KERNEL AND MULTI-CLASS CLASSIFIERS FOR MULTI-FLOOR WLAN LOCALISATION

By

MOHD AMIRUDDIN BIN ABD RAHMAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, & University of Sheffield in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

June 2016

To my lovely wife,

Siti Zaleha,

and children,

Amsyar, Sidrah, Aakif, Safaa and Addin.

Abstract of thesis submitted to the Senate of Universiti Putra Malaysia and University of Sheffield in fulfilment of the requirement for the degree of Doctor of Philosophy

KERNEL AND MULTI-CLASS CLASSIFIERS FOR MULTI-FLOOR WLAN LOCALISATION

By

MOHD AMIRUDDIN BIN ABD RAHMAN

June 2016

Chairman: Associate Professor Zulkifly Abbas, PhD

Faculty: Science

Indoor localisation techniques in multi-floor environments are emerging for location based service applications. Developing an accurate location determination and time-efficient technique is crucial for online location estimation of the multi-floor localisation system. The localisation accuracy and computational complexity of the localisation system mainly relies on the performance of the algorithms embedded with the system. Unfortunately, existing algorithms are either time-consuming or inaccurate for simultaneous determination of floor and horizontal locations in multi-floor environment. This thesis proposes an improved multi-floor localisation technique by integrating three important elements of the system; radio map fingerprint database optimisation, floor or vertical localisation, and horizontal localisation. The main focus of this work is to extend the kernel density approach and implement multi-class machine learning classifiers to improve the localisation accuracy and processing time of the each and overall elements of the proposed technique.

For fingerprint database optimisation, novel access point (AP) selection algorithms which are based on variant AP selection are investigated to improve computational accuracy compared to existing AP selection algorithms such as Max-Mean and InfoGain. The variant AP selection is further improved by grouping AP based on signal distribution. In this work, two AP selection algorithms are proposed which are Max Kernel and Kernel Logistic Discriminant that implement the knowledge of kernel density estimate and logistic regression machine learning classification.

For floor localisation, the strategy is based on developing the algorithm to determine the floor by utilising fingerprint clustering technique. The clustering method is based on simple signal strength clustering which sorts the signals of APs in each fingerprint according to the strongest value. Two new floor localisation algorithms namely Averaged Kernel Floor (AKF) and Kernel Logistic Floor (KLF) are studied. The former is based on modification of univariate kernel algorithm which is proposed for single-floor localisation, while the latter applies the theory kernel logistic regression which is similar to AP selection approach but for classification purpose.

For horizontal localisation, different algorithm based on multi-class k-nearest neighbour (kNN) classifiers with optimisation parameter is presented. Unlike the classical kNN algorithm which is a regression type algorithm, the proposed localisation algorithms utilise machine learning classification for both linear and kernel types. The multi-class classification strategy is used to ensure quick estimation of the multi-class kNN algorithms.

The proposed algorithms are compared and analysed with existing algorithms to confirm reliability and robustness. Additionally, the algorithms are evaluated using six multi-floor and single-floor datasets to validate the proposed algorithms. In database optimisation, the proposed AP selection technique using Max Kernel could reduce as high as 77.8% APs compared to existing approaches while retaining similar accuracy as localisation algorithm utilising all APs in the database. In floor localisation, the proposed KLF algorithm at one time could demonstrate 93.4% correct determination of floor level based on the measured dataset. In horizontal localisation, the multi-class kNN classifier algorithm could improve 19.3% of accuracy within fingerprint spacing of 2 meters compared to existing algorithms.

All of the algorithms are later combined to provide device location estimation for multi-floor environment. Improvement of 43.5% of within 2 meters location accuracy and reduction of 15.2 times computational time are seen as compared to existing multi-floor localisation techniques by Gansemer and Marques. The improved accuracy is due to better performance of proposed floor and horizontal localisation algorithm while the computational time is reduced due to introduction of AP selection algorithm.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia dan University of Sheffield sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KERNEL DAN KLASIFIKASI KELAS PELBAGAI UNTUK LOKALISASI BANGUNAN BERTINGKAT WLAN

Oleh

MOHD AMIRUDDIN BIN ABD RAHMAN

Jun 2016

Pengerusi: Profesor Madya Zulkifly Abbas, PhD

Fakulti: Sains

Teknik lokalisasi dalam bangunan untuk bangunan bertingkat sedang memuncul untuk aplikasi berdasarkan servis lokasi. Membangunkan sistem yang tepat dan cekap masa penting untuk anggaran lokasi semasa dalam bangunan bertingkat. Ketepatan dan kecekapan masa system bergantung terutamanya kepada prestasi algorithma yang terbenam dalam sistem. Namun, algoritma sedia ada adalah kurang cekap atau kurang tepat untuk dibangunkan dalam bangunan bertingkat. Tesis ini mencadangkan system lokalisasi bertingkat yang di tambah baik. Tiga elemen penting sistem iaitu pengoptimuman pangkalan data, lokalisasi lantai atau menegak, dan lokalisasi mendatar disepadukan dalam teknik lokalisasi bertingkat. Fokus utama kerja ini ialah untuk menambah baik ketepatan dan kecekapan pengiraan algoritma dengan mengambil kira setiap elemen tersebut.

Untuk pengoptimuman pangkalan data, teknik pemilihan titik akses (TA) yang baharu berdasarkan pemilihan TA berbeza dikaji untuk menambah baik ketepatan pengiraan berbanding teknik pemilihan TA yang lepas. Permilihan TA berbeza seterusnya ditambah baik dengan mengumpul TA berdasarkan ciri signal. Berdasarkan aspek ini, dua algoritma pemilihan TA dicadangkan iaitu algoritma Max Kernel dan Kernel Logistic Discriminant yang menggunapakai ilmu anggaran ketumpatan kernel dan klasifikasi pembelajaran mesin regresi logistik.

Untuk lokalisasi lantai, strategi adalah berdasarkan menggabungkan teknik pengklusteran fingerprint dengan algoritma lokalisasi lantai. Kaedah pengklusteran adalah berdasarkan pengklusteran kekuatan signal mudah dengan menyusun signal TA di setiap fingerprint berdasarkan nilai paling kuat. Dua algoritma lokalisasi lantai dinamakan algoritma Averaged Kernel Floor dan Kernel Logistic Floor dikaji. Algoritma pertama adalah berdasarkan pengubahsuaian algoritma kernel univariate yang digunakan untuk lokalisasi satu aras. Algoritma kedua menggunakan teori kernel regresi logistic yang sama dengan teknik pemilihan TA tetapi untuk tujuan klasifikasi.

Untuk lokalisasi mendatar, algoritma lokalisasi berbeza berdasarkan algoritma klasifikasi kelas pelbagai k -nearest neighbour (k NN) dengan parameter pengoptimum dicadangkan. Tidak sama seperti algoritma k NN klasik yang merupakan algoritma jenis regresi, algoritma lokalisasi yang dicadangkan juga berdasarkan pengklasifikasi pembelajaran mesin. Algoritma tersebut dicadangkan dalam dua versi iaitu linear dan kernel. Strategi pelbagai-kelas untuk klasifikasi digunakan untuk memastikan anggaran pantas algoritma kNN pelbagai-kelas.

Kesemua algoritma yang dibangunkan dibandingkan secara kendiri dan dianalisa dengan algoritma terdahulu untuk mengesahkan kejituan dan kemapanan. Di samping itu, penilaian algoritma dibuat dengan pelbagai pangkalan data bertingkat dan satu aras untuk memastikan kebolehgunapakaian algoritma yang dicadangkan. Dalam pengoptimuman pangkalan data, algoritma pemilihan TA yang dicadangkan boleh mencapai sehingga 91.3% ketepatan di antara lokasi 2 meter dan pada masa yang sama menurunkan 17.7% kerumitan pengiraan. Dalam lokalisasi lantai, algoritma lantai yang dicadangkan menujukkan sehingga 96.8% ketepatan lantai. Dalam lokalisasi mendatar, algoritma yang dibangunkan mencapai sehingga 93.7% ketepatan di antara lokasi 2 meter.

Algoritma tersebut kemudian digabungkan untuk menganggarkan lokasi peranti untuk bangunan bertingkat. Keputusan purata 73.6% ketepatan di antara lokasi 2 meter dan 93.4% ketepatan lantai menunjukkan peningkatan berbanding teknik terdahulu oleh Gansemer dan Marques. Penambahbaikan kejituan disebabkan oleh prestasi algoritma lokalisasi lantai dan mendatar yang lebih baik manakala pengurangan kerumitan pengiraan disebabkan oleh pengenalan algoritma pemilihan TA.

ACKNOWLEDGEMENTS

Thanks to God! Many people have helped me in succession of this project by supporting me with different level of assistance. I by my deepest gratitude would like to thank them for their contributions.

First of all, I would like to thank my both main supervisors in UPM and University of Sheffield, Associate Professor Dr. Zulkifly Abbas and Professor Dr. Jie Zhang. Dr. Abbas has been excellent in helping me finishing this thesis in both technical and non-technical matters. Professor Zhang has given continuous support and guidance during this study. Also I would like to thank my supervisory committee, Dr. Xiaoli Chu (Sheffield) and Associate Professor Dr. Alyani Ismail (UPM) for their advices. My special gratitude goes to Dr. Marzieh Dashti (Nokia Bell Labs, Dublin) who has helped me in developing my technical skills in indoor localisation area. Also my appreciation goes to Dr. Holger Claussen, the manager in Nokia Bell Labs Dublin, and Dr. Fahim Kawsar, Director Internet of Things in Nokia Bell Labs Antwerp, for their trust in me to demonstrate the fingerprinting indoor positioning research work during my short period of time over there. Last but not least, my wonderful family for their continuous support from the beginning of this journey till now. Special gratitude goes to my wonderful wife, Siti Zeleha, for her never-ending love, patience, and encouragement in all these years.

I certify that a Thesis Examination Committee has met on **28 June 2016** to conduct the final examination of Mohd Amiruddin Bin Abd Rahman on his thesis entitled **"Kernel and Multi-Class Classifiers For Multi-Floor WLAN Indoor Localisation"** in accordance with the Universities and University Colleges Act 1971 and the Constitution of the University Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Mohamad Rushdan Md Said, PhD
Associate Professor
Institute of Mathematical Research
Universiti Putra Malaysia
(Chairman)

Aduwati Sali, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia

Salam Khamas, PhD
Lecturer
Department of Electronic and Electrical Engineering
University of Sheffield

Josaphat Tetuko Sri Sumantyo, PhD
Professor
Centre for Enviromental Remote Sensing
Chiba University

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of **Doctor of Philosophy**. The members of the Supervisory Committee were as follows:

Zulkifly Bin Abbas, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Jie Zhang, PhD
Professor
Department of Electronic and Electrical Engineering
University of Sheffield
(Member)

Alyani Binti Ismail, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

Xiaoli Chu, PhD
Lecturer
Department of Electronic and Electrical Engineering
University of Sheffield
(Member)

BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly acknowledged;
- ownership of intellectual property from the thesis is as stipulated in the Memorandum of Agreement (MoA), or as according to the Universiti Putra Malaysia (Research) Rules 2012, in the event where the MoA is absent;
- permission from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) are required prior to publishing it (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: __________________

Name and Matric No: MOHD AMIRUDDIN BIN ABD RAHMAN (GS35311JD)

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:	____________________	Signature:	____________________
Name of Chairman of Supervisory Committee:	____________________	Name of Member of Supervisory Committee:	____________________

Signature:	____________________	Signature:	____________________
Name of Member of Supervisory Committee:	____________________	Name of Member of Supervisory Committee:	____________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AKF	Averaged Kernel Floor
ANN	Artificial Neural Network
AP	Access Point
CDF	Cumulative Density Function
FAF	Floor Attenuation Factor
GNSS	Global Navigation Satellite Service
IRLS	Iterative Reweighted Least Square
KDE	Kernel Density Estimate
KLF	Kernel Logistic Floor
KLPD	Kernel Logistic Pairwise Discriminant
KLR	Kernel Logistic Regression
*k*NN	k-Nearest Neighbour
LBS	Location-Based Service
LDA	Linear Discriminant Analysis
MAC	Media Access Control
MAP	*Maximum A Posteriori*
NLL	Negative Likelihood

NLOS	Non-Line of Sight
OVA	One-Versus-All
OVO	One-Versus-One
RBF	Radial Basis Function
RF	Radio Frequency
RSS	Received Signal Strength
SVM	Support Vector Machine
UPM	Universiti Putra Malaysia
WLAN	Wireless Local Area Network

LIST OF SYMBOLS

$\arg\max_{\xi} p(.)$	*Maximum A Posteriori* estimate of ξ
$\mathbf{A}_v$	Pairwise APs signal vector of v-th combination in the radio map database
$\mathbf{A}_\beta$	Pairwise APs signal vector of β-th combination in the online sample
AP_l	l-th access point
$\hat{c}$	Estimated cluster/floor
$\boldsymbol{c}_f$	f-th cluster
D_i	Signal distance of i-th fingerprint
$D_{\beta i}$	Nearest signal distance of kNN classifier at i-th fingerprint location
$diag[.]$	Diagonal matrix function
$\mathbf{F}(\mathbf{p}_i)$	Fingerprint element at i-th location
h	Adjustable similarity parameter of Kernel Density Estimate
i	Fingerprint/Location index
k	Number of nearest estimated location
K(.)	Kernel function
$NLL(.)$	Negative likelihood function of logistic regression

$\hat{\mathrm{p}}_{x\xi}$	Estimated x-coordinate location
$\hat{\mathrm{p}}_{y\xi}$	Estimated y-coordinate location
$\hat{\mathbf{p}}$	Estimated location
$\mathbf{p}_i$	Location of i-th fingerprint
$\mathrm{p}(.)$	Probability function
$\mathrm{P}_r(d_0)$	RSS at reference distance d_0 from an AP
$\mathrm{p}_{x\xi}$	Actual x-coordinate location
$\mathrm{p}_{y\xi}$	Actual y-coordinate location
$P_r(d)$	RSS at a fingerprint location of d distance from an AP
q	Location index of the AP l signal grouped in the cluster $\boldsymbol{c}_f$
$\boldsymbol{r}$	Vector of pairwise APs signal data
$RSS_{AP_l}(d_{AP_l})$	Received Signal Strength of AP_l at a distance d from the AP
$RSS_{AP_l}(d_0)$	Received Signal Strength of AP_l at a reference distance d_0
$S_{\beta i}$	Averaged of ϖ minimum $D_\beta(\psi_{\omega i})$ signal distance of i-th fingerprint location of kNN classifier
T	Transpose function
T	Total number of samples measured at one fingerprint location
U	Total number of points of signal range to determine the kernel in density estimate

$\boldsymbol{w}$	Model parameter vector of the logistic function
y	Class/Location of vector pairwise signal data $\boldsymbol{r}$
Ω_i	i-th class/location averaged $D_{\beta i}$or $S_{\beta i}$ signal distance of test sample AP combination
$\boldsymbol{\varphi}_i$	Offline signal vector of i-th fingerprint
Δf_{neigh}	Difference between two neighbouring kernel u and $u+1$
Ξ	Total number of location estimation
Φ	Total class labels or total number of vector signal data existed in both classes
ξ	Index of location estimation
η_{AP_l}	Path Loss Exponent referring to AP_l signals
η_{SF}	Similar floor path loss exponent
ρ	Number of class label from total of Φ
ϖ	Number of nearest minimum signal distance of each fingerprint location of kNN classifier
$\boldsymbol{\varphi}$	Online measured signal vector
γ	Adjustable similarity parameter of kernel function

CHAPTER 1

INTRODUCTION

1.1 Overview

Location is one of the most valuable information in mobile communication nowadays. Today’s mobile devices are designed and programmed to have location features so it can be complemented with Location-Based Service (LBS) applications (Schiller and Voisard 2004). The location is important because it reflects interaction and context of the user based on the location of the device. In past times, location is mainly used to guide users to move from one place to another by giving the best possible route to reach the destination. However currently, the location information is used in much wider context. For example, by using smartphone a user can locate user’s current position and share the location with his or her friends on the social network. Also, a user can book a taxi service or finding the nearest restaurants or cash machines by considering user’s current location.

Unfortunately, all of these applications infer the device location mainly based on its position in outdoor environment which mainly depends on the information provided by Global Navigation Satellite System (GNSS) receiver integrated with the mobile device. However, with only information of outdoor location, further development or enhancement of LBS applications is restricted. In near future, LBS applications are designed and developed to work for indoor-based services. For example, to assist shoppers to find items that they want to purchase by locating the exact aisle of the item in a hypermarket, to help drivers find their car in a multi-story indoor airport car park, and to supply information for smart building administrators to monitor temperature, room availability, and lightings. Therefore it is a requirement to know accurate indoor location to achieve these objectives. GNSS receiver however is generally not suitable to provide indoor location due to blockage and attenuation of the signals by roofs, walls and other objects.

Researchers have been working to find alternative technologies to obtain accurate indoor location information. Some solutions includes Wireless Local Area Network (WLAN) infrastructure (Fang and Lin 2010, Prieto et al. 2012, Mirowski et al. 2014, Wang et al. 2015, Liang, Zhang, and Peng 2015), infrared (Petrellis, Konofaos, and Alexiou 2006, Tao et al. 2014), and Bluetooth (Hossain and Soh 2007, Jianyong et al. 2014, Gu and Ren 2015). Among the solutions, one of the most promising solutions is WLAN as its signal coverage is available almost anywhere in urban environments. Indoor localisation methods based on WLAN are largely documented in the literature and surpass any other indoor localisation technologies.

1.2 WLAN Indoor Localisation

Indoor localisation based on WLAN was pioneered by Bahl and Padmanabhan (2000). WLAN based localisation is established by associating received Radio Frequency (RF) signals with physical location. The received RF signals or also known as Received Signal Strength (RSS) could characterise different locations as the propagated signals are location dependent. To localise unique location, RSS is measured throughout the floor area as combination of multiple signals from multiple Access Points (APs).

The location estimation technique could be classified by two methods; radio propagation based model and fingerprinting method. In radio propagation based model, the location is estimated by triangulation where the location of three or more access points must be known and the path loss model such as log-distance, which described the environment dependent relationship of the distance between the transmitter (AP) and receiver (device) according to variation of signal strength value, of the APs are determined. During position request by the device, the signal vector measured by the device at an unknown location is used as the input to the path loss model to determine the distance of the device from the APs which translates the location of the device. On the other hand, the fingerprinting method first requires real surveying by collecting the signal signature at every unique physical location which is also called as fingerprint location. The collection of multiple signal signatures associated with the physical locations are stored in the database as radio map and during the location request by the device, the signal vector of the device is compared its similarity with the one in the database to determine the location.

Between the two methods, the latter technique, fingerprinting, is preferred. This is because higher positioning accuracy could be achieved compared to radio propagation based model. Radio propagation based model could not provide finer accuracy due to inability of the model to characterise complex multipath signals received at each specific locations. However, fingerprinting technique comes with the cost of high processing time of localisation algorithm due to large amount of signal signatures in the radio map. In today's application, indoor localisation system should be embedded in mobile device such as smartphone which has small computing capability, the localisation algorithm must be designed and developed to utilise as small processing power as possible and at the same time retain good positioning accuracy. Some examples of good and robust classical localisation algorithms are k-Nearest Neighbour (kNN) (Bahl and Padmanabhan 2000), univariate kernel (Roos et al. 2002), and multivariate kernel (Kushki et al. 2007). Additionally, the localisation algorithm must be designed so that it can work in multiple indoor environments especially in urban area where the application is demanded. Generally, these areas are occupied with various multi-floor constructions. Therefore, the indoor localisation system must be designed and developed for this kind of infrastructure.

1.3 Multi-Floor Localisation

Numerous studies can be found on development of indoor localisation system. However, majority of them are focussing on single-floor localisation (Wu et al. 2013, Sorour et al. 2015, Chen and Wang 2015). It is investigated that the research on multi-floor localisation receives less attention is mainly due to two reasons. First, large radio map datasets is required as multi-floor data must be collected e.g. large fingerprint dataset for fingerprinting method, or large AP location dataset for AP based method. Second, the perception that development of multi-floor localisation could be easily extended from single-floor localisation technique. However, multi-floor WLAN localisation is actually much more challenging compared to single floor localisation. Figure 1.1 illustrates the comparison between single-floor localisation and multi-floor localisation. It is understood that multi-floor localisation challenges comes additional floor environment which increases the complexity of localisation in multi-floor setting.

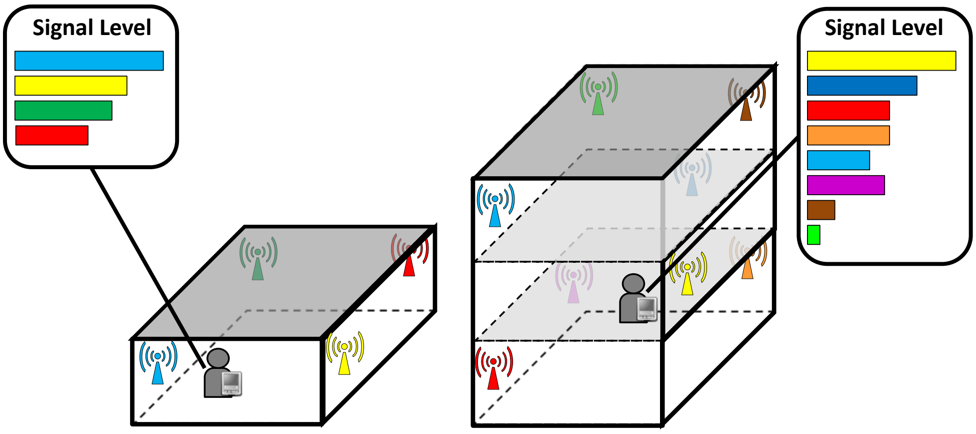


Figure 1.1. WLAN indoor localisation in single-floor and multi-floor settings

There are significant differences between multi-floor and single-floor WLAN localisation. First, the datasets of the collected signal data must be characterised by three-dimensional position, which includes floor level of the building and horizontal positions of the data compared to only horizontal positions required for single-floor localisation. As the amount of entries in the radio map varies according to the number of fingerprint locations, therefore the amount of entries of the datasets for multi-floor localisation is generally in multiplication of the number of fingerprint locations and number of floor level exists within the building. Second, the number of APs increases proportionally with the number of floor level in multi-floor building. This causes the dimensionality of AP during online phase increase. Additionally, during the signal measurement process, each signal vector for multi-floor localisation will be added with multiple APs signal from other floors in addition to signal from the APs on the floor itself compared to single-floor where the AP signals mostly come from the APs installed on the floor. The computation of the localisation algorithms depends on the three factors which are number of fingerprint locations, number of APs within the environment and the size of signal vector and these are the elements that are mentioned in first, second, and third comparisons. Therefore the computational complexity or processing time of the algorithm in multi-floor setting will be much higher compared to single-floor localisation. The fourth difference is the estimated locations in multi-floor environment require additional coordinate of the floor level or the z-coordinate compared to single-floor environment which is described by only x and y-coordinate. Lastly, the probability of error in multi-floor localisation is generally higher because of possibility that the estimated location is in different floor level than expected. All of these comparisons are summarised in Table 1.1.

Multi-floor localisation system could be divided into two main problems which are to locate the floor level of the device and to position the device on the chosen floor level which determines the horizontal location. The algorithms should be accurate and quickly processed the radio map database to give estimation of the location. Therefore, three important categories that should be investigated in order to produce an efficient multi-floor location are radio map fingerprint database optimisation, floor localisation algorithm, and horizontal localisation algorithm.

1.4 Kernel and Multi-Class Classifier Approach for Multi-Floor Localisation

In particular, the technique to estimate the location in multi-floor environment has been focusing on similar type of algorithm. The algorithm also is the extension of previously developed algorithm for single-floor localisation. Multi-floor localisation involves processing larger database compared to single-floor and therefore the extended algorithm is no longer suitable to be applied for multi-floor case. The usage of classical algorithms in multi-floor environment also leads to increasing computational complexity as related to increasing amount of database element. To solve the problem, this thesis proposes new multi-floor localisation technique based on kernel and multi-class classifier. The techniques implements kernel density estimate and multi-class classifier based on logistic regression as tools for AP selection, and floor localisation algorithm. The kNN multi-class classifier is applied for new horizontal localisation algorithm. Theoretical foundation and algorithm implementation of the technique is described in details in Chapter 3 and 4 respectively.

Table 1.1. Summary of differences between single-floor WLAN localisation versus multi-floor WLAN localisation

Issue	Single Floor Localisation	Multi-Floor localisation
• Radio map database	• Radio map database is for single floor and the quantity of the entries of the database depends on the number of fingerprint locations.	• Radio map database is in multiple number of floors exist in the building and the entries generally consists of multiplication of number of fingerprint locations and number of floor level.
• Dimensions of AP	• Dimensionality of AP depends on the number of APs installed in single floor.	• Dimensionality of APs increase with increasing number of APs installed on every floor level of the building.
• Fingerprint signal vector	• Signal vector at each fingerprint location is majorly from multiple APs that are installed on the floor level and the signal follows normal path loss model.	• Signal vector at each fingerprint location consists of signal from APs within similar floor level and also from other floor level within the building and the signal follows multi-floor path loss model.
• Estimated location coordinate dimension	• Estimated location is in 2 dimensional coordinate (x and y) which is the horizontal location.	• Estimated location is in 3 dimensional coordinate (x, y, and z) which includes floor location and horizontal location.
• Probability of location error	• The probability of location error is only within horizontal locations.	• The probability of location error may increase due to possibility of estimated location is located on different floor level than expected.

1.5 Problem Statement

To develop an efficient multi-floor localisation system, all of the elements of multi-floor localisation in Table 1.1 should be analysed according to categories mentioned above. There have been some developments of multi-floor WLAN localisation system and comprehensive review on the topic is written in Chapter 2 of this thesis. From the review, it is indicated that majority of multi-floor WLAN localisation algorithm are developed based on extension of single-floor localisation algorithms. Generally, the problem with this kind of system did not consider computational complexity of the algorithms implemented in multi-floor environment. Also, the algorithm is not optimised for simultaneous estimation of both floor and horizontal locations. Specifically, in order to develop an efficient and robust multi-floor indoor localisation system, this thesis investigates the following problems:

1. Majority of the proposed floor localisation algorithm is still based on classical similarity measure algorithm which is extended for multi-floor localisation. This means the developed floor localisation algorithm requires calculating every single entry of fingerprint to perform floor estimation. As discussed in Section 1.3, the multi-floor building problem involves the number of fingerprints is in the multiple of number of floor level exists inside the building. Therefore the computational complexity increases as the number of floor increases.

2. Considering all APs for localisation may degrade the performance of the localisation system. The number of APs installed within building increases as the number of floor level increases so that the coverage of the signal is enough for localisation system to work. This leads to AP dimensionality in multi-floor environment is much higher compared to single floor.

3. The horizontal localisation algorithm in multi-floor setting is mainly implemented based on previously developed algorithm for single-floor problem. However in multi-floor building, additional AP signals are measured from other floor levels which degrade the performance of the algorithm. Consequently, the location estimation error in multi-floor location could not be minimised compared to single-floor location if similar algorithm is implemented for both environments.

4. The validity of some existing multi-floor localisation algorithm is questionable as the algorithms are only tested in limited testing area such as one or two buildings and the buildings are low rise which contains less than five floor levels, and it is not guaranteed that the proposed algorithm will produce the similar performance as in different environments.

5. The work on combining radio map database optimisation, floor localisation and horizontal localisation is not well studied to improve the multi-floor localisation system. Existing techniques are based solely on either improving floor localisation only or combination of database optimisation and floor localisation. The performance of combining all of the techniques is unknown.

1.6 Objective of the Research

The aim of this thesis is to develop a robust and efficient multi-floor localisation system emphasizing on: i) the accuracy of the localisation algorithms in both vertical and horizontal position which are characterised by estimated location error of the algorithms, and ii) the computational complexity of localisation algorithm which is to reduce the processing time of the developed algorithms. To achieve the aim, detail objectives are given as follows:

1. To optimise the radio map database by implementing AP selection technique to limit or reduce the number of required APs information to perform localisation and at the same time retain similar accuracy with using all APs information. Two novels AP selection algorithms are introduced to improve the selection of APs in optimising the database. The performances of the proposed AP selection schemes are compared with existing AP selection technique to evaluate the performance of the proposed algorithms based on three classical localisation algorithms of kNN, univariate kernel, and multivariate kernel.

2. To reduce the processing complexity of floor localisation algorithm and at the same time to improve the accuracy of the estimated floor level. Proposed two new floor localisation algorithms based on clustered multi-floor radio map. The performances of the new floor localisation algorithms are compared with existing floor localisation algorithms to evaluate the effectiveness of the algorithm.

3. To improve the location estimation error of horizontal localisation algorithm by introducing novel localisation algorithms for both multi-floor and single-floor environments. The horizontal localisation algorithms are compared with classical single-floor localisation algorithms (kNN, univariate kernel, and multivariate kernel) and the performances of all algorithms are analysed and discussed.

4. To combine the proposed AP selection technique, floor localisation algorithm and horizontal localisation algorithms to evaluate performance in multi-floor environment. The test is in multiple multi-floor environments with different number of floor levels to verify the performance of the proposed algorithm. The performance is also compared with existing multi-floor localisation algorithm.

1.7 Organisation of the Thesis

In Chapter 2, review of the literature on the multi-floor localisation system is given. The review presents in depth study on problems exist in existing multi-floor localisation system which leads to the development of problem statement in Section 1.5. The review identifies the gaps in current research particularly in multiple scenarios of multi-floor e.g. validity of chosen environment for testing the multi-floor localisation system, radio map database optimisation techniques, floor localisation algorithms, and lastly the horizontal localisation algorithms.

Chapter 3 first presents the background on WLAN fingerprint localisation and theory that is related to localisation algorithm used in this thesis and second describes the novel algorithms developed for the multi-floor localisation. The state-of-the-art WLAN fingerprint localisation mechanism is introduced. The related improvement components of the multi-floor localisation are presented which involves the database optimisation, floor localisation, and horizontal localisation. The theory of three popular classical algorithms which are used as benchmark for proposed algorithms are also discussed. The theory on kernel density estimate and machine learning multiclass classification using kNN and logistic regression are explained. The theories are applied for the following proposed algorithms. The database optimisation algorithms implement Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD). The floor localisation comprises of Averaged Kernel Floor and Kernel Logistic Floor algorithms. Normal and kernel multi-class kNN algorithms are used for horizontal localisation

Chapter 4 explains the measurement setup of collecting the RSS signal data. This includes the measurement tools, floor map, and specification of the measurement. Also the method to extract the path loss parameters of the measured signal data to be tested with propagation model is shown. The details of measured fingerprint database specification and evaluation of path loss model using extracted path loss parameter of the measured signal data from the database are described. The performance metrics to evaluate the developed algorithms are discussed. Additionally, the method to determined number of k used for kNN algorithm is presented.

Chapter 5 to 8 discuss the results related to the developed multi-floor system. The results explain the performance of the proposed AP Selection (Chapter 5), floor localisation (Chapter 6), and horizontal localisation algorithms (Chapter 7) of which results are evaluated and discussed. The results of combining the three proposed algorithms for multi-floor localisation are explained in Chapter 8. The results are mainly focusing on accuracy and computational complexity of the algorithms.

Lastly, Chapter 9 draws conclusion on the proposed multi-floor system. The work presented in the thesis is summarised. The contributions of the thesis are highlighted. Also, further research directions are suggested.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Existing multi-floor localisation techniques still face problems as discussed in previous chapter. This chapter explores the problems in details and reviews the techniques of Wireless Local Area Network (WLAN) localisation in multi-floor building. The trends and development of existing multi-floor localisation systems are discussed. Section 2.2 discusses the two categories of multi-floor localisation systems which are the Access Point (AP)-based and fingerprint. Section 2.3 explains the inconsistency in preparing the radio map database for multi-floor localisation and the resulting performance in existing works. Section 2.4, 2.5 and 2.6 highlights the main issues in existing multi-floor localisation system which are optimisation of radio map database technique, floor localisation algorithm, and horizontal localisation algorithm. In Section 2.7, the existing techniques for multi-floor localisation are discussed.

2.2 Multi-floor WLAN Localisation

Multi-floor WLAN localisation determines simultaneously both the vertical which the floor level and horizontal positions of the device compared to single-floor localisation which only requires determination of horizontal location. Up to date, most of the work in WLAN localisation is focused on single floor horizontal localisation and less attention is given to multi-floor localisation. Excellent reviews on single-floor WLAN localisation technique or mainly for single-floor localisation techniques and systems can be found in Liu et al. (2007), Honkavirta et al. (2009), Farid, Nordin, and Ismail (2013), and He and Chan (2015). However, comprehensive review on the multi-floor localisation techniques has not been found. Therefore, the review on the existing development of the multi-floor localisation technique is provided here.

In multi-floor environment, the localisation strategy is generally described by two techniques which are single-stage and two-stage estimations. In single-stage estimation, the localisation of floor level and horizontal location are simultaneously determined such as demonstrated in Al-Ahmadi et al. (2010) and Lohan et al. (2015). On the other hand, the two-stage localisation, which is more popular approach, first estimates the floor level and second performs horizontal localisation. Example of the works with two-stage localisation could be found in Liu and Yang (2011), Marques, Meneses, and Moreira (2012), Campos, Lovisolo, and de Campos (2014). The two-stage approach is more flexible and usually requires less dataset during localisation because the dataset of unrelated floor levels is filtered after floor level determination process is done. Also the two-stage approach gives flexibility in terms of developing two different algorithms for both floor classification and horizontal position estimation.

In multi-floor localisation system, the location of the measured test sample is assumed known in order to describe the accuracy of the multi-floor localisation algorithm by comparing it with the output location. In general, three types of accuracies are reported which are the floor or vertical accuracy (z), horizontal accuracy (x and y), and the multi-floor accuracy which calculates the location accuracy based on difference of considering both vertical and horizontal location of estimated and actual location (Cheng et al. 2014, Campos, Lovisolo, and de Campos 2014). The floor accuracy is the correctness of floor level estimated by the algorithm given the current position of the device. The horizontal accuracy is defined as the distance error of estimated horizontal position to actual horizontal position of the device. The multi-floor accuracy is determined based on combination of floor and horizontal location error of the estimated position from the actual position of the device.

Figure 2.1 shows graphical summary of the works on multi-floor localisation system. The blue writings with arrows show that the technique used at each sub component of multi-floor localisation system. The numbers associated with the arrows refers to the reference of the work. For example, path loss model technique is used for floor and horizontal localisation in multi-floor WLAN AP based localisation system by Gupta et al. (2014) and Bhargava, Krishnamoorthy, and Karkada (2013) and is also used in single-stage multi-floor WLAN fingerprint localisation system by Lohan et al. (2015) . The detail description of each sub component in the black circles is discussed in the following sections.

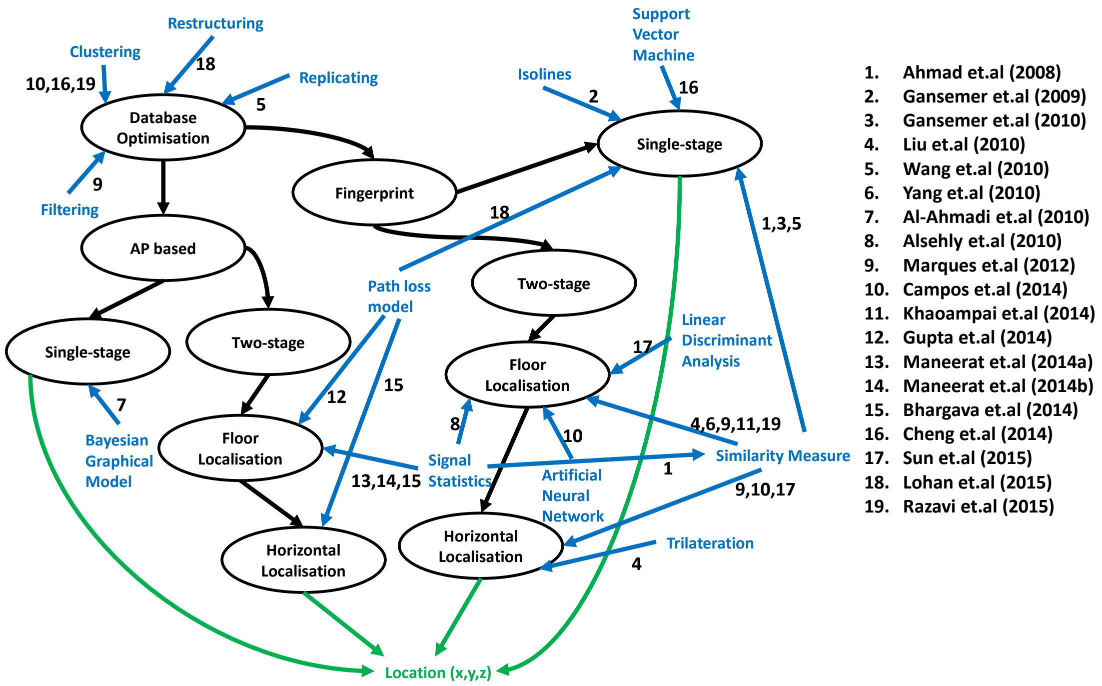


Figure 2.1. Graphical summary of existing works in multi-floor indoor localisation system

2.3 Preparation of Datasets for Multi-floor Localisation

First and foremost element in indoor localisation is the preparation of dataset which includes both measurement samples and test samples. Preparing the correct set of database ensures evaluation of localisation algorithm is not biased and thus reporting true performance of the algorithm. Some studies has reported very accurate floor localisation algorithm because utilising small test samples measurement dataset. For example, Liu and Yang (2011) reported that the proposed floor localisation algorithm achieves 100% accuracy based on three test samples on each floor of the investigated buildings. However later the author reworked on similar algorithm by using larger test set shows that the accuracy of floor detection is reduced to 88% (Liu, Liao, and Lo 2015). Additionally, it could be predicted that high accuracy could be achieved for tests in a low-rise building e.g. floor localisation accuracy is 99.5% in three-story building as presented by Marques, Meneses, and Moreira (2012). Comparison can be observed in different studies where large dataset is used in high-rise building. For example, algorithm proposed by Campos, Lovisolo, and de Campos (2014) accurately estimate floor at 91% by using 924 samples for both fingerprints and testing. Razavi, Valkama, and Lohan (2015) also reported comparable accuracy of 92% in six-story mall building tested with Nearest Neighbour (NN) algorithm using 1633 fingerprints and 3503 test samples.

2.4 Multi-floor Localisation Fingerprint Database

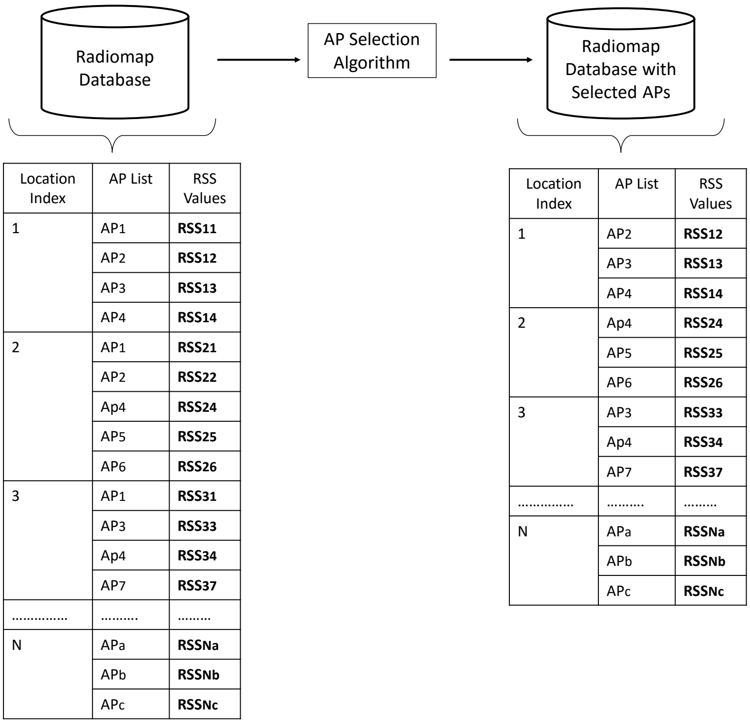
Upon receiving localisation request from a mobile device, the first element that is processed by localisation algorithm is the radio map database (refer to Figure 3.1). To ensure efficient localisation, the radio map could be optimised. The main aims of optimising radio map are two-fold which are to reduce computational complexity of the localisation algorithm and to achieve improved accuracy in estimated location. The techniques includes fingerprint clustering (Campos, Lovisolo, and de Campos 2014, Cheng et al. 2014, Razavi, Valkama, and Lohan 2015), replicating the fingerprint elements (Wang et al. 2010), database filtering (Marques, Meneses, and Moreira 2012), and restructuring the database into different form (Lohan et al. 2015).The database optimisation algorithms could be combined together or perform individually. The optimised database later could be used to reduce complexity of the processing algorithm. Even though the aim of database optimisation is to lower computational complexity and at least retain similar accuracy as existing database, some works reported mixed results when using an optimisation technique by reforming the database. For example, Lohan et al. (2015) replace the radio map dataset with analytical path loss model to reduce the size of the database and it is reported that improved floor accuracy is obtained but at the same time the 3D localisation error is larger than using normal radio map. For clustering technique, it has been reported that improved localisation accuracy is acquired e.g. kohonen layer with conscience clustering improves floor accuracy and horizontal localisation error compared to using normal signal similarity measure method from 78% to 85% and from 7.9 m to 5.0 m respectively (Campos, Lovisolo, and de Campos 2014), and k-means clustering offers slight improvement over NN in floor accuracy (Razavi, Valkama, and Lohan 2015). Although database optimisation technique gains popularity in WLAN single-floor localisation, the research on the topic is still limited for multi-floor application despite some few techniques mentioned above. For instance, AP selection and database filtering technique has been paid little attention even though many methods exist in single-floor case such proposed in Kushki et al. (2007), Park et al. (2010), Chintalapudi, Padmanabha Iyer, and Padmanabhan (2010), Lin et al. (2014). In this work, AP selection is proposed for optimising the fingerprint database and description of the AP selection mechanism is described in the following section.

2.4.1 AP Selection Mechanism

During fingerprint data collection in WLAN environment, multiple AP signals are detected at the receiver and it is common to use as much signals as possible to improve the accuracy of the system. However, such method suffers high computational time, complexity, and processing power due to large computational variables in large dimensions. The AP list which is one part of the dimensions could contain redundant or useless information for positioning and in some cases may degrade the accuracy of the localisation algorithm. Therefore, AP selection mechanism has been proposed to reduce dimensionality of APs during positioning to overcome such problems (Fang and Lin 2010, Fang and Lin 2012, Lin et al. 2014).

In real indoor localisation problem, the amount of APs found at every scanned location of the signal is usually more than three APs which are more than needed to perform localisation. The approach of AP selection algorithm is to select subset of APs from observed APs based on some selection criteria. Figure 2.2 shows overview of the working mechanism of AP selection using subset of three APs.

In previous single-floor localisation researches, AP selection is approached in two ways. First, the AP is selected based on distribution of signal of the APs within each similar fingerprint. For example, the AP selection in this category is Max Mean which is proposed by Youssef, Agrawala, and Udaya Shankar (2003), and Signal Divergence which was investigated by Kushki et al. (2007). Second, the APs are chosen according to variation of the AP signals across the floor or the whole fingerprint locations. In other words, the APs are selected based on discrimination of its signals between each fingerprint location. Example of this type of AP selection are InfoGain (Chen et al. 2006) and Group Discriminant (Lin et al. 2014). Generally, former type of AP selection is computationally efficient compared to the latter. But in terms of performance wise, it was reported that the latter algorithm could provide better estimation. However, since the AP selection is generally determined in offline phase, the computation complexity could be marginalised. In this work, two novel AP selection algorithms respectively based on information of AP signal distribution and also AP signal variation across fingerprint are proposed to improve the performance of existing techniques for multi-floor localisation. The proposed algorithms are described in details in Chapter 4.



