and G E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[47] S George K and S Joseph. Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature. *IOSR Journal of Computer Engineering*, 16(1):34–38, 1 2014. ISSN 22788727. doi: 10.9790/0661-16153438.

[48] Y Yoshikawa, T Iwata, and H Sawada. Latent Support Measure Machines for Bag-of-Words Data Classification. In *Advances in Neural Information Processing Systems*, pages 1961–1969, 2014.

[49] X Glorot, A Bordes, and Y Bengio. Deep Sparse Rectifier Neural Networks, 2011. ISSN 1938-7228. URL `http://proceedings.mlr.press/v15/glorot11a.html`.

[50] A Tuerk and S J Young. Polynomial Softmax Functions for Pattern Classification. URL `http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=8B36B7AA89620BF16805335E3DA5F727?doi=10.1.1.26.6075`.

[51] D Kingma and J Ba. Adam: A Method for Stochastic Optimization. 2014.

[52] K Demirli and B Türkşen. Sonar based mobile robot localization by using fuzzy triangulation. *Robotics and Autonomous Systems*, 33(2-3):109–123, 11 2000. ISSN 09218890. doi: 10.1016/S0921-8890(00)00082-8.

[53] J Torres-Sospedra, R Montoliu, S Trilles, Belmonte, and J Huerta. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, 42(23):9263–9278, 12 2015. ISSN 09574174. doi: 10.1016/j.eswa.2015.08.013.

[54] A N Raghavan, H Ananthapadmanaban, M S Sivamurugan, and B Ravindran. Accurate mobile robot localization in indoor environments using bluetooth. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4391–4396, 5 2010. doi: 10.1109/ROBOT.2010.5509232.

[55] M H Habaebi, R O Khamis, A Zyoud, and M R Islam. RSS Based Localization Techniques for ZigBee Wireless Sensor Network. In *Computer and Communication Engineering (ICCCE), 2014 International Conference on*, pages 72–75, 9 2014. doi: 10.1109/ICCCE.2014.32.

[56] L Calderoni, M Ferrara, A Franco, and D Maio. Indoor localization in a hospital environment using Random Forest classifiers. *Expert Systems with Applications*, 42 (1):125–134, 1 2015. ISSN 09574174. doi: 10.1016/j.eswa.2014.07.042.

[57] L M Ni, Y Liu, Y C Lau, and A P Patil. LANDMARC: indoor location sensing using active RFID. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 407–415, 3 2003. doi: 10.1109/PERCOM.2003.1192765.

[58] J Wang, Y Wang, and X Guan. An indoor localization system based on backscatter RFID tag. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6, 4 2016. doi: 10.1109/WCNC.2016.7565023.

[59] A Meghdari, K Kobravi, H Safyallah, M Moeeni, Y Khatami, and H Khasteh. A New Approach to Sonar Based Indoor Mapping Localization. In *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, California, 2005.

[60] J Kolakowski, A Consoli, V Djaja-Josko, J Ayadi, L Morrigia, and F Piazza. UWB localization in EIGER indoor/outdoor positioning system. In *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems:*

Technology and Applications (IDAACS), pages 845–849. IEEE, 9 2015. ISBN 978-1-4673-8359-2. doi: 10.1109/IDAACS.2015.7341422.

[61] L Zwirello, T Schipper, M Harter, and T Zwick. UWB Localization System for Indoor Applications: Concept, Realization and Analysis. *Journal of Electrical and Computer Engineering*, 2012:1–11, 2012. ISSN 2090-0147. doi: 10.1155/2012/849638.

[62] M Shchekotov. Indoor Localization method Based on Wi-Fi Trilateration Technique. In *Proceeding of the 16th Conference of FRUCT association*.

[63] C A Balanis. *Antenna theory : analysis and design*. Wiley & Sons, Incorporated, John, first edition, 1982. ISBN 9780471603528.

[64] J S Seybold. *Introduction to RF propagation*. Wiley, 2005. ISBN 0471655961.

[65] S Y Seidel and T S Rappaport. 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40(2):207–217, 2 1992. ISSN 0018-926X. doi: 10.1109/8.127405.

[66] V Bahl and V Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System. Institute of Electrical and Electronics Engineers, Inc., 3 2000.

[67] Rajat Agarwal Y. Shreyas, Shivam Seth. Wireless Network Visualization and Indoor Empirical Propagation Model for a Campus WI-FI Network. *World Academy of Science, Engineering and Technology*, 2008.