Location Index	AP List	RSS Values
1	AP1	**RSS11**
	AP2	**RSS12**
	AP3	**RSS13**
	AP4	**RSS14**
2	AP1	**RSS21**
	AP2	**RSS22**
	Ap4	**RSS24**
	AP5	**RSS25**
	AP6	**RSS26**
3	AP1	**RSS31**
	AP3	**RSS33**
	Ap4	**RSS34**
	AP7	**RSS37**
....	....	....
N	APa	**RSSNa**
	APb	**RSSNb**
	APc	**RSSNc**

Location Index	AP List	RSS Values
1	AP2	**RSS12**
	AP3	**RSS13**
	AP4	**RSS14**
2	Ap4	**RSS24**
	AP5	**RSS25**
	AP6	**RSS26**
3	AP3	**RSS33**
	Ap4	**RSS34**
	AP7	**RSS37**
....	....	....
N	APa	**RSSNa**
	APb	**RSSNb**
	APc	**RSSNc**

Figure 2.2. AP Selection mechanism with subset of three APs (Chen et al. 2006)

2.5 Fingerprint Versus AP-based Technique

The technique to localise a mobile device in multi-floor building implementing AP-based technique is demonstrated in Alsehly, Arslan, and Sevak (2011), Bhargava, Krishnamoorthy, and Karkada (2013), Gupta et al. (2014), Maneerat and Prommak (2014), Maneerat, Prommak, and Kaemarungsi (2014). The steps to determine the multi-floor location is done as follows. First, the system must installed new APs within the environment or must locate the location of previously installed APs within the building. Second, enough signal measurement should be measured at different locations throughout the building. Third, the localisation algorithm is based on predicting the output of path loss model as estimated location (Bhargava, Krishnamoorthy, and Karkada 2013, Gupta et al. 2014) or statistical signal analysis to determine the estimated location (Alsehly, Arslan, and Sevak 2011, Maneerat and

Prommak 2014, Maneerat, Prommak, and Kaemarungsi 2014). The floor accuracy reported using AP-based technique is quite high e.g. 99% in Bhargava, Krishnamoorthy, and Karkada (2013), 100% in Gupta et al. (2014), and also 100% in Maneerat and Prommak (2014). However, the horizontal location accuracy is undesirable e.g. horizontal mean error of 9.4 m as reported in Locus. The source of the error is mainly contributed from sub-optimal path loss model parameters that could not well-predict the variation of RSS at measured locations. Because of this issue, many works on WLAN multi-floor system are based on fingerprint method.

2.6 Multi-floor WLAN Fingerprint Localisation Algorithms

The localisation algorithm is the heart of an indoor localisation system. It describes overall performance of a localisation system. The study on multi-floor localisation algorithm is the most reported research findings in multi-floor indoor localisation. The main motivation to design an algorithm for indoor localisation is to improve the existing algorithm performance especially in terms of minimising the mean location error and improving the precision of the estimated locations. The techniques to perform localisation in multi-story building could be categorised into two which is single-stage and two-stage localisation as discussed in Section 2.2 and detailed review of both techniques are given in subsections below.

2.6.1 Single-stage Multi-floor Localisation Algorithm

The algorithm search space of the database for single-stage localization proportionally increases with increasing number of floor level inside the building.

This leads to slow output from the algorithm e.g. (Razavi, Valkama, and Lohan 2015) demonstrate that processing of NN algorithm on a large four-floor buildings using single-stage localization could take up to 32.6 seconds. Because of such reason, only few works implements single-phase method e.g. Ahmad et al. (2008), Gansemer et al. (2009), Gansemer, Grossmann, and Hakobyan (2010), Cheng et al. (2014). The proposed localisation algorithms for single-stage multi-floor localisation are focusing on measuring similarity between fingerprint samples and test samples (Gansemer, Grossmann, and Hakobyan 2010) with addition of signal statistics (Ahmad et al. 2008), isolines (Gansemer et al. 2009), and support vector machine (Cheng et al. 2014).

2.6.2 Two-stage Multi-floor Localisation Algorithm

The two-stage localisation algorithm is more efficient and is the trend in multi-floor localisation. The two-stage localisation could be described by two different algorithms which are floor localisation algorithm and horizontal localisation algorithm.

2.6.2.1 Floor Localisation Algorithm

There are quite a number of algorithms that have been proposed to determine the floor using WLAN fingerprints. The floor localisation algorithms have been approached using techniques such as similarity measures (Liu and Yang 2011, Razavi, Valkama, and Lohan 2015, Marques, Meneses, and Moreira 2012), artificial neural network (Campos, Lovisolo, and de Campos 2014), and linear discriminant analysis (LDA) (Sun et al. 2015). However the techniques have not been compared to each other except in Sun et al. (2015). Therefore, the performance of existing floor localisation algorithm could not be well-summarised. However, it is known that feature extraction algorithm such as LDA could provide better estimate to floor location by choosing the most distinguishable signal pattern compared to normal similarity measure techniques. Table 2.1 summarises the current available floor localisation algorithm based on WLAN. It could be noted that the proposed algorithms mainly originating from single-floor localisation algorithm which is modified to estimate the floor location (Liu and Yang 2011, Khaoampai, Nakorn, and Rojviboonchai 2014, Razavi, Valkama, and Lohan 2015). According to author's knowledge, other algorithms which has proven to give high localisation accuracy in single-floor case such as probabilistic algorithms, support vector machines, and some other types of machine learning classification algorithms has not been investigated for floor localisation.

Table 2.1. Comparison of WLAN fingerprinting floor localisation algorithms

Reference	Algorithm Name	How floor is determined?	Number of floor level during test	Reported Floor Accuracy
Liu (2010) Liu (2014)	Grouped nearest neighbour	The lowest signal distance of averaged grouped fingerprint on every floor	8	100.0% (using small test samples (2010)) 88.2% (after considering larger test samples (2014))
Campos (2014)	Kohonen layer clustering with Backpropagation Artificial Neural Network	Majority voting of binary classifier of related floor	13	91.0%
Khaoampai (2014)	AP fingerprint similarity	Matching similarity order of AP ID for each floor	4	90.8%
Razavi (2015)	k-means fingerprint clustering with nearest neighbour	Lowest Euclidean distance of fingerprints within chosen cluster	Univ-1 – 4 Univ-2 - 3 Mall - 6 Office – 4	90.4% (Building Univ-1), 98.7% (Building Univ-2), 93.7% (Mall), 80.4% (Office)
Sun (2015)	Linear discriminant analysis	Most discriminative floor with largest eigenvalue	6	94.3%

2.6.2.2 Horizontal Localisation Algorithm

There are diverse types of algorithms available for single-floor localisation that has been demonstrated in literature e.g. probabilistic approach (Mirowski et al. 2014, Mu et al. 2014), Support Vector Machine (SVM) (Tran and Thinh 2008), and Artificial Neural Network (ANN) (Ahmad et al. 2006, Shih-Hau and Tsung-Nan 2008, Laoudias, Kemppi, and Panayiotou 2009, Genming et al. 2013) . However in multi-floor environment, it is observed that the most of the horizontal localisation problems have been solved using similar type of algorithm which is the similarity function. For example, Marques, Meneses, and Moreira (2012) test multiple similarity functions which are Euclidean, Manhattan, and Tanimoto distance to determine horizontal locations. Campos, Lovisolo, and de Campos (2014) implements non-square root Euclidean distance similarity function for horizontal localisation after performing clustering technique to the database. Recently, Sun et al. (2015) implements weighted version of similarity function called as weighted-RSS algorithm. On the other hand, only one different type of algorithm based on trilateration approach which pre-estimates virtual AP location to estimate the horizontal locations as implemented in Liu and Yang (2011). It can be noted that all of these algorithms are based on classical approach which has been applied in single-floor localisation. Moreover, the performance of these algorithms is unknown as the performance has not been compared. Unlike single-floor localisation technique, normally the performance of proposed algorithm is compared with established algorithm such as k-Nearest Neighbour (Bahl and Padmanabhan 2000), univariate kernel (Roos et al. 2002), and multivariate kernel algorithm (Kushki et al. 2007). Other interesting algorithms such as probabilistic approach e.g. Gaussian and kernel density estimates, classification analysis, and neural networks, and their performances have not been tested. This is similar case to the floor localisation algorithm as discussed above in Section 2.6.2.1.

2.7 System Performance of Existing WLAN Fingerprint Multi-floor Indoor Localisation

Table 2.2 lists the existing multi-floor localisation systems with their performance-wise. The first three rows list the single-stage localisation system and others are the two-stage localisation system. The performance of the system is usually described in terms of localisation errors which are the floor error, horizontal distance error and multi-floor location error. The errors are the difference between estimated location and actual location presents accuracy of the location information. However, it is noted that the error parameters are inconsistently reported. For example, all of single-phase systems use different error measures e.g. Gansemer et al. (2009) and Cheng et al. (2014) use precision measure by reporting the error in percentile while Lohan et al. (2015) report the error in root mean square error. This is different from single-floor localisation system where usually mean error is used as indicator to measure accuracy of the system which follows the guidelines given in Liu et al. (2007). Additionally, multi-floor location error is not reported in existing two-stage localisation systems. From the table, it could be assumed that two-stage system is better for implementation in order to achieve flexibility in computation and better average location accuracy. The mean error near to 1 m could be achieved with proper development of the system such as in Sun et al. (2015).

Table 2.2. Comparison of multi-floor localisation system

System	Database Optimisation (Category and Name)	Algorithm	Floor Error (%)	Mean Error (Horizontal) (m)	Mean Error (Multi-floor) (m)	Limitation of the System
Gansemer (2009)	-	Trilateration (Isolines)	13.3 (university), 3.2 (museum)	4.2 (university), 5.2 (museum) (50th percentile error)	-	Lower accuracy of estimated locations compared to popular nearest neighbour algorithm
Cheng (2014)	K-medoids (clustering)	Machine learning classification (Support vector machine)	-	-	6.0 (50th percentile error)	Accuracy of the location estimation is not confirmed with established algorithms
Lohan (2015)	AP-based dataset (restructuring)	Path loss model	3.1	-	9.9 (root mean square error)	Less accurate due to over-fitting of model to measured data
Gansemer (2010)	-	Similarity function (with separated single fingerprint element to multiple elements according to different directions)	0.6	2.0 (50th percentile error)	-	Extensive computation due to addition of data at each fingerprint dataset
Liu (2011)	-	Floor localisation - Similarity Function (grouping fingerprint according to floor), Horizontal localisation - Trilateration (weighted screening)	0.0	1.7	-	Works only on small dataset and validity of the test samples is questionable
Marques (2012)	Strongest signal strength (filtering)	Floor localisation and Horizontal localisation - Similarity function (with majority rule threshold)	0.5	3.4 (Mahhattan Distance)	-	Extensive computation due to repetition of filtering process
Campos (2014)	Principal component analysis (filtering) & Kohonen Layer (clustering)	Floor localisation - Artificial Neural Network (Backpropagation), Horizontal localisation - Similarity function (Non-Squared Euclidean Distance)	9.0	4.5	-	Requires averaged periodic online test samples to achieve high accuracy. Measurement fingerprint grid spacing is not mentioned.
Sun (2015)	-	Floor localisation - Machine learning classification (Linear Discriminant Analysis) Horizontal localisation - Similarity function (Weighted-RSS)	5.7	1.2 (assisted with accelerometer)	-	Requires combination with other sensor data to localise

From the table, the existing algorithms used in the multi-floor localisation system could be divided and explained as in the following sections.

2.7.1 Trilateration

The method of trilateration finds absolute location of measured online signal using the measurement of distances. The measurement of distance could be calculated using geometry shapes such as circles, spheres or triangles. Gansemer et al. (2009) develop isolines technique based on trilateration using triangles. A number of triangles are drawn within the floor plan of the measurement and measurements of offline signal strength are made at every triangle node. An interpolation technique called isolines is used to estimate the signal coverage of an online sample within a matching triangle. Multiple isolines are drawn if the signals from multiple APs matched between online sample and offline. The number of isolines intersections within a triangle determines the location of the device. On the other hand, Liu and Yang (2011) proposes weighted screening technique which implements trilateration using circles. The technique involves two steps. In the first step, circles are drawn based on offline measured signal strength of the device at three different locations to estimate location of AP from the signals. The locations are chosen based on the most similar signal strength in offline and online sample. In the second steps, the estimated AP locations are used to draw circles to find intersection points between the circles which determine the location of the device. The distance of the circles are calculated based on path loss model. Interpolation technique requires numerous geometry shapes to be drawn and accordingly multiple signal strength measurement are required to obtain higher accuracy. Gansemer et al. (2009) reported that the performance of the isolines technique with multi-floor mean error up to 5.4 m is less accurate compared to nearest neighbour algorithm. Liu and Yang (2011) stated that the weighted screening method offers slight improvement compared to nearest neighbour technique.

2.7.2 Similarity Function

The similarity function technique calculates the signal distance between online and offline samples. The signal distance could be computed according to standard distance measure such as Euclidean and Manhattan, or to a modified distance measure. The location is estimated based on the nearest signal distance or averaged of multiple nearest signal distance. Gansemer, Grossmann, and Hakobyan (2010) calculate similarity function of adapter Euclidean distance based on single period of measured offline signals. However, the method suffers from high computational complexity to be applied for multi-floor localisation due to increasing number of computation elements. Liu and Yang (2011) implements similarity function to estimate the floor by grouping the signal distance according to the floor level and it was reported that high accuracy could be obtained. Marques, Meneses, and Moreira (2012) proposed calculation of signal distance of each fingerprint location on every floor and adds two filtering technique to remove non-related data to reduce computational time. Marques, Meneses, and Moreira (2012) also reported high accuracy of floor estimation of 99.5 % in two three-floor buildings and the mean error of horizontal location is 3.4 m. However, the result published is questionable due to small sample dataset.

2.7.3 Machine Learning Classification

In machine learning classification technique, the location is estimated based on determining the test sample signal in correct class. The class is the location of a pairwise signal of related AP combination and trained according to certain classification technique. Figure 2.3 shows example of pairwise RSS signal of three different classes (locations) available for AP1 and AP2 to be trained according to a classification technique. The most popular technique is Support Vector Machine (SVM). In multi-floor localisation, classification using SVM was demonstrated in Cheng et al. (2014). Another classification technique by Linear Discriminant Analysis has also been implemented for floor localisation as described in Sun et al. (2015). The machine learning classification technique could establish good localisation accuracy if the class could be uniquely characterised by the classification technique. However, the computational complexity of the algorithm increases with increasing amount of APs. Despite the two found techniques, the localisation approach using machine learning has not been extensively studied particularly for multi-floor localisation. There are other machine learning techniques that have not been implemented for localisation as found in Murphy (2012), Marsland (2014), Kung (2014). This thesis proposes logistic regression classification approach for multi-floor localisation. The technique is implemented as AP selection algorithm and floor localisation algorithm.

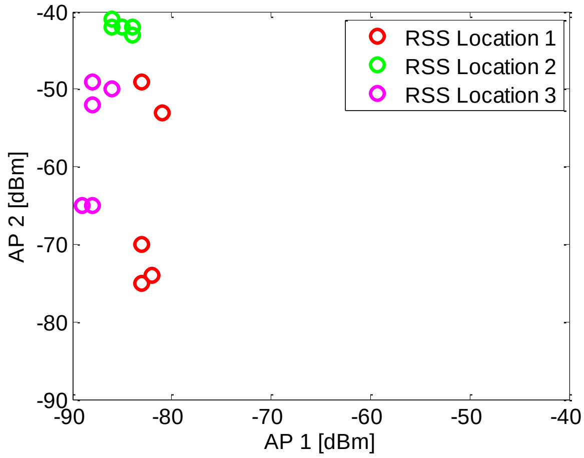


Figure 2.3. Pairwise RSS of AP1 and AP2 in three different classes (locations) for classification based on machine learning (Bolliger 2011)

2.7.4 Artificial Neural Network

Artificial Neural Network (ANN) technique processes the measured the fingerprint RSS signal and the associated location as the input and output in the training phase to develop a network. The network developed by ANN finds the relationship or weighting function to determine associated fingerprint RSS signal with the correct location. The way to find the relationship is by solving the network using directed graph structure. The graph structure contains vertices (neurons) and edges (synapses). The vertices are connected with the edges and categorised according to layers which are inputs, hidden layer, and outputs. An illustration of ANN scheme for multi-floor location is given in Figure 2.4. During online phase, the test sample signal (online signal) is used as the input and the output (estimated location) is determined based on calculation of established weighting function. Campos, Lovisolo, and de Campos (2014) utilise ANN to determine floor location. However the performance of the algorithm is comparable to similarity function algorithm as shown in Table 2.2.

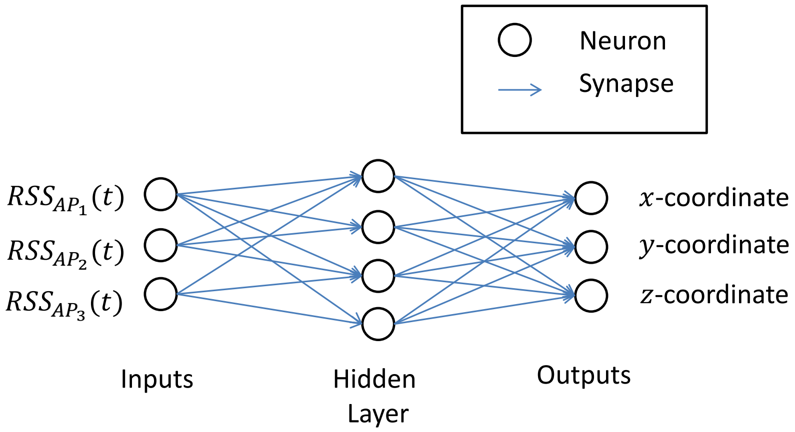


Figure 2.4. ANN scheme for multi-floor localisation where the inputs consists of RSS of multiple APs and the outputs is the x, y, and z location coordinate

2.7.5 Path Loss Model

To develop a path loss model, a particular AP signal is extracted from fingerprint elements in the database. The basic path loss model that is commonly used to evaluate signal in a wireless environment is given as (Rappaport 1996):

$$RSS_{AP_l}(d_{AP_l})[\text{dBm}] = RSS_{AP_l}(d_0)[\text{dBm}] - 10\eta_{AP_l}\log\left(\frac{d_{AP_l}}{d_0}\right) + \chi_\sigma \qquad (2.1)$$

where $RSS_{AP_l}(d_{AP_l})$ is the measured RSS of a particular AP_l, $RSS_{AP_l}(d_0)$ is the RSS at a reference distance from AP_l where $d_0 = 1$, η_{AP_l} is the path loss exponent related to AP_l measured signal, d_{AP_l} is the distance between fingerprint location of measured AP_l, RSS signal and location of AP_l, and χ_σ is the random variable which models the shadowing of measured RSS signal. The location of AP_l is first estimated based on measured RSS of the AP, RSS_{AP_l}, at multiple fingerprint locations using a weighting function. Knowing the location of AP, the distance of measured RSS fingerprint location to the AP, d_{AP_l}, could be calculated. The n_{AP_l}, and $RSS_{AP_l}(d_0)$ are calculated using optimisation algorithm. The complete path loss model is used to estimate the location based on RSS of test sample during online phase based on estimating the distance of RSS from the respected AP. The path loss model technique was implemented by Lohan et al. (2015) to localise the device in multi-floor environment. However, the accuracy of the localisation using path loss model is poor due to high attenuation of the signal which contributes to high uncertainties of the predicted location.

2.8 Summary of WLAN Multi-floor Localisation

This chapter reviews the available multi-floor localisation technique and its limitation. There are two types of multi-floor localisation technique which are single-stage and two-stage. Single-stage determines the floor and horizontal location simultaneously. Two-stage estimates the floor level and then horizontal location. The latter technique is more practical because of low computational complexity. Additionally, floor estimation and horizontal localisation could be independently optimised to provide efficient multi-floor localisation system. WLAN fingerprint technique is more reliable compared to AP based technique. This is due to better performance could be achieved especially for horizontal localisation accuracy. Datasets of the multi-floor localisation in some works is not reliable due to low number of test samples and low height of the tested building environment. Usually, this kind of datasets provides bias estimates of the true localisation performance. Database optimisation method is important in multi-floor localisation problem. This is especially to solve problems in high floor buildings. However the currently available fingerprint database optimisation technique is limited. There are other kinds of optimisation techniques that have not been investigated for multi-floor localisation which has shown good performance in single-floor localisation. In current research, all works is focused on improving localisation accuracy while neglected important aspects in terms of computational complexity of the algorithm and processing time of the algorithm to perform localisation. Although accuracy is important, localisation requires real time result in order to be practical for real life implementation and therefore the algorithm should be able to produce immediate result. Critical comparison and analysis between developed algorithms does not exist except in Sun et al. (2015) which compares the floor localisation accuracy. Some comparisons of the algorithms are made indirect and only by summarising the work done by other researchers such as in Campos, Lovisolo, and de Campos (2014). However, comparison of full WLAN multi-floor localisation algorithm has not been found in any publications. This is contrary with single-floor case where the developed localisation algorithm are usually compared and analysed against each other. This indicates that the validity of the developed algorithm is questionable. Most of the developed floor and horizontal localisation algorithms is based on similarity measures and less attention is given to other types of localisation algorithm. This is different case to single-floor where plethora types of algorithms have been investigated. The review has shown that current multi-floor systems still faces problems particularly in terms of its localisation performance. The localisation performance is mainly dependent on the algorithms developed within the system. Therefore, it is necessary to propose improved algorithms for the multi-floor localisation problem. Three main performance parameters investigated in this thesis are the accuracy, precision, and computational complexity of the algorithms which are explained in details in Chapter 4. The next chapter describes about the underlying theory of the proposed algorithms and the details of proposed algorithms.

CHAPTER 3

KERNEL AND MULTI-CLASS CLASSIFIERS: THEORY AND ALGORITHMS

3.1 Introduction

As discussed in previous chapter, the solution of multi-floor localisation has not been properly addressed. The available algorithms are limited where they mainly focussed on similar kind of algorithms. Also, the optimisation of the radio map fingerprint database has not been fully investigated to improve time efficiency of the system. Therefore a new multi-floor localisation technique is required. Before an improved multi-floor localisation system could be developed, understanding on the basics of the Wireless Local Area Network (WLAN) localisation and the established localisation algorithms is needed. This chapter discusses about the topics in Section 3.2 and 3.3. Once the topics are understood, overview of the main processing blocks of the proposed multi-floor localisation system is presented as in Section 3.4. The block points out the sub-blocks to be improved for development of novel algorithms within the system. The investigated sub-blocks are the database optimisation, floor and horizontal localisation algorithms. The discussion on the theory used for the development of the improved multi-floor localisation system is presented in Section 3.5 to 3.6. Section 3.5 introduces theory of kernel density estimates which is used as the basis of the proposed Access Point (AP) selection algorithm and floor localisation algorithm. Additionally, the proposed AP selection and floor algorithms and together with new horizontal localisation algorithm utilise the theory of machine learning classification based on multi-class k-Nearest Neighbour (kNN) and logistic regression of which details is given in Section 3.6. Then, the implementation of the algorithms using the theory is described in details. Section 3.7 presents the database optimisation using AP selection of Max Kernel and Kernel Logistic Pairwise Discriminant algorithm. Section 3.8 explains floor localisation algorithm using Averaged Kernel and Kernel Logistic Floor localisation algorithms. Finally Section 3.9 gives the details on horizontal localisation algorithm using multi-class kNN classifiers.

3.2 Basics of WLAN Fingerprinting Localisation System

The process of development of a WLAN fingerprint indoor localisation system generally consists of two phases (Youssef 2004) which are offline and offline phases. The fingerprint term refers to the location of the receiving antenna or mobile device capturing the WLAN signal. Figure 3.1 summarizes the general architecture of the fingerprint-based indoor localisation system.

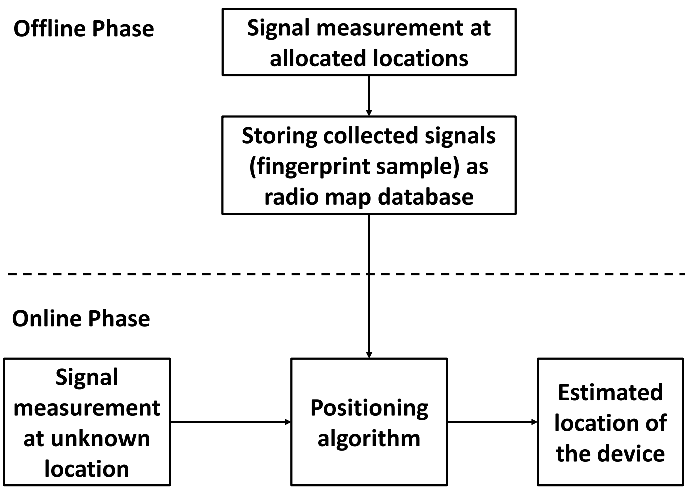


Figure 3.1. Basic architecture of WLAN fingerprint indoor localisation system

In the offline phase, the signal strength from multiple APs is measured by surveyor using a mobile device at each pre-defined fingerprint locations which is illustrated as in Figure 3.2. The pre-defined fingerprint locations within the area of the floor are marked with points in rectangular grids. Distance between grid points is referred as fingerprint spacing which is normally around 1 to 5 meters (Fang and Lin 2010, Yijun, Zuo, and Bang 2012, Redzic, Brennan, and O'Connor 2014). For inaccessible locations, the points are usually omitted. At each fingerprint location, the measurement of signal is taken for a period of time in order to capture ample variation of the propagating signals. The information of the location fingerprint together with the measured signal values associated with the Media Access Control (MAC) ID of the AP is then stored in a database called radio map. The database can either be stored in the mobile device or in a cloud server. Every element or entry in the radio map is usually sorted according to the location fingerprint which is associated with the vector of signal data. Figure 3.3 shows example of one element of the radio map database. Mathematically, each fingerprint element is vector of signal in $L \times T$ matrix (Kushki et al. 2007) where:

$$\mathbf{F}(\mathbf{p}_i) \triangleq \begin{pmatrix} \varphi_i^1(1) & \cdots & \varphi_i^1(T) \\ \vdots & \ddots & \vdots \\ \varphi_i^L(1) & \cdots & \varphi_i^L(T) \end{pmatrix} \tag{3.1}$$

where $\mathbf{F}(\mathbf{p}_i) = [\boldsymbol{\varphi}_i(1); \dots; \boldsymbol{\varphi}_i(T)]$ refers to each fingerprint element or entry in the database, $\boldsymbol{\varphi}_i(t) = [\varphi_i^1(t), \dots, \varphi_i^L(t)]$ means the column of the RSS vector taken at period t that contains reading from L APs when measurement is made at point $\mathbf{p}_i$. Sometimes signals from certain APs are missing which is due to multipath propagation of the signals which especially happens when the APs are located far away from the fingerprint location or there is error of the receiver to capture the signal. If any of the signals is missing, the value is replaced with a low signal value which is usually equivalent to the sensitivity level of the receiver (Arya et al. 2013) e.g. -100 dBm. Additionally, the number of L can vary at different $\mathbf{p}_i$. This is according to the availability of the signal when the measurement is made at respected location. For multi-floor problem, the measurement process will be done in all floors of the building and the location will be added with the floor number or the z location in addition to horizontal x and y location.

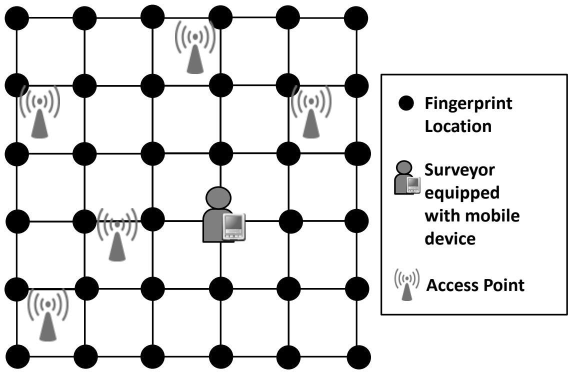


Figure 3.2. Signal measurement process by measuring the signal strength of multiple APs at allocated locations

Floor Level (if multi-floor)	Horizontal Location	AP MAC ID	RSS [dBm]
5	(1,2)	00:04:e2:ff:6a:98	(-35,-34,-35,-35,-37,-36,-38,-33,-36,-37)
		00:1e:31:de:17:22	(-60,-58,-64,-61,-61,-62,-60,-64,-60,-60)
		f8:d1:11:44:9b:8c	(-88,-87,-86,-87,-87,-88,-88,-87,-85,-87)

Figure 3.3. Entry of a radio map fingerprint database of a fifth floor level in a building

The online phase is triggered when there is request of location. This could be either from a cloud server that wants to communicate with the mobile device or a user equipped with any type of mobile device wants to locate him or herself on the floor map. In the second case, one requirement for the user is the user must be able to access the indoor localisation application for the building which should be installed on the mobile device to use the service. In this thesis, the second case is investigated. The online phase process is initiated when the mobile device measures the vector of signal strength from multiple APs at an unknown location inside the building measure. Then, the measured signal vector is sent to the positioning or localisation algorithm for processing and later produces the estimated location. The processing of the algorithm may be executed either on the mobile or the server. There are many types of WLAN fingerprint localisation algorithm that has been proposed. Mainly, the research in indoor localisation is concerned on improving the localisation algorithm. Detail discussion on the localisation algorithm is given in the next section.

3.3 Classical WLAN Fingerprint Localisation Algorithms

The three popular positioning algorithms mentioned in Section 2.6.2.2 are normally used as benchmark to verify developed positioning algorithm is discussed. These algorithms are implemented in the Chapter 5 and 7 to serve as comparison for the proposed multi-floor positioning algorithms. The algorithms are k-neareast neighbour (kNN) which is introduced by Bahl and Padmanabhan (2000), and two probabilistic methods which are univariate kernel that is proposed by Roos et al. (2002) and multivariate kernel which is implemented by Kushki et al. (2007). To mention the popularity of these algorithms, the citations received by these documents are more than 7000, 800, and 200 for Bahl and Padmanabhan (2000), Roos et al. (2002), and Kushki et al. (2007) respectively at the time of writing of this thesis.

3.3.1 $\boldsymbol{k}$-Nearest Neighbour ($\boldsymbol{k}$NN)

The Nearest Neighbour (NN) algorithm is based on finding the minimum signal distance between the RSS vector captured during online phase and the measured signal at every fingerprint location stored in the radio map database. The time varying RSS vectors in the radio map database is first averaged and the averaged signal is compared to the online measured signal. If the all subset of APs in online RSS vector is matched to the all AP subset of offline signal of a fingerprint location in the database, the signal distance is calculated as follows:

$$D_i = \|\boldsymbol{\varphi} - \boldsymbol{\varphi}_{ui}\| \tag{3.2}$$

where $\boldsymbol{\varphi}$ is the current online measured signal vector and $\boldsymbol{\varphi}_{\mu i}$ is averaged signal vector of every i-th fingerprint entry with matched AP list of online sample given $\boldsymbol{\varphi}_{\mu i} = [\boldsymbol{\varphi}_{\mu i}^1, \dots, \boldsymbol{\varphi}_{\mu i}^l \dots, \boldsymbol{\varphi}_{\mu i}^{L_m}]$ where L_m is the number of matching APs of online and offline sample and $\boldsymbol{\varphi}_{\mu i}^l$ is given as:

$$\boldsymbol{\varphi}_{\mu i}^l = \frac{1}{T}\sum_{t=1}^{T} \varphi_i^l(t) \tag{3.3}$$

where $\varphi_i^l(t)$ is the signal of AP_l collected at time t at i-th location. The signal distance D_i is sorted according to ascending order and the fingerprint location obtains the lowest D_i is the estimated location. For kNN, the top k signal distances are chosen, and the average of related k fingerprint locations are the estimated location, $\hat{\mathbf{p}}$ which is calculated as follows:

$$\hat{\mathbf{p}} = \frac{1}{k}\sum_{\kappa=1}^{k} \hat{\mathbf{p}}_\kappa \tag{3.4}$$

where κ is total of k nearest estimated location.

3.3.2 Probabilistic Method

In probabilistic method, the estimated location is determined according to Bayes rule which finds the *Maximum A Posteriori* (MAP) estimate according to Roos et al. (2002):

$$\hat{\mathbf{p}} = \arg\max_{p_i} \mathrm{p}\,(\mathbf{p}_i |\, \boldsymbol{\varphi}) \tag{3.5}$$

and

$$\mathrm{p}(\mathbf{p}_i |\, \boldsymbol{\varphi}) = \frac{\mathrm{p}(\boldsymbol{\varphi}|\mathbf{p}_i)\mathrm{p}(\mathbf{p}_i)}{\mathrm{p}(\boldsymbol{\varphi})} \tag{3.6}$$

where $\mathrm{p}(\mathbf{p}_i)$ is the prior probability of a device to be at location $\mathbf{p}_i$ without knowing the value of observation variable, and $\mathrm{p}(\boldsymbol{\varphi}) = \sum_{\mathbf{p}_{i'}}^{\mathbf{p}_{I'}} \mathrm{p}(\boldsymbol{\varphi}|\mathbf{p}_{i'})\, \mathrm{p}(\mathbf{p}_{i'})$ is the summation of all possible estimated fingerprint location available on the floor space. Usually uniform prior is used to determine the value of $\mathrm{p}(\mathbf{p}_i)$ and $\mathrm{p}(\boldsymbol{\varphi})$ is treated as normalising constant, and therefore the estimate of location $\hat{\mathbf{p}}$ depends only on $\mathrm{p}(\boldsymbol{\varphi}|\mathbf{p}_i)$.

3.3.2.1 Univariate Kernel

Univariate kernel algorithm was introduced in Roos et al. (2002). Instead of using the straight forward signal strength values for computation, univariate kernel transposes the RSS vectors into statistical patterns. The pattern is described in terms of probability estimate for each RSS values. In order to implement kernel density estimate, the signals of different APs are assumed to be uncorrelated and independent. Detail description of kernel density estimates is given in Section 3.5.

Here the univariate kernel for localisation algorithm is introduced. In univariate case, kernel density estimate of an online observation at every fingerprint location i is given as:

$$\mathrm{p}(\boldsymbol{\varphi}|\mathbf{p}_i)=\prod_{l=1}^{L_m}\mathrm{p}(\varphi_l|\mathbf{p}_i) \tag{3.7}$$

where $\boldsymbol{\varphi}$ is the online signal vector and φ_l is signal value of AP l presents during online observation. Kernel is a type of probability mass, and density estimate of an AP l at fingerprint location i is defined as:

$$\mathrm{p}(\varphi_l|\mathbf{p}_i) = \frac{1}{T}\sum_{t=1}^{T} K(\varphi_l; \varphi_{il}(t)) \tag{3.8}$$

where $K(\varphi_l; \varphi_i(t))$ is the kernel function and $\varphi_{il}(t) = \varphi_i^l(t)$ is respective offline RSS value at time t. The commonly used kernel function is the Gaussian kernel which is defined as follows:

$$K(\varphi_l; \varphi_{il}(t)) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(\varphi_l-\varphi_{il}(t))^2}{2{\sigma_i}^2}\right) \tag{3.9}$$

where σ_i is the adjustable parameter that describes the kernel width. In this thesis, σ_i is determined as follows (Scott 2015):

$$\sigma_i = \left(\frac{4}{3}\right)^{\frac{1}{5}}\sigma T^{-\frac{1}{5}} \tag{3.10}$$

where σ is the standard deviation of the T total amount of collected signal at location i.

3.3.2.2 Multivariate Kernel

The version of univariate kernel is extended to multivariate by Kushki et al. (2007). Multivariate is extended from univariate by dimension of the kernel L (where $L > 1$ which is the number of available APs in online sample). The multivariate kernel function is given as:

$$\mathrm{p}(\boldsymbol{\varphi}|\mathbf{p}_i) = \frac{1}{T(\sqrt{2\pi}\sigma_i)^d} \sum_{t=1}^{T} \exp\left(\frac{-\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_i(t)\|^2}{2\sigma_i{}^2}\right) \tag{3.11}$$

where $\boldsymbol{\varphi}_i(t)$ is the offline signal vector at time t. The adjustable parameter σ_i is determined following the guideline in (Scott 2015) as follows:

$$\sigma_i = \left(\frac{4}{2L_m+1}\right)^{\frac{1}{L_m+4}} \sigma T^{-\frac{1}{L_m+4}} \tag{3.12}$$

The difference of calculation of univariate kernel and multivariate kernel is presented is graphical figure as shown in Figure 3.4 for two APs case. The calculation of univariate kernel involves multiplication of kernel density estimate of both AP 1 and AP 2. On the other hand, the calculation of multivariate kernel density estimate already considers dimension of the APs *in situ* as given by Equation 3.11.

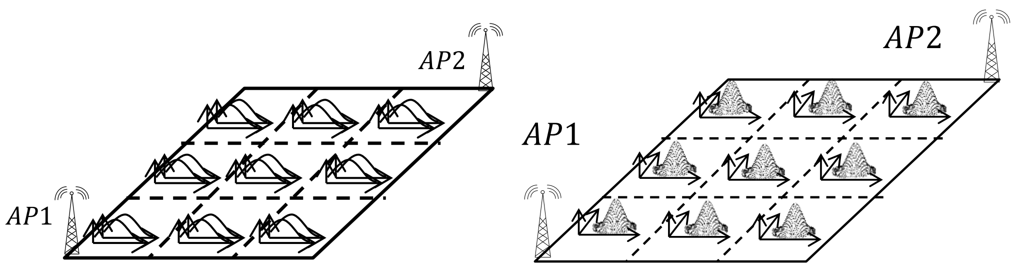


Figure 3.4. Univariate kernel versus multivariate kernel for two APs case

3.4 Main Processing Blocks of Proposed WLAN Multi-floor Localisation System

The multi-floor WLAN fingerprint indoor positioning system involves two phases: offline and online. Figure 3.5 shows the architecture of the proposed positioning system used in this work. The offline phase involves measurement of signal strength signature at each allocated fingerprint location and storing the collected data as a database. Signal strength measurement collects Radio Frequency (RF) signal of available WLAN at each allocated location at the test site. The measurement of the data is done by mobile device that is embedded with WLAN signal receiver. Detailed specification of the measurement tools, test site of the measurement, and the scale of the measurements are discussed in the Chapter 4. Once the signals are measured, they are processed and stored as radio map database. The database contains information of the location where the signals are measured which includes the floor number and horizontal location on the floor, the ID of the signals which is the MAC ID of the AP, and also RSS values. An example of the multi-floor radio map database table could be viewed in Figure 3.6.

Additionally in offline phase, optimisation of the radio map database (as highlighted in the upper green box) is applied to ensure time efficiency of positioning. The radio map database is optimised in database optimisation block by AP selection. The optimised radio map database is served as the input for two-stage system involving floor localisation and horizontal localisation. For horizontal localisation, the radio map is optimised by AP selection technique before processing by localisation algorithm (orange flow lines). For floor localisation, the database is optimised two stages where first by a signal strength clustering technique and second by the AP selection which selects the APs within the cluster (purple flow lines). In this work, two novel AP selection algorithms are proposed which are Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD). For clustering technique, the floor signal strength clustering algorithm is introduced. The detail applications of these algorithms are described in the following sections.

In online phase, the location of the device is estimated using new floor and horizontal localisation algorithms (lower green box). Two floor localisation algorithms are investigated namely Averaged Kernel Floor (AKF) and Kernel Logistic Floor (KLF). Once the floor is estimated, the horizontal localisation is calculated using new multi-class kNN classification algorithm. Details of these algorithms are described in the following sections. The combination of the algorithms processed the unknown online signal by predicting the multi-floor location of the device.

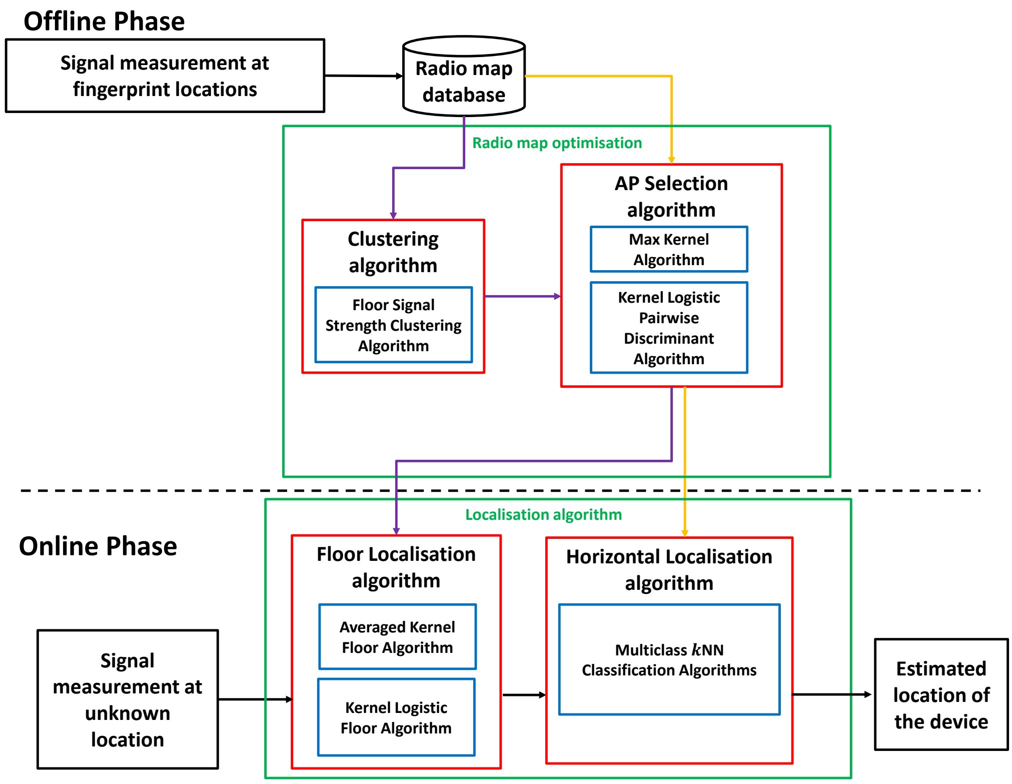


Figure 3.5. Architecture of proposed WLAN fingerprint multi-floor localisation system

Floor	Location	AP ID	RSS (dBm)
1	(1,1)	AP 1	{-80, -81, -78, -80, -79}
		AP 2	{-40, -39, -39, -41, -43}
		AP 3	{-67, -67, -70, -68, -65}
⋮	⋮	⋮	⋮
2	(1,1)	AP 4	{-88, -87, -90, -88, -86}
		AP 5	{-95, -95, -97, -96, -92}
		AP 2	{-62, -58, -58, -63, -66}

Figure 3.6. Example of multi-floor radio map database layout

3.5 Kernel Density Estimate (KDE) as a Tool for AP Selection and Floor Localisation Algorithm

Kernel Density Estimate (KDE) is a non-parametric estimation of an independent and identically distributed random variable distribution. In case of indoor positioning, the random variable is the collected RSS vector at varying time period. KDE is a smooth version of histogram estimate which is defined by a specific kernel function. It is different from parametric density estimation such as Gaussian or log normal where the statistic function is determined based on distribution of the whole dataset. KDE could be illustrated as superposition of multiple kernel functions at each related data entries which effect could be seen as "bumps" to these functions. Some popular kernel functions are Gaussian, Uniform, and Epichenkov (Martinez and Martinez 2007). Among the kernel functions, Gaussian kernel is used in this thesis.