[68] P Tarrío Alonso, H Martín Rodríguez, and A M Bernardos Barbolla. Enhancing the Performance of Propagation Model-Based Positioning Algorithms, 1 2009. URL `http://oa.upm.es/5601/1/INVE_MEM_2009_68030.pdf`.

[69] Cisco. Location Tracking Approaches. In *Wi-Fi Location-Based Services 4.1 Design Guide*. https://www.cisco.com, 2014.

[70] M Kotaru, K Joshi, D Bharadia, and S Katti. SpotFi: Decimeter Level Localization Using WiFi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication - SIGCOMM '15*, volume 45, pages 269–282, New York, New York, USA, 8 2015. ACM Press. ISBN 9781450335423. doi: 10.1145/2785956.2787487.

[71] J Xiong and K Jamieson. ArrayTrack: A Fine-Grained Indoor Location System. In *Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, pages 71–84, Lombard, IL, 2013. USENIX. ISBN 978-1-931971-00-3.

[72] S Kumar, S Gil, D Katabi, and D Rus. Accurate indoor localization with zero start-up cost. In *Proceedings of the 20th annual international conference on Mobile computing and networking - MobiCom '14*, pages 483–494, New York, New York, USA, 9 2014. ACM Press. ISBN 9781450327831. doi: 10.1145/2639108.2639142.

[73] A Marcaletti, M Rea, D Giustiniano, V Lenders, and A Fakhreddine. Filtering Noisy 802.11 Time-of-Flight Ranging Measurements. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies - CoNEXT '14*, pages 13–20, New York, New York, USA, 12 2014. ACM Press. ISBN 9781450332798. doi: 10.1145/2674005.2674998.

[74] M Ciurana, F Barcelo-Arroyo, and F Izquierdo. A ranging system with IEEE 802.11 data frames. In *2007 IEEE Radio and Wireless Symposium*, pages 133–136, 1 2007. doi: 10.1109/RWS.2007.351785.

[75] P Torteeka and X Chundi. Indoor positioning based on Wi-Fi Fingerprint Technique using Fuzzy K-Nearest Neighbor. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences Technology (IBCAST) Islamabad, Pakistan, 14th - 18th January, 2014*, pages 461–465, 1 2014. doi: 10.1109/IBCAST.2014.6778188.

[76] P Jiang, Y Zhang, W Fu, H Liu, and X Su. Indoor Mobile Localization Based on Wi-Fi Fingerprints Important Access Point. *International Journal of Distributed Sensor Networks*, 2015:1–8, 1 2015. ISSN 1550-1329. doi: 10.1155/2015/429104.

[77] C Laoudias, D G Eliades, P Kemppi, C G Panayiotou, and M M Polycarpou. Indoor Localization Using Neural Networks with Location Fingerprints. In Cesare Alippi, Marios Polycarpou, Christos Panayiotou, and Georgios Ellinas, editors, *Artificial Neural Networks – ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II*, pages 954–963. Springer Berlin

Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-642-04277-5. doi: 10.1007/978-3-642-04277-5_96.

[78] E Mok and B Cheung. An Improved Neural Network Training Algorithm for Wi-Fi Fingerprinting Positioning. *ISPRS International Journal of Geo-Information*, 2(3): 854–868, 9 2013. ISSN 2220-9964. doi: 10.3390/ijgi2030854.

[79] E Jedari, Z Wu, R Rashidzadeh, and M Saif. Wi-Fi based indoor location positioning employing random forest classifier. In *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, pages 1–5, 10 2015. doi: 10.1109/IPIN.2015.7346754.

[80] Y Feng, J Minghua, L Jing, Q Xiao, H Ming, P Tao, and H Xinrong. An Improved Indoor Localization of WiFi Based on Support Vector Machines. *International Journal of Future Generation Communication and Networking*, 7.

[81] C Song, J Wang, and G Yuan. Hidden Naive Bayes Indoor Fingerprinting Localization Based on Best-Discriminating AP Selection. *ISPRS International Journal of Geo-Information*, 5(10):189, 10 2016. ISSN 2220-9964. doi: 10.3390/ijgi5100189.

[82] L Dantanarayana. *Navigation and control for assistive robotics*. PhD thesis, University of Technology Sydney, 2016. URL `https://opus.lib.uts.edu.au/handle/10453/62982`.