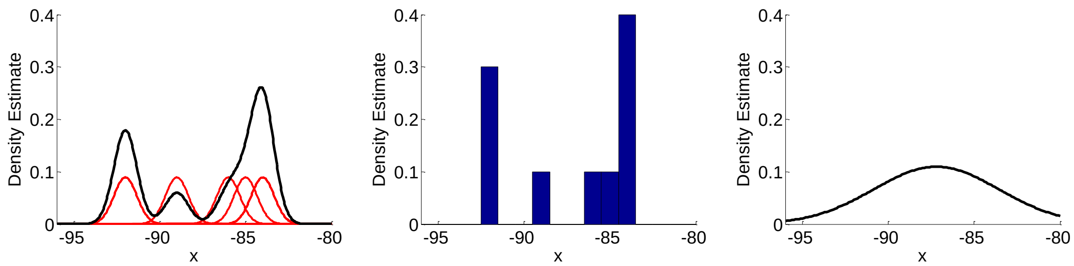
Assume that a signal data distribution is $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_T]$, the KDE is calculated as follows:

$$\hat{f}(x) = \frac{1}{Th} \sum_{t=1}^{T} K\left(\frac{\varphi - \varphi_t}{h}\right) \tag{3.13}$$

where $K(.)$ is the kernel function and h is the adjustable parameter dependent on standard deviation of the signal distribution as mentioned in Equation 3.10. The guideline to select h is given in Equation 3.9. The Gaussian kernel function is defined as:

$$K(\mathcal{H}) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\mathcal{H}^2}{2}\right) \tag{3.14}$$

where $\mathcal{H} = \frac{\varphi - \varphi_t}{h}$. To give an example of KDE, assume that a distribution consists of 10 point of signal data in dBm given $\varphi_1 = -84$, $\varphi_2 = -84$, $\varphi_3 = -84$, $\varphi_4 = -84$, $\varphi_5 = -85$, $\varphi_6 = -86$, $\varphi_7 = -89$, $\varphi_8 = -92$, $\varphi_9 = -92$, and $\varphi_{10} = -92$. The KDE is illustrated in Figure 3.7(a) and comparison with histogram estimate (Figure 3.7(b)) and Gaussian density estimate (Figure 3.7(c)) is given. From the figure, it can be seen that at each point where exist a data point, a Gaussian kernel (red line) is plotted with the mean value of the kernel is at the centre of the data point. If there are more points at the same value, the Gaussian kernel is overlapped. The overall KDE (black line) is based on summation of these kernels and the number of overlapping Gaussian kernels will determine the bumps height. It is also noted that the shape of the KDE is similar to histogram estimate with addition of smoothing effect to the graph. Additionally, probability estimate in KDE is distributed across whole data distribution compared to histogram estimate where zero probability is obtained when a data is missing within the distribution. In comparison to Gaussian density estimate, the plot of Gaussian is centred on the average value of the distribution thus provide optimistic probability estimate of the data.



(a) (b) (c)

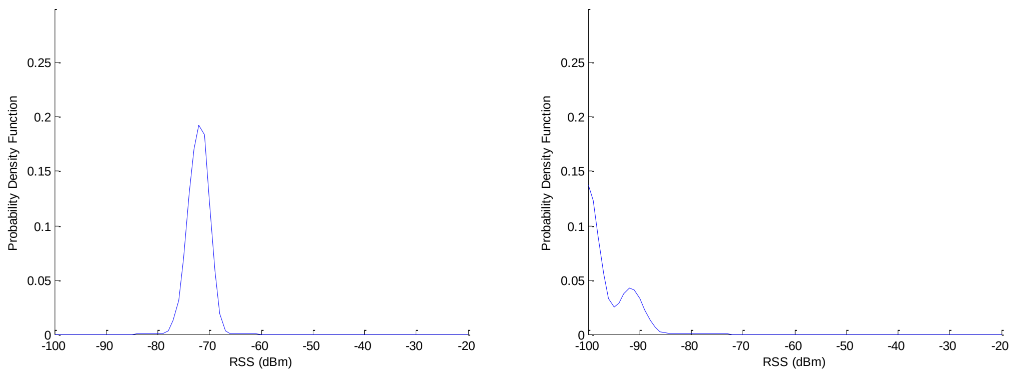
Figure 3.7. Comparison of (a) kernel density estimate, (b) histogram and (c) Gaussian density estimate

For AP selection purpose, the APs are grouped according to the KDE distribution before development of the AP selection algorithm. The process to group the APs are explained in Section 3.5.1 and the grouped APs used to develop AP selection algorithm described in the Section 3.7.2. On the other hand, the floor localisation algorithm is developed based on utilising Equation 3.13 to characterise different floor locations. Detailed development of the floor localisation algorithm is explained in Section 3.8.2.

3.5.1 Characterising AP Signals Based on KDE of Signal Distribution for AP Selection

It could be determined that distribution WLAN signals of an AP at each fingerprint location could be characterised either as strong or weak signal. The strong signal is defined as the signal that is mostly available of which most of the time could be captured by the WLAN receiver of the device during the scanning period of the signal. This signal usually comes from the AP that locates near the fingerprint location. On the other hand, the weak signal could be described as the signal that during period of scanning at most of the time could not be detected by the receiver. This could be due to the signal level of the AP is too low compared to the sensitivity level of the AP. The signal value is low because the AP locations of these signals are located far from the receiver location. Additionally, another problem could be related to hardware error in the receiver at any period during collection of the signal. The strong and weak AP could be distinguished by the distribution of the RSS each fingerprint location and the idea is to exploit the knowledge of probability distribution of each AP signals at each fingerprint location for selecting appropriate APs.

Normally during preparation of offline fingerprint database, the missing signal of the APs at any period of scanning time is replaced with weak signal value of the AP in the database. The weak signal value is given as -100 dBm (as mentioned in Section 3.2). To determine if the general signal distribution of an AP is either strong or weak, KDE of every signals using Gaussian kernel function at every fingerprint location are calculated. Probability density function of kernel forms multiple peaks if the distribution contains multiple high probability of receiving multiple signals as could be seen in Figure 3.8(b). Additionally, the distribution indicates true distribution of the signals which is almost similar to histogram. This is different from Gaussian density function where only single peak is observed for the distribution. Therefore signals group could be categorised based on the KDE. From kernel density estimates of each AP signal, the strong AP is categorised based on the maximum peak of the KDE of the AP signal and the maximum peak of the signal is not equal to the missing signal value. The signal at the maximum peak is the true signal value. Oppositely, the weak AP is determined based on the maximum peak of the AP signal is possessed by the missing signal value and the peak does not present the true value of the signal. The true value of the signal is presented by the second highest peak of the density function. Example of the both of these APs according to its KDE is shown as in Figure 3.11. In Figure 3.8(a), the strong AP has the true value of the signal at -72 dBm while in Figure 3.8(b), the weak AP signal is owned by second highest peak of the KDE at -92 dBm.



(a) (b)

Figure 3.8. Kernel density estimate of (a) Strong AP and (b) Weak AP where the missing signal is determined as -100 dBm

Once strong and weak APs are categorised, the AP list at each fingerprint data is prepared for selection. APs are selected by giving priority to the strong APs. For example, if number of AP to be selected is three and a fingerprint data consists of five APs, and there are two strong APs and four weak APs. Then, the first two selected APs are the two strong APs and the remaining AP is chosen from those four weak APs. However if there exists four strong APs, the three APs is only selected from those four strong APs and neglecting the weak APs.

In the next stage, the APs within each group, strong or weak, is chosen according to proposed AP selection algorithm. Two new AP selection algorithms are introduced: maximum kernel density or simply Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD). In Max Kernel algorithm (discussed in Section 3.7.2), the classification of strong and weak APs only involves classification of AP signals within AP list of fingerprint data per location. The selection of APs is dependent on the MAP estimate of each AP list within each fingerprint elements. The technique is simple and computationally efficient. On the other hand, KLPD algorithm (discussed in Section 3.7.3) is based on discrimination of each AP signals across fingerprint locations as discussed above. It analyses separability of AP signals at different fingerprint locations and classifies them in different classes. The way to classify the locations is by transforming the fingerprint print elements to a feature space of which the data is the pairwise signal of related APs found at any fingerprint locations. Example of the transformation of the fingerprint elements in feature space is illustrated in Figure 3.9. The discriminability of each pairwise AP is quantified using the negative log-likelihood of the kernel logistic function of Kernel Logistic Regression (KLR). Theory of the logistic regression is discussed in Section 3.6.3 and details of implementation of KLR for KLPD AP selection algorithm is discussed in Section 3.7.3.

3.6 Machine Learning Classification As a Tool For AP Selection, Floor and Horizontal Localisation

Machine learning classification is the problem of identifying a new set of data belongs to which category of prior data. The algorithm that implements classification task is the classifier. The task of a classifier is to map a new data to the prior data according to their category (Murphy 2012). The working principle of a classifier is first extracting the data into feature space and then determines optimal decision boundary or separating hyperplane to maximize the margin between classes of data and at the same time to reduce the misclassification of the data. Generally, the classification is done for two dimensional datasets. The top figure in Figure 3.9

shows the projection of fingerprint data in feature space for classification. Some example of popular classifiers in machine learning is support vector machine (SVM), decision trees, and k-Nearest Neighbour (kNN).

In classification technique, a classifier needs to map an input data r to output y which is a category of prior data given $y \in \{1,2,\ldots,Y\}$ where Y is the total number of categories or classes. For $Y = 2$, the problem is solved by binary classification. For more than two classes, the problem becomes multiclass classification. In indoor positioning, the problem is the multiclass classification because there are more than two fingerprint locations (classes). There are three ways to solve multiclass problems: one-versus-all (OVA) or one-versus-one (OVO) scheme which is by binary classification, or by multiclass classification. Figure 3.9 compares graphically the entire scheme for three class problem. Each data of blue, red or black markers could be imagined as the pairwise data.

In OVA scheme, every available class is compared to the rest of other classes and the resulting class is the class that obtains the highest vote. For OVO, each available class is compared to another class by one to one. On the other hand, the multiclass classification classifies right away the online sample to appropriate class according to the classifier. In terms of computational time, the multiclass classification is the lightest and followed by OVA scheme and OVO scheme is the slowest. However, the process to determine decision boundary for multiclass classification is hardest compared to both OVA and OVO scheme. Additionally, only some classifiers could work with multiclass classification.

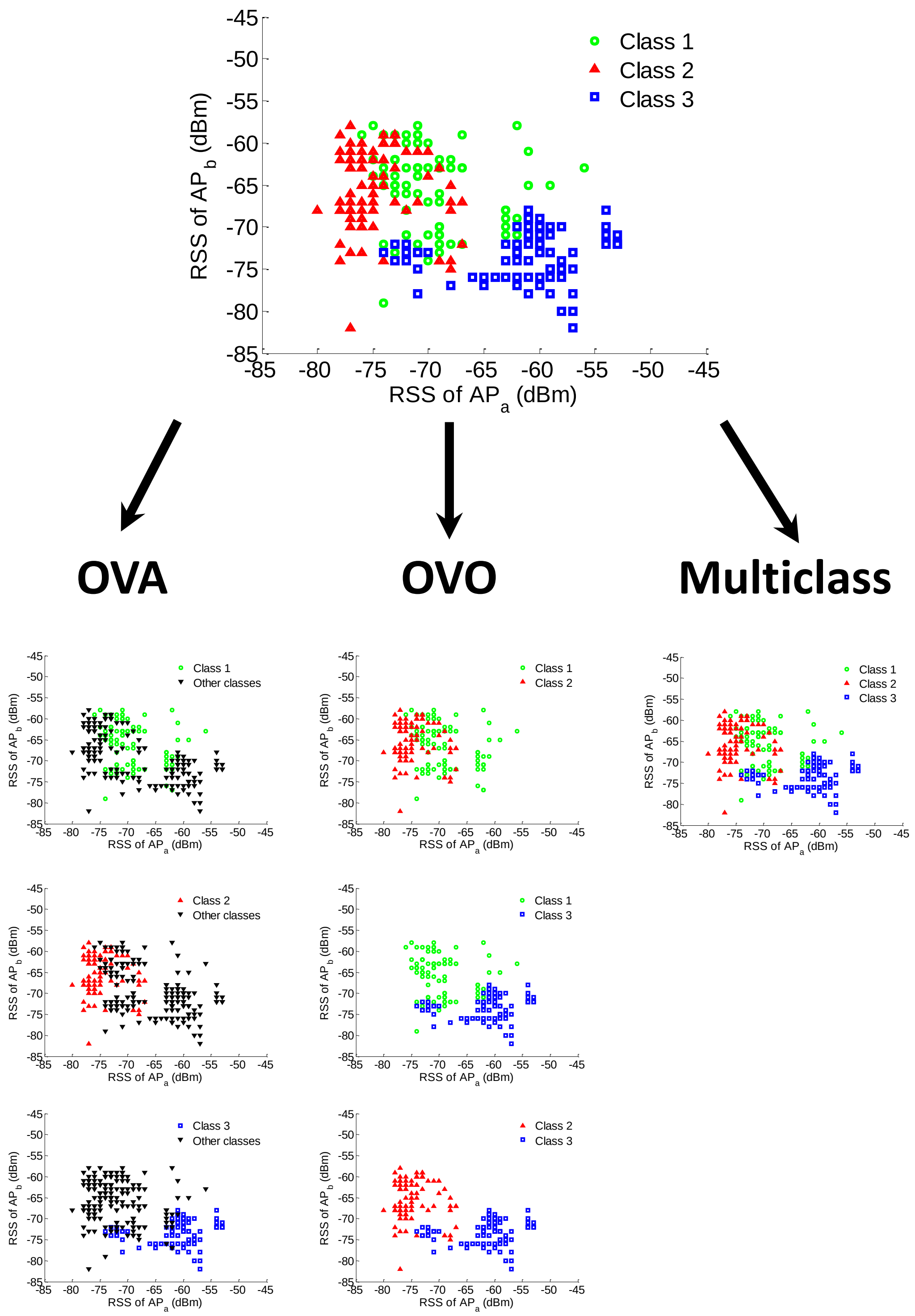


Figure 3.9. Top figure shows projection of fingerprint dataset into feature space for classification. The AP pair data is found at three different locations (classes) and OVA, OVO, and multiclass scheme are compared

3.6.1 Kernel Trick for Classifiers

Kernel trick is different from the kernel density estimate which is discussed in Section 3.4. Kernel trick is used in machine learning classification to map non-linear data to a high-dimensional space so that the data is linearly separable. The mapping process is specified according to kernel function. If the mapping function is given as Ψ, the kernel function is denoted as: $\boldsymbol{K}(\boldsymbol{r}_p, \boldsymbol{r}_q) = \Psi(\boldsymbol{r}_p)'\Psi(\boldsymbol{r}_q)$ where $\boldsymbol{r}$ is a data point vector in the feature space. In this thesis, a specific choice of kernel function is chosen to serve as the kernel function which is the radial basis function (RBF) kernel that is defined as:

$$\boldsymbol{K}(\boldsymbol{r}_p, \boldsymbol{r}_q) = e^{-\gamma\|r_p - r_q\|} \tag{3.15}$$

where $\gamma = \frac{1}{2\sigma^2}$ is an adjustable similarity parameter of $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ with σ standard deviation.

3.6.2 Logistic Regression and k-Nearest Neighbour Classifiers for Multi-Floor Localisation

In indoor positioning, the new data could be defined as the online sample or measurement and the input data is the fingerprint dataset. The fingerprint location is the category of the input data. Therefore in indoor positioning, the fingerprint dataset should be represented as vectors of data according to each pairwise AP set. For L APs presents for a fingerprint dataset, the number of classification to be considered is $(L(L-1)/2)$. To solve the localization problem using classification, the fingerprint entries are reconstructed as AP pairwise dataset which its element is defined as follows:

$$\mathbf{A}_v(\mathbf{p}_i) \triangleq \left(\begin{bmatrix} r_i^{v_1}(1) \\ r_i^{v_2}(1) \end{bmatrix}; \quad \cdots \quad ; \begin{bmatrix} r_i^{v_1}(\tau) \\ r_i^{v_2}(\tau) \end{bmatrix}\right) \tag{3.16}$$

where ($v = 1, \dots, \frac{V(V-1)}{2}$) and V is the total number of unique APs in the fingerprint database) is the pairwise AP combination index at location $\mathbf{p}_i$ and τ is the number of matching pairwise signal over duration of the fingerprint collection time. Here, for non-matching pairwise signal, the signals are neglected and thus τ will vary according to available signal at every $\mathbf{p}_i$. The header of the new AP pairwise based-database could be defined as collection of similar group of pairwise AP at multiple locations given $\mathbf{A}_v = [\mathbf{A}_v(\mathbf{p}_1); \dots; \mathbf{A}_v(\mathbf{p}_I)]$ where I is the total number of fingerprint locations.

As mentioned above, indoor positioning problem generally involves more than two fingerprint locations and thus the problem of classification is multi-class. The choice of a scheme as discussed above to solve multi-class classification depends on the choice of classifier. However, generally computational complexity of the schemes is related to the number of classes and the number of pair AP sets.

This thesis investigates two types of classifiers which are logistic regression and kNN. Please note that kNN classifier is different from kNN algorithm which is discussed in Section 3.3.1 where the prior kNN is a regression type algorithm. Here, the kNN of machine learning classification is distinguished from previous kNN by the term “classifier”. Details of the theory of the logistic regression and kNN classifier are explained in Section 3.6.3 and 3.6.4 respectively.

3.6.3 Logistic Regression Classifier

The working principle of logistic regression is similar to most other machine learning classification which seeks for boundary function for each class to define separability of data between classes. Once the signal data is transformed into feature space as in Figure 3.9, the separation of data between two classes could be defined by construction of sigmoid function where:

$$\mathrm{p}(y|\boldsymbol{r},\boldsymbol{w}) = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{r}}} \tag{3.17}$$

where $\boldsymbol{r}$ is the vector a pairwise AP signal data, $\boldsymbol{w}$ is the model parameter vector of the logistic function, T is the transpose matrix, and probability function $\mathrm{p}(.)$ which lies within $[0,1]$ infers the probability of a signal vector presence within the class of the signal data with y class label. For kernel version of the logistic regression, the vector of a pairwise signal data $\boldsymbol{x}$ is mapped into higher dimension and it is translated according to kernel function where: $\boldsymbol{K} = \left[\boldsymbol{K}\left(\boldsymbol{r}_p,\boldsymbol{r}_q\right)\right]_{\Phi\times\Phi} = \left[\Psi\left(\boldsymbol{r}_p\right)'\Psi\left(\boldsymbol{r}_q\right)\right]_{\Phi\times\Phi}$ with Φ is total number of vector signal data existed in both classes. Therefore Equation 3.17 becomes:

$$\mathrm{p}(y|\boldsymbol{K},\boldsymbol{w}) = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{K}}} \tag{3.18}$$

To determine the optimal value of $\boldsymbol{w}$, sigmoid function of kernel logistic regression in Equation 3.18 is solved by minimising the objective function which is represented by negative log-likelihood (NLL) function as follows (Zhu and Hastie 2005):

$$NLL(\boldsymbol{w}) = -\boldsymbol{y}^T(\boldsymbol{K}\boldsymbol{w}) + \boldsymbol{1}^T \ln(1+\exp(\boldsymbol{K}\boldsymbol{w})) + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{K}\boldsymbol{w} \tag{3.19}$$

where $\boldsymbol{y} = y_{1:\Phi} \in \{0, 1\}$ is the vector of class label of $\boldsymbol{r}_{1:\Phi}$ data, and the right most term is the equation is a regularisation term to avoid over fitting of NLL estimation.

To solve the NLL function of kernel logistic regression, an optimisation algorithm is used. A common optimisation approach is based on iterative reweighted least square (IRLS). The details of optimisation algorithm of the IRLS to find the optimised $\boldsymbol{w}$ parameter for NLL function in Equation 3.19 are given in Appendix A1.

The NLL function is used to develop AP selection algorithm as discussed in Section 3.7.3 while the sigmoid function is used for floor determination as described in details in Section 3.8.3.

3.6.4 $\boldsymbol{k}$-Nearest Neighbour Classifier

For every location request, the matching pairwise AP signal in the online sample is matched to the similar pairwise AP in the database. Since at every pairwise AP, $\mathbf{A}_u$, exists numerous signals of multiple classes (fingerprint locations), the classification problem becomes a multiclass problem. For indoor positioning applications, the kNN classifier efficiently works by multiclass classification scheme compared to OVO and OVA. Since the kNN is an unsupervised classifier, the prediction of the location could be done by finding the nearest signal distance between the vector signals of the related pairwise AP. The distance of online sample to each data in each class for every online sample AP combination $\mathbf{A}_\beta$ is defined as follows:

$$D_\beta(\psi_{\Theta u}) = \left\| \boldsymbol{r}_o^\beta - \boldsymbol{r}_i^\beta(\psi_{\Theta i}) \right\| \tag{3.20}$$

where $\psi_{\Theta i}$ is each data sample labelled Θ of class or location i, β is the number of matching pairwise APs, and $\boldsymbol{r}_i^{\beta}(\psi_{\Theta i}) \in \boldsymbol{r}_i^{v}$ is the signal vector of the data sample in the related class u. In this work, the signal vector that contains the weakest signals is neglected and therefore the number of $\psi_{\Theta i}$ in different class may vary. The signal vector $\boldsymbol{r}_i^{\beta}(\psi_{\Theta i})$ could be visualised in Figure 3.19(b) as the red, green, and magenta circles that indicates signal vectors in three classes. The signal distance $D_{\beta}(\psi_{\Theta i})$ is implemented to estimate the horizontal location in this work. The implementation detail is discussed in Section 3.9.1.

3.7 Radio Map Fingerprint Database Optimisation By Max Kernel and Kernel Logistic Discriminant Pairwise AP Selection

In existing AP selection algorithm strategy, the work is mainly improve the AP selection algorithm by introducing new selection scheme by directly process the fingerprint signal data. Here, a new concept is introduced by first pre-determining the AP group based on the signal distribution of the AP at each fingerprint location and then processing the AP within each group to determine the selected APs (as discussed in Section 3.5.1). Additionally, the selected subset of APs varies from each fingerprint to another which is unlike classical approach that uses fixed AP numbers at each fingerprint. The Max Kernel algorithm is developed based on exploiting the characteristic of kernel density estimates theory while the KLPD algorithm exploits the objective function or Negative Likelihood (NLL) function of machine learning classification algorithm - the kernel logistic regression. The first algorithm evaluates the importance of APs based on individual fingerprint data while the second algorithm computes signal discriminability between APs exists in the whole dataset.

Details of the implementation of the proposed AP selection techniques are explained in the following sections.

3.7.1 Variant AP Selection Technique

In classical AP selection technique (Chen et al. 2006, Lin et al. 2014), the number of selected AP ID for each fingerprint element in the database is fixed and the order of the selected APs is also uniform. This sometimes increases the computational time of the localisation algorithm because introduction of additional AP to some fingerprint elements. Figure 3.10 shows the output fingerprint element using classical AP selection. In classical approach, if the number of selected AP is four, then four APs is placed at each fingerprint elements. This thesis presents different AP selection method called as variant AP selection. The variant AP selection considers different APs to be selected at different fingerprint elements which are according to available AP ID at the element. Example of variant AP selection technique is given in Figure 3.10. Comparing the variant and classical AP selection technique, it is clearly indicates that using variant AP selection could further reduce the computational time of the localisation algorithm because lower number of APs at each fingerprint could be processed. The selection of APs by variant AP selection at each fingerprint element is determined according to the selection criteria using kernel density information with theory of kernel logistic regression which is discussed in further subsections.

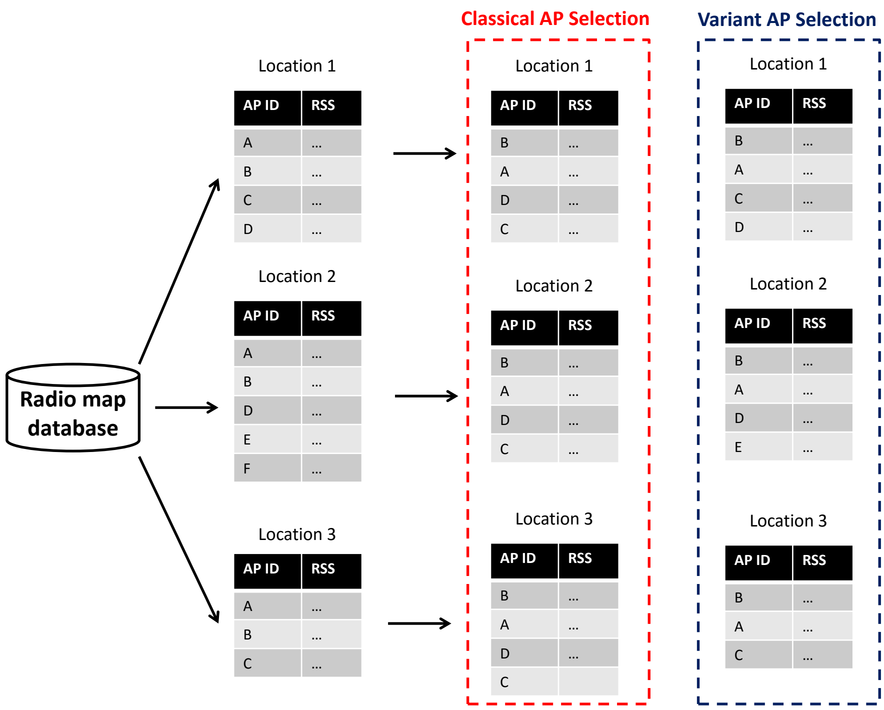


Figure 3.10. Classical versus Variant AP Selection Technique for four selected APs

3.7.2 Max Kernel

As explained in Section 3.5.1, the signal of each AP of strong and weak measured at a fingerprint location is calculated using KDE function. The technique to choose APs according to Max Kernel algorithm is based on sorting the APs in terms of value of the maximum kernel density function at each fingerprint location. The method is relatively simple but has not been proposed in prior researches. The APs are selected according to maximum value of the probability within both strong and weak AP groups with selection are prioritising the strong signal group.

For strong signal APs, the higher peak in density distribution the sharper peak. The AP with high peak value has its signal data distributed in small scale of variance. Otherwise, the AP with low distribution peak value has spread signals values. This type of signals is not unique in presenting the fingerprint location and the AP with this kind of signal distribution should be given lower priority compared to APs with higher distribution peak value. The AP with higher probability distribution value could characterise the location better and should be prioritised in the selection. Example of the difference between these kinds of AP signals could be observed in Figure 3.11. The AP 2 has sharper peak and the probability distribution is higher compared to AP 1 of which peak is lower. The distribution of signals of in AP 1 (between -57 dBm to -83 dBm) is in larger range compared to AP 2 (between -57 dBm to -77 dBm) and AP 2 is the more preferable AP during selection process. Additionally, the received signal of an AP could contain multiple peaks as shown in the figure. This is due to at that peaks, the most of the signals lies at that values. However, the higher peak within those peaks shows that the signal mostly lays at the value of the higher peak compared to lower peak and the density function at the higher distribution peak signal is considered in AP selection process. For example, AP 2 has two peaks -63 dBm and -69 dBm. This means most of the signals collected during the scanning period are characterised by either values. However, the peak at -69 dBm is higher which describes the maximum kernel density function is chosen for determining the density function.

However for weak signal APs, the maximum value of kernel density function reflects the missing signal value as previously shown in Fig 3.8(b). To determine the true value of density function in weak APs, the value of second highest peak of the distribution is chosen e.g. in Figure 3.11 the maximum value of AP 5 should be 0.02 at signal of -77 dBm. To find such signal peak, a simple peak finding algorithm based on difference between neighbouring signals of kernel density function is proposed. Since the maximum peak is owned by the undetected signal, the first step is to filter out this signal by finding the nearest valley of that signal in the density function. The nearest valley is found by calculating difference between two neighbouring kernel:

$$\Delta f_{neigh} = f_u - f_{u+1} \ \ (u = 1, 2, 3, \dots, U - 1) \tag{3.21}$$

where U is the total number of points of signal range to determine the kernel e.g. 81 points for signal range of -20 dBm to -100 dBm. The point where first positive different is achieved is the point of the nearest valley. Once the valley point is found, the value of density function before the valley is filtered out and the next step is to find the maximum peak among the remaining density function.

To give an example, Figure 3.11 shows the kernel density function of six APs exists at one fingerprint location. For selection of three APs, three out of five strong APs which are AP 2, AP 4, AP 1, AP 3 and AP 5 are first separated from the weak AP (AP 6) and according to Max Kernel, AP 2, AP 4, and AP 1 will be the chosen APs. The flow chart of Max Kernel algorithm is given as Figure 3.12.

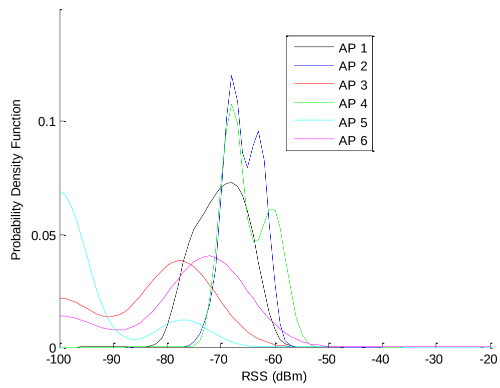


Figure 3.11. Example of kernel density estimate of six APs at one fingerprint location

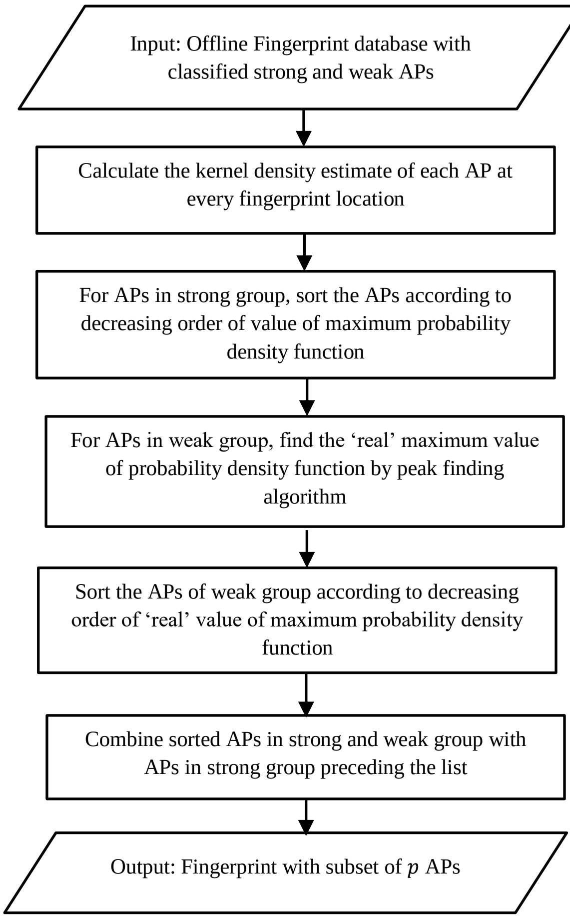


Figure 3.12. Flow chart of Max Kernel AP selection algorithm

3.7.3 Kernel Logistic Pairwise Discriminant (KLPD)

Although Max Kernel is computationally simple and efficient and may work better than other AP selection algorithm within the same AP selection category, it does not consider discrimination of signals between different fingerprint locations. Therefore Max Kernel algorithm may perform less in this situation e.g. environment that have multiple APs that are located within good visibility from fingerprint locations. Thus another AP selection which is based on selection of AP signals which discriminates between locations is investigated. The idea is to adapt the knowledge of machine learning classification approach to define the discrimination of AP signals across fingerprint locations. Here, kernel logistic regression approach is studied to address the problem.

The value of NLL function of KLR in Equation 3.19 describes the discriminability of the datasets between two classes. If the value of NLL is nearly zero, the datasets between classes are well separated and the locations are distinguished better. On the other hand, high value of NLL indicates that the datasets between classes are close and the locations of the dataset may be misinterpreted.

In indoor positioning problem, multiple locations exists which means the problem must be solved using multinomial version of KLR. For simplicity, OVO scheme of multinomial KLR is implemented to find discriminability of every pairwise AP available in the fingerprint dataset. For every pairwise AP at two locations combination, a single NLL value is contributed. Every location combination is noted by $\left(m = 1, \dots, \frac{I(I-1)}{2}\right)$ where m is the index of the combination of the pairwise locations (classes) and I is the total available fingerprint locations. In order to determine the most discriminative pairwise AP, the NLL at each location combination is summed. This could be noted as a score function which is given as:

$$\Upsilon_v = \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} NLL_{jk} \tag{3.22}$$

where j and k is the index of location refers to index combination b e.g. exists three locations ($m = 3$), the location $j = 1$, and $k = 2$ refers to $v = 1$, location $j = 1$, and $k = 3$ refers to $v = 2$, and location $j = 2$, and $k = 3$ refers to $v = 3$. The v notation refers to the index of combination of each pairwise AP given $\left(v = 1, \dots, \frac{V(V-1)}{2}\right)$ with V is the number of APs exists in the database (refer to Section 3.6.2). The most discriminative AP pair is the one that obtains minimum Υ_v and for every fingerprint locations, the APs are sorted to increasing value of Υ_v. Similarly as in Max Kernel algorithm, the sorting of AP according to Υ_v is done within each strong and weak groups of AP. This means only APs within each related group are sorted against each other. The flowchart of the kernel logistic pairwise discriminant algorithm is given in Figure 3.13.

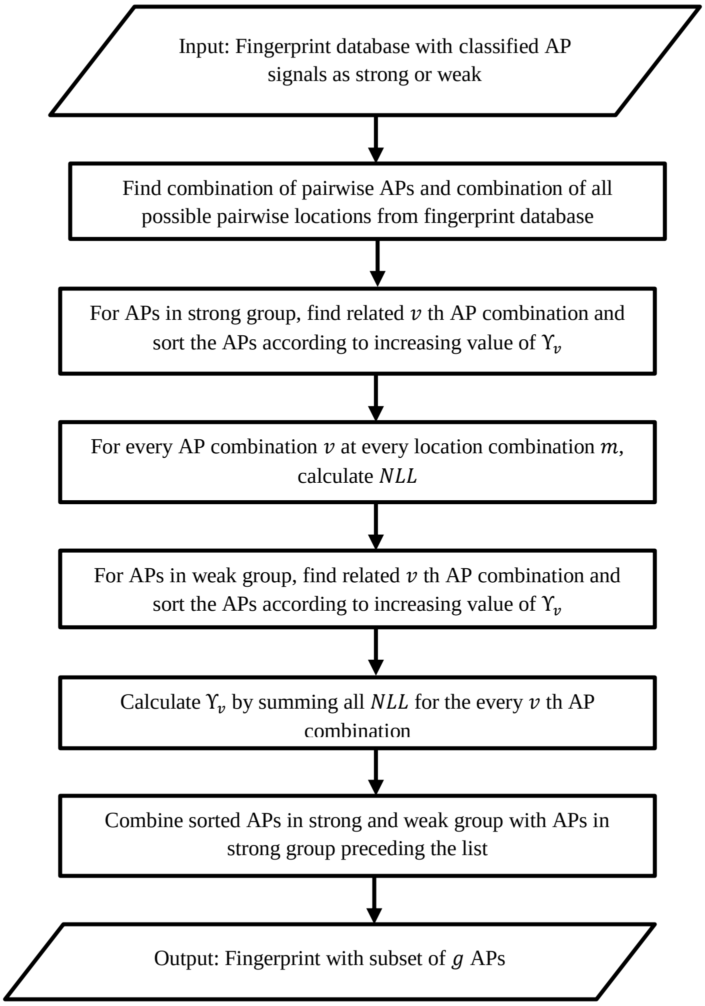


Figure 3.13. Flowchart of Kernel Logistic Pairwise Discriminant AP selection algorithm

3.8 Averaged Kernel and Kernel Logistic Floor Localisation Algorithms

In two stage multi-floor localisation technique, vertical or floor location estimation is done priori before horizontal location is determined. Estimating correct floor location of the device is important to ensure proper horizontal database selection. Additionally, it is important to reduce computation time in estimating the floor since estimation of horizontal location requires more time to give accurate location information by processing multiple fingerprint data entries. Here, two stage strategies are proposed in order to reduce computational time of the floor algorithm and at the same retain the accuracy of estimated floor. Firstly, the fingerprint database is clustered. Then, the clustered fingerprint data is served as the input to the new floor localisation algorithm. The clustering technique is based on simple AP grouping based on the APs signal strength value. Meanwhile, for floor localisation algorithm, two algorithms are proposed which are Averaged Kernel Floor algorithm and Kernel Logistic Floor algorithm. The clustering technique and developed floor localisation algorithms are discussed in following sections.

3.8.1 Floor Signal Strength Clustering

To cluster the fingerprint dataset by signal strength, first the APs in each fingerprint data is sorted according to maximum average RSS. Then the clustering processing is started with selecting the top AP from each of the fingerprint as the cluster head. The fingerprints with similar cluster head are merged together. Grouping of the fingerprint according to the dataset is done according to number of available fingerprint locations of each floor. Mathematically, a cluster is defined as follows:

$$\boldsymbol{c}_f \triangleq \{\boldsymbol{M}_{l_{z1}}, \dots, \boldsymbol{M}_{l_{zb}}, \dots, \boldsymbol{M}_{l_{zB}}\} \tag{3.23}$$

where $\boldsymbol{M}_{FP_{bz}}$ refers to a fingerprint vector with cluster head of AP l on floor z of b-th fingerprint sample of similar cluster head. For example, if there are 20 fingerprint data available on the floor, the clustering algorithm will group only these 20 fingerprint data and the process is repeated for another floor. Figure 3.14 shows example of signal strength clustering done on one floor level for five fingerprint entries that are grouped as three clusters. Additionally, the cluster that locates on the same floor is similarly labelled according to the floor number. The flow chart of the clustering technique is given as in Figure 3.15.

In the next two sections, two floor localisation algorithms are proposed i.e. Averaged Kernel Floor algorithm and Kernel Logistic Floor algorithm which use the cluster as the input to the algorithms.

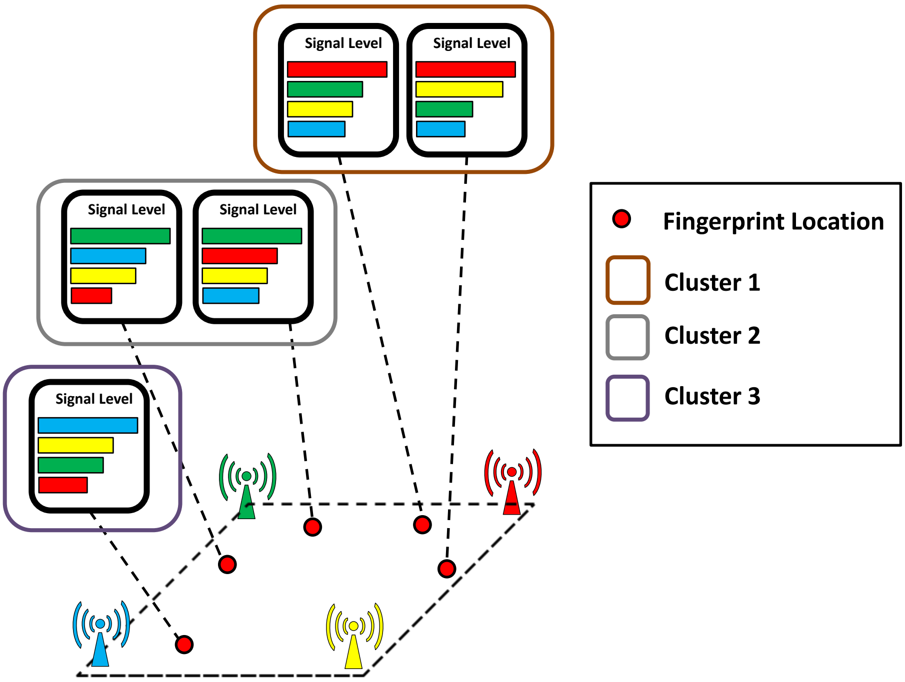


Figure 3.14. Floor signal strength clustering of five fingerprint entries which groups as three clusters

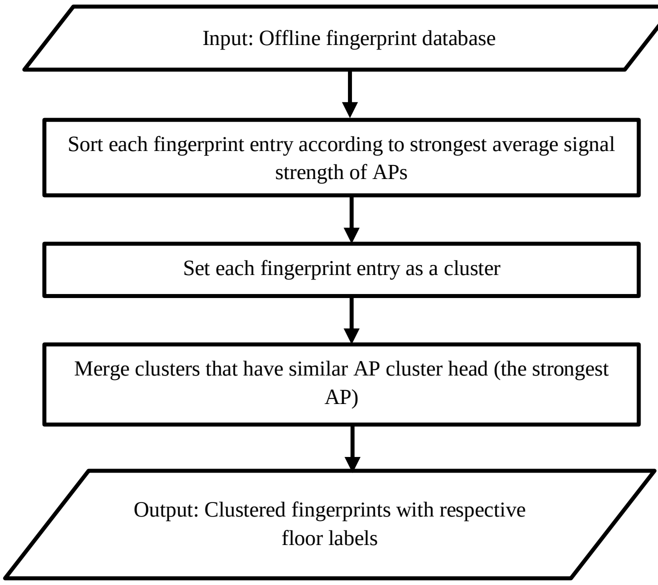


Figure 3.15. Flow chart of the proposed floor signal strength clustering

3.8.2 Averaged Kernel Floor Algorithm

Averaged Kernel Floor (AKF) algorithm is based on the theory of kernel density estimates as discussed in Section 3.5. Extending Equation 3.13, the average kernel density estimate of AP l in a cluster $\boldsymbol{c}_f$ could be defined as:

$$\mathrm{p}\left(\varphi_l|\boldsymbol{c}_f\right) = \frac{1}{Th_cQ}\sum_{q=1}^{Q}\sum_{t=1}^{T} K\left(\frac{\varphi_l - \varphi_{lq}(t)}{h_c}\right) \qquad (3.24)$$

where Q is the total number of q locations of the AP l signal grouped in the cluster $\boldsymbol{c}_f$, φ_l is the online signal of AP l, $\varphi_{ql}(t)$ is the offline signal of AP l at period t of q labelled location in cluster $\boldsymbol{c}_f$, and $K(.)$ is the kernel function which is similarly defined as in Equation 3.14. Determination of floor is made according to MAP of the density estimate of the cluster $\boldsymbol{c}$ where:

$$\hat{\boldsymbol{c}}_{MAP} = \arg\max_{c_j} \mathrm{p}\left(\boldsymbol{c}_f \,|\, \boldsymbol{\varphi}\right) = \arg\max_{F_j} \mathrm{p}\left(\boldsymbol{c}_f \,|\, \boldsymbol{\varphi}\right) \qquad (3.25)$$

and

$$\hat{p}_z = z(\hat{\boldsymbol{c}}_{MAP}) \tag{3.26}$$

where $\boldsymbol{\varphi}$ is the online signal vector and according to Bayes rule as discussed in Section 3.3.2, the $p(\boldsymbol{c}_f|\,\boldsymbol{\varphi})$ could be assumed similar to $p(\boldsymbol{\varphi}|\boldsymbol{c}_f)$ which is calculated as follows:

$$p(\boldsymbol{\varphi}|\boldsymbol{c}_f) = \prod_{l=1}^{L} p\left(\varphi_l|\boldsymbol{c}_f\right) \tag{3.27}$$

The flow chart of the AKF algorithm is illustrated in Figure 3.16.

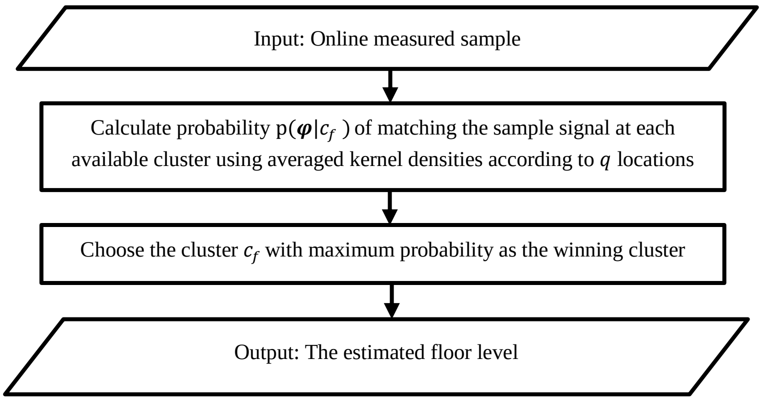


Figure 3.16. Flow chart of Averaged Kernel Floor algorithm

3.8.3 Kernel Logistic Floor Algorithm

Another proposed floor localisation is the Kernel Logistic Floor (KLF) algorithm which implements kernel logistic regression machine learning classification to solve floor localisation problem. To implement kernel logistic regression in floor localisation, the online signal sample vector, $\boldsymbol{\varphi}$, is transformed into multiple pairwise signal vector $\boldsymbol{r}_o^v$ where $v = 1, 2, \ldots, \frac{L(L-1)}{2}$ (refer to Section 3.6.2). Then every signal vector $\boldsymbol{r}_o^v$ is evaluated using sigmoid function to determine the class of the input data. Given the clusters signal transformed into pairwise signal vector as the input data, the online signal vector could be classify to which cluster according to modified sigmoid function of Equation 3.18 given as:

$$\mathrm{p}(y|\boldsymbol{K}_v, \boldsymbol{w}) = \frac{1}{1+e^{-w^T K_v}} \tag{3.28}$$

where $\boldsymbol{K}_v$ is the radial basis function (RBF) kernel function of AP combination index v given as:

$$\boldsymbol{K}_v = \boldsymbol{K}\left(\boldsymbol{r}_o^v, \boldsymbol{r}_\varsigma^v\right) = e^{-\gamma_{LR}\|r_o^v - r_\varsigma^v\|} \tag{3.29}$$

where $\boldsymbol{r}_o^v$ is the online vector signal data of related AP combination index b, $\boldsymbol{r}_\varsigma^v$ is the clusters vector signal data of AP combination index v of ς found cluster, and $\gamma_{LR} = \frac{1}{2\sigma_{LR}{}^2}$ is adjustable similarity measure of $\boldsymbol{r}_o^v$ and $\boldsymbol{r}_\varsigma^v$. The number of ς clusters found of every AP combination index v varies from one AP combination to another. This is related to sensitivity of the receiver which can only capture the signal level as low as -100 dBm. The value of σ_{LR} is fixed as 10. The vector $\boldsymbol{w}$ is calculated in the training stage based on finding the minimum possible NLL using optimisation technique by Equation A1.1 to A1.6 in Appendix A1. The decision that the one pairwise online signal vector $\boldsymbol{r}_o^v$ belongs to cluster $\boldsymbol{c}_f$ ($\boldsymbol{c}_f \in \varsigma$) is made if $\mathrm{p}(y = \boldsymbol{c}_f|\boldsymbol{K}_v, \boldsymbol{w}) \geq 0.5$.

Since multiple online signal vectors exist in one online sample, two further steps are involved to determine the chosen floor. In the first step, the single online signal vector $\boldsymbol{r}_o^v$ belongs to which cluster must be determined. OVO scheme is implemented to determine the cluster which is chosen according to the cluster $\boldsymbol{c}_f$ that receives majority votes. As mentioned above, the number of available clusters of every online signal vector $\boldsymbol{r}_o^v$ is ς, and the majority votes is made based on $\frac{\varsigma(\varsigma-1)}{2}$

binary kernel logistic regression classification. The total count cluster $\boldsymbol{c}_f$ out of $\frac{\varsigma(\varsigma-1)}{2}$ classifications is determined as the winning cluster for each online signal vector $\boldsymbol{r}_o^v$. The process of determining the winning cluster using OVO scheme is repeated for another online signal vector until $\frac{L(L-1)}{2}$ round. Figure 3.17 shows the example of architecture of majority voting using OVO scheme to determine the winning cluster for each round of signal vector $\boldsymbol{r}_o^v$ classification. In the second step, the cluster that defines every online signal vector $\boldsymbol{r}_o^v$ is counted to determine the floor. The maximum counted cluster decides the floor where:

$$\hat{\boldsymbol{c}}_{KLR} = \arg\max_{\boldsymbol{c}_f}\{\, \boldsymbol{c}_f \; count\} \tag{3.30}$$

and the estimated floor is given as:

$$\hat{\mathrm{p}}_z = z(\hat{\boldsymbol{c}}_{KLR}) \tag{3.31}$$

The flow chart of proposed KLF algorithm is illustrated in Figure 3.18.

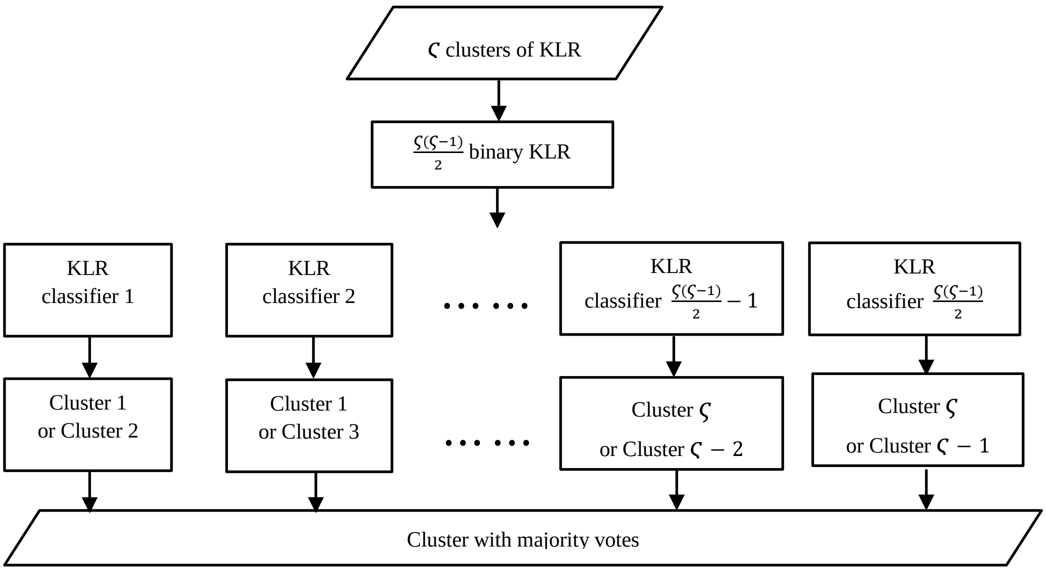


Figure 3.17. OVO scheme of each round of online signal vector $\boldsymbol{r}_o^v$ in KLF algorithm

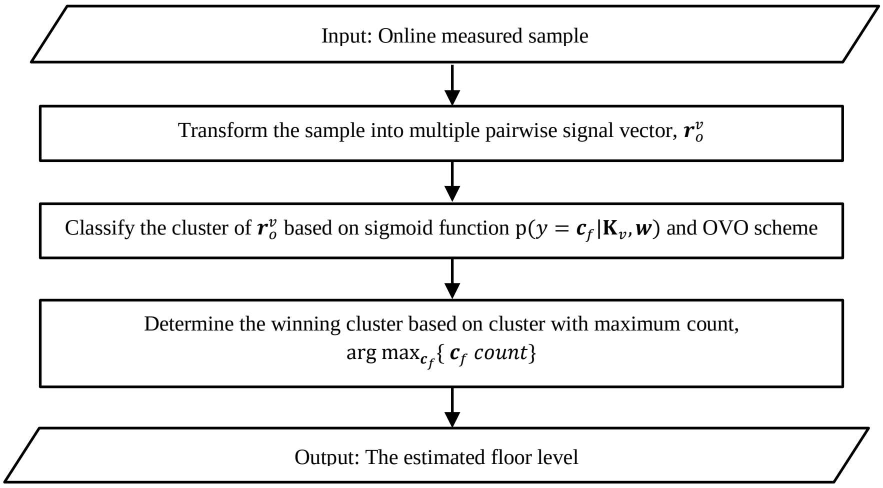


Figure 3.18. Flow chart of the proposed Kernel Logistic Floor algorithm

3.9 Horizontal Localisation Algorithm

Once the floor is estimated, the horizontal localisation algorithm is processed. Horizontal localisation estimates precise x and y coordinate of the device on the floor level. Detailed of the proposed horizontal localisation algorithm developed in this work is presented in the following sub sections.

3.9.1 Multi-class k-Nearest Neighbour Classifiers as Horizontal Localisation Algorithm

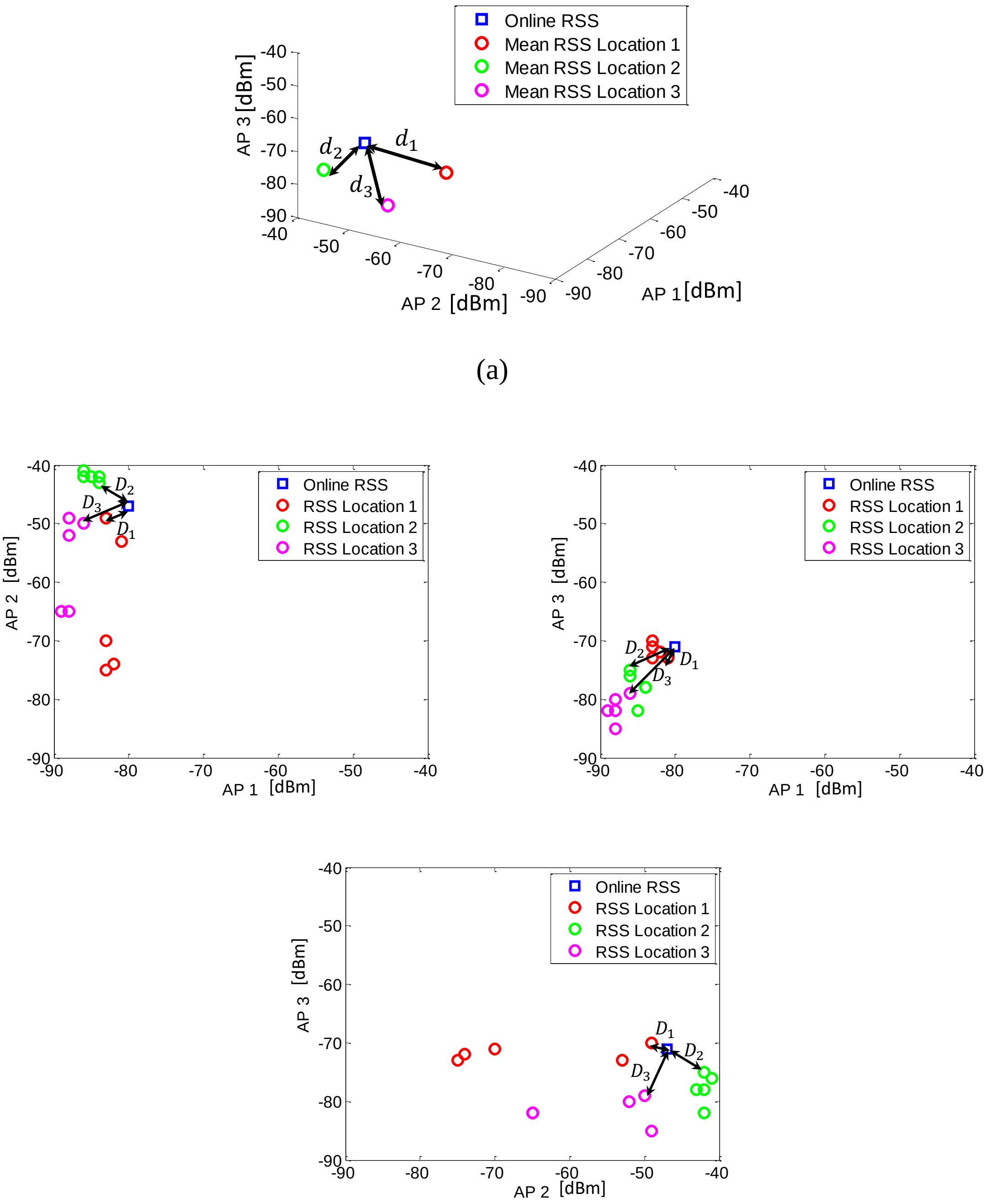
The k nearest neighbour classifier used in this work is a non-parametric supervised learning algorithm. The algorithm does not require a pre-determined decision boundary to classify a test sample. The algorithm is a different type of algorithm from the classical kNN algorithm as discussed in Section 3.3.1. The existing kNN algorithm is a regression type algorithm which finds the closest neighbour based on straight line distance between the test sample and the average value of training sample at each specific fingerprint in the dataset. On the other hand, the kNN classifier works by extracting the test sample signal data into multiple vectors and classify the vectors based on closest distance to the fingerprint data vectors in feature space. The difference between kNN regression and kNN classifier is illustrated in Figure 3.19 for three AP case. In classical kNN algorithm (Figure 3.19(a)), the signal distance between signal of online sample and mean signal of offline samples are simultaneously calculated for the three APs and the k offline samples with shortest distances are chosen to determine the estimated location. Meanwhile in kNN classifier algorithm (Figure 3.19(b)), the online and offline signal data are extracted into three pairwise AP feature space i.e. [AP 1, AP 2], [AP 1, AP 3], and [AP 2, AP 3] in order to find the offline sample with shortest distance to the online sample.

Using signal distance between pairwise AP signal of test sample (online) and fingerprint sample (offline) as in Equation 3.20, the location could be estimated by taking the nearest data $\psi_{\Theta i}$ of each class i that gives the minimum $D_\beta(\psi_{\Theta i})$ is chosen where:

$$D_{\beta i} = \arg\min D_\beta(\psi_{\Theta i}) \tag{3.32}$$

Since there are multiple combinations exists for each online sample $\mathbf{A}_\beta$, the decisive signal distance for every potential estimated class i is calculated as the average value of each class minimum distance for every online sample AP combination given as follows:

$$\Omega_i = \frac{1}{\mathrm{B}}\sum_{\beta=1}^{\mathrm{B}} D_{\beta i} \tag{3.33}$$



(a)

(b)

Figure 3.19. Comparison of (a) classical *k*NN algorithm and (b) *k*NN classifier for 3 AP case

The distance Ω_i is sorted in ascending order and first k classes are used to determine the estimated class (location), $\hat{\mathbf{p}}$ which is calculated as follows:

$$\hat{\mathbf{p}} = \frac{1}{k}\sum_{\kappa=1}^{k}\hat{\mathbf{p}}_{\kappa} \tag{3.34}$$

The scheme is called as multiple multi-classes kNN classification. Figure 3.19(b) shows the example of multiple multi-class kNN classification using 3 APs dataset. First the average distance of D_1, D_2, and D_3 for three AP combinations are computed. Next, the average distance is sorted in ascending order. If the required k is 2, the top two classes with lowest average distance are chosen and the estimated distance is the average class of the two classes (locations).

3.9.2 Optimisation of the Multi-Class kNN Classifier

Instead of using only nearest single grain of RSS vector data as in Equation 3.32, the second, third, fourth and etc. nearest data could also be considered in the distance $D_\beta(\psi_{\Theta i})$ calculation. The purpose of using additional data is to overcome noisy estimation of the location which is normally experienced if single RSS vector data is used. For example as in Figure 3.20, three nearest distances are considered for each class and the resulting estimated location is the Location 2 (which belongs to the green circles signal data). However, if only the nearest distance is considered, the estimated location is Location 1 (which belongs to the red circles signal data). The real location of the online sample (blue square) is Location 2 and this means if multiple nearest distance is considered, the location estimation could be improved. The total number of nearest signal distance is noted as ϖ. The Equation 3.32 could be replaced with:

$$S_{\beta i} = \frac{1}{\varpi}\sum_{\omega=1}^{\varpi} \min_{\omega} D_\beta(\psi_{\omega i}) \qquad (3.35)$$

where ϖ is the total of ω nearest data in each class i. The term $D_{\beta i}$ in Equation 3.32 could be replaced with $S_{\beta i}$ and the equation becomes:

$$\Omega_i = \frac{1}{B}\sum_{\beta=1}^{B} S_{\beta i} \tag{3.36}$$

The effect of optimisation parameter ϖ to the localisation performance is discussed in Chapter 7.

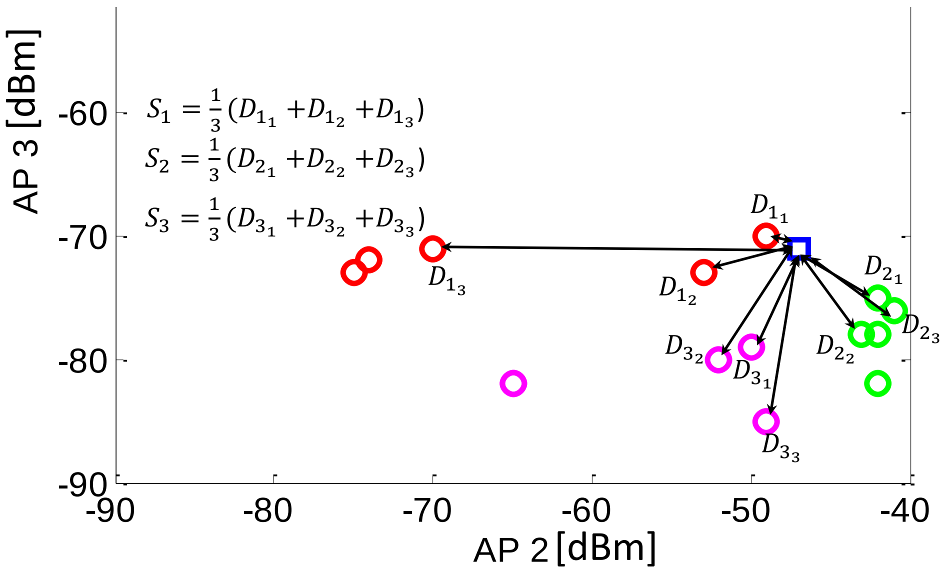


Figure 3.20. Example of optimisation parameter used in kNN classification with $\varpi = 3$ to calculate averaged signal distance S using example in the bottom most figure in Figure 3.19(b)

3.9.3 Kernel Multi-class kNN Classifier

The kernel version of kNN classifier finds the distance between kernel function of the online sample and the kernel function of the AP pair signal datasets. Therefore, the Equation 3.20 could be rewritten as follows:

$$D_\beta = \left\| \boldsymbol{r}_o^\beta - \boldsymbol{r}_i^\beta(\psi_{\Theta i}) \right\|$$

$$= \boldsymbol{K}\left(\boldsymbol{r}_o^\beta, \boldsymbol{r}_i^\beta(\psi_{\Theta i})\right) - 2\boldsymbol{K}\left(\boldsymbol{r}_o^\beta, \boldsymbol{r}_i^\beta(\psi_{\Theta i})\right) + \boldsymbol{K}\left(\boldsymbol{r}_i^\beta(\psi_{\Theta i}), \boldsymbol{r}_i^\beta(\psi_{\Theta i})\right) \tag{3.37}$$

Then, similarly the steps to determine the estimated location are followed as linear kNN classifier.

3.10 Summary

This chapter presents the related theory of proposed indoor localization system. The basics of WLAN fingerprint localisation are described where two phases: offline and online involved in developing a localisation system. In offline phase, fingerprint radio map is developed based on measurement of signal strength at multiple locations in indoor environment. In online phase, the location of the device is estimated by running the localisation algorithm which processes the radio map to infer the location of the device. Descriptions of three popular localisation algorithms are given which are kNN, univariate kernel, and multivariate kernel. The algorithms are used for comparison with proposed algorithm in Chapter 5 and 7. Then, detailed theory on kernel density estimates and machine learning classification involving logistic regression and kNN classifiers are presented which is used for proposed multi-floor localisation technique. The kernel density estimates and logistic regression classifier is applied for AP selection approach and floor localisation algorithm while kNN classifier is implemented for horizontal localisation algorithm. Based on the described theory, the proposed kernel and multi-class classifiers algorithm for multi-floor localisation is presented. For AP selection technique, two methods are described based on kernel density estimates and KLR. Kernel density estimates based on Gaussian kernel function which finds the *Maximum A Posteriori* estimate to choose the APs while NLL function of the KLR classifier is used to find the best AP combination to serve as the basis to select the AP. To determine the floor location, fingerprint elements are grouped as clusters according to proposed signal strength clustering technique. The signal strength clustering technique groups the APs of similar ID with strongest signal of AP at each fingerprint location. The floor is determined based on kernel density estimates which average the density estimates of fingerprint elements within similar cluster. The floor location algorithm is also proposed based on classification technique using KLR by counting the decision of the sigmoid function. The horizontal location algorithm is proposed based on kNN classifier optimised with additional parameter ϖ. Additional kernel kNN classifier is also investigated. All of the proposed algorithms with related objectives mentioned in Chapter 1 as given in Table 3.1.

Table 3.1. Summary of proposed algorithms with related objectives

Algorithm Type	Algorithm Name	Objective Number in Chapter 1
AP Selection	1. Max Kernel 2. Kernel Logisitic Pairwise Dicriminant	1
Vertical or Floor Localisation	1. Average Kernel Floor 2. Kernel Logistic Floor	2
Horizontal Localisation	1. Normal Multi-Class kNN Classifier 2. Kernel Multi-Class kNN Classifier	3

CHAPTER 4

MEASUREMENT AND METHODS

4.1 Introduction

The detailed implementation of the proposed algorithms for multi-floor localisation system was discussed in previous chapter. This chapter presents the fingerprint measurement procedure of WLAN signal data and related methods in order to test the validity and performance of the developed algorithms. Details of measurement and computation apparatus are explained in Section 4.2. The floor plans of the measurement location are described in Section 4.3. Descriptions on additional datasets used to test the algorithms are provided in Section 4.4. Section 4.6 explains the method to verify the measured WLAN signal data is justified. Section 4.7 describes about the classification of the dataset as the test and measured samples for the algorithm is mentioned. Section 4.8 discusses the method to verify the measured signals for the radio map database based on the path loss model and the result of extracted path loss parameters from measured fingerprint signal data is analysed. The performance metrics used to evaluate the output from the algorithms are given in details in Section 4.9. Lastly, the method to choose value of k in classical kNN algorithm is defined.

4.2 Measurement and Computation Tools

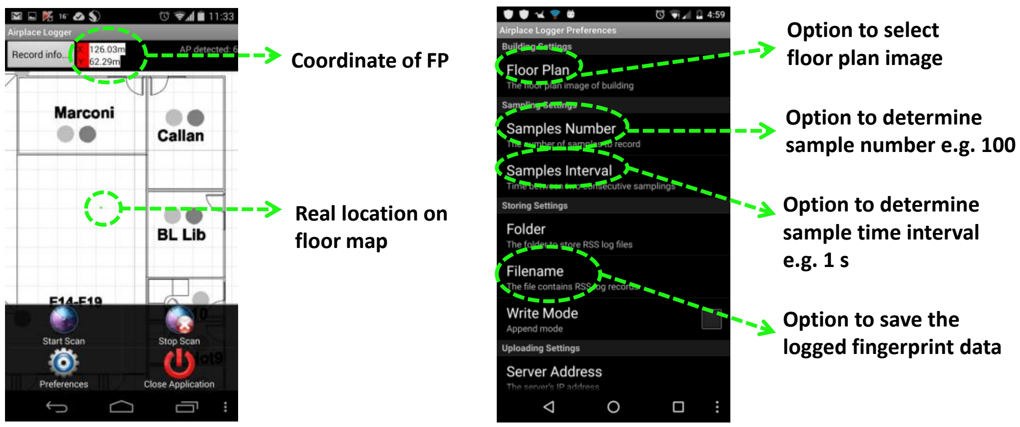
The measurement of the WLAN fingerprint signal data is done using by a smartphone with Android signal strength application (Wang et al. 2015). The smartphone is Google Nexus 5 and the specification which includes information of the WLAN receiver is given in Table 4.1.

Table 4.1. Specification of the Google Nexus 5 mobile phone as the measurement device

Model Number	LG-D821
Processor	2.26 GHz Quad-core Snapdragon 800
Random Access Memory (RAM)	2 GB
GPU	450 MHz Adreno 330
WLAN Frequency Band	2.4 GHz
Operating System	Android 4.4
WLAN Technology Connectivity	IEEE 802.11 a, b, g, n, ac
Embedded WLAN Receiver	Broadcom BCM4339 5G Wi-Fi combo chip
WLAN Antenna Type	Flexible Printed Circuit Board Antenna
WLAN Antenna Gain	-1.96 dBi

The device to measure the signal strength information for indoor localisation is mainly based on mobile device such as laptop equipped with WLAN card (Kushki et al. 2007, Fang and Lin 2010) or smartphone (Wang et al. 2015, Liang, Zhang, and Peng 2015). The variability of measurement devices are due to the fact that up to now there is no standardised measurement guideline has been published for indoor localisation system. Since this thesis is mainly focused on the localisation algorithms for the system, it is enough to limit to one measurement device in order to verify the performance of the algorithms. Moreover, the measured signal data is confirmed follows the propagation rule by extracting the path loss parameters of the signals according to path loss model which is described in Section 4.8.

To measure the WLAN signal using the smartphone, an application interface must be installed within the operating system of the smartphone. The application that is used in this work is the Airplace Logger (Laoudias et al. 2012) which was developed by University of Cyprus research group. The interface of the application is illustrated as in Figure 4.1. To start measuring the WLAN signal, the user is required to upload respective floor plan map within the software and set the configuration of the required fingerprint data logging process. Then, the user chooses the location on the loaded map which matches the location of the device to record the WLAN signal data. The process is repeated by the user moves the device into different locations to record the data. During the signal recording, the phone is held in on a tripod where the height of the phone is about 1 m from the ground. The posture of the user with the smartphone placed on a tripod recording the signal is illustrated in Figure 4.2. Similar measurement procedure was practiced in Wang et al. (2015).



(a) (b)

Figure 4.1. The interface of the Android Airplace Logger application which shows (a) the main screen where the user can record the fingerprint data, and (b) the configuration screen where the user can set the fingerprint collection procedure

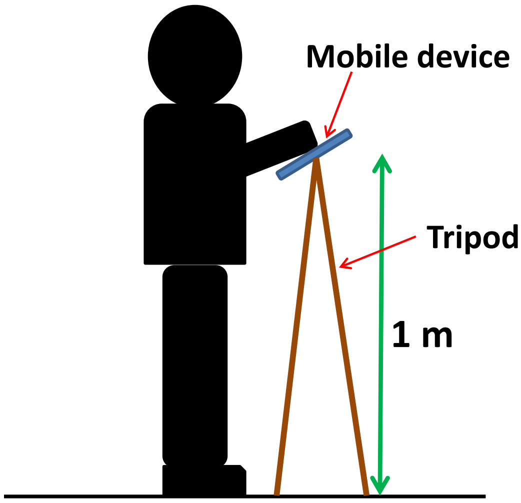


Figure 4.2. The posture of user measure the signal strength at a fingerprint location with a smartphone (mobile device) held by a tripod 1 m above the ground for measuring the fingerprint data at every allocated position

The recorded signal data is served as offline and online samples of the developed algorithm as discussed in Chapter 3 to calculate estimated location and evaluate the results. All of the algorithm computations were realised using Mathworks MATLAB 2013a software. The MATLAB is chosen because it offers some advantages such as easier to implemented mathematical computation as the language provides many predefined function to execute the computation process, and also the storing and processing of datasets could be done in one platform. Additionally, the performance of the different algorithms could be easily analysed and compared by graph plotting which is built within the software. On the other hand, the main limitation of the MATLAB is slow processing compared to other low-level language such as C or C++. The MATLAB software is run on the desktop computer with specification as listed in Table 4.2. The output which is the results of the algorithms computation is discussed in Chapter 5 to 8.

Table 4.2. Main specification of the desktop computer running the MATLAB software

Processor	Intel Core-i5-4460 3.20GHz
Memory	8 GB DDR3 RAM
Operating System	Microsoft Windows 8.1

4.3 Measurement Environments and Locations

The measurement is done in three buildings of different scenario. In existing works, mainly only one building is used for measurements (Alsehly, Arslan, and Sevak 2011, Bhargava, Krishnamoorthy, and Karkada 2013, Campos, Lovisolo, and de Campos 2014). Thus by using three different buildings in this work is good enough to confirm the results of the developed algorithms. The buildings are located within Faculty of Science (Building A), Faculty of Engineering (Building B), and students residential college (Building C) of Universiti Putra Malaysia (UPM) campus in Selangor, Malaysia. The measurement is done in the month of October 2014. The measurement is made within 8 AM to 5 PM during weekdays. The measured signal could be interfered from nearby ad-hoc mobile hotspots deployment, co-channel interference of the APs, the movement of people, and door openings and closings. Additionally, the measurement of signals for the fingerprint dataset and the test samples are made in different days. The building environment details are summarised in Table 4.3. The real building locations are referred to their GPS coordinates. Each building comprises different number of floor levels i.e. Building A has 5 floors, Building B has 11 floors, and Building C has 8 floors, and also has different floor plan dimensions. The vertical height of each floor level of all buildings is 3 m. The fingerprint locations on the ground floor plan of the three buildings are shown in the Figure 4.3, 4.4, and 4.5. The photos of actual buildings are shown in Appendix B1. The walls of the buildings are made of concrete. The doors are finished with solid wood while the windows are constructed from glasses. The floors are made of reinforced concrete covered with floor tiles. In Building B, there are open spaces between second to ninth floors. In Building C, there are hollows between outside walls on every level. The locations of fingerprint measured on first floor and above for all of the buildings are shown in Appendix B2. The floor plan is obtained from the office of development and asset management of the university. The dimension of each floor plan is stated in the figures.

As could be seen from the figure, the measurement locations between floor specifically in Building B and C are slightly different. Building B floor plan from third to tenth floor (Figure B1.2 (b) to (j)) contains open spaces between these floors as marked as "VOID" in the figures compared to first and second floor. On the other hand, the areas of measurement for Building C from first to the highest floor are only on the right wing of the building due to restriction of the area on the left wing. The fingerprint locations on each floor plan of Building A, B, and C are the non-restricted area of the buildings. The figures also plot the location of the APs installed within the buildings (with green square markers in Figure 4.3, 4.4, and 4.5). The total fingerprint location numbers of each floor of each building are different due to the different dimension of the floor plans, different structure of each floor, and also the restriction of access to the area. The measured fingerprint location spacing is determined as 2 m for all buildings. The fingerprint spacing has been reported in existing work being measured between 1m to 5m (Kushki et al. 2007, Redzic, Brennan, and O'Connor 2014, Wang et al. 2015). The measured signal data is stored as radio map database which will be input of the proposed localisation algorithms and also used as online samples to verify the performance of the algorithms. Description of the database, online samples, and the statistics of the measured signals are given in the following section.

Table 4.3. Building A, B, and C physical specifications

Building Reference	Building Type	GPS coordinate location (Latitude, Longitude)	Number of Floor Levels
A	Academic Building	3.00071, 101.705241	5
B	Laboratories and Offices Tower Building	3.00870, 101.721399	11
C	Student Residential Building	3.01006, 101.720102	8

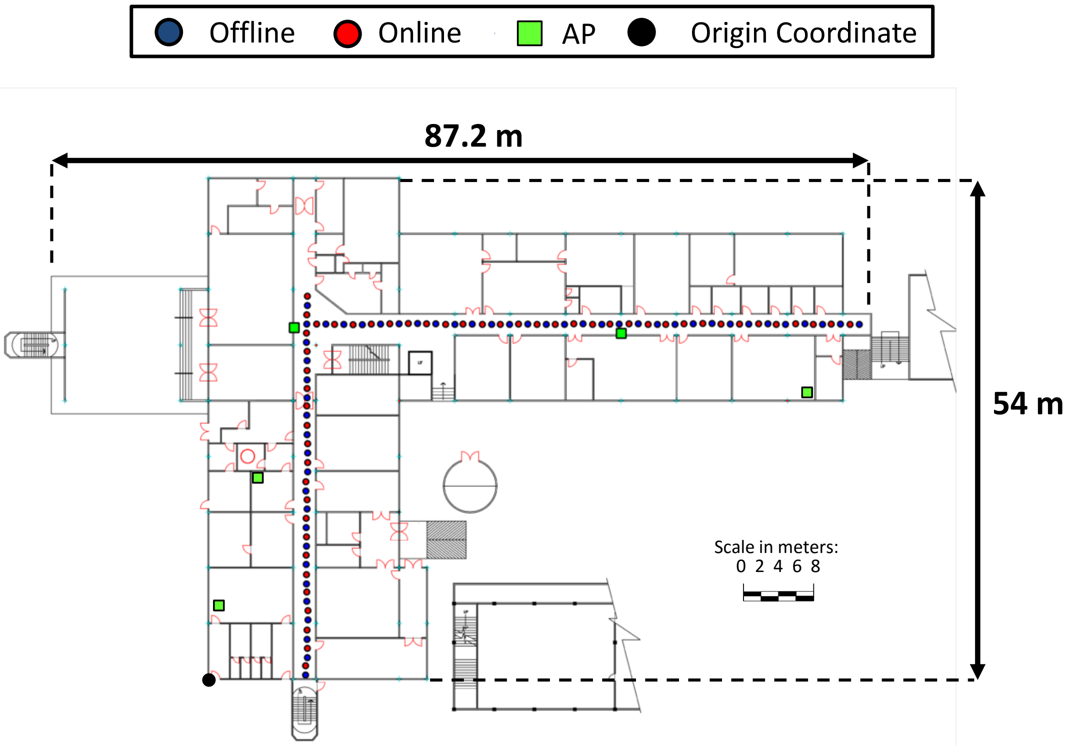


Figure 4.3. Ground floor plan of Building A where WLAN signals are measured. Location of fingerprint, measured online samples, AP locations, and origin measurement location coordinate are marked as shown in the legend.

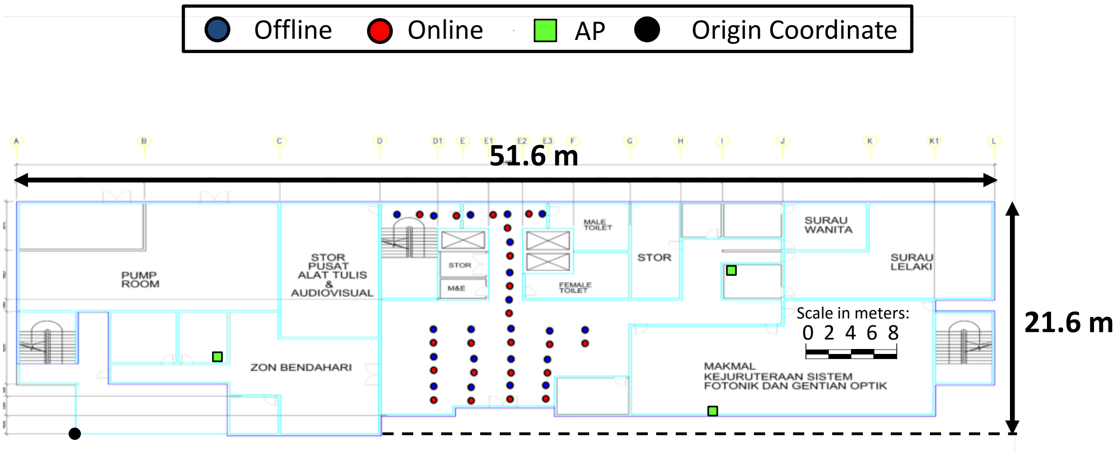


Figure 4.4. Ground floor plan of Building B where WLAN signals are measured. Location of fingerprint, measured online samples, AP locations, and origin measurement location coordinate are marked as shown in the legend.

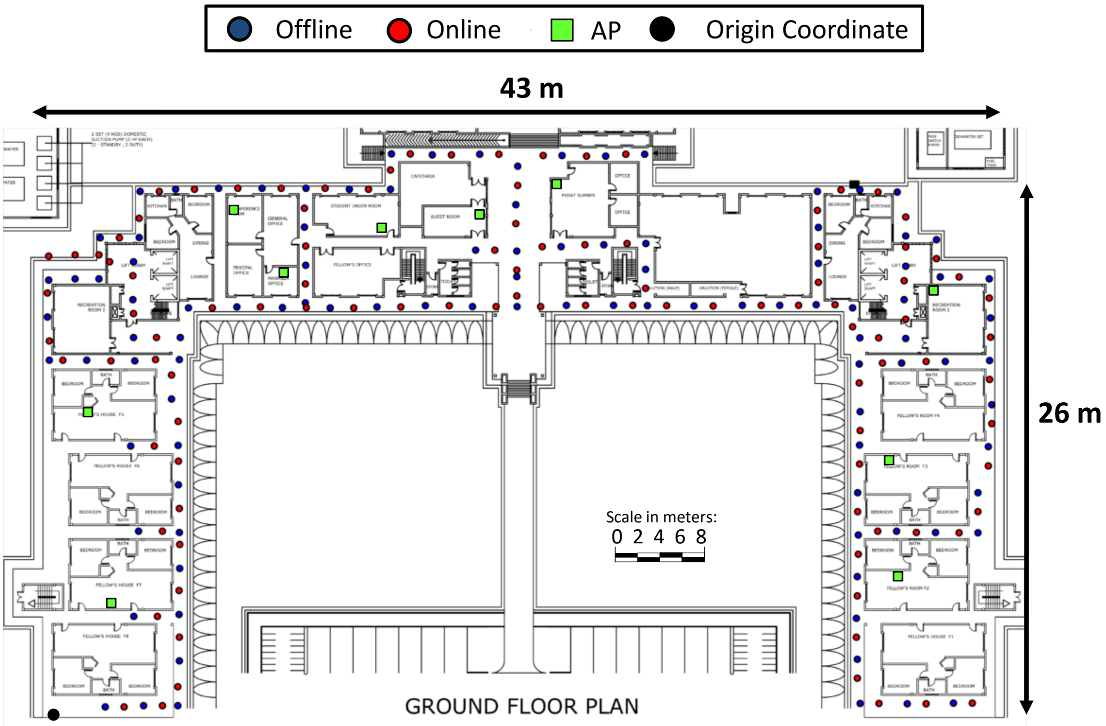


Figure 4.5. Ground floor plan of Building C where WLAN signals are measured. Location of fingerprint, measured online samples, AP locations, and origin measurement location coordinate are marked as shown in the legend.

4.4 Radio Map Database and The Measured Signal Statistics

Each entry of radio map database is referred by the fingerprint location on specific floor location of the building. The example of the entry was given in Figure 3.3. At each fingerprint location, a total of 100 measurement samples are taken at duration of 1 s which follows the procedure of Lin et al. (2014). The samples are recorded in four directions of the user facing north, east, south and west of the floor plan. This means there are 25 signal samples for each direction at one fingerprint location. In total there are 73300 of fingerprint measurement samples in all buildings of which 25400, 19600, and 28300 are in Building A, B, and C, respectively. Comprehensive statistics of the measured data for the radio map is discussed in the Section 4.6.

4.5 Additional Datasets

In this thesis, three additional test datasets are used to confirm the performance of the algorithms in multiple buildings. Additionally, the datasets could verify the validity of the measured datasets in Building A, B, and C to evaluate the algorithms. The difference between these dataset with previously discussed dataset is the measurement is only done on single-floor level compared to multi-floor level in previous dataset. Those three datasets are names as CRAWDAD, Antwerp, and Dublin. The detailed description the datasets is as follows:

4.5.1 CRAWDAD

CRAWDAD is the fingerprint dataset of University of Mannheim, Germany which could be downloaded from CRAWDAD Community webpage (Thomas et al. 2008).

CRAWDAD has also been used previously in indoor localisation works e.g. Fang and Lin (2009), Fang and Lin (2012). The measurement tools include IBM ThinkPad R51 laptop running Linux kernel 2.6.13 and Wireless Tools 28pre software. The WLAN signal receiver of the laptop is the Lucent ORiNOCO Silver PCMCIA network card supporting 802.11b connectivity.

The dataset is based on single-floor measurement. The spacing between fingerprint locations is about 1 m. The measurement locations of offline and online sample, and location of APs on the floor map are shown in Figure 4.6. The offline and online measurements are done at 166 and 60 locations respectively. At every fingerprint location, 880 samples are collected where 110 samples are collected in one of eight direction of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° from north. For this work, the offline dataset is filtered which takes only random 20 random samples in each direction as suggested by the original author in King et al. (2006)]. Thus each fingerprint element consists of 160 samples. The total of filtered offline samples is 26560. For online sample, the direction of each sample is randomly picked between those eight directions at each measured location.

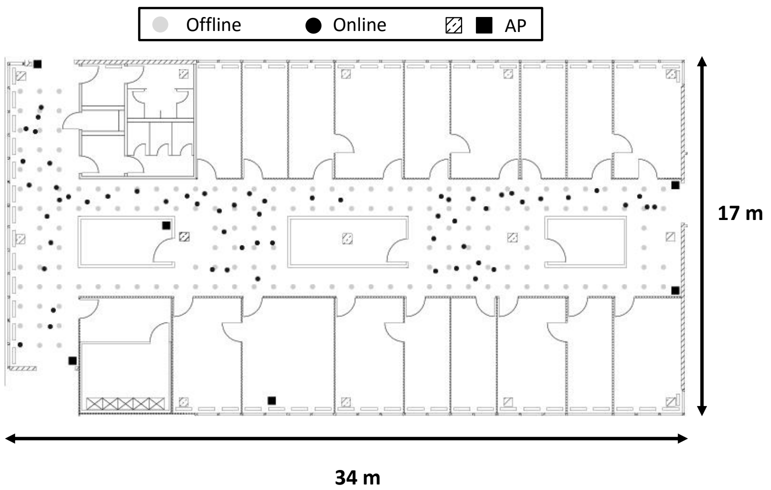


Figure 4.6. Floor plan of CRAWDAD dataset. The location of measurements and APs are marked according to the legend.

4.5.2 Antwerp

Antwerp is the dataset measured at Alcatel-Lucent Building in Antwerp, Belgium. The measurement is done on single-floor level of seventh floor of the building which is the highest floor level. The measurement is done also using Google Nexus 5 smartphone using Airplace logger software. The floor map is obtained from the building administrator. The outside walls are made of concrete and the walls between rooms and the doors are made of plasterboard. The windows are fully covered by glasses. There are soft partitions of working spaces within the main area of the measurement for both environments.

Fingerprint spacing is 2 m. The offline and online measured samples and location of APs are shown in Figure 4.7. There are 145 of offline fingerprint and online sample locations. The distance of online sample location to the offline fingerprint location is 1 m. For offline fingerprint data, the number of measurement sample per location is 20. The measurement sample consists of four directions of north, east, south, and west. For each direction, 5 samples are taken per fingerprint at interval of 1 s. The total of measured offline samples are 2900. For online sample, one measurement sample is taken per location of which direction is randomly chosen.

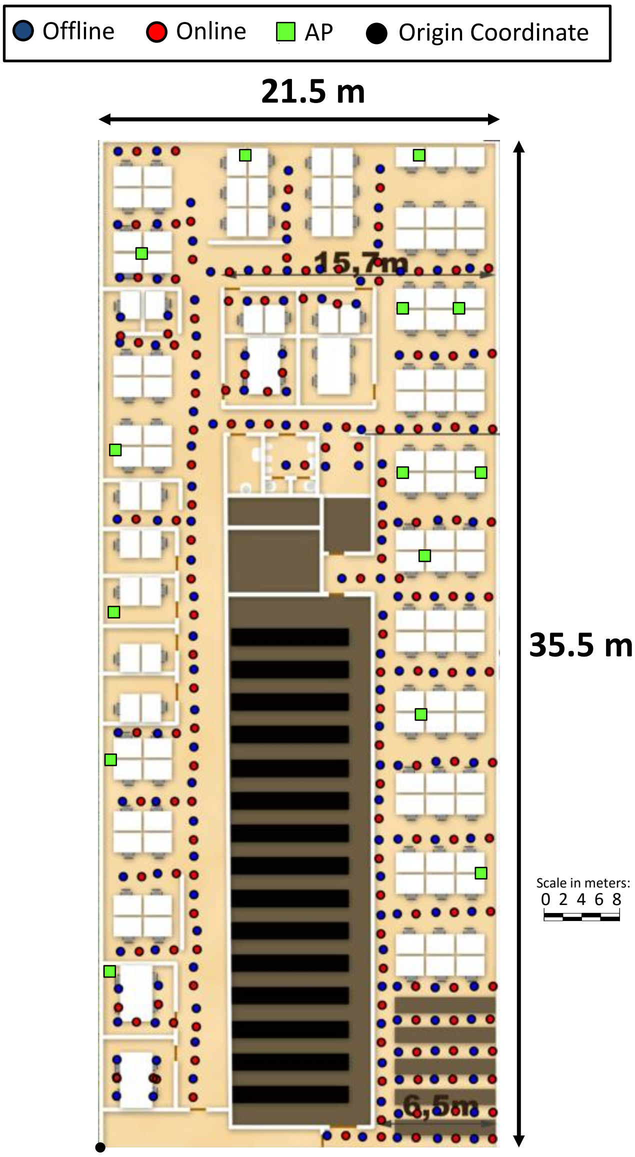


Figure 4.7. Floor plan of Antwerp dataset with measured offline and online samples, AP locations, and origin coordinate of measurement locations.

4.5.3 Dublin

Measurement of Dublin WLAN signal fingerprint dataset took place in Bell-Laboratories building in Blanchardstown, Dublin, Ireland. The dataset is also a single floor dataset. The building is one-floor level. Building description is similar as Antwerp where the outside walls are made of concrete and the walls between rooms and the doors are made of plasterboard. The windows are made by glasses. The measurement is completed using Google Nexus 7 tablet with Airplace Logger software. The antenna type is Planar Inverted-F Antenna with -0.541 dBi gain. The tablet is equipped with 1.5 GHz Qualcomm Snapdragon S4 Pro processor, 2 GB of RAM, and Adreno 320 GPU. The WLAN receiver embedded within the tablet is Qualcomm Atheros WCN3660 which supports WLAN a, b, g, and n connectivity.

The measurement is done in open-space area. The offline fingerprint data is measured at 79 locations with spacing of 2 m as could be observed in Figure 4.8. At each location, 4 samples are taken where each sample is according to north, east, south, and west directions at interval of 1 s. The total amount of offline measurement samples collected in Dublin building is 316. Amount of measured location of online samples is collected at collected at 249 locations which are about three times larger than offline samples. The direction of user taking the online sample at each fingerprint is randomly chosen between the four directions.

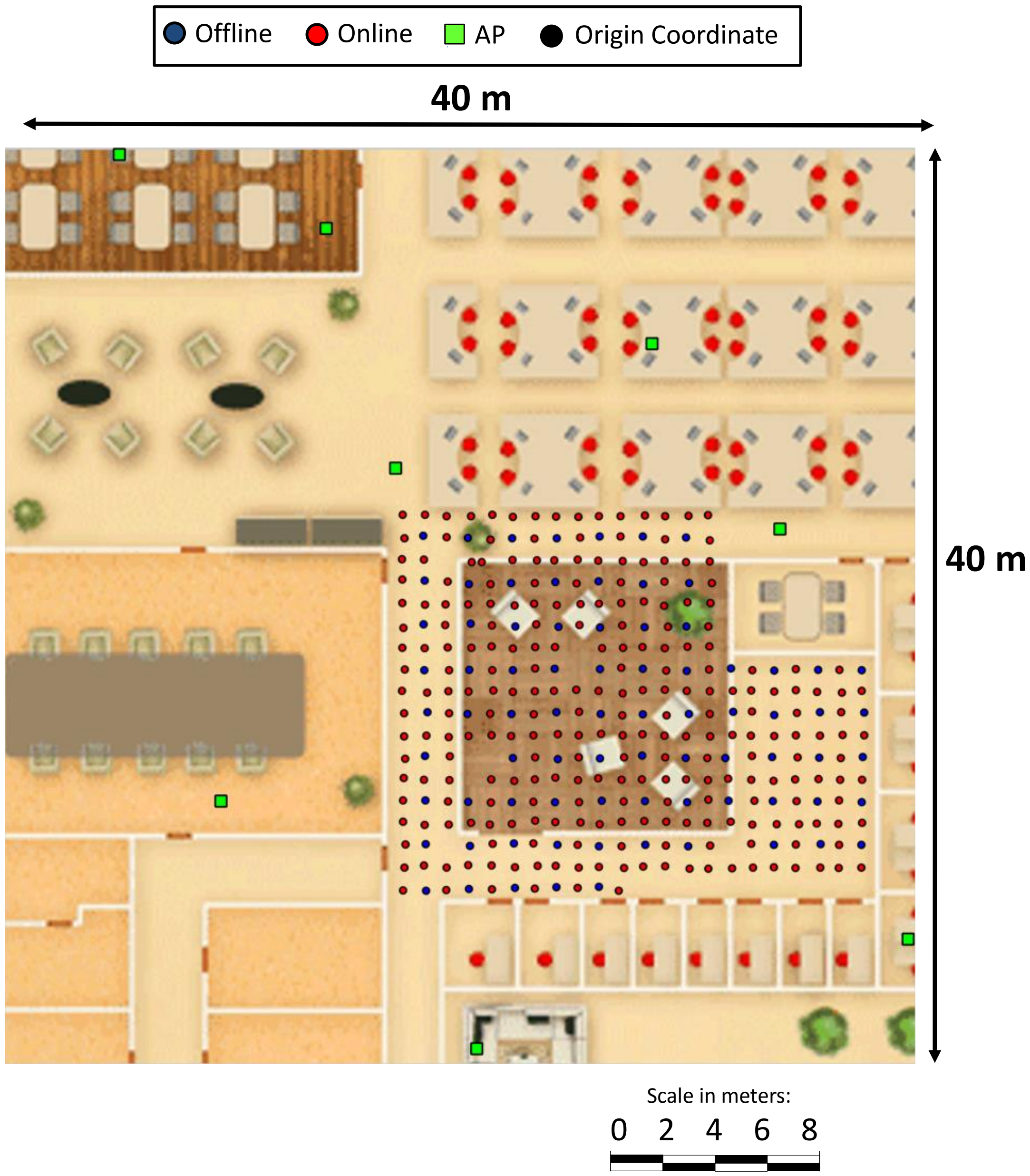


Figure 4.8. Floor plan of Dublin dataset. The offline and online measurement and AP locations are as marked.

4.6 Statistics of Radio Map Database of Several UPM's, Bell Lab's Buildings, and CRAWDAD

Floor plan layouts of UPM's have been described in Section 4.3. Also, the Bell Lab's buildings and CRAWDAD floor plans have been discussed in Section 4.5. The building layouts of all UPM buildings are described follows. The walls are made of concrete. The doors are made of solid wood. The windows are made of glasses. The floors are made of reinforced concrete finished with tiles. In Building B, there are open spaces between second to ninth floors as shown in Figure 4.9. In Building C, there are hollows between outside walls on every level. On the other hand, the building description of both Bell Labs environments is different. The outside walls are made of concrete and the walls between rooms and the doors are made of plasterboard. The windows are fully covered by glasses. There are soft partitions of working spaces within the main area of the measurement for both environments. The measurements of UPM buildings are made in every floor level while the measurement in both Bell Labs buildings and are done on single-floor. The CRAWDAD dataset consists of single-floor measurement. The localisation algorithm applied in multi-floor buildings is expected to perform less compared to single-floor buildings because of introduction of additional signals from additional APs and fingerprint samples as discussed in Chapter 1.

Figure 4.9. Open space between floors in Building B

Table 4.4 shows the statistics of the multi-floor radio map database of every floor level for all buildings in UPM. In Table 4.4, the total number of fingerprints on each floor level is similarly as marked on the floor map shown in Section 4.3. The number of fingerprints varies between floors according to accessibility of spaces on the floor. It can be seen from the table that each floor of all the buildings has almost similar number of fingerprints except for ground floor Building C which has a larger area of measurement. As could be seen in the Table 4.4, the number of AP measured in this work is more than the minimum number of AP required to obtain distinctive fingerprint signature according to trilateration rule (Kushki et al. 2007) which is three. Minimum and maximum number of detected APs per fingerprint location in UPM's Building A, B, and C are 3 and 14, 3 and 39, and 5 and 36 which is given in column 3 and 4 in Table 4.4. On average, the number of detected APs per fingerprint location in the respectively buildings on one floor level are at least 6.6 (fourth floor), 8.1 (ground floor), and 17.1 (first floor). This indicates that number of APs detected at each fingerprint location is not uniform. In the dataset of each building, the number of unique AP ID found is at least 24, 37, and 56 respectively in Building A, B, and C. The minimum value of strongest signal level of -46 dBm, -66dBm, and -37dBm in Buildings A, B and C respectively show that the location of APs is near to the location of fingerprint measurement.

Table 4.5 gives the statistics for single-floor dataset for Bell Labs (Antwerp and Dublin) and CRAWDAD. The minimum and maximum number of APs detected within the floor is 9 and 13, 7 and 20, and 27 and 38 respectively in CRAWDAD, Antwerp, and Dublin environments. The average number of APs detected is 10.1, 16.8, and 33.0. The number of unique AP IDs detected in the respective buildings is 21, 20, and 39. Also the strongest signal level of -45 dBm, -20 dBm, and -49 dBm shows proximity of the measurement locations with the location of APs. The values of the signal level vary from Table 4.4 because they depend on the distance between location of the APs and the fingerprint measurement location. The number of fingerprints is larger than the UPM buildings on the floor level because the measurements are carried out at large open-space area of the building.

Table 4.4. Radio map fingerprint database signal statistics per floor level of multi-floor dataset

Building	Floor Level	Number of fingerprints	Number of detected APs per fingerprint			Total unique AP IDs detected	Strongest signal level observed [dBm]
			Minimum	Maximum	Average		
A	0	51	5	12	7.4	37	-36
	1	51	5	13	8.4	28	-34
	2	51	6	11	8.5	24	-46
	3	51	4	12	7.0	24	-34
	4	50	3	14	6.6	37	-34
B	0	21	3	16	8.1	38	-66
	1	15	6	30	15.4	37	-43
	2	19	10	31	18.5	40	-53
	3	19	12	30	17.5	40	-47
	4	19	11	31	18.0	49	-32
	5	19	13	34	19.9	50	-37
	6	19	14	35	20.8	61	-34
	7	19	14	38	19.2	53	-48
	8	24	13	38	20.6	98	-45
	9	11	13	39	20.7	49	-55
	10	11	10	28	17.9	45	-60
C	0	102	5	36	19.4	118	-35
	1	27	8	30	17.1	56	-33
	2	27	8	25	18.0	60	-36
	3	27	8	32	20.4	66	-34
	4	27	5	29	18.3	68	-33
	5	26	12	35	20.5	83	-31
	6	23	9	32	19.7	82	-32
	7	24	9	29	17.8	77	-37

Note:
Column 3: Number of fingerprint elements which equals to the number of measurement location.
Column 4: Minimum number of APs detected in one fingerprint elements.
Column 5: Maximum number of APs detected in one fingerprint elements.
Column 6: Average number of APs detected in every fingerprint elements on each floor level.
Column 7: Unique ID of the APs detected observed in all fingerprint elements on each floor level.
Column 8: The best signal level detected from the fingerprint elements on each floor level.

Table 4.5. Radio map fingerprint database signal statistics for single-floor dataset

Building	Number of fingerprints	Number of detected APs per fingerprint			Total number of unique AP IDs detected	Strongest signal level observed on the floor [dBm]
		Minimum	Maximum	Average		
CRAWDAD	166	9	13	10.1	21	-45
Antwerp	145	7	20	16.8	20	-20
Dublin	78	27	38	33.0	39	-49

4.7 Online Test Samples

The number of online test samples is chosen of similar amount as number of fingerprints locations on each floor of the buildings to ensure realistic results. The direction of user at each location of the tested samples is randomly determined. The measurement of the test samples is done one day after the fingerprint measurement. The locations of the samples are distinctly measured from the collected fingerprints of around 1 m. The red filled circle markers in Figure 4.3,4,4, and 4.5 shows the measured locations of online sample in Building A, B, and C respectively.

4.8 Extraction of the Path Loss Parameters of Measured Signal of Radio Map Database

The measured signal strength values at multiple fingerprint locations could be validated using log-distance path loss model. The validation could be done by extracting the path loss parameters of the path loss model and apply the extracted parameters to the model. According to the path loss model in Seidel and Rappaport (1992), the measured signal strength values at each fingerprint should follow the log-distance signal rule. The model employs the signal at reference location from the AP and the related signal strength at varying separation distance of the AP and measurement locations. At different separation distance of AP and the fingerprint location, the measured signal at the fingerprint location should change according logarithmic values. For single-floor case, the most common path loss model (Rappaport 1996) is based on the following equation:

$$P_r(d)[\text{dBm}] = \mathrm{P}_r(d_0)[\text{dBm}] - 10\eta \log\left(\frac{d}{d_0}\right) + \chi_\sigma \tag{4.1}$$

where $P_r(d)$ is the RSS value observed at respected fingerprint location which is distance of the location from the AP is d, $\mathrm{P}_r(d_0)$ is the RSS value at a reference distance d_0 from the AP (usually chosen as 1 m), η is the path loss exponent, and χ_σ is the Normal random variable having standard deviation of σ. The path loss model is developed using linear regression where the term n and σ could be computed as suggested in Rappaport (1996). For multi-floor case, the Equation 4.1 is used together with another path loss model to describe the measured signal at different floor level. The measured RSS at different floor is described as follows (Seidel and Rappaport 1992):

$$P_r(d)[\text{dBm}] = \mathrm{P}_r(d_0)[\text{dBm}] - 10\eta_{SF} \log\left(\frac{d}{d_0}\right) + FAF[\text{dBm}] \tag{4.2}$$

where η_{SF} is the path loss exponent computed based on the same floor measurement of the AP signal, and FAF is the floor attenuation factor which describes the attenuation of the signal travelling across floor. For example if the AP is located at Floor 3 of the building, the measured RSS at different fingerprint locations is verified using Equation 4.1. Additionally, to verify the signal from the same AP which is measured on Floor 2, Equation 4.2 is used.

4.8.1 Path Loss Parameters Extraction Using Seidel's Model

Empirical path loss parameters are obtained by applying linear regression of signal strengths at each fingerprint on the log of separation distance between the AP and the fingerprints. According to Seidel and Rappaport (1992), the gradient (m) and intercept (c) of the line ($y = mx + c$) correspond to path loss exponent, η, and

RSS value at reference location d_0 for the same-floor locations of AP and the fingerprint measurements. The RSS value at reference location $d_0 = 1$ (meter) is measured by the device. Table 4.6 in column 5 and 6 lists the values of η, and σ obtained from the regression line according to measured RSS values (blue cross markers) at different fingerprint locations of three chosen APs as shown in Figure 4.10 to 4.12. It can be seen that the signal strength at d_0 is around -40 dBm for all buildings. The values of η is within the expected values which is identical to previously reported values for indoor locations. The η and σ values in Building A is around 1.8 to 1.9 and 5.7 to 6.0 respectively which are comparable to 1.57 and 4.02 reported in Akl, Tummala, and Li (2006) of which the measurement is also made in a closed corridor. The values obtained in this work are slightly higher because the signals found fingerprint locations in corridor on the east side are non-line of sight (NLOS) to the APs. The values of η in Building B and C are marginally larger because the NLOS fingerprint locations increases compared to Building A. The empirical values of of η, and standard deviation, σ, from the linear regression are used as the parameters of path loss models defined in Section 4.8. Substituting η and σ into Equation 4.1 will give the path loss model (blue line) as plotted in Figure 4.10 to 4.11 for APs in Building A, B, and C respectively. The models shows good fit with the measured RSS values in tested environments.

For multi-floor signal propagation, floor attenuation factor (FAF) must be calculated. According to Seidel and Rappaport (1992), the floor attenuation factor is calculated based on the average of the difference between predicted RSS by path loss model (blue line) plotted in Figure 4.10 to 4.12 and the mean measured RSS value on the multi-floor locations. The value of calculated FAF is tabulated in column 7 and 9 of Table 4.6 for one and two floor difference respectively. The trends of FAF on increasing floor for all buildings are similar where the higher the floor the higher the FAF as mentioned in Seidel and Rappaport (1992). The results of FAF for Building A almost similar as reported in Seidel and Rappaport (1992) for Office Building 2 environment i.e. the value of FAF though one and two floors of Building A between 13.7 to 14.8 and 20.0 and 24.8, are close to 16.2 and 27.5 of Office Building 2. The value of FAF depends on structures within the building which affect the attenuation of the signal. The FAF values in Building A increase higher compared to Building B and C because the environment is closed-space compared to the other two buildings which the floor structures are partly open e.g. Building B has hollows between floors and ceilings and Building C has hollows between walls. The multi-floor path loss model is plotted in Figure 4.10 to 4.12 for though one floor (red line) and two floors (blue line) path loss. The model shows good agreement with the measured RSS value of the fingerprints.

The plotted single-floor and multi-floor path loss model show that RSS value of a specific AP changes at different fingerprint locations according to propagation rule. This means a fingerprint location could be uniquely characterised by RSS value of AP. Therefore, localisation algorithms could be developed to estimate the location according to measured RSS value by the device. The results of localisation performance by the developed algorithms are discussed in the rest of following sections

Table 4.6. Path loss model parameters for chosen APs in Building A, B, and C

Building	AP Location	Floor Level	RSS at Reference Location, (d0 = 1) (Averaged)	Path loss Exponent , η	Standard Deviation, σ	One Floor FAF [dBm]	One Floor Standard Deviation, σ	Two Floors FAF [dBm]	Two Floors Standard Deviation, σ
A	(71.3,37.7)	1	-41.0	1.9	5.8	12.7	4.2	20.0	3.6
	(9.4,36.9)	3	-40.8	1.8	6.0	14.8	4.1	24.8	3.3
	(68.4,37.7)	4	-40.4	1.9	5.7	13.7	3.8	22.5	3.3
B	(16.5,5.1)	3	-40.5	3.2	4.6	5.2	4.5	8.7	3.6
	(28.3,5.2)	3	-40.3	3.6	4.9	2.5	4.6	5.7	3.6
	(14.6,5.1)	4	-40.1	2.5	4.8	9.2	3.9	14.9	3.7
C	(5.3,13.6)	5	-40.0	2.6	5.3	6.8	4.8	12.8	3.5
	(32.6,47.1)	6	-39.8	3.5	5.5	3.6	4.9	8.3	4.8
	(5.9,18.7)	7	-41.7	2.4	6.6	12.4	5.4	16.0	3.4

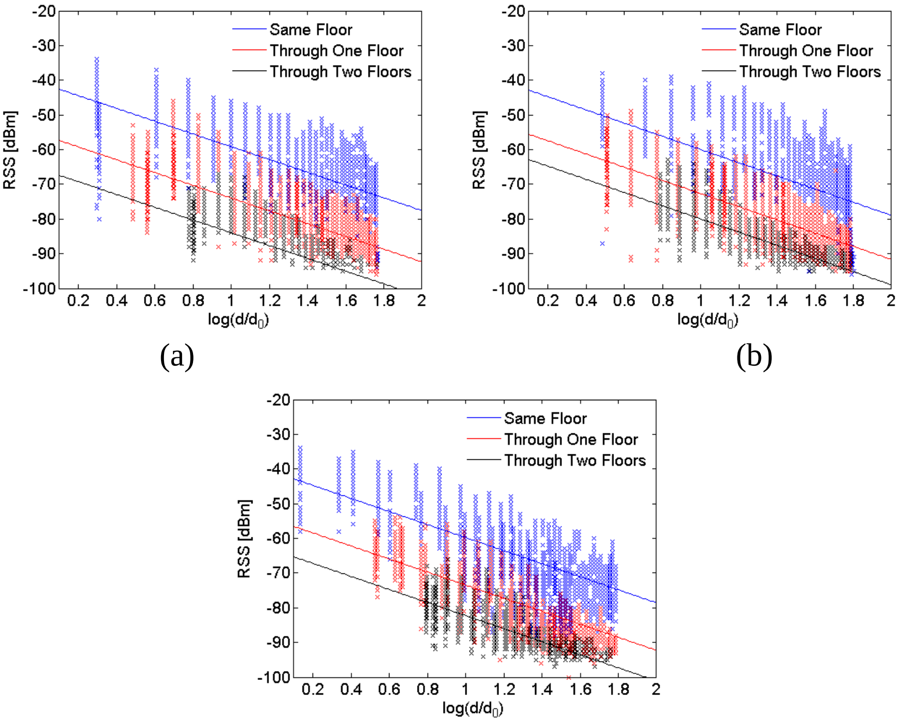


Figure 4.10. Path loss model of APs in Building A where (a) is the AP at Location (71.3, 37.7) in Floor 1, (b) is the AP at Location (9.4, 36.9) in Floor 3, and (c) is the AP at Location (68.4, 37.7) in Floor 4

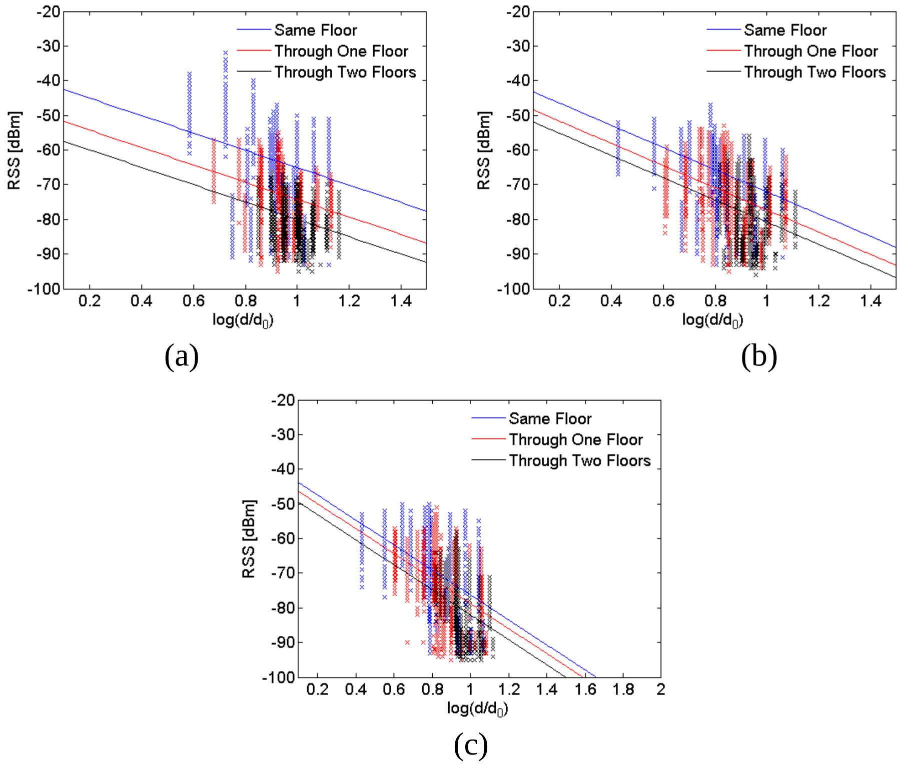


Figure 4.11. Path loss model of APs in Building B where (a) is the AP at Location (16.5, 5.1) in Floor 3, (b) is the AP at Location (28.3, 5.2) in Floor 3, and (c) is the AP at Location (14.6, 5.1) in Floor 4

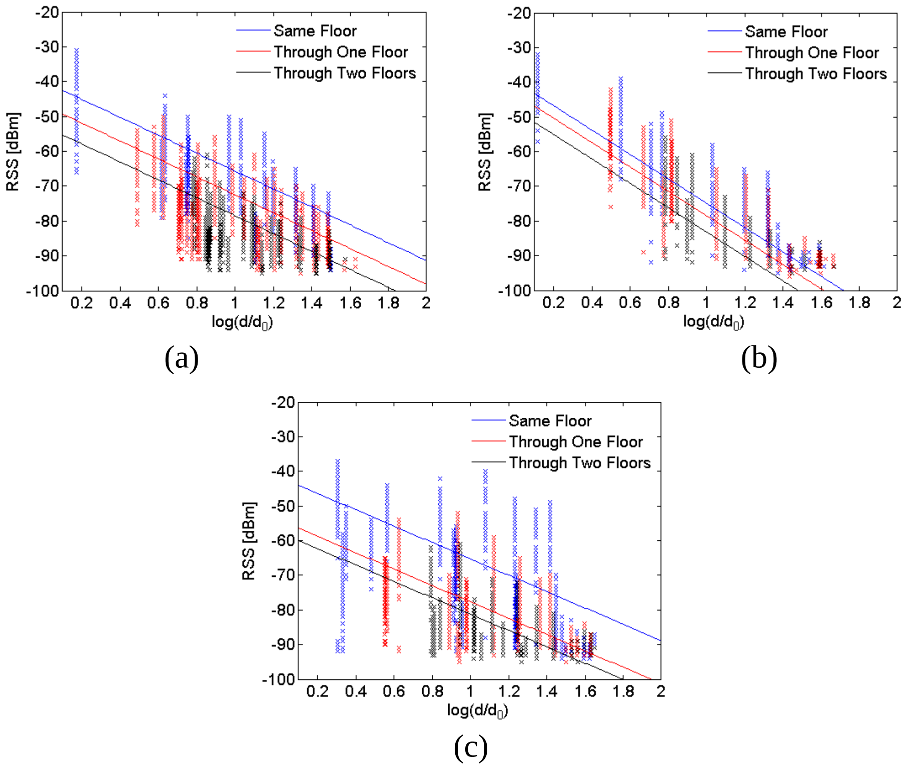


Figure 4.12. Path loss model of APs in Building C where (a) is the AP at Location (5.3, 13.6) in Floor 5, (b) is the AP at Location (32.6, 47.1) in Floor 6, and (c) is the AP at Location (5.9, 18.7) in Floor 7

4.9 Performance Metrics of Multi-Floor Localisation System

To evaluate performance of developed algorithms of multi-floor WLAN localisation system, the metrics used are accuracy, precision, and complexity. In fact, the accuracy and precision are mainly used to evaluate the algorithm is existing indoor localisation works. Other performance metrics that could be used is mentioned in Liu et al. (2007). The additional metrics are robustness, scalability, and cost. However, the comparison of robustness, scalability and cost are benchmarked for different type of technology used for localisation system e.g. WLAN, RFID, gyroscope, and sensor fusion. This thesis evaluates the performance of the algorithms presented in previous chapter using the three metrics i.e. accuracy, precision, and complexity because the comparison is made with other algorithms implementing similar technology – the WLAN.

For the algorithms proposed in Chapter 3, the performance of floor and horizontal localisation algorithms could be evaluated directly using the discussed metrics. However for AP selection algorithm, the algorithm is first paired with localisation algorithm to evaluate the performance of the algorithm. Similarly, the clustering algorithm is combined with floor localisation algorithm to determine the performance of the algorithm. The results presented in Chapter 5 to 9 are mainly according to the performance metrics discussed here. The method to evaluate the algorithms according to the metrics is outlined below.

4.9.1 Accuracy

The accuracy is measured by mean distance error or simply mean error (Liu et al. 2007). The mean error is the average Euclidean distance between the locations estimated by the algorithm and the actual measurement locations. Mathematically, the mean error is defined as follows:

$$Mean\ Error\ [m] = \frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\sqrt{\left|\mathrm{p}_{x\xi} - \hat{\mathrm{p}}_{x\xi}\right|^2 + \left|\mathrm{p}_{y\xi} - \hat{\mathrm{p}}_{y\xi}\right|^2} \quad (4.3)$$

where $\mathrm{p}_{x\xi}$ and $\hat{\mathrm{p}}_{x\xi}$ and $\mathrm{p}_{y\xi}$ and $\hat{\mathrm{p}}_{y\xi}$ are the actual and estimated x and y coordinate respectively of number ξ estimation of the total Ξ estimations. Each of the estimated location $\hat{\mathbf{p}} = (\hat{\mathrm{p}}_{x\xi}, \hat{\mathrm{p}}_{y\xi})$ is calculated according to the type of algorithm used. The related equations of $\hat{\mathbf{p}}$ could be found in Equation 3.4 for kNN algorithm, Equation 3.5 for univariate kernel and multivariate kernel algorithm, and Equation 3.34 for normal and kernel multi-class kNN algorithm. The lower mean error, the better the system. However mean error could potentially be biased as many of error differences may deviate from the mean. Therefore, to investigate the detail of the positioning capability, the precision metric is used.

For multi-floor case, the estimated locations will have additional floor level information given $\hat{\mathbf{p}} = \left(\hat{\mathrm{p}}_{x\xi}, \hat{\mathrm{p}}_{y\xi}, \hat{\mathrm{p}}_{z}\right)$ where $\hat{\mathrm{p}}_{z}$ is the estimated floor level given by Equation 3.26 for AKF algorithm and Equation 3.31 for KLF algorithm. The calculation of $\hat{\mathbf{p}}$ is similarly done according to single-floor case according to type of the localisation algorithm used but with multi-floor database. To measure the performance of algorithm in multi-floor, two extended metrics are included which are multi-floor mean error and floor accuracy. The multi-floor mean error calculates the average of Euclidean distance of estimated location and actual measurement considering floor level of both locations. The computation of multi-floor (MF) mean error is done as follows:

$MF\ Mean\ Error\ [m] =$

$$\frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\begin{cases}\sqrt{\left(\sqrt{\left|\mathrm{p}_{x\xi}-\hat{\mathrm{p}}_{x\xi}\right|^2+\left|\mathrm{p}_{y\xi}-\hat{\mathrm{p}}_{y\xi}\right|^2}\right)+|h_F|^2}, & F>0\\ \sqrt{\left|\mathrm{p}_{x\xi}-\hat{\mathrm{p}}_{x\xi}\right|^2+\left|\mathrm{p}_{y\xi}-\hat{\mathrm{p}}_{y\xi}\right|^2}, & F=0\end{cases} \tag{4.4}$$

where h_F is the height according to the number of difference of floor level between estimated and actual location. As mentioned in Section 4.3, the vertical height of between floors of the buildings is 3 m and thus $h_1 = 3$, $h_2 = 6$, $h_3 = 9$, and so on. If $F = 0$, there is no difference of floor level. On the other hand, the accuracy of the floor is calculated in terms of percentage of the difference between correctly estimated floors from the actual measured floors. The percentage could be calculated as follows:

$$Floor\ Accuracy\ [\%] = \frac{Number\ of\ correct\ floor\ estimated}{Number\ of\ estimation}\ \mathrm{x}\ 100 \tag{4.5}$$

4.9.2 Precision

The precision describes the detail of the estimation error compared to mean error which evaluates the error in average. Using precision metric, one could determine the percentage of probability at specific distance error. This means the variation of the error could be observed at specific distance. To determine the precision of algorithm, cumulative density function (CFD) is plotted. The CDF could be plotted according to:

$$CDF(\phi) = \frac{Counts\ of\ distance\ error \le \phi}{Counts\ of\ all\ distance\ estimation} \tag{4.6}$$

where x is the distance error. The distance error could be calculated for each test sample using Equation 4.3 without averaging term on the left side of the equation. For example, if the percentage of error at 2.5 m is 90 %, the CDF value is 0.9 for 2.5 m distance error. The CDF graph is plotted according to specified range of increasing ϕ e.g. given $\phi = [0{:}\,0.5{:}\,8]$. If the accuracies of two compared algorithms are similar, the algorithm of the CDF graph which attains high probability values quicker is chosen. This is because the distance error is concentrated in small error values indicate better precision. As suggested in Liu et al. (2007) , the algorithm is accessed at 90% and 95% of the CDF value to evaluate its precision.

From CDF plot, the percentage of distance error within fingerprint spacing could also be extracted. This means the CDF plot is read with the value of percentage when the distance reached the fingerprint spacing distance e.g. at 1 m or 2 m. The percentage of distance error within fingerprint spacing is defined as:

$$\%\ dist\ error\ within\ FP\ spacing = \frac{Counts\ of\ distance\ error \leq FP\ spacing}{Counts\ of\ all\ distance\ estimation} \times 100\% \tag{4.7}$$

The algorithm that achieves higher percentage of distance error within the investigated fingerprint spacing is the more precise algorithm.

For floor localisation as discussed in Section 3.8, the CDF is plotted to investigate the precision of predicted floor. The precision of predicted floor is gives as the CDF of the floor errors which is calculated follows:

$$CDF_{Floor}(\zeta) = \frac{Floor\ error\ counts \leq \zeta}{All\ floor\ estimation\ counts} \tag{4.8}$$

where ζ is the floor error . For instance, if the algorithm estimates the test sample to be on Floor 4 but actual location of the sample is on Floor 1, then the floors errors is equal to 3. The CDF at floor errors of 3 is calculated by summing the entire floor errors that are equal to 3 or below and divided by the number of estimation. The CDF at floor errors of zero indicates the floor accuracy which is similarly expressed as Equation 4.5 where $Floor\ Accuracy\ = CDF_{Floor}(0)$.

4.9.3 Computational Complexity

The computational complexity of the algorithm is evaluated according to processing time of the algorithm to produce the result. The algorithm is accessed according to its ability to perform localisation by lowest processing time. The speed of computing is dependent on the how the algorithm processes the database. Introduction of compressed database may help in reducing the computation complexity of the algorithm. A better system has the algorithm that requires lower processing time.

4.10 Choosing Value of k in kNN Localisation Algorithm

The classical kNN algorithm presented in Chapter 3 requires the value of k evaluated by experiments in order to have the best version of the algorithm. To determine the best k, the kNN localisation algorithm is tested using the fingerprint dataset of three buildings and the additional dataset by varying the k from 1 to 10. Figure 4.13 shows the mean distance error results for different k in every building.

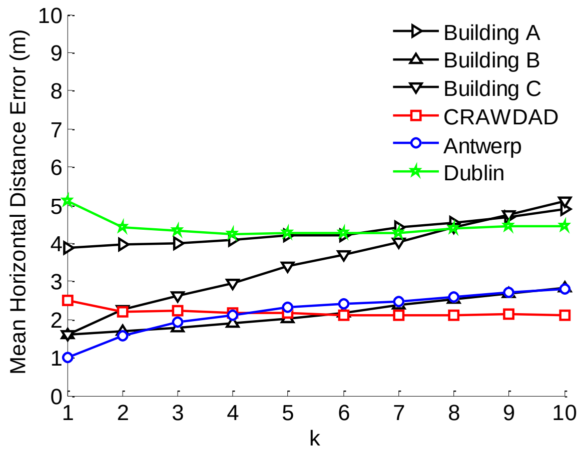


Figure 4.13. Mean distance error plot of different k values of kNN algorithm in different buildings

The result shows that for multi-floor environments (Building A, B, C and Antwerp dataset), the best k is 1. Contrarily in CRAWDAD data set where the fingerprint is collected in single floor environment, the best k is 6. This application k in multi-floor environment has yet to be reported in previous works. It is in agreement with existing approach in multi-floor case that mainly utilise nearest neighbour algorithm (1-NN) rather than using k value above one (Gansemer et al. 2009, Razavi, Valkama, and Lohan 2015). On the other hand, for single floor environment existing works reported that the best k is either 3 or 4. More k is needed ($k = 6$) for CRAWDAD environment could be related to placement of APs and the open-space scenario. Furthermore to confirm the result, it has been reported in original publication (King et al. 2006) which implements the randomised CRAWDAD data, the mean error of positioning by kNN using best k is 2.2 m which is comparable to this result which is 2.1 m. In Dublin, similar scenario as CRAWDAD are observed and ($k = 4$) is used to achieve lowest mean error.

4.11 Summary

This chapter presents the measurement procedure to collect fingerprint data and related methods. The measurement tool is based on smartphone's WLAN receiver and Airplace Android application. The computation of algorithms is done using MATLAB software and an Intel's i5 desktop computer. The measurement locations consists of three multi-storey buildings of different heights of five, eleven, and eight located in UPM's campus. Additional fingerprint datasets of single-floor environment to test the performance of the algorithms are also presented which are CRAWDAD, Antwerp, and Dublin. The CRAWDAD is a downloaded dataset measured by researcher in University of Mannheim, Germany while Antwerp and Dublin datasets are measured in Bell Labs buildings in the respective locations. The radio map database is served as the input to the proposed algorithms for multi-floor localisation which involves AP selection, floor localisation, and horizontal localisation. The method to extract the path loss parameters from the measured signal to check the propagation characteristic of the signals is also explained. The statistic of the radio maps database has shown the measured signal strength data is suitable for development of multi-floor localisation in the environment. Additionally, the extracted path loss model parameters according to Seidel's model has shown that variation of signal strength is observed across horizontal and vertical locations of the building which is one of the important factor to distinguish different fingerprint locations. Metrics to describe the performance of the developed localisation algorithm which consists of accuracy, precision, and computational complexity for both horizontal and multi-floor locations are explained in detail. Lastly, the method to choose value of k for classical kNN localisation algorithm by finding the lowest mean error at a range of k values is described.

CHAPTER 5

RESULTS AND DISCUSSION I: PERFORMANCE OF MAX KERNEL AND KERNEL LOGISTIC PAIRWISE DISCRIMINANT AP SELECTION ALGORITHMS

5.1 Introduction

Previous chapter discussed the extraction of path loss parameters of the measured fingerprint samples to ensure that it can be used for testing the algorithms for localisation. In this chapter, the measured fingerprint samples integrated as radio map database is optimised according to two proposed AP selection algorithms namely Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD). The AP selection algorithms will optimise each of the fingerprint samples by finding the right combination of APs by utilising lower number of APs as explained in Section 2.4.1. The output of the AP selection algorithm is the optimised radio map fingerprint database containing lower number of APs at every fingerprint element. Therefore to test whether the optimised radio map fingerprint could produce similar location error as non-optimised radio map, the radio map is paired with a localisation algorithm to obtain the results. Here, three localisation described in Section 3.3 is used which are univariate kernel, multivariate kernel and kNN. The distance and vertical accuracy in the determination of the number of APs proposed Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD) AP selection algorithms in Section 3.7 are evaluated through comparison with two commonly used AP selection techniques: the Max Mean (Youssef, Agrawala, Shankar, & Noh, 2002), and InfoGain (Chen, Yan, Yin, & Chai, 2006). Description of Max Mean and InfoGain AP selection algorithm and the differences of the techniques compared to proposed algorithms are described in Appendix C1. The following sections describe the performance of the AP selection algorithms in details.

5.2 Accuracy of Localisation of Different Number of Selected APs By Pairing Max Kernel and KLPD with Univariate Kernel, Multivariate Kernel and *k*NN Localisation Algorithms

The most important criterion in accessing the performance of AP selection algorithm is to determine if the algorithm is able to produce the lowest distance error in the degree of independence of number of APs. The results of distance error are presented in details in terms of mean distance error based on both the single-floor and multi-floor locations. The single-floor performance of multi-floor dataset (Buildings A, B, and C) corresponds to similar floor level of test (online) sample and measured fingerprint (offline) samples (refer to Section 3.2) inside the building. On the other hand, the multi-floor performance considers processing the offline samples of all locations in all floors. The performance of single-floor and multi-floor of univariate kernel, multivariate kernel and kNN is described in Section 5.2.1 and 5.2.2, 5.2.3 and 5.2.4, and 5.2.5 and 5.2.6 respectively.

5.2.1 Single-floor Performance Using Univariate Kernel Algorithm

Figure 5.1(a) to 5.1(e) compare the plot of mean error versus number of selected APs using the proposed Max Kernel and Kernel Logistic Pairwise Discriminant (KLPD) AP selection algorithms with the Max Mean and InfoGain algorithms. Two types of results are investigated for single-floor performance i.e. the mean distance error and percentage of accuracy within fingerprint spacing which are calculated using Equation 4.3 and 4.7 respectively.

As can be seen in Figure 5.1, the proposed algorithms shows excellent performance compared to existing AP selection algorithms except for Dublin environment. In Building A, B, C and Antwerp (Figure 5.1(a), 5.1(b), 5.1(c), and 5.1(e)), the proposed Max Kernel is much better compared to Max Mean and InfoGain algorithm. All of these environments are multi-floor including Antwerp. As mentioned in Chapter 4, Antwerp measurement is done on seventh floor of the building. In Building A, the Max Kernel algorithm found that six selected APs contains the most information for positioning compared to Max Mean and InfoGain which requires 14 APs to achieve lower positioning error. Similarly in Building C, Max Kernel only requires 13 APs to obtain good positioning accuracy compared to 25 and 29 APs required by Max Mean and InfoGain algorithms respectively. In Building B, all algorithms give almost similar performance due to presence of the open spaces between the floors which causes multiple AP combinations could be used to achieve similar positioning accuracy. In Antwerp, the Max Mean and InfoGain achieve nearly accuracy of Max Kernel algorithm by using lower subset of APs (6 and 12 APs respectively) due to presence of many APs installed within the floor which improves the signal strength level. Moreover, Max Kernel algorithm also performs better compared to KLPD algorithm in multi-floor environments and this shows that using probability information is better compared to using discriminative AP information in multi-floor buildings. In all multi-floor buildings Max Kernel algorithm saturates at 9, 20, 27 and 20 APs in Building A, B, C and Antwerp respectively where the lowest mean error is achieved.

In single-floor CRAWDAD environment (Figure 5.1(d)), the Max Mean and InfoGain algorithms quickly reduce the mean error at small subset of APs. However, the lowest mean errors are achieved at 11 APs by both algorithms compared to mean errors of proposed Max Kernel and KLPD algorithms which are achieved at 10 APs. Additionally, the mean errors of both proposed algorithms does not deviates too far from the error achieved by Max Mean and InfoGain algorithms at low subset of APs compared to the error obtained by Max Mean and InfoGain algorithms at same small subset of APs in multi-floor environments. The accuracy of Max Mean and InfoGain in CRAWDAD environment could be similarly explained as Antwerp environment where strong AP signals presence due to multiple placements of the APs within the floor. Additionally, the performance of the algorithms is helped by absence of multi-floor signals from other floor levels.

Poor performance in single-floor Dublin environment (Figure 5.1(f)) could be described by inadequate RSS information. As mentioned in Chapter 4, Dublin datasets only contains RSS values of $T = 4$ at each fingerprint position compared to other datasets which use $T = 100$ of RSS at each fingerprint position. The poor performance of Max Kernel algorithm is expected in this case as Max Kernel utilises probability distribution of the signal. It is because statistical information of the maximum probability information is improperly interpreted due to low number of signal to estimate the probability density. This leads to the information of probability distribution marginally matches the collected online measured RSS during localisation. On the other hand, KLPD algorithm which uses information of all distributed RSS value offers better results compared to Max Kernel especially at low subset of APs e.g. 16 APs and below. This is because at these subsets of APs, the subsets contain most discriminative pairwise AP signals for localisation. MaxMean and InfoGain perform better because the algorithms use average value of RSS and therefore the algorithms is more robust even if only small distribution of RSS exists.

It could be noticed that in certain plots e.g. in CRAWDAD and Dublin there are “bump” effects by certain algorithms. In CRAWDAD, it is seen at 6 AP selections in Max Kernel. In Dublin, it is seen at 22 to 32 selected APs in KLPD, at 24 APs in Max Mean, and at 25 to 33 APs in InfoGain. This means at the selected APs number, the introduction of additional AP to the subset decrease the accuracy of the localisation. This emphasises that only at certain proper selected AP numbers the AP selection algorithm could achieve the highest accuracy as possible. The effects is also could be seen in other figures in later sections.

Extending the results in Figure 5.1, the accuracy of location estimation that locates within fingerprint spacing using different number of AP subset is studied. The fingerprint spacing for all environments is 2 m except CRAWDAD which is 1 m. The plots of the result are given in Figure 5.2. The algorithm is better if the percentage of accuracy is higher. The figure shows that accuracy gradually rises to the peak as the subset of APs increases. The graphs show that the results are almost in agreement with the results presented in Figure 5.1. In Building A and C, the Max Kernel and KLPD algorithms outperforms the Max Mean and InfoGain in all investigated subset of APs. The highest accuracy obtained by Max Kernel algorithm in Building A and C is at 9 and 27 APs respectively. The KLPD algorithm reaches highest accuracy at 11 and 27 APs respectively. This shows reduction of 62.5% and 54.2% (15 and 13 out of 24) of required APs by implementing Max Kernel and KLPD algorithm respectively in Building A and 51.8% (29 out of 56) in Building C by using both algorithms compared to using all detected APs on the floor of the building. On the other hand, the accuracy of Max Mean and InfoGain algorithm in

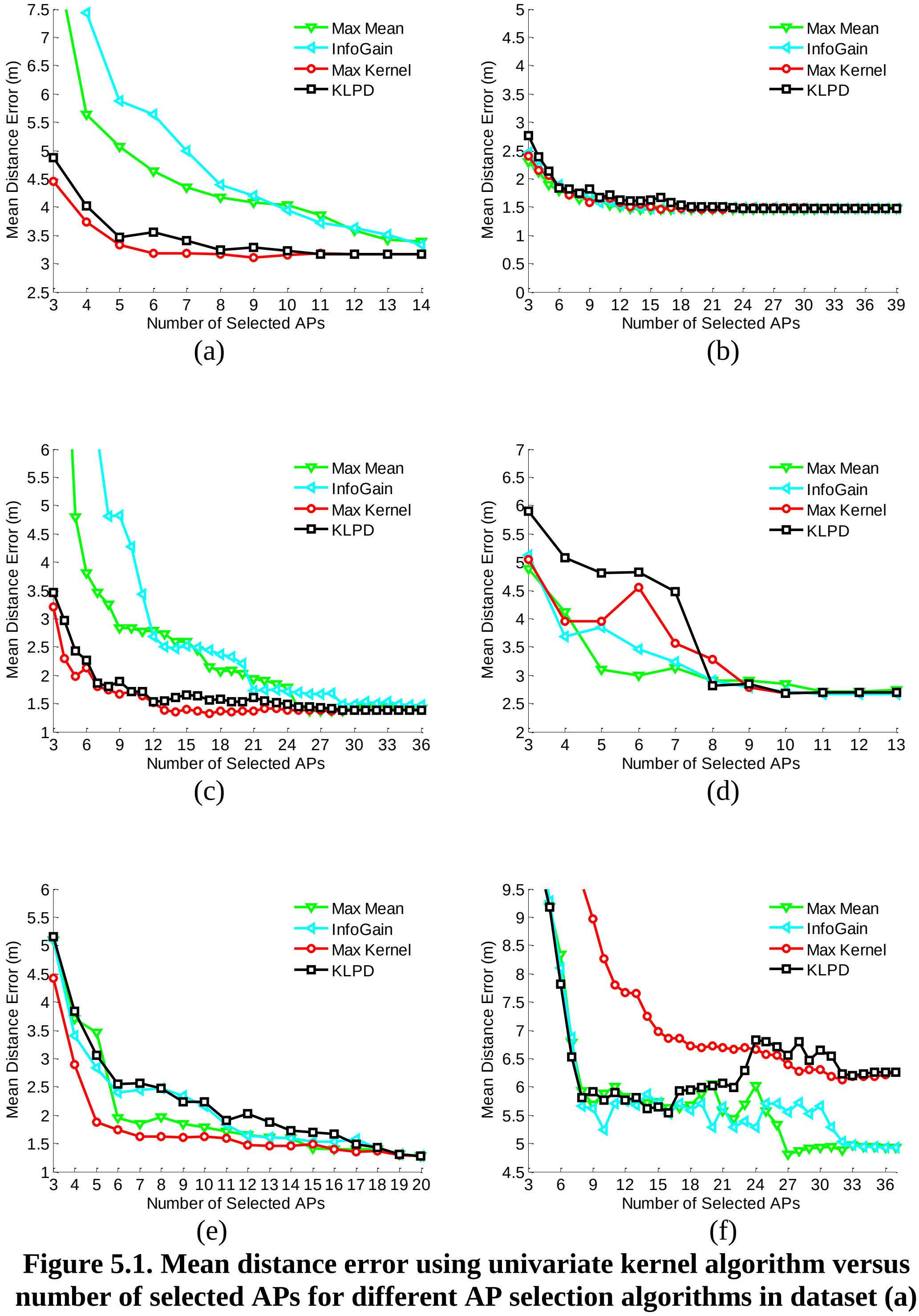


Figure 5.1. Mean distance error using univariate kernel algorithm versus number of selected APs for different AP selection algorithms in dataset (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

Building B (Figure 5.2(b)) increases compared to proposed algorithms at subset of 14 to 32 APs. However the results are still comparable for most subset of APs. In Antwerp, the results are in agreement with Figure 5.1(e) where the highest accuracy is obtained by Max Kernel, and followed by both Max Mean and InfoGain, and lastly by KLPD algorithm. The results in Dublin dataset also shows similar tendency, the highest accuracy is demonstrated by Max Mean and InfoGain algorithms and is followed by KLPD and Max Kernel algorithms. In CRAWDAD, both Max Kernel and KLPD algorithms achieve the highest accuracy using 9 APs subset, while Max Mean and InfoGain obtain highest accuracy at lower subset of APs which are 5 and 8 APs respectively, but at lower percentage.

5.2.2 Multi-floor Performance Using Univariate Kernel Algorithm

Multi-floor accuracy is investigated for Building A, B, and C as illustrated in Figure 5.3 (a), (b), and (c) respectively. The results are described in terms of mean distance error (Equation 4.4), percentage of error within fingerprint spacing (Equation 4.7), and floor accuracy (Equation 4.5). Compared to single-floor results in Section 5.2.1, additional accuracy measure of floor accuracy is investigated in multi-floor building because accurate floor determination reflects the results of multi-floor mean distance error. From the figure, it can be observed that the mean error of all tested AP selection algorithms increases at every subset of APs as a result of introduction of additional floor dataset. The figure also suggests that the accuracy using selected APs by Max Mean and InfoGain are much more influenced by multi-floor dataset compared to proposed Max Kernel and KLPD algorithms. In Building A, for example at 8 APs subset, the mean error of both Max Kernel and KLPD algorithms

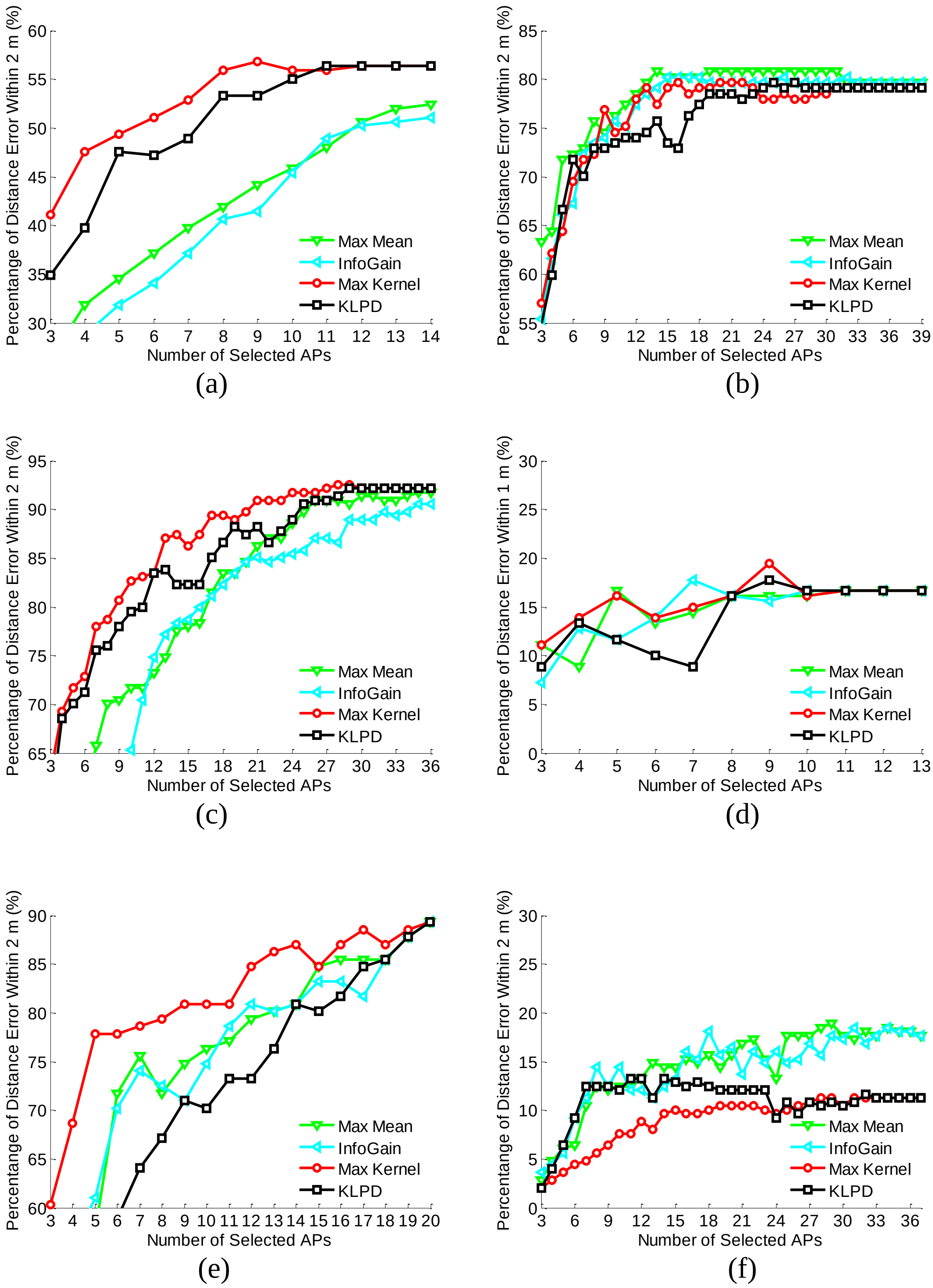


Figure 5.2. Percentage of distance errors within fingerprint spaces (two meters for Building A, B, C and Antwerp and one meter for CRAWDAD and Dublin dataset) using univariate kernel algorithm versus number of selected APs in dataset (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

only increases 15.6% (3.2 m to 3.7 m) compared to Max Mean and InfoGain algorithms which the mean error increases 35.7% (4.2 m to 5.7 m) and 34.1% (4.4 m to 5.9 m) respectively. The effect could also similarly be seen in Building B (Figure 5.3(b)) and C (Figure 5.3(c)). In single-floor performance of Building B which is discussed in Section 5.2.1, all of the tested AP selection algorithms perform almost equally in all subset of APs. However with introduction of multi-floor dataset, the performance of existing AP selection algorithms gets worse compared to proposed algorithms especially between selections of 5 to 32 APs. The results show that the proposed algorithms could perform well with presence of noise due to additional fingerprint entries which contains multi-floor AP signals.

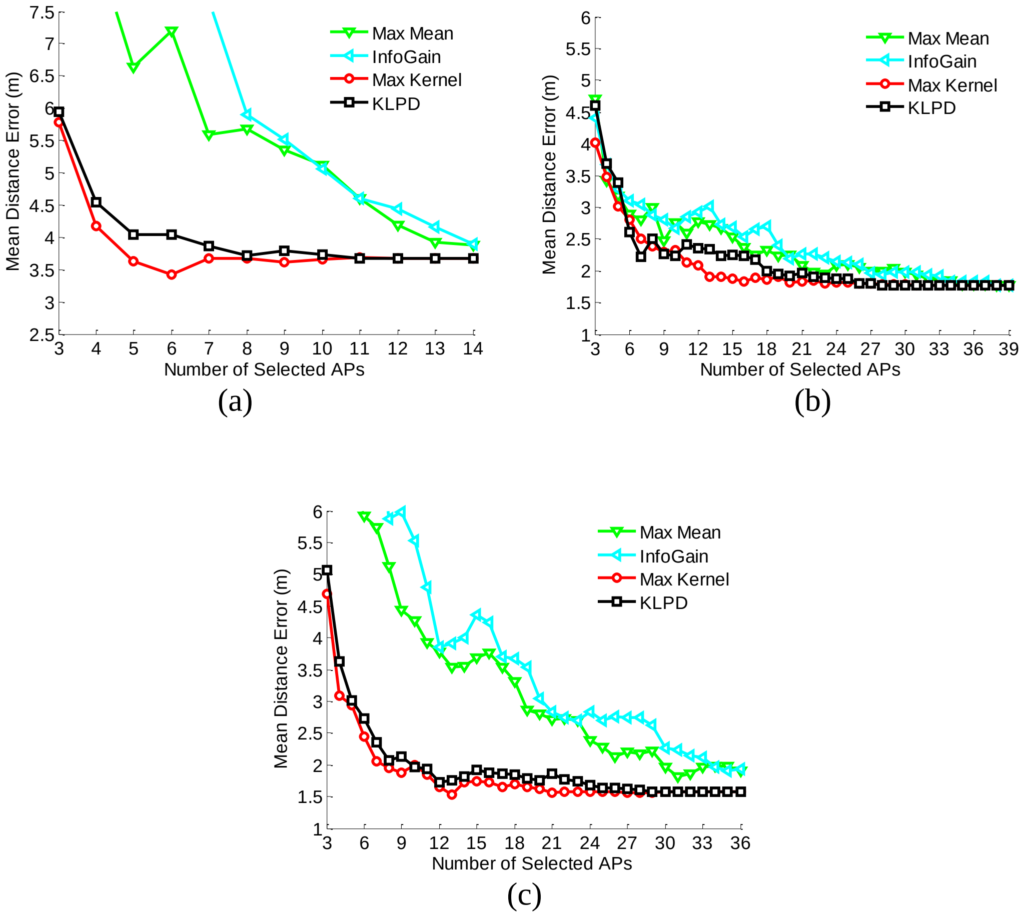


Figure 5.3. Mean of multi-floor distance error using univariate kernel algorithm in multi-floor building environment versus number of selected APs for (a) Building A, (b) Building B, and (c) Building C

Similar trends is also seen in terms of estimated location accuracy within fingerprint spacing in all building as depicted in Figure 5.4(a) to 5.4(c). In all buildings, the peak is reached quicker at lower subset of APs for both proposed AP selection algorithms compared to Max Mean and InfoGain. Moreover, in Building A and C the Max Mean and InfoGain algorithm could not matched the accuracy obtained by the proposed algorithms as the selection of AP increases. In agreement with results presented in Figure 5.3, the Max Kernel algorithm performs better in most selection of APs in all buildings. In Building B however the highest accuracy of KLPD algorithm is better than Max Kernel which can be seen at 28 AP selections.

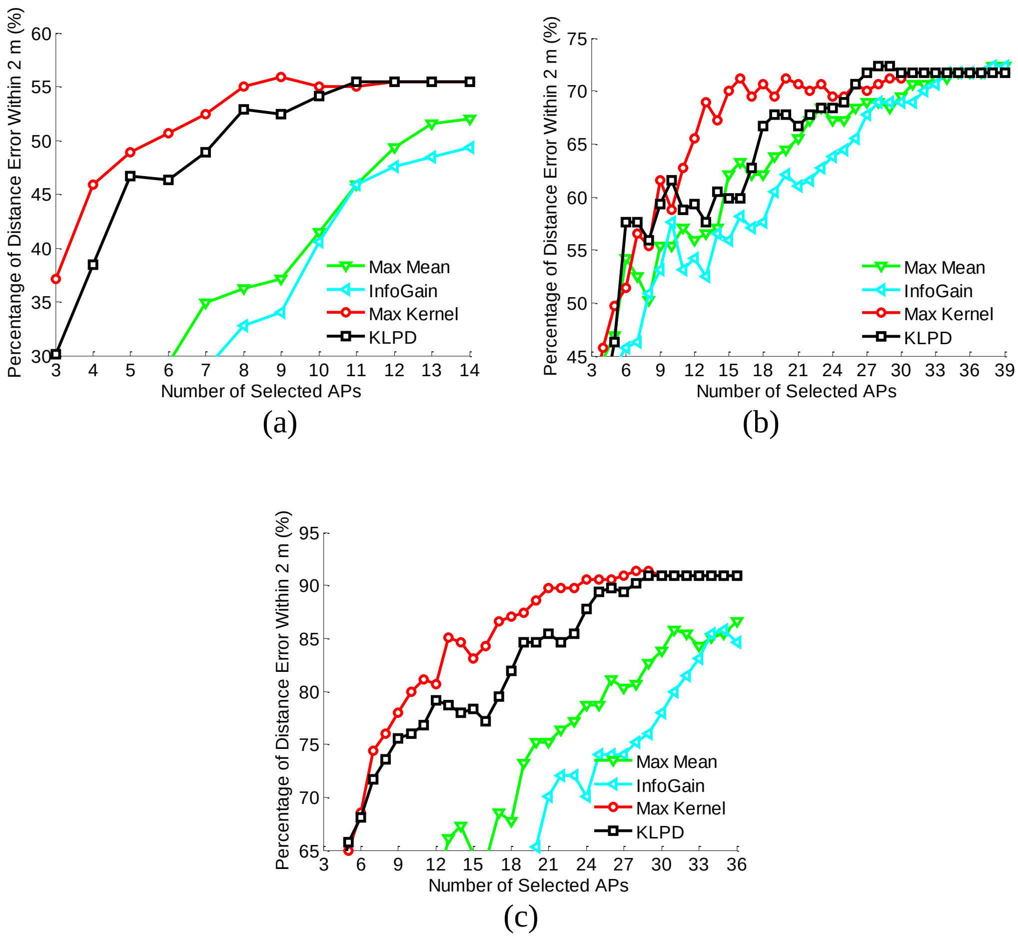


Figure 5.4. Percentage of multi-floor location accuracy using univariate kernel algorithm within fingerprint spacing versus number of selected APs for (a) Building A, (b) Building B, and (c) Building C

The multi-floor performance of the algorithms is also characterised by accuracy of the correct floor level. Figure 5.5 plots the result of floor accuracy versus number of selected APs in Building A, B, and C. It can be seen that both proposed algorithms show excellent performance in all buildings. By selecting only 6 APs in Building A, the Max Kernel algorithm obtains highest floor accuracy of 96.9% compared to other algorithms. This is followed by KLPD algorithm which gives 95.6% of floor accuracy using 10 APs subset. Similar floor accuracy is obtained by Max Mean algorithm only at 13 selected APs. The performance of the proposed algorithms in Building B is better compared to existing AP selection algorithms. The KLPD algorithms perform well compared to other algorithms by showing 91.5% accurate floor level estimation using subset of 28 APs. At similar subset of APs, Max Kernel algorithm also gives high floor accuracy of 91.0%. To achieve similar performance as Max Kernel, the Max Mean and InfoGain require 38 APs which is addition of 40.7% more APs. In Building C, the performance of proposed algorithms could not be matched by Max Mean and InfoGain algorithms even though after selecting 36 APs which is the maximum total number of APs present when estimation the location without implementing AP selection algorithm. Additionally, it could be observed that the floor accuracy of both proposed algorithms is almost stable even at low subset of APs especially in Building A and C which show robustness of the algorithm in estimating floor level of the building. The performance of the algorithms in Building B is different because existence of open spaces between floors raising uncertainties in estimating correct floor level using low subset of APs. Additionally, it could be noted that the selected APs to determine best floor accuracy is in parallel with the results achieved for multi-floor mean error as discussed above. For example, by selection of 6 APs for Max Kernel, the best mean error and floor accuracy is obtained as can be notified in Figure 5.3(a) and 5.5(a) respectively. The results in all tested buildings indicates that using Max Kernel and KLPD for determining the floor level requires less AP dimensions compared to existing AP selection approaches.

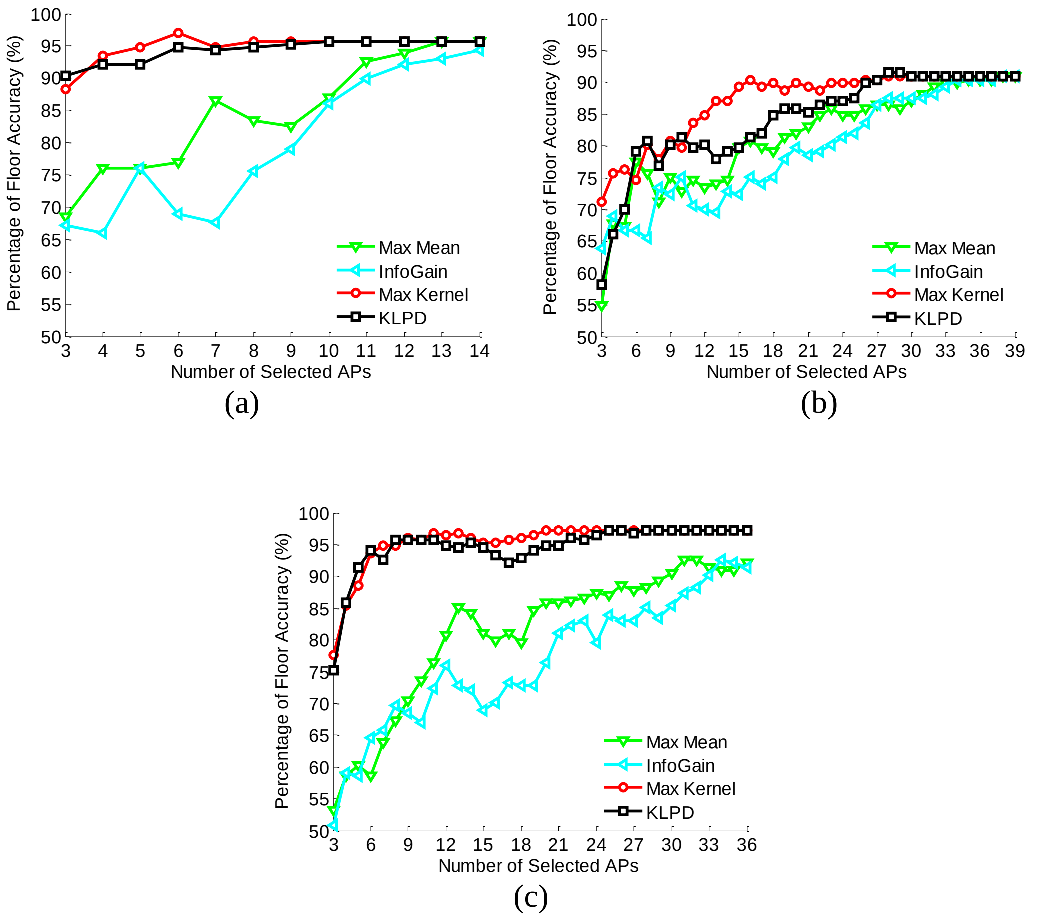


Figure 5.5. Floor accuracy of multi-floor building using univariate kernel algorithm versus number of APs of AP selection algorithms in (a) Building A, (b) Building B, and (c) Building C

5.2.3 Single-Floor Performance Using Multivariate Kernel Algorithm

The performance of the AP selection algorithm is also tested with multivariate kernel localisation algorithm. As discussed in Chapter 3, the multivariate kernel algorithm is extension of univariate kernel in multi-dimensional problem. Similarly as univariate kernel algorithm (Section 5.2.1 and 5.2.2), the results is discussed in terms of single-floor and multi-floor performance. This section discusses the performance of single-floor and the next section describes the multi-floor performance of AP selection techniques utilising multivariate kernel algorithm. Figure 5.6 shows the results of mean error at every subset of APs using using AP selection algorithms with multivariate kernel localisation algorithm in all tested buildings. Comparing with the result of univariate kernel, the multivariate kernel offers better accuracy in almost all environments. The performance of all AP selection algorithms in all environments is almost similar. The result of Max Kernel algorithm is also improved mainly in single-floor dataset i.e. CRAWDAD and Dublin. For example, at 3 AP selection in CRAWDAD, the Max Kernel could obtain the lowest mean error compared to other algorithhms. In Dublin dataset, the accuracy of Max Kernel is better as mean error at many subset of APs are closer to Max Mean and InfoGain. Otherwise, the result of KLPD algorithm is mostly similar to the results obtained in univariate kernel.

Figure 5.7 shows the results of accuracy within fingerprint spacing of AP selection algorithms in all tested dataset. All of the plotted results almost reflect the results of mean error which was discussed above. However, It could be highlighted that in CRAWDAD dataset the performance of proposed algorithms are better compared to Max Mean and InfoGain. This shows improvement of the accuracy between proposed and existing algorithms compared to their performance in univariate kernel. The KLPD algorithm attains the highest accuracy within 1 m location of 20.0% at subset of 8 APs. Also it could be noticed that in Building B the accuracy of Max Mean and InfoGain within 3 to 20 subset of APs are better than the proposed algorithm even though almost similar mean error is obtained by all algorithms as shown in Figure 5.7(b).

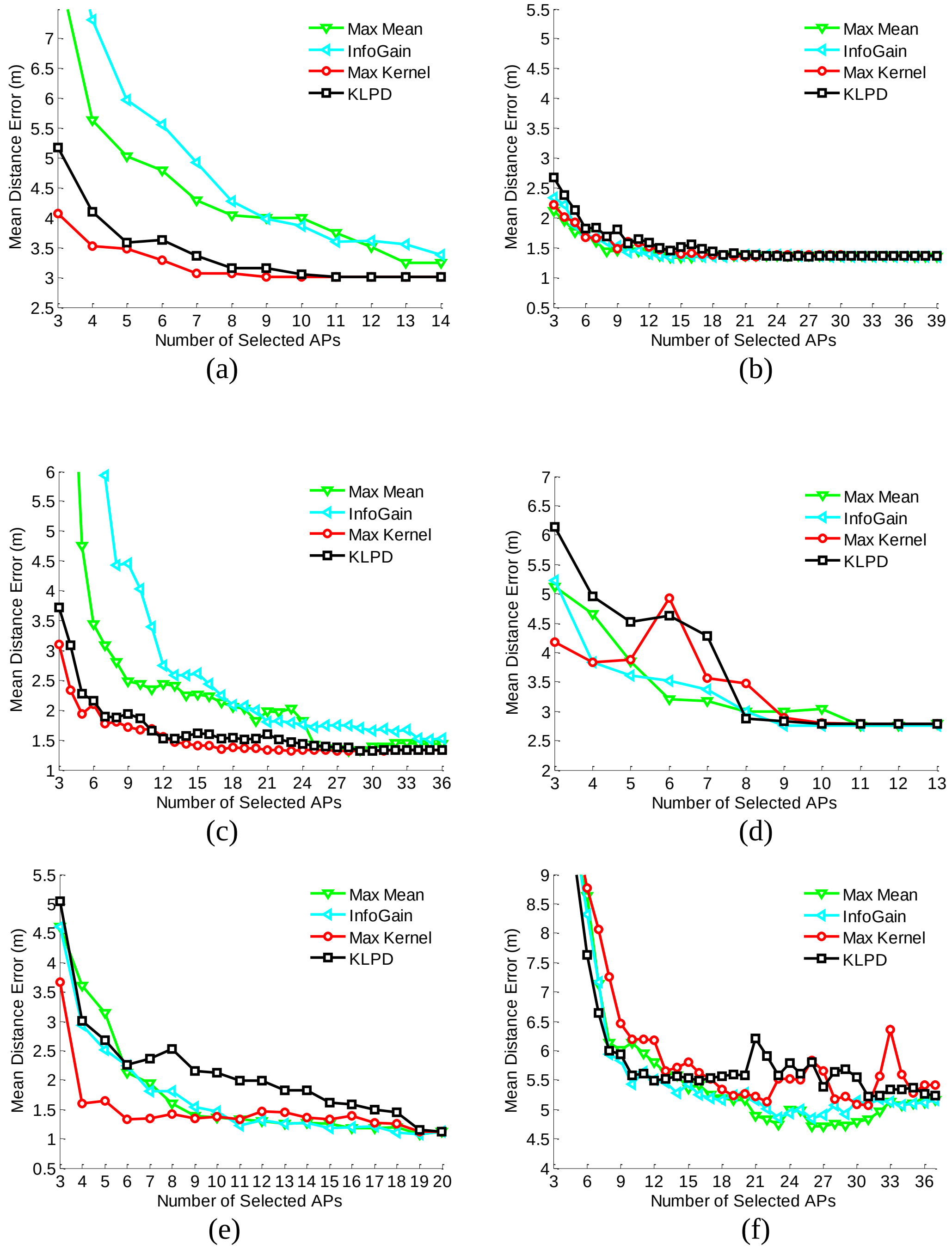


Figure 5.6. Mean error versus number of selected APs for single-floor localisation using different AP selection algorithms paired with multivariate kernel algorithm in (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

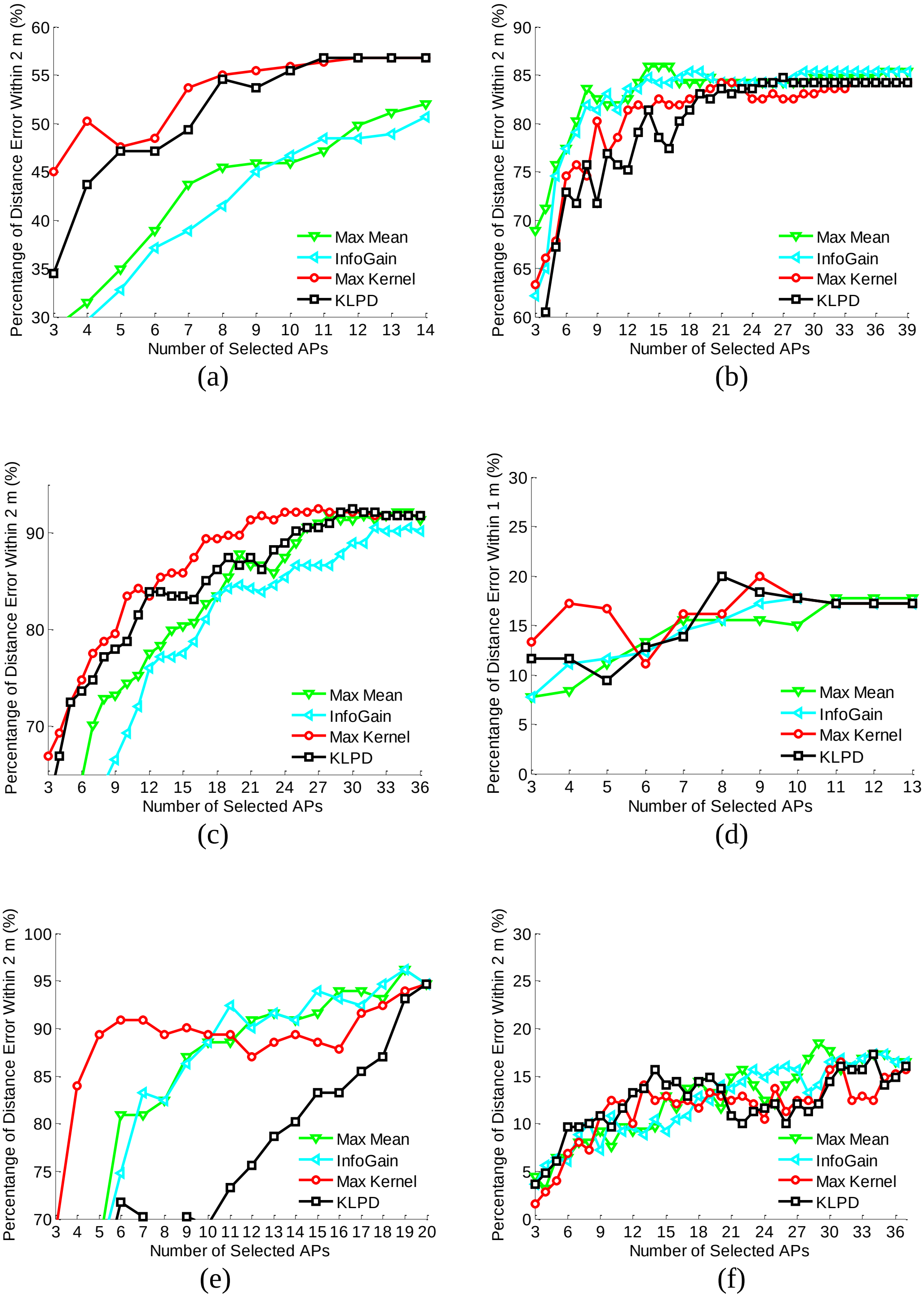


Figure 5.7. Accuracy within fingerprint spacing versus number of selected APs combined with multivariate kernel localisation algorithm in multi-floor environments of (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

5.2.4 Multi-Floor Performance Using Multivariate Kernel Algorithm

The multifloor performance of AP selection algorithms paired with multivariate kernel algorithm is also examined. Figure 5.8 illustrates the mean error of each AP selection algorithm versus the number of selected AP of multi-floor dataset in Building A, B, and C. The results show the proposed algorithm still outperforms Max Mean and InfoGain algorithms in all buildings. There is not much difference seen between the performance of proposed algorithms and existing algorithms in all buildings compared to using univariate kernel and kNN algorithm as the backend algorithm.

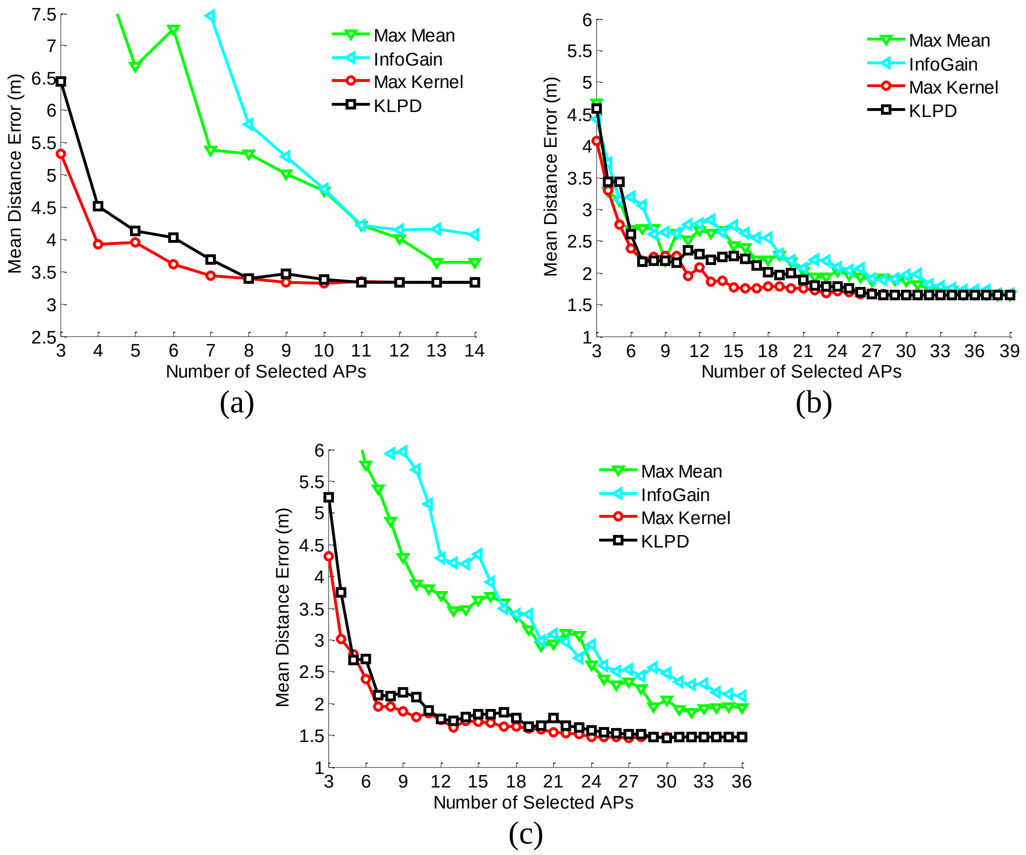


Figure 5.8. Mean error versus number of selected APs for multi-floor localisation using different AP selection algorithms paired with multivariate kernel algorithm in (a) Building A, (b) Building B, and (c) Building C

Figure 5.9 presents the accuracy of estimated location between fingerprint spacing at varying selected APs. It could be observed that the error difference between proposed and existing AP selection algorithms increases. For example, Max Mean requires 11 APs to match the accuracy of Max Kernel algorithm at 4 APs. However, it only requires 8 and 9 APs in kNN and univariate kernel respectively to match similar subset of AP of Max Kernel.

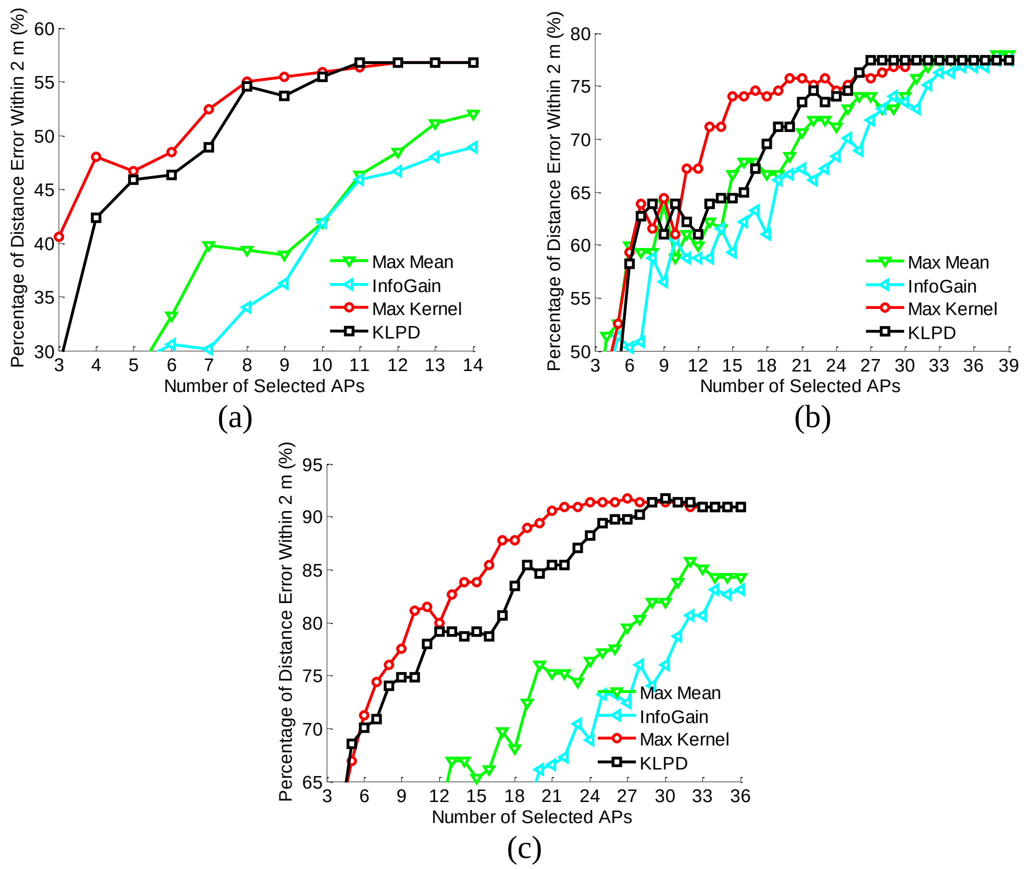


Figure 5.9. Accuracy within fingerprint spacing versus number of selected APs of AP selection algorithms paired with multivariate kernel algorithm of multi-floor settings in (a) Building A, (b) Building B, and (c) Building C

The results of floor accuracy in Figure 5.10 also show advantage of the proposed algorithms. It can be seen that in Building A and C the graph is almost ‘flat’ for large range of selected APs for both Max Kernel and KLPD algorithms. This means both algorithms already achieve high floor accuracy at low number of selected APs. The results show that both algorithms could perform better if multivariate kernel is used as the localisation algorithm.

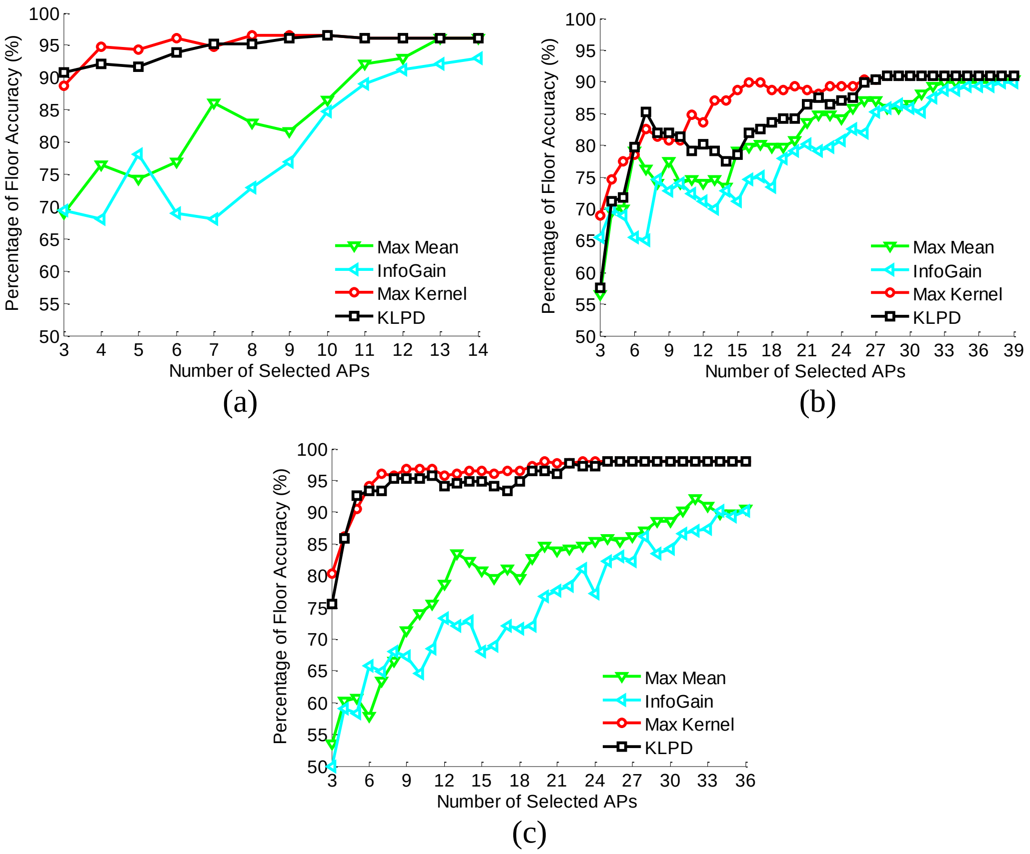


Figure 5.10. Floor accuracy versus number of selected APs of AP selection algorithms paired with multivariate kernel algorithm of multi-floor settings in (a) Building A, (b) Building B, and (c) Building C

5.2.5 Single-floor Performance Using *k*NN algorithm

The performance of AP selection technique paired with kNN localisation algorithm is also investigated. Similar to univariate kernel and multivariate kernel algorithm as discussed in Section 5.2.1 to 5.2.4, the result is described in terms of single (this section) and multi-floor (next section). Figure 5.11 shows the accuracy of every AP selection algorithms paired with *k*NN algorithm in different buildings at different selection of APs. The results shows almost similar trend as univiariate kernel in

Building A, B, C and CRAWDAD. However, for Antwerp and Dublin datasets, the proposed AP selection underperforms compared to existing localistion algorithm. This is mainly due to the way of kNN algorithm works which is based on considering average signal of each AP at each fingerprint location which gives advantage to Max Mean and InfoGain which is based on the average signal of the AP. On the other hand, the proposed algorithm works well with probabilistic algorithm e.g. univariate kernel as presented in previous sections because the theory of the algorithm also relies on probability calculation. However, the superior performance of the proposed algorithms at lower subset of APs in Building A, B, and C could also be explained by the presence of weak AP signals at multiple fingerprint locations exceeds strong AP signals in those environments which gives less advantage to Max Mean and InfoGain algorithms in those environment.

The percentage of accuracy within fingerprint spacing for every algorithm is also plotted in all environments as shown in Figure 5.12. The results shows highest accuracy could be obtained in Building A, B, C, and CRAWDAD. In Building A, B, and C, 41.5%, 79.7%, 88.6% accuracy is achieved by Max Kernel at 9, 19, and 22 APs selection. This shows that in Building B the performance of Max Kernel is better especially within 2 m accuracy compared to other algorithms even though generally the mean error is almost similar for all algorithms as could be observed in Figure 5.12. Interestingly, KLPD algorithm could achieve highest accuracy of 27.8% with 9 subset of APs in CRAWDAD compared to other algorithm despite low estimation accuracy at lower subset of APs compared to Max Mean and InfoGain. This is in parallel with the lowest mean error obtained by KLPD algorithm in similar dataset in Figure 5.12(d). Also as expected, the results in Antwerp and Dublin shows poor performance of the proposed algorithm which is in agreement with the result of mean error

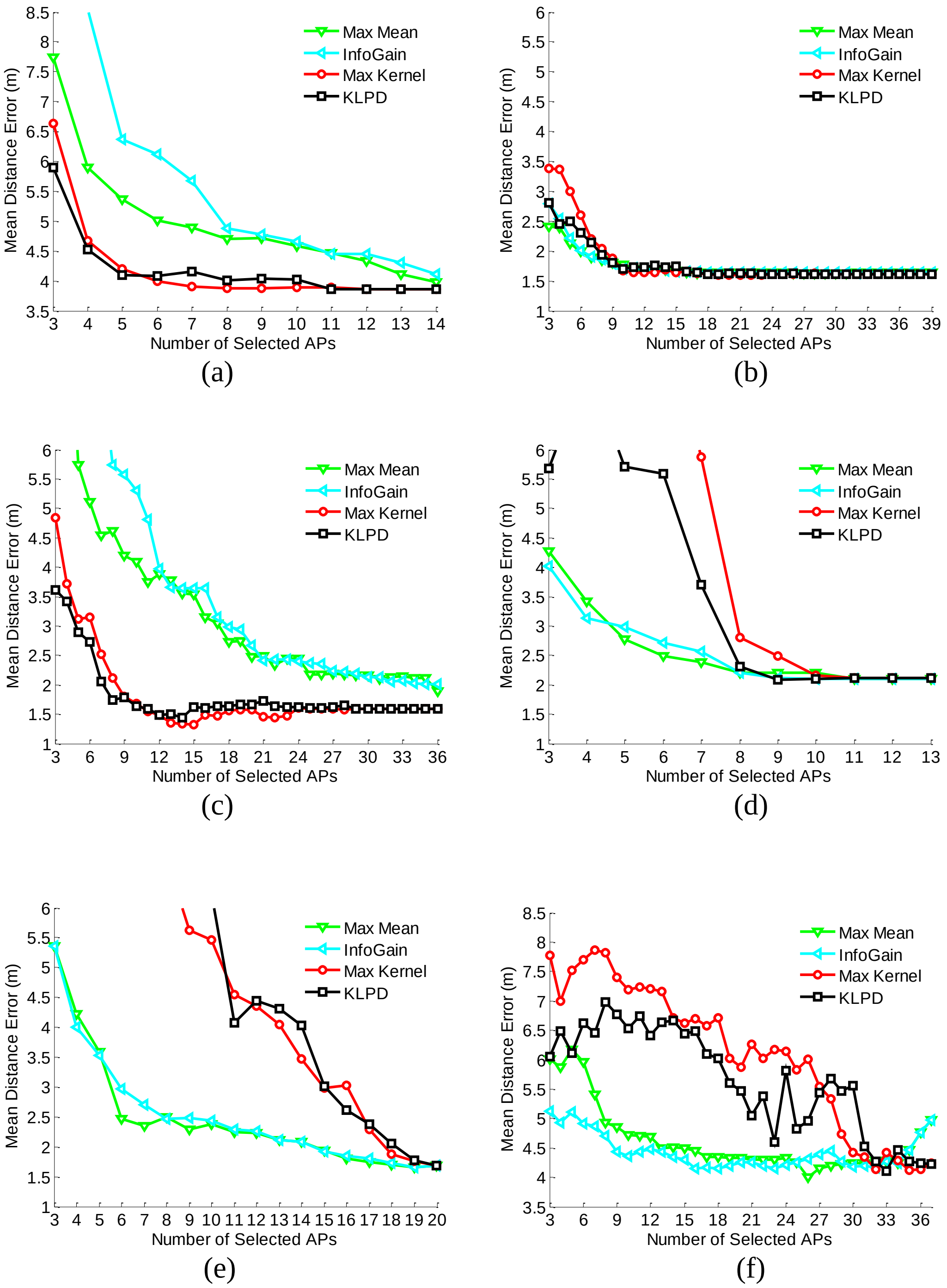


Figure 5.11. Mean error versus number of selected APs for single-floor localisation using different AP selection algorithms paired with kNN algorithm in (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

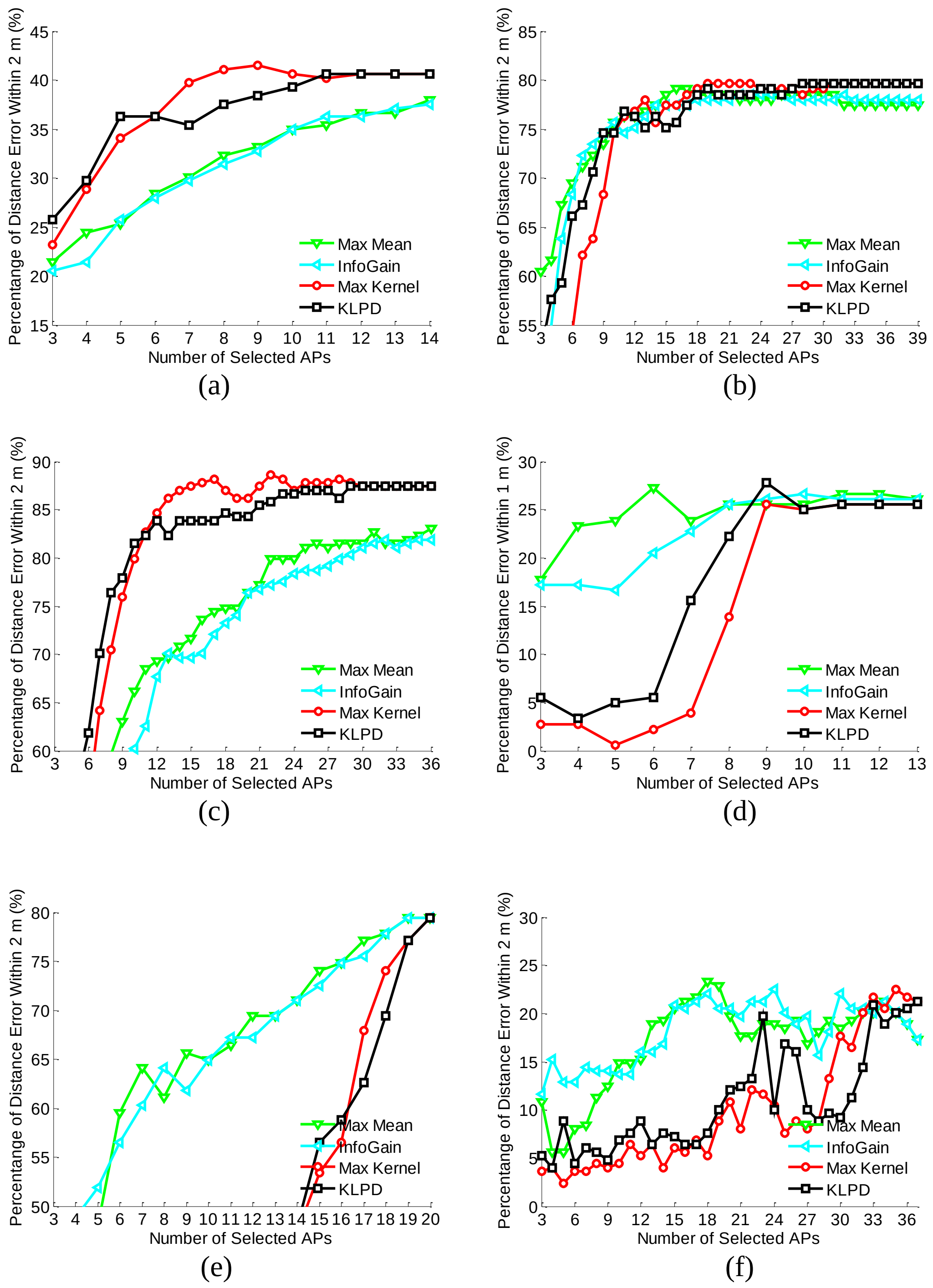
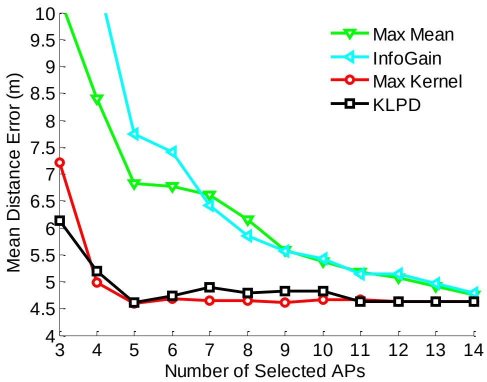


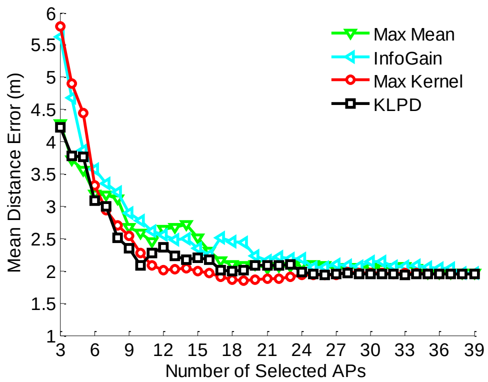
Figure 5.12. Accuracy within fingerprint spacing versus number of selected APs combined with *k*NN positioning algorithm in (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

5.2.6 Multi-floor Performance Using kNN Algorithm

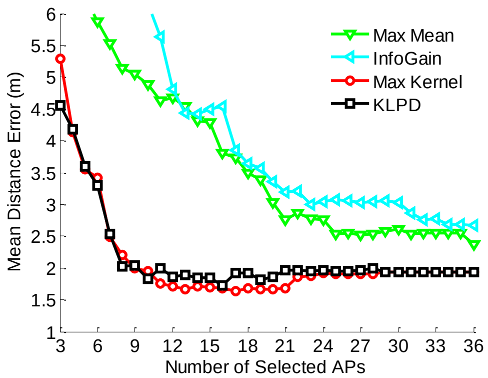
The results of multi-floor performance of using AP selection technique with kNN localisation algorithm is shown in Figure 5.13. Pairing the AP selection algorithms with kNN localisation algorithm in multi-floor environments gives similar trends of results as obtained using univariate kernel algorithm in Section 5.2.2. It is observed that the number of selected APs required to achieve similar localisation accuracy by Max Mean and InfoGain compared to proposed AP selection algorithms which is reduced compared to univariate kernel as in Figure 5.3. For example, in Building A, the number of selected APs required by Max Mean and InfoGain to achieve similar accuracy as KLPD algorithm by kNN at 4 APs is 11 while by univariate kernel is 12. Also in Building C, 18 APs required in kNN compared to 19 in univariate kernel to achieve similar accuracy as KLPD algorithms with 6 selected APs. Futhermore in Building B, the range of the subset of APs of the proposed AP selection algorithms have better accuracy in kNN decreases compared to univariate kernel. Even though the error difference of Max Mean and InfoGain and the proposed algorithms is better in kNN than in univariate kernel case, the results still show that the backend of the AP selection algorithm (which is localisation algorithm) in multi-floor case does not affect the performance of the proposed AP selection algorithm. On the other hand, the performance of KLPD algorithm is noted better than Max Kernel algorithm at very low number of selected APs as can be observed in all buildings compared to the univariate kernel case.



(a)



(b)



(c)

Figure 5.13. Mean error versus number of selected APs for multi-floor localisation using different AP selection algorithms paired with kNN algorithm in (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

Figure 5.14 shows the accuracy of the AP selection algortihms of location estimation within fingerprint spacing in Building A, B, and C. Similar trends of mean error results is also observed where Max Kernel algorithm outperforms other AP selection algorithm in all buildings. Highest accuracy is offered by Max Kernel at 9, 19, and 22 selected AP which indicates 38.9% ,72.3%, and 86.6% accuracy respectively in Building A, B, and C.

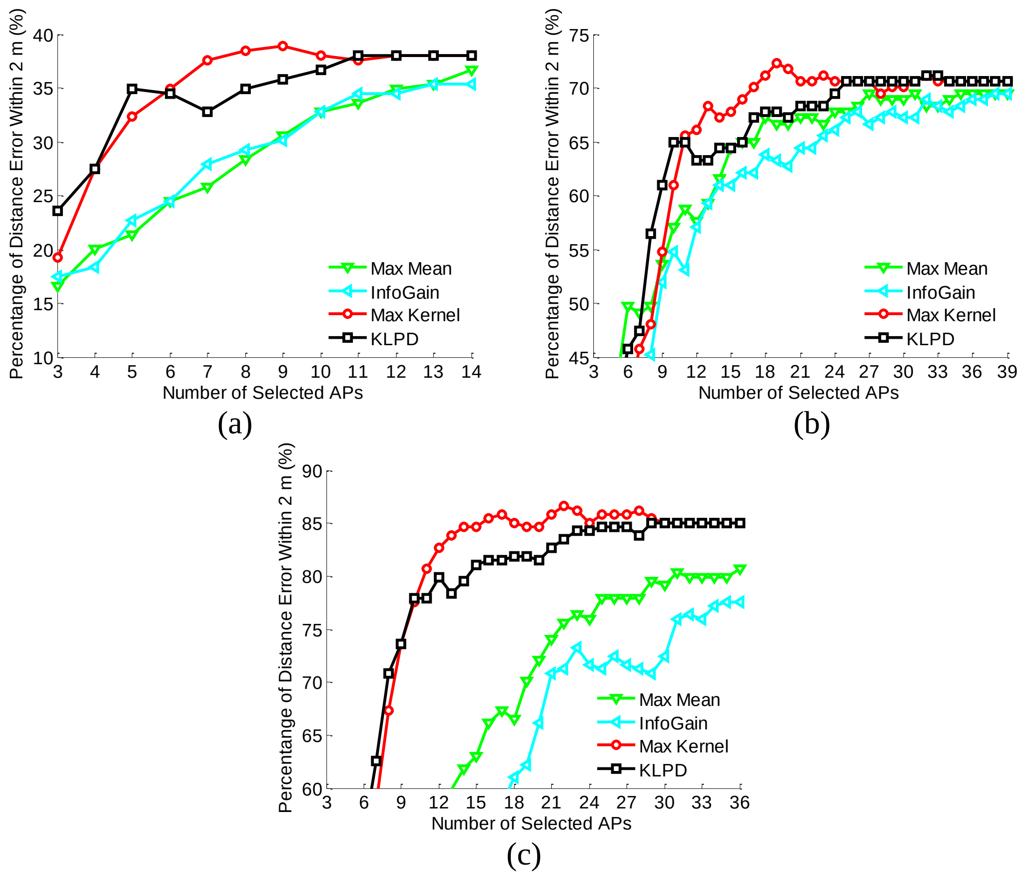


Figure 5.14. Accuracy within fingerprint spacing versus number of selected APs combined with *k*NN localisation algorithm in multi-floor environments of (a) Building A, (b) Building B, and (c) Building C

The results of floor accuracy by using *k*NN algorithm also shows superiority of the proposed algorithms in all environments as shown in Figure 5.15. However, the Max Mean and InfoGain performance is also closer which follows the results of multi-floor mean error. In Building A, the floor accuracy of Max Mean is already comparable to proposed algorithms at 8 APs in *k*NN while at 12 APs in univariate kernel. In Building B, similar trend is also observed where the existing algorithms requires 26 selected APs which is lesser than univariate kernel case where 31 selected APs are needed. In Building C, Max Mean floor accuracy is comparable to Max Kernel and KLPD algorithms at 24 selected APs using kNN but in univariate kernel, the algorithm could not match the accuracy of proposed algorithms.

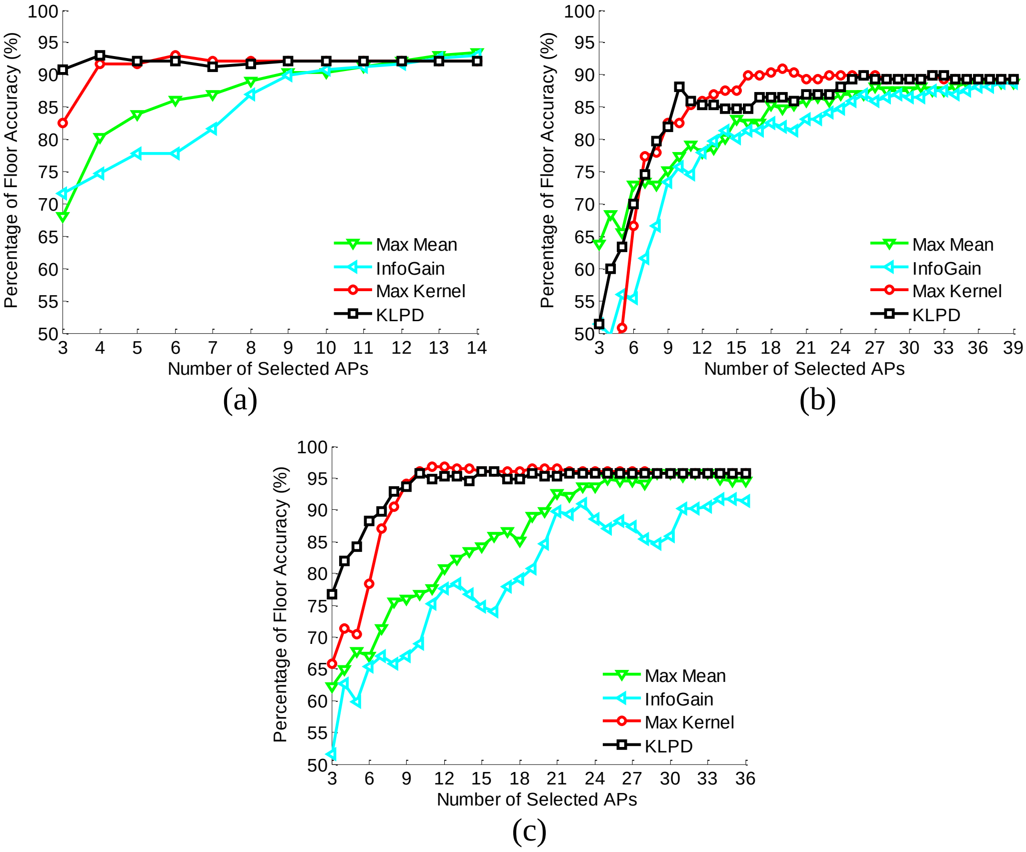


Figure 5.15. Floor accuracy of multi-floor building versus number of APs used by AP selection algorithms paired with *k*NN localisation algorithm in (a) Building A, (b) Building B, and (c) Building C

5.3 Effect of Reducing the Number of Fingerprint Measurement Samples, *T* to Max Kernel and KLPD AP Selection Paired with Univariate Kernel, Multivariate Kernel, and *k*NN Localisation Algorithm

Section 5.2 identifies the optimal number of APs found by Max Kernel and KLPD algorithm. The results in the section utilise 100 training samples ($T = 100$) at each fingerprint location. This section investigates the performance of the AP selection algorithm using the optimal AP numbers while reducing the number of fingerprint measurement (training) samples. Table 5.1 shows the optimal number of APs of either Max Kernel or KLPD paired with different localisation algorithm as investigated in Section 5.2 for both single-floor and multi-floor localisation in

Building A, B, and C. The optimal number of APs is chosen based on lowest mean error obtained by each localisation algorithm utilising the minimum APs as possible. Using the optimal number of APs, the training sample T is varied between 4 to 100 samples for each different localisation algorithm as illustrated in Figure 5.16 to 5.21 for single-floor and multi-floor localisation.

Table 5.1. Number of optimal APs of Max Kernel or KLPD AP Selection paired with localisation algorithm in Building A, B, and C for single-floor and multi-floor dataset

Localisation Algorithm	Number of Optimal APs in each building (Left Column: Single-Floor dataset, Right Column: Multi-floor dataset)					
	A		B		C	
Univariate kernel	9	6	20	28	17	13
Multivariate kernel	11	10	21	30	27	27
kNN	11	5	19	19	15	17

Figure 5.16 and 5.17 shows the results of reduction of the number of APs on the mean error of based on using univariate kernel localisation algorithm for single-floor and multi-floor localisation respectively. It could be seen that at any number of training sample, the performance of both proposed Max Kernel and KLPD are superior to Max Mean and InfoGain in all of the tested buildings for both single and multi-floor localisation. In single-floor localisation, it is noticed that the decrement of training sample as low as 20 samples does not affect the accuracy of both proposed algorithm. The mean distance error of Max Kernel is slightly lower between 20 to 100 samples compared to KLPD algorithm and similar mean distance error is obtained from both algorithms for less than 20 samples in all buildings. Similar trend of result is also seen for multi-floor case in Figure 5.16 where the mean distance error is also similar by using 20 to 100 samples. Performance of Max Mean and InfoGain however provides more uncertainties in multi-floor localisation compared to single-floor as seen in the figure.

Figure 5.18 and 5.19 shows the performance of the AP selection algorithm tested with multivariate kernel. Again, the Max Kernel and KLPD outperform existing algorithm. However, it could be seen that at least 60 samples are required to achieve convergence of mean distance error as seen for Building A and C. Higher samples increasing the accuracy of the estimated location by multivariate kernel because the dimensionality of the kernel (Equation 3.11) representing a location will be more unique compared to others. The performance in Building B however is similar to those of univariate kernel due to signal propagation within the building is affecting the results compared to the dimensionality of the kernel. It is also noticed that the error of Max Mean and InfoGain are closer to proposed algorithms because the number of optimal APs is larger compared to the one in univariate kernel case as could be seen in Table 5.1.

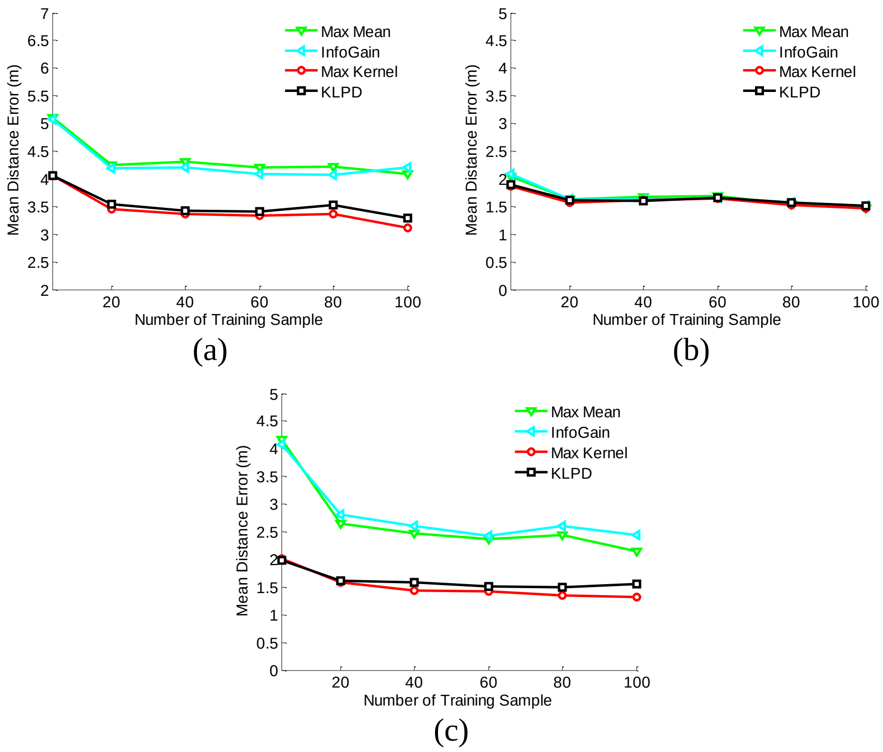


Figure 5.16. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with univariate kernel algorithm in single-floor environment of (a) Building A, (b) Building B, and (c) Building C

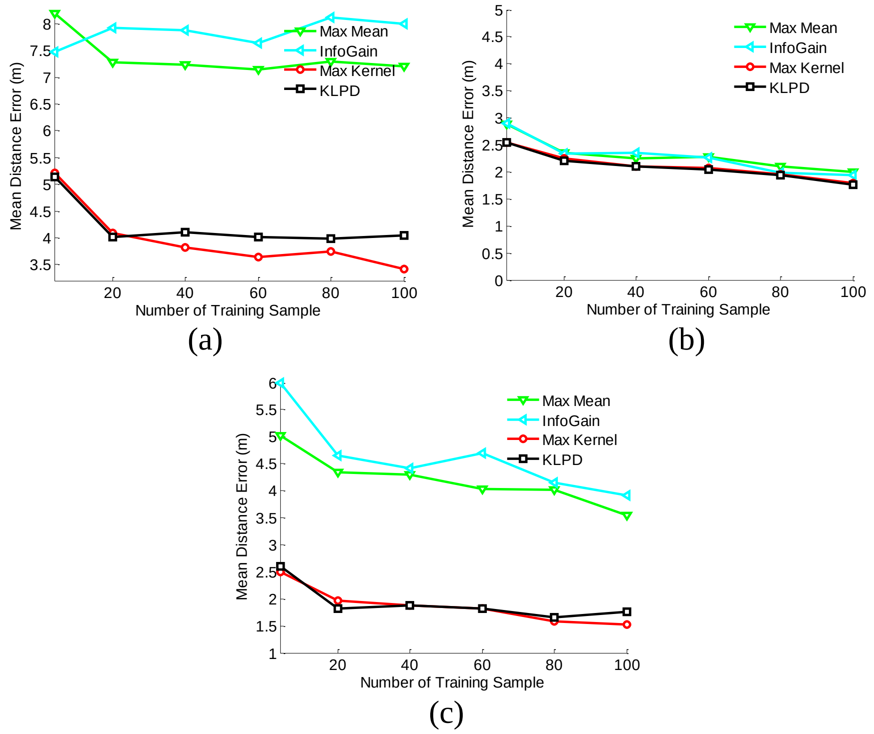


Figure 5.17. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with univariate kernel algorithm in multi-floor environment of (a) Building A, (b) Building B, and (c) Building C

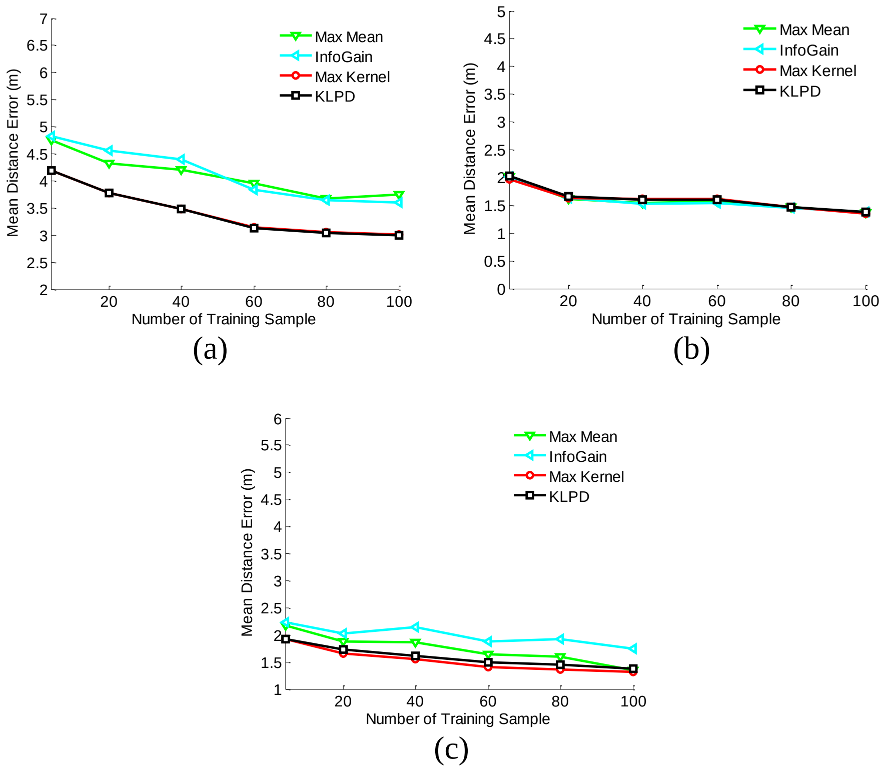


Figure 5.18. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with multivariate kernel algorithm in single-floor environment of (a) Building A, (b) Building B, and (c) Building C

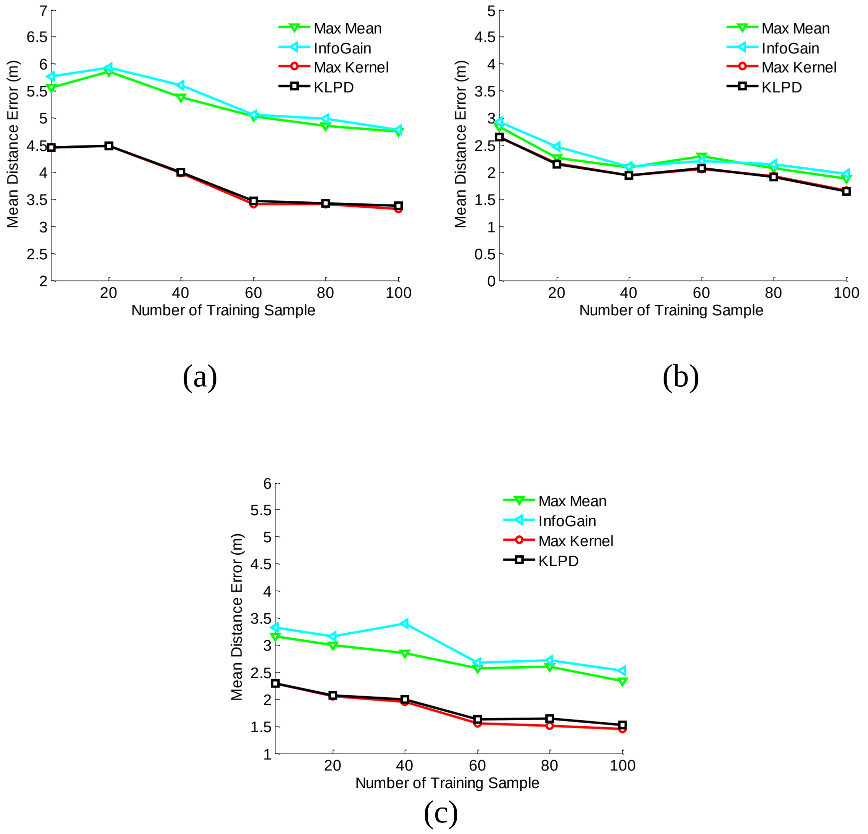


Figure 5.19. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with multivariate kernel algorithm in multi-floor environment of (a) Building A, (b) Building B, and (c) Building C

Figure 5.20 and 5.21 plots the performance of AP selection technique with kNN localisation algorithm. The performance of proposed Max Kernel and KLPD are still better than existing algorithms for all samples in all buildings for both single and multi-floor localisation. However, unlike univariate and multivariate kernel, the accuracy of the proposed algorithms is similar for all tested samples of 4 to 100. Similar result is also seen for existing Max Mean and InfoGain algorithm. This shows that the averaged of RSS (Equation 3.3) implemented in kNN for small sample size does not much change if larger sample size is used. This leads to kNN producing similar results for any sample number used.

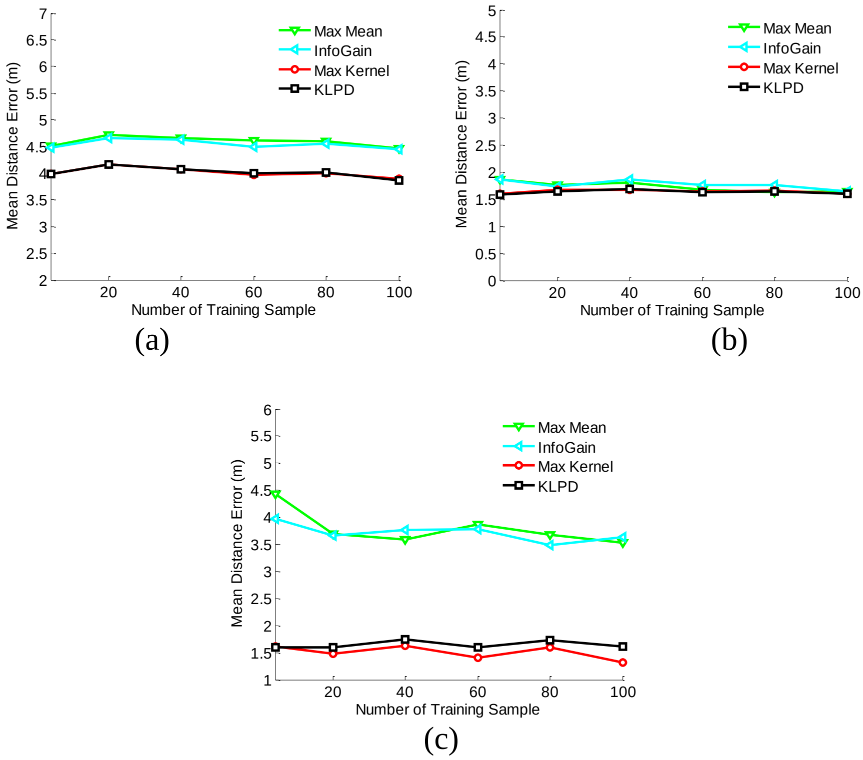


Figure 5.20. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with *k*NN algorithm in single-floor environment of (a) Building A, (b) Building B, and (c) Building C

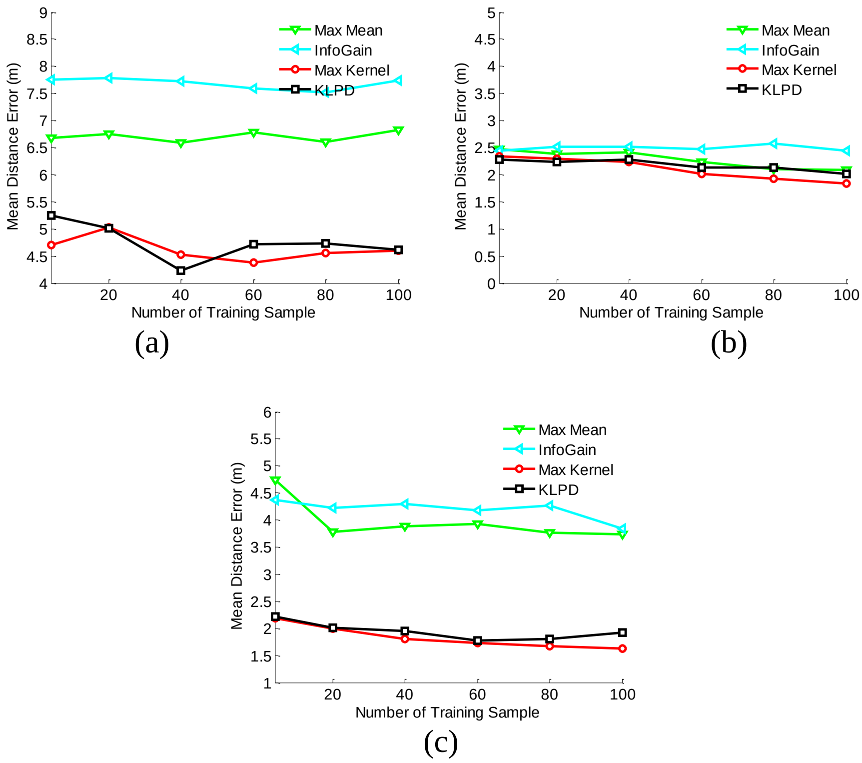


Figure 5.21. Mean distance error versus number of training sample at optimal AP numbers of the radio map database tested with *k*NN algorithm in multi-floor environment of (a) Building A, (b) Building B, and (c) Building C

From the above discussion, it could be summarised that the proposed AP selection algorithm could give similar performance using lower number of samples. This further reduced the computational complexity of the localisation algorithms which already obtained reduction of complexity by using lower number of APs at each fingerprint elements. The rate of AP reduction obtained by the AP selection algorithm is discussed in the next section.

5.4 APs Reduction By Different AP selection Techniques on Univariate Kernel, Multivariate Kernel, and *k*NN Localisation Algorithm

Section 5.2 discussed about the accuracy of the localisation as AP selection is implemented while Section 5.3 discussed the effect of reducing the training samples at optimal AP numbers identified by either Max Kernel or KLPD algorithm. Differently in this section, the reduction of APs in the optimised database produced by AP selection technique on the localisation algorithm of univariate kernel, multivariate kernel, and *k*NN is investigated. The analysis is done by comparing the mean error produced when optimised database of AP selection applied to the localisation algorithm with non-optimised database used with the localisation algorithm. The related error measures of univariate kernel localisation algorithm with no AP selection are tabulated in Table 5.2 and 5.3 for both single-floor and multi-floor case. The single-floor and multi-floor mean error is calculated using Equation 4.3 and 4.4 respectively while the percentage of distance error within fingerprint spacing is calculated from Equation 4.7.

Table 5.2. Single-floor performance metrics of univariate kernel, multivariate kernel, and *k*NN localisation algorithm without using optimised fingerprint database

Localisation Algorithm	Building	Mean Error (m)
Univariate Kernel	A	3.2
	B	1.5
	C	1.3
	CRAWDAD	2.7
	Antwerp	1.3
	Dublin	6.3
Multivariate Kernel	A	3.0
	B	1.4
	C	1.3
	CRAWDAD	2.8
	Antwerp	1.1
	Dublin	5.4
*k*NN	A	3.9
	B	1.6
	C	1.6
	CRAWDAD	2.2
	Antwerp	1.7
	Dublin	4.3

Table 5.3. Multi-floor performance metrics of univariate kernel, multivariate kernel, and *k*NN localisation algorithm without using optimised fingerprint database

Localisation Algorithm	Building	Mean Error (m)
Univariate Kernel	A	3.7
	B	1.8
	C	1.6
Multivariate Kernel	A	3.3
	B	1.7
	C	1.5
*k*NN	A	4.6
	B	2.0
	C	1.9

To compare univariate kernel localisation algorithm with and without AP selection, the number of selected APs of proposed AP selection algorithms is chosen according to minimum subset of APs having 95% upper confidence limit of the mean error (Table 5.2 and 5.3) of the algorithm without AP selection. The 95% upper confidence limit is used because the 95% bound of the accuracy is generally acceptable measure if the algorithm is applied using larger number datasets which may increase the localisation errors. The 95% upper confidence limit has also been used in describing the mean distance error of localisation algorithm e.g. in (Letchner, Fox, & LaMarca, 2005). An example of the plot of upper 95% of confidence limit of mean error of the algorithm for Building A dataset is given as in Figure 5.22. The blue area in the figure shows the errors that are within the 95% upper limit of the mean error of univariate kernel localisation algorithm without AP selection. The calculation of 95% confidence limit of mean distance error is given as follows (Snedecor & Cochran, 1989):

$$Upper\ 95\%\ Confidence\ Limit = \bar{\mu} + \Delta_{1-0.025,\Xi-1}\frac{\sigma_{ME}}{\sqrt{\Xi}} \tag{5.1}$$

where $\bar{\mu}$ is the mean error, σ_{ME} is the standard deviation of the samples, Ξ is the total number of samples which equals to the total number of estimated locations (refer to Equation 5.3), and $\Delta_{1-0.025,\Xi-1}$ is the $100(1-\alpha/2)$ percentile (with $\alpha = 0.05$) of the t-distribution with $\Delta - 1$ degrees of freedom. The resulting upper 95% confidence limit of the mean error is shown in the second column of Table 5.4 and 5.5 for single-floor and multi-floor dataset respectively for univariate kernel algorithm, Table 5.6 and 5.7 for multivariate kernel algorithm, and Table 5.8 and 5.9 for kNN algorithm. Also in the tables, the number of APs required to achieve the limit by each AP selection algorithms is presented in column 3 to 6. It can be observed from the table that low number of APs subset is needed to achieve almost similar positioning accuracy as localisation algorithm without AP selection.

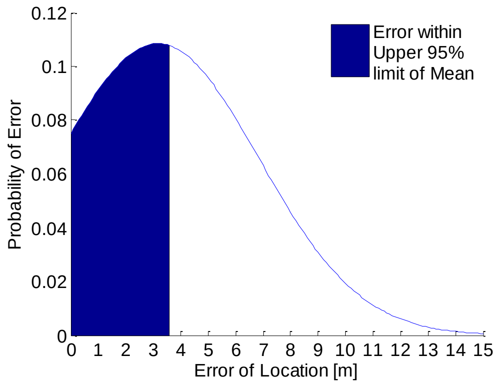


Figure 5.22. The plot of error within upper 95% of mean error confidence limit (error ≤ 3.7 m) of univariate kernel of single-floor localisation in Building A

5.4.1 Single-Floor Performance Using Univariate Kernel Algorithm

Single-floor results are presented in Table 5.4. It is observed that Max Kernel algorithm requires only 5 APs compared to at least 11 APs by InfoGain which is reduced by 54.6% in Building A. Similar reductions are also seen in Building C. In other environments i.e. Building B, CRAWDAD, and Antwerp, the number of required APs is comparable except in Dublin as previously discussed due to misinterpreted probability estimate. However it could be found that KLPD algorithm performs well compared to Max Kernel algorithm in single floor environments such as Dublin and CRAWDAD.

Table 5.4. Upper 95% confidence limit of single-floor mean error of original kernel localisation algorithm and the number of selected APs required by localisation algorithm with different AP selection algorithm to reach the confidence limit mean error

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Number of selected APs required to achieve upper 95% confidence limit according to different algorithms			
		Max Mean	InfoGain	Max Kernel	KLPD
A	3.7	12	11	5	7
B	1.6	10	10	9	12
C	1.6	25	29	12	12
CRAWDAD	3.0	6	8	9	8
Antwerp	1.4	15	18	16	18
Dublin	6.7	8	8	18	7

*Max Mean refers to Youssef et al. (2002) and InfoGain refers to Chen et al. (2006)

The number of required APs in table by each algorithm is transformed as percentage of AP reduction compared to the number of APs required in localisation without AP selection as shown in Figure 5.23. From the figure, it could be observed that up to 81.8% of APs could be reduced (KLPD algorithm in Dublin dataset) if AP selection is used with localisation algorithm. This shows that implementing AP selection could greatly reduce the computational complexity of the localisation algorithm. Performance of the proposed AP selections is generally good compared to Max Mean and InfoGain algorithms in all environments. The Max Kernel algorithm shows excellent performance in reducing the APs compared to other algorithms in Building A, B, C. In CRAWDAD, Antwerp and Dublin environments, the percentage of reduced APs is close between proposed algorithms and Max Mean and InfoGain. It could also be noted that KLPD algorithm better in AP reduction in Dublin dataset compared to Max Mean and InfoGain even though the mean error could be reduced as much as Max Mean and InfoGain as shown in Figure 5.23(f).

5.4.2 Multi-Floor Performance Using Univariate Kernel Algorithm

Table 5.5 presents the results of the required AP numbers in the multi-floor datasets (Building A, B, and C). It is found that the number of required APs for proposed algorithms in multi-floor environment is comparable with single-floor case. However, it is noted that mean error increases in multi-floor setting (Figure 5.3) at the same subset of APs compared to single-floor (Figure 5.1). For example, at 8 selected APs for Max Kernel algorithm, the mean error in multi-floor setting in Building C is 1.9 m but the mean-error in single-floor is 1.7 m. Therefore the number of selected APs in multi-floor settings and single-floor environment could not be

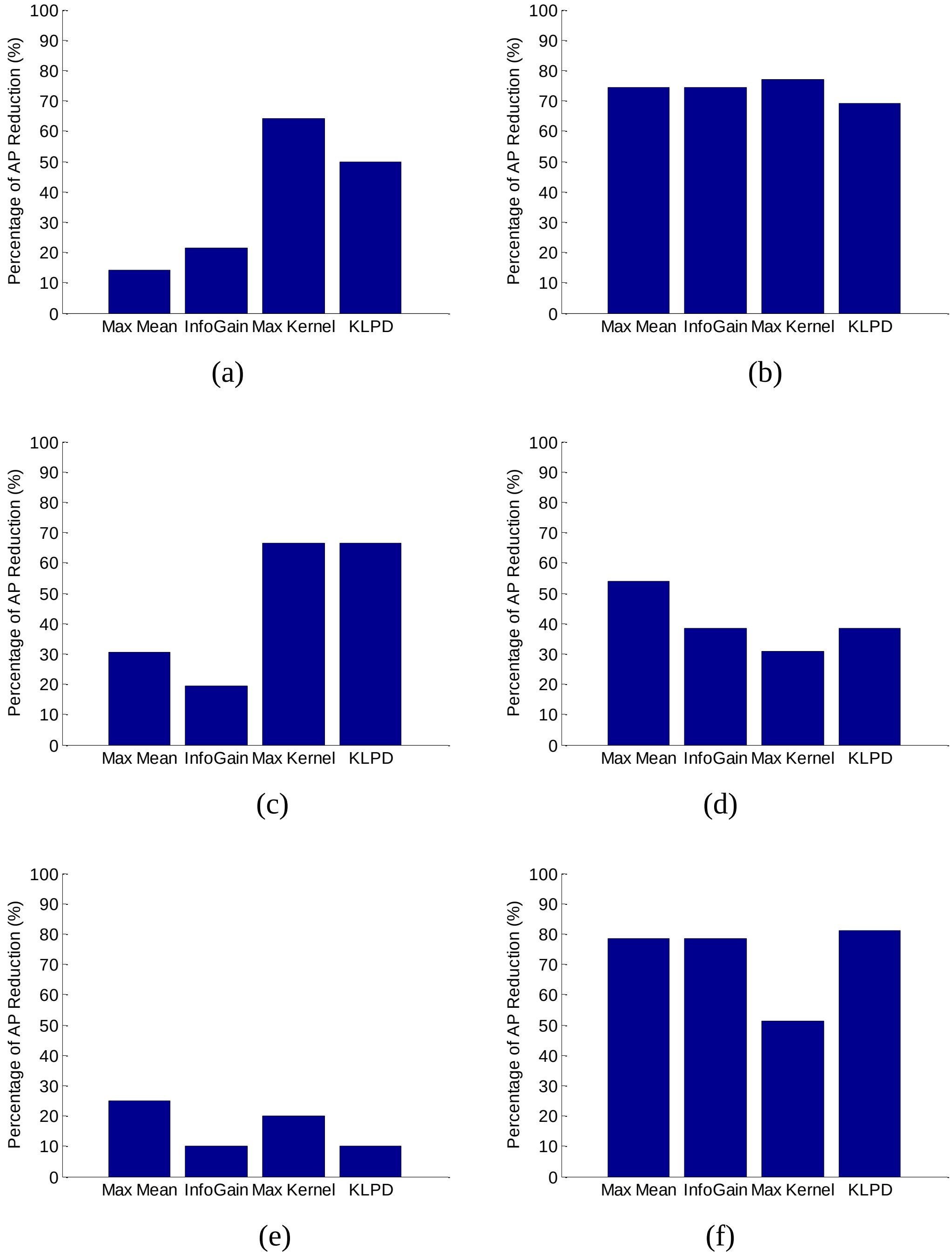


Figure 5.23. Percentage of AP reduction compared to by AP selection algorithms compared to localisation algorithm without AP selection in single-floor setting

comparable. The important observation is the difference of between number of selected APs required to achieve comparable estimate with localisation algorithm without AP selection between proposed algorithms and existing algorithms in these buildings. In Building A, Max Mean requires 200.0% (8 out of 4) more APs than Max Kernel in multi-floor environment compared to 140.0% (7 out of 5) in single-floor. In Building C, larger AP selection is needed by Max Mean compared to Max Kernel with 275.0% (22 out of 8) increment in multi-floor setting compared to 108.3% (13 out of 12) in single-floor. Additionally, in single-floor case in Building B, the subset of APs is almost comparable between Max Kernel and Max Mean but in multi-floor case the performance of Max Mean gets worse. Max Mean requires 69.2% (9 out of 13 APs) more APs than Max Kernel to achieve confidence interval mean error of localisation algorithm without AP selection. In all buildings, the performance of KLPD is near to Max Kernel algorithm and the results of InfoGain are slightly worse than Max Mean.

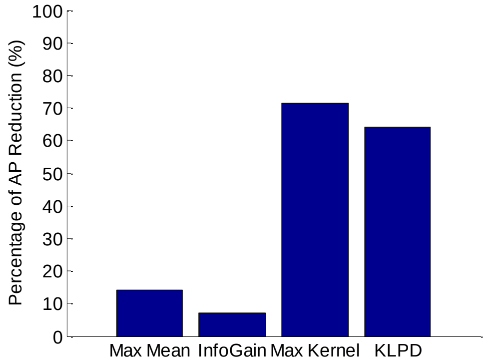
Table 5.5. Upper 95% confidence limit of multi-floor mean error of original kernel localisation algorithm and the number of selected APs required by localisation algorithm with different AP selection algorithm to reach the confidence limit mean error

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Number of selected APs required to achieve upper 95% confidence limit according to different algorithms			
		Max Mean	InfoGain	Max Kernel	KLPD
A	4.4	12	13	4	5
B	2.0	22	27	13	18
C	2.0	30	34	8	10

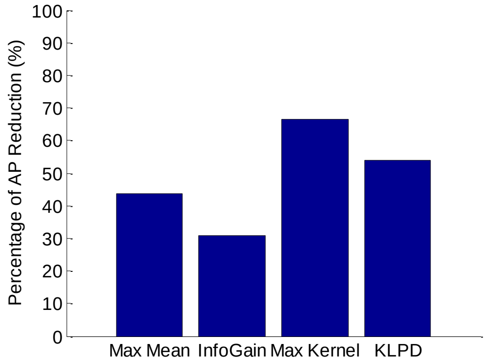
Similar to single-floor case, the percentage of reduction of APs of every AP selection algorithm in multi-floor settings compared to localisation algorithm without AP selection is illustrated as in Figure 5.24(a) to 5.24(e). In agreement with results shown in Table 5.5, the proposed Max Kernel and KLPD algorithms outperforms existing Max Mean and InfoGain algorithms in all datasets. Highest reduction of APs is obtained by Max Kernel algorithm in Building C which 77.8% of APs are reduced compared to APs required by localisation algorithm without AP selection. Compared to single-floor case, the performance of Max Mean and InfoGain gets worse. The main reason is the Max Mean and InfoGain are based on traditional AP selection technique which considers fixed APs for any floor level as described in Section 3.7.1. This means several inaccurate estimations could happens because the signal of the APs in the test sample in a floor level matches the elements of database on different floor and consequently increasing the mean distance error. The results presented in Table 5.5 and Figure 5.24 show that the proposed algorithms with variant AP selection could better reduce the number of required APs in multi-floor environment compared to existing algorithms which depends on fixed selected APs at each subset of APs for all fingerprint dataset.

5.4.3 Single-Floor Performance Using Multivariate Kernel Algorithm

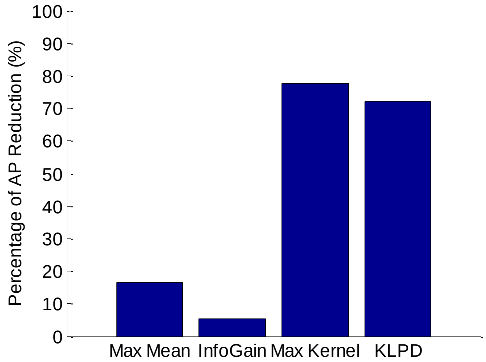
Table 5.6 shows the number of APs to reach upper 95% confidence limit of mean error when AP selection is paired multivariate kernel localisation algorithm. Similar trends of results as in univariate kernel case are observed. Also, the number of selected APs is almost similar for all algorithms in all environments if comparing Table 5.4 of univariate kernel and Table 5.6 of multivariate kernel. This is because the multivariate kernel algorithm is the extension of univariate kernel which considering the simultaneous computation of multiple dimension kernel density estimate as discussed in Chapter 3. The minimum percentage of AP reduction of 5.0% (Antwerp using Max Kernel and KLPD) for multivariate kernel algorithm is



(a)



(b)



(c)

Figure 5.24. Percentage of AP reduction compared to by AP selection algorithms compared to localisation algorithm without AP selection in multi-floor setting given (a) Building A, (b) Building B, and (c) Building C

Table 5.6. Number of APs required achieving upper 95% confidence limit and percentage of AP reduction by different AP selection algorithm paired with multivariate kernel algorithm in single-floor environments

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Left Column: Number of selected APs required to achieve upper 95% confidence limit according to different algorithms Right Column: Percentage of AP Reduction (%) Max Mean		InfoGain		Max Kernel		KLPD	
A	3.4	13	7.1	14	0.0	6	57.1	7	50.0
B	1.5	8	79.5	10	74.4	9	76.9	13	66.7
C	1.6	25	30.6	34	5.6	12	66.7	12	66.7
CRAWDAD	3.1	8	38.5	8	38.5	9	30.8	8	38.5
Antwerp	1.2	18	10.0	18	10.0	19	5.0	19	5.0
Dublin	5.9	12	67.6	9	75.7	13	64.9	10	73.0

comparable to univariate kernel with minimum reduction of 10.0% (Antwerp using KLPD). Highest reduction is obtained by Max Kernel algorithm in Building where 76.9% of AP usage could be reduced.

5.4.4 Multi-Floor Performance Using Multivariate Kernel Algorithm

The multi-floor results using multivariate kernel is tabulated in Table 5.7. Comparable performance of proposed Max Kernel and KLPD AP section as in Section 5.4.3 is also observed. Highest reduction of AP is obtained by Max Kernel algorithm in Building C at 72.2% which is comparable to 77.8% reduction using similar AP selection paired with univariate kernel algorithm in same environment. The minimum reduction of 46.2% is obtained by using KLPD algorithm in Building B. However, the performance of existing AP selection (Max Mean and InfoGain) gets worse in multi-floor estimation. No reduction of APs could be seen if using the algorithms in Building C. The error using Max Mean and InfoGain are much higher for multivariate kernel compared to univariate kernel because in multivariate kernel the kernel calculation is in multi-dimensional multiplication compared to univariate kernel which based on per AP signal multiplication as given in Equation 3.11. This indicates that the freedom of selected APs using traditional AP selection to find correct location estimate is much better by univariate kernel algorithm compared to multivariate kernel and thus causing the lower mean error for univariate kernel. This reflects the percentage of AP reduction of univariate and multivariate kernel algorithms.

Table 5.7. Number of APs required achieving upper 95% confidence limit and percentage of AP reduction by different AP selection algorithm paired with multivariate kernel algorithm in multi-floor environments

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Left Column: Number of selected APs required to achieve upper 95% confidence limit according to different algorithms Right Column: Percentage of AP Reduction (%) Max Mean		InfoGain		Max Kernel		KLPD	
A	3.9	13	7.1	15	7.1	6	57.1	7	50.0
B	1.9	27	30.8	32	18.0	13	66.7	21	46.2
C	1.8	51	0.0	48	0.0	10	72.2	12	66.7

5.4.5 Single-Floor Performance Using *k*NN Algorithm

The performance of AP selection algorithms tested with *k*NN localisation algorithm in single-floor environments is presented in Table 5.8. Minimum percentage of 10.0% AP reduction by the proposed AP selection algorithm is obtained in Antwerp using either Max Kernel or KLPD. Highest percentage of reduction at 77.8% is obtained by KLPD algorithm in Building C. Even though the upper 95% confidence limit of mean error in all environments is higher compared to both univariate and multivariate kernel algorithm, the number of selected APs for any environment is comparable to univariate kernel and multivariate as discussed in Section 5.4.1 and 5.4.3.

Table 5.8. Number of APs required achieving upper 95% confidence limit and percentage of AP reduction by different AP selection algorithm paired with *k*NN algorithm in single-floor environments

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Left Column: Number of selected APs required to achieve upper 95% confidence limit according to different algorithms Right Column: Percentage of AP Reduction (%) Max Mean		InfoGain		Max Kernel		KLPD	
A	4.3	13	7.1	13	7.1	5	64.2	5	64.2
B	1.8	10	74.4	10	74.4	10	74.4	10	74.4
C	2.0	36	0.0	36	0.0	9	75.0	8	77.8
CRAWDAD	2.3	8	38.5	8	38.5	10	23.1	8	38.5
Antwerp	2.0	15	25.0	15	25.0	18	10.0	18	10.0
Dublin	4.6	13	64.9	9	75.7	30	18.9	23	37.8

5.4.6 Multi-Floor Performance Using kNN Algorithm

The performance of the AP selection paired with kNN localisation algorithm in multi-floor case is presented in Table 5.9. The results also show similarity with the performance of AP selection using univariate kernel and multivariate kernel algorithm. The AP could be reduced up to 77.8% as demonstrate by proposed algorithm of Max Kernel or KLPD algorithm in Building C while minimum reduction by either algorithm is 71.4% in Building A. It is noticed that the number of APs to achieve upper 95% confidence limit is slightly lower compared to univariate and multivariate kernel case. This is because higher error tolerance of 5% (difference of 95% to 100% confidence limit) of the mean error is bigger compared in kNN to univariate kernel and multivariate kernel which is lower due to smaller mean error possess by both algorithms.

Table 5.9. Number of APs required achieving upper 95% confidence limit and percentage of AP reduction by different AP selection algorithm paired with kNN algorithm in multi-floor environments

Environment	Upper 95% Confidence Limit of Mean Error of Localisation Algorithm without AP selection (m)	Left Column: Number of selected APs required to achieve upper 95% confidence limit according to different algorithms Right Column: Percentage of AP Reduction (%)							
		Max Mean		InfoGain		Max Kernel		KLPD	
A	5.3	11	21.4	11	21.4	4	71.4	4	71.4
B	2.2	17	56.4	21	46.2	11	71.8	10	74.4
C	2.5	37	0.0	44	0.0	8	77.8	8	77.8

5.5 Comparison of Computation Time of using Different AP Selection Techniques

At mentioned in Section 3.7, the proposed AP selection algorithm i.e. Max Kernel and KLPD is based on variant AP selection which means at every fingerprint location the number of APs varies which according to original fingerprint dataset.

Therefore the computation time is expected lower than the existing AP selection approach i.e. Max Mean and InfoGain which use fixed amount of APs at every fingerprint location. To test the computational time of the AP selection techniques, the time required to perform each online location request by related technique is calculated. The calculation is done for each subset of AP and is repeated 100 times and averaged to confirm the result. Figure 5.25 and 5.26 shows the result comparing computation time required by the proposed variant AP selection of Max Kernel and KLPD and classical AP selection of Max Mean and InfoGain at every subset of APs in Building A, B, and C based on pairing with univariate kernel localisation algorithm for single and multi-floor dataset. It is noticed that the variant AP selection technique of Max Kernel and KLPD effectively reduces the computational complexity in Building A and C. The KLPD algorithm achieves lowest maximum computational time in all buildings and followed by Max Kernel. KLPD obtains better computational time of 2.0% and 5.9%, and 1.0% and 4.9% compared to classical AP selection algorithm (Max Mean or InfoGain) for single-floor and multi-floor localisation in Building A and C respectively when paired with univariate kernel algorithm. In Building B, the result is comparable for both variant and classical AP selection algorithms because of two reasons. First reason is the number of fingerprints is small so the difference is not much seen as the number of processed data is low. It could be seen that in Figure 5.25 in Building B around 4 ms computation time is required compared to 9.8 ms and 10.0 ms in Building A and C respectively. Second reason is the minimum numbers of APs available at each fingerprint elements is large (refer Table 4.4). This causes all AP selection algorithms select similar amount of APs for each investigated AP selection number. Additionally, Table 5.10 and 5.11 show the maximum computational time required by univariate kernel, multivariate kernel and kNN algorithm by each of the AP selection algorithm. It could be seen that computational time of KLPD algorithm could be improved further when multivariate kernel or kNN is used. The computational time of KLPD is reduced 4.1% and 11.6% (Building A), and 10.1% and 14.2% (Building C) by multivariate kernel and 17.2% and 13.8% (Building A), 21.6% and 14.6% (Building C) by kNN when compared with existing algorithm based on single-floor and multi-floor localisation. In general, the results show that the proposed variant AP selection technique could offer quicker computational time compared to classical AP selection technique.

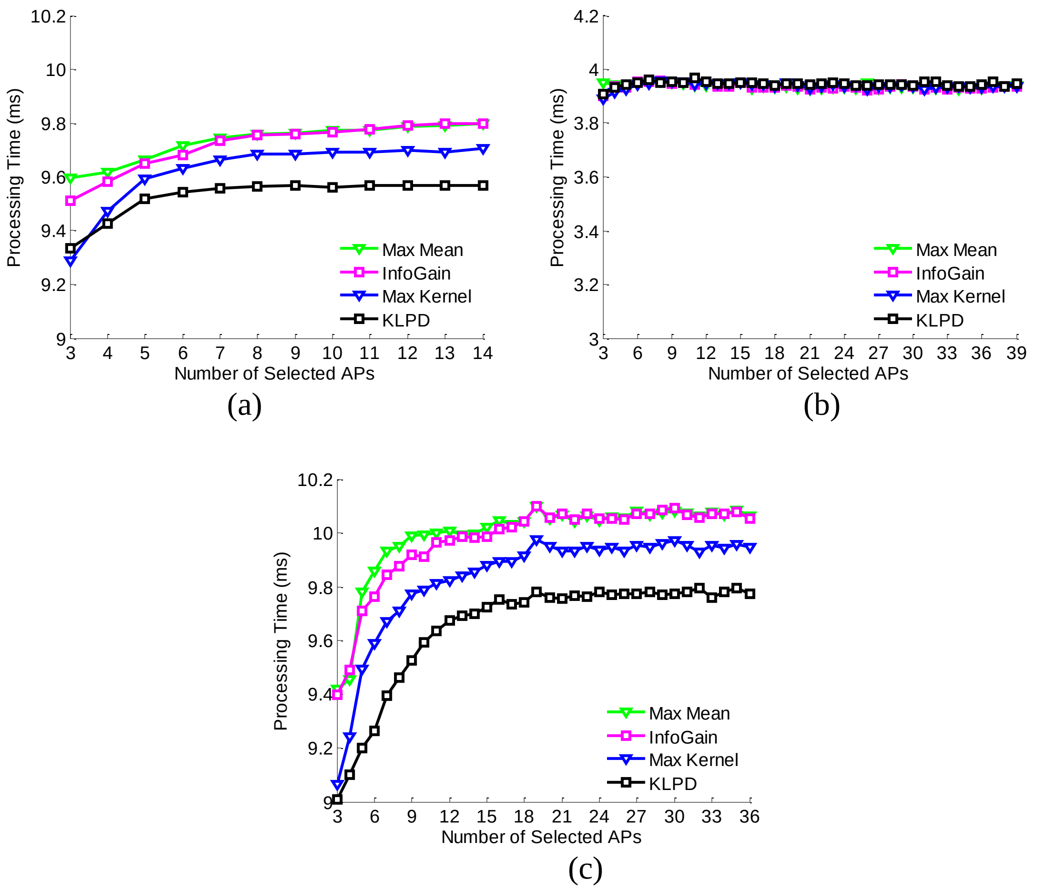
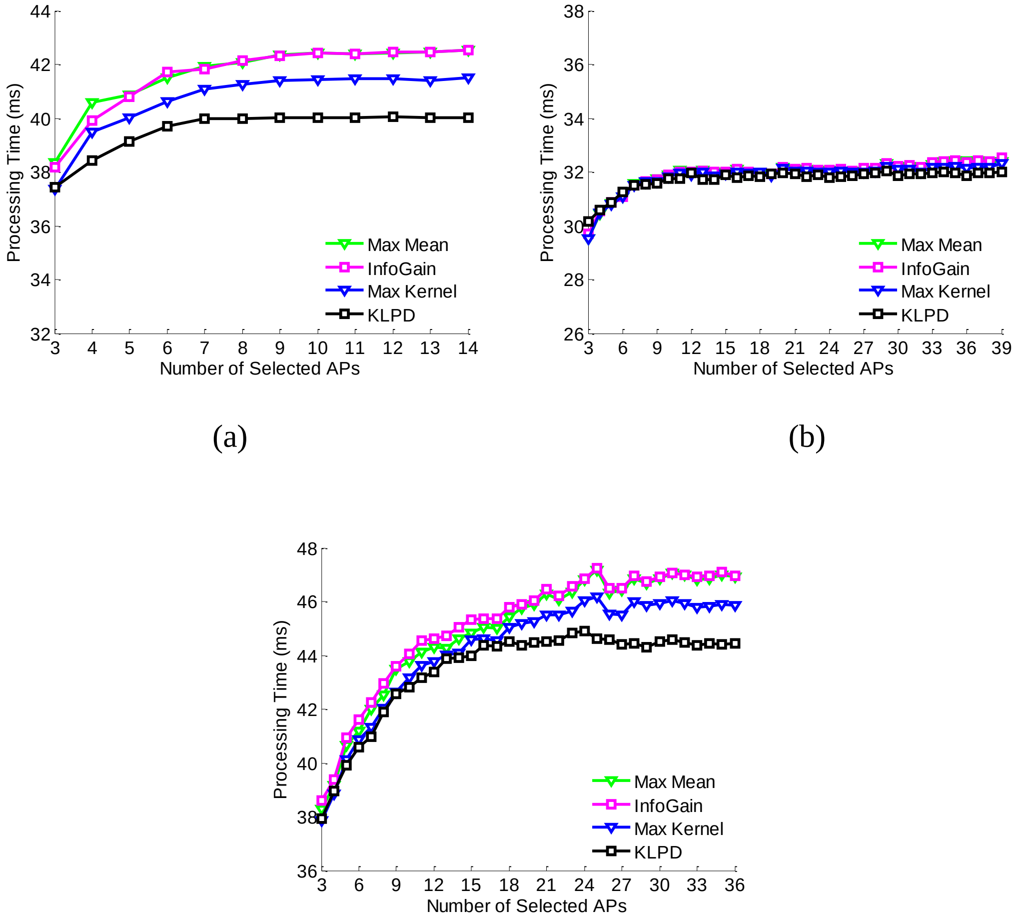


Figure 5.25. Processing time of AP techniques versus number of selected APs in single-floor dataset of (a) Building A, (b) Building B, and (c) Building C. The selected APs are processed with univariate kernel localisation algorithm



(a) (b)

(c)

Figure 5.26. Processing time of AP techniques versus number of selected APs in multi-floor dataset of (a) Building A, (b) Building B, and (c) Building C. The selected APs are processed with univariate kernel localisation algorithm

Table 5.10. Computation times of univariate kernel, multivariate kernel, and *k*NN using different AP selection technique in Building A, B, and C for single-floor dataset

Localisation Algorithm	AP Selection	Maximum Computation Time in Different Buildings (ms)		
		A	B	C
Univariate Kernel	Max Mean	9.8	4.0	10.1
	InfoGain	9.8	4.0	10.1
	Max Kernel	9.7	4.0	10.0
	KLPD	9.6	4.0	9.8
Multivariate Kernel	Max Mean	16.8	7.1	18.8
	InfoGain	16.8	7.1	18.8
	Max Kernel	16.5	7.1	18.0
	KLPD	16.1	7.0	16.9
*k*NN	Max Mean	22.7	10.5	26.9
	InfoGain	22.6	10.5	27.0
	Max Kernel	21.1	10.0	24.6
	KLPD	18.8	9.2	21.1

Table 5.11. Computation times of univariate kernel, multivariate kernel, and *k*NN using different AP selection technique in Building A, B, and C for multi-floor dataset

Localisation Algorithm	AP Selection	Maximum Computation Time in Different Buildings (ms)		
		A	B	C
Univariate Kernel	Max Mean	42.5	32.4	47.2
	InfoGain	42.5	32.5	47.2
	Max Kernel	41.5	32.3	46.2
	KLPD	40.0	32.0	44.9
Multivariate Kernel	Max Mean	75.7	64.4	86.8
	InfoGain	75.8	64.4	86.4
	Max Kernel	72.5	62.8	81.6
	KLPD	66.9	60.4	74.5
*k*NN	Max Mean	102.6	101.5	122.0
	InfoGain	102.6	101.8	122.7
	Max Kernel	96.7	96.3	114.8
	KLPD	88.4	88.3	104.2

5.6 Summary

The proposed AP selection algorithms of Max Kernel and KLPD show excellent performance in multiple scenarios. More importantly the proposed algorithms achieve good localisation accuracy in multi-story building. The dimensionality of APs could be reduced as high as 77.8% to perform similarly as utilising all available APs for positioning. In some cases, positioning accuracy obtained by using AP selection algorithms could be improved. Between two proposed AP selection algorithms, the Max Kernel is generally better compared to KLPD. Additionally the proposed AP selection algorithms could further reduce computational time as high as 21.6% compared to classical AP selection approach to estimate the location in multi-floor building.

CHAPTER 6

RESULTS AND DISCUSSION II: PERFORMANCE OF AVERAGE KERNEL FLOOR AND KERNEL LOGISTIC FLOOR LOCALISATION ALGORITHMS

6.1 Introduction

The previous section discussed the performance of radio map fingerprint database optimisation using AP selection technique of Max Kernel and KLPD algorithm. Here, the performance of AKF and KLF algorithms which are proposed for vertical or floor localisation is presented. The proposed AKF and KLF localisation algorithms have been discussed in Section 3.8. The algorithms apply clustered (optimised) fingerprints based on signal strength compared to non-optimised (original) fingerprints as described in Section 3.8.1. In the following sections, the results of floor localisation are discussed as follows. The resulting number of cluster after computation by the proposed signal strength clustering technique is discussed in Section 6.2. In the subsequent sections, the accuracy and precision of AKF and KLF were compared with other floor algorithms (Liu and Marques (refer Appendix D1 for details of the algorithm and comparison)) in Section 6.3, comparison with original results of Liu and Marques algorithms (Section 6.4), per floor (Section 6.5), computational complexity (Section 6.6), and reducing the number of fingerprint samples (Section 6.7). Section 6.8 describes the results of combining the proposed Max Kernel AP selection with the proposed AKF and KLF floor algorithms.

6.2 Number of Clusters Per Dataset of The AKF and KLF Algorithms

Table 6.1 shows the number of clusters obtained after signal strength clustering algorithm is performed on the dataset of Building A, B, and C. The clustering algorithm (Equation 3.23) was executed before running the AKF and KLF floor localisation algorithms as described in Chapter 3. The number of available clusters presents the number of elements that is processed by the floor algorithms. Compared to processing large fingerprint elements in original database such as Liu algorithm, the AKF and KLF algorithms only process smaller fingerprint elements to determine the floor. The criteria to group the fingerprint elements as cluster are according to AP that has strongest signal in each of the fingerprint. From the table, it is noticed that number of cluster is very small compared to number of original fingerprint elements in the buildings. The reduction between 75.5% and 91.2% could be obtained by grouping the fingerprints as cluster. The compressed number of elements indicates that the processing time of the algorithms could be lowered. Even though large database is used the result of location may not improve if the database is not optimised. Clustering technique has been shown effective to reduce to and at the same time yield improved localisation performance to the algorithm as demonstrated in Razavi, Valkama, and Lohan (2015). The novel AKF and KLF algorithm utilised with proposed signal strength clustering also demonstrated similar result as presented in Section 6.3. Detail of the result is discussed in the section.

Table 6.1. Comparison of number of clusters processed by signal strength clustering algorithm and number of original fingerprints in each tested building

Building	Number of Clusters	Number of Original Fingerprints	Percentage of Reduction of Elements Using Clusters (%)
A	20	254	92.1
B	48	196	75.5
C	59	283	79.2

6.3 Accuracy of Floor Estimation Using AKF and KLF Algorithms

The performance of the AKF and KLF algorithm alongside with other floor localization is analysed by cumulative density function (CDF) of the floor error (Equation 4.8). Figure 6.1 shows the graph of CDF versus the floor error for Liu, Marques, AKF, and KLF algorithms in all three tested buildings. The CDF value at zero floor error indicates the estimated floor accuracy. The value of CDF at 1, 2, 3, and etc. floor error indicates the floor accuracy considering $\pm$ 1, 2, 3, and etc. wrong floor estimated from actual (floor level of particular test sample) floor level.

The results from the Figure 6.1 indicate that the proposed floor localisation algorithm performs better than existing floor localization algorithms. The KLF algorithm achieves 93.4%, 72.9%, and 91.7% accuracy in Building A, B, and C respectively. On the other hand, the AKF algorithm obtains accuracy of 90.0%, 80.2% and 91.7% for similar buildings. The results of both of the algorithms shows better accuracy of 33.7%, 13.0% and 27.5% in Building A, B, and C compared to either Liu or Marques algorithms. This shows that using kernel density and logistic regression techniques are better than using similarity functions which are implemented by Liu and Marques algorithms to estimate the floor level of the building. It is also noted that in much challenging environments e.g. Building B and C where signal of APs is seen in multiple floors, the AKF algorithm performs better than KLF algorithm.

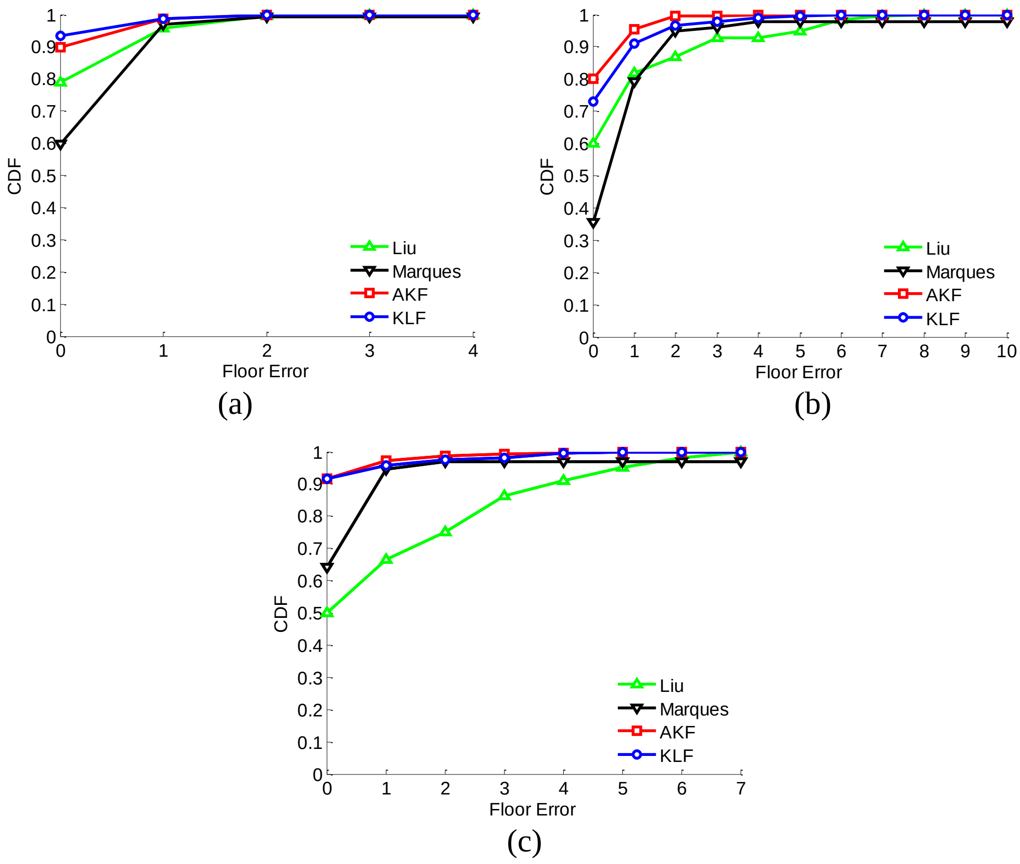


Figure 6.1. CDF of number of floor error on each tested building of (a) Building A, (b) Building B, and (c) Building C

It could also be observed that the maximum floor errors of the proposed algorithms are smaller compared to existing floor localisation algorithms. For example, the maximum floor error of KLF algorithm is 1, 5, and 4 compared to Liu algorithm which obtains maximum of 2, 7, and 6 floor errors in Building A, B, and C respectively. The maximum floor error of AKF algorithm is similar to KLF algorithm in all buildings except in Building B where only maximum of 3 floor errors observed. However, the maximum floor errors of Marques algorithm are equal to the number of available floor level in the buildings. This is because the algorithm is unable to find matching strongest APs of online test sample in the offline database as required in the first filtering step of the algorithm. Therefore, the estimation of the floor is not performed for the related online test sample. The performance of Marques algorithm could be improved if the AP filtering considers more than 2 AP as the filter heads.

Table 6.2 shows the numerical results of the CDF plot in Figure 6.1. The error measures reported is the 90% and 95% percentile of the CDF. The proposed algorithms shows excellent performance compared to Liu and Marques algorithms of at least above 1 floor errors seen in all buildings at 90% percentile. For example, AKF and KLF algorithm obtain no floor error at 90% percentile of CDF in Building A compared to one floor error produces by Liu and Marques algorithms. The performance of Liu algorithm is the worst in Building B and C. This describes that combining the signal distance of fingerprints in one floor level in one group could not provide good estimation of floor location because less distinguishable signal pattern is observed. Almost similar tendency is also seen for result of 95% percentile for all algorithms.

Table 6.2. Numerical results of cumulative density function plot of Figure 6.1

Building	Algorithm	Floor Error (rounded)	
		90% percentile	95% percentile
A	Liu	1	1
	Marques	1	1
	AKF	0	1
	KLF	0	1
B	Liu	3	6
	Marques	2	3
	AKF	1	1
	KLF	1	2
C	Liu	4	5
	Marques	1	2
	AKF	0	1
	KLF	0	1

6.4 Comparison of Floor Localisation Algorithms with Original Findings

Table 6.3 and 6.4 compares the floor accuracy between published results in original publications and this work results. Different buildings consist of different dimensions as listed in Table 6.3 and thus different database setup. The accuracies of the Liu and Marques algorithms can be examined by comparing the results with measured fingerprint data. It should be noted that originally Liu utilised only three test sample position for each floor while Marques' results were based on only three floors measurement data as given in Table 6.3. The Liu , Marques, AKF and KLF algorithms are together tested on three different buildings with different floors (five, eleven and eight floors) as represented by the uniform database setup represented by the 2nd to 5th column in Table 6.4. The results are listed in the last column of Table 6.4. Comparing the results of Table 6.3 and 6.4, it is noted that optimistic result is obtained if small test sample is used or if the height of building is low. Detail analysis of floor error according to every floor level for all algorithms is analysed in the next section.

Table 6.3. Tested floor localization algorithms with reported results

	Original Database Setup				
Algorithm	**Orientation per Location Fingerprint**	**Height of the Tested Building (Floor Numbers)**	**Duration of Scanned Fingerprint Data per Location, (sec)**	**Total number of collected RF Fingerprints**	**Reported Result**
Liu	Not mentioned	8 Floors	60	3420 (57 fingerprint locations)	100% for 3 test samples for each floors with tested on 1st, 2nd, and 5th floor of eight floors building. 88.2% (using larger test samples (2014))
Marques	Not mentioned	3 Floors each for 2 buildings	Not mentioned	9358 (392 fingerprint locations)*	99.5% for total of 1416 test samples in both buildings

*Reference: Marques, Meneses, and Moreira (2012)

Table 6.4: Floor localisation algorithms with this work database setup and reported result

Algorithm	This Work Database Setup: Orientation per Location Fingerprint & Spacing Between Locations	Height of the Tested Building (Floor Numbers)	Length of Scanned Fingerprint Data per Location	Total number of collected RF Fingerprints	This Work Result (Using 254 Test Samples Building A, 196 Test Samples Building B, and 283 Test Samples Building C)
Liu	4 orientations (North, East, South, and West) & 2m spacing between locations	5 Floors (Building A), 11 Floors (Building B), and 8 Floors (Building C)	100	25400 (254 Fingerprint Locations on Building A), 19600 (196 Fingerprint Locations on Building B), and 28300 (283 Fingerprint Locations on Building C)	79.0 % (Building A), 59.9 % (Building B), and 50.0 % (Building C)
Marques					59.8 % (Building A), 35.6 % (Building B), and 64.2 % (Building C)
AKF					90.0 % (Building A), 80.2 % (Building B), 91.7 % (Building C)
KLF					93.4 % (Building A), 72.9 % (Building B), 91.7 % (Building C)

6.5 Per Floor Analysis of AKF and KLF Localisation Algorithms

Per floor analysis as illustrated in Figure 6.2 investigates performance of the algorithms according to each of individual floors of the building. The graph of correct floor estimated (in percentage) or floor accuracy versus every floor level of the building is presented (Equation 4.5). The full bar (100%) e.g. the AKF algorithm on second floor of Building B (Figure 6.2(b)) shows that the related algorithm correctly estimated all the test samples to their actual floor position and the missing bar e.g. Marques algorithm on ninth and tenth floor of similar building means that all of the estimated floors of the test samples are wrongly classified. The result of per floor analysis is important to determine if each of algorithm performs robustly on each floor of each building by which the CDF presents in previous section only represents the algorithm performance in general.

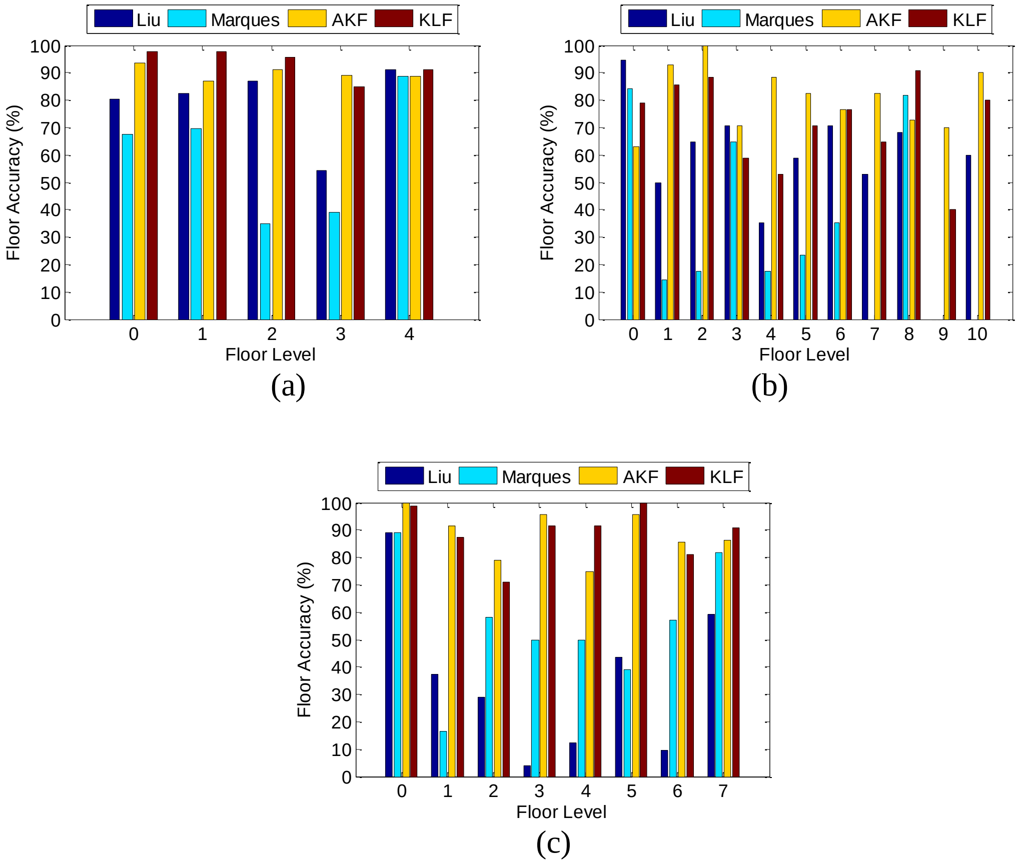


Figure 6.2. Per floor analysis of all tested algorithms in all buildings of (a) Building A, (b) Building B, and (c) Building C

From the figure, it is noticed that the proposed algorithms could perfectly estimate certain floor level of the buildings. For example, AKF algorithm could perfectly estimate second floor of Building B, and ground floor of Building C, and KLF algorithm correctly evaluate the fifth floor of Building C. Additionally, the algorithms could provide good estimates for each floor level in all buildings with the lowest estimation is 40.0% by KLF algorithm of ninth floor of Building B. Less variation of accuracy in different floor levels in different buildings is seen for estimation by the proposed algorithms especially using AKF algorithm. On the other hand, the accuracy of 59.9% of Liu algorithm in Building B is helped by the high accuracy of ground floor estimation of 94.7% and at the same time the all-floor performance is degraded by all wrong estimations of ninth floor. Similarly, the localisation accuracy of Marques algorithm in Building C is 63.0% but on the second floor of the building the accuracy of the floor level is only 20.0%. This shows that the all-floor performance of Liu and Marques algorithms in Building B and C does not indicate true performance of the algorithms since high variation of estimation accuracy in different floor exists. The poor performance of Marques of Liu on certain floors e.g. Floor 2 and 3 in Building A, and Floor 1, 2, 4, 5, 6, 9 and 10 in Building B, is due to generalised threshold setting of the algorithm which is used to improve the floor accuracy of the all floors of building but not per floor. On the other hand, Liu's algorithm perform less especially in certain floors of Building C (Floor 2, 3, 4, and 6) because the values of grouped signal distance (Equation D1.1) between adjacent floors are closed which could give misperception to the algorithm to locate the correct floor. To investigate further, the variance of floor estimation accuracy for all floor level of each building is tabulated in Table 6.5 for every algorithms. The table shows that the proposed algorithms of AKF and KLF obtain low variance compared to Liu and Marques algorithms in all buildings. This indicates that proposed algorithms are more robust in estimation every single floor level of every building compared to existing algorithms. The variance of all floor localisation algorithms is highest on Building B mainly due to two reasons: physical open spaces between floors and fingerprint and test sample locations are very close to the stairs and metal-walled elevators.

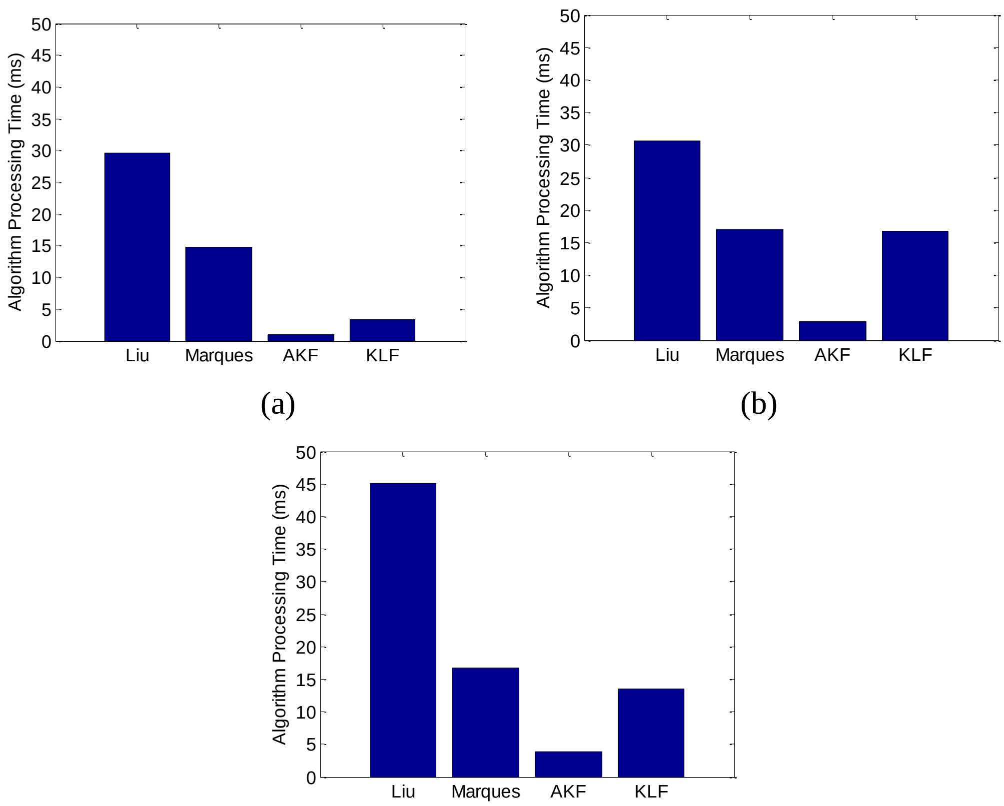
Table 6.5. Variance of estimation accuracy of each algorithm in every building

Algorithm	Variance of Correct Floor ID Estimated		
	Building A	Building B	Building C
Liu	208.2	577.6	818.0
Marques	513.0	1014.3	524.4
AKF	6.3	127.5	75.5
KLF	30.9	255.6	90.6

6.6 Computational Complexity of AKF and KLF Localisation Algorithms

To test the computation complexity of each algorithm, each algorithm is run for 100 times using Intel-i5 3.2GHz desktop computer (described in Chapter 4) and the processing time is averaged. Processing time is linearly related to power consumption required by a device to run the algorithm. Also processing time determines capability of an algorithm to provide location estimation in real time. Less processing time means less consumption of power by the device and hence reduces energy required to run the algorithm. Figure 6.3 shows the averaged processing time of the floor localization algorithms in three tested buildings. It is seen that floor localization algorithms of Liu and Marques requires long time to complete each floor location request at the expense of higher accuracy. The Liu and Marques algorithm processing times are 29 and 14 times higher compared to proposed AKF algorithm respectively as demonstrated in Building A. This is because the processing time of Liu and Marques algorithms is linear in the number of location fingerprint in the database. It can be explained by the search step of those two algorithms which compares each location fingerprint data existed in the database (the measurement sample) to the one that is currently measured (the test sample). Additionally, at each location fingerprint the algorithms need to search similarity of ID of the AP of the test sample to the measurement sample in the database. Therefore the processing time could be roughly estimated according to number of fingerprint elements times the average number of APs available in each fingerprint element. Referring to Table 6.2, the number of fingerprint times average number of APs in increasing order is Building A, B, and C and processing time of Liu and Marques algorithms shows similar trend as observed in Figure 6.3.

The proposed algorithms require low processing time because the search steps are according to the number of cluster which is much lower than amount of fingerprint location in normal database. Between AKF and KLF algorithms, AKF processing time is quicker because the RSS of test samples only needs to be matched with the maximum of the averaged kernel of the RSS of related APs before maximum likelihood calculation is performed. KLF algorithm on the other hand has higher computation time because the computation of each online sample depends on the size of vector $\boldsymbol{w}$ which the size varies according to availability of the pairwise AP signals between multiple clusters. This means if the signal of a certain pairwise AP is available in more than two clusters, the size of vector $\boldsymbol{w}$ is larger compared to the one of two clusters.



(c)

Figure 6.3. Average algorithm processing time for each tested floor detection algorithm in (a) Building A, (b) Building B, and (c) Building C

6.7 Effect of Number of Fingerprint Measurement Samples, *T* on AKF and KLF Localisation Algorithms

The performance of all floor localisation algorithms in terms of varying the fingerprint measurement samples at each location is investigated. Using similar test samples as in previous sections, the number of offline measurement samples is plotted against floor accuracy of each algorithm for all three tested buildings as shown in Figure 6.4. It can be observed that in Figure 6.4 the performance of the proposed AKF algorithm is unmatched by any other algorithm from 4 to 100 measurement samples except by the proposed KLF algorithm at 100 measurement samples in Building A. It is also seen that all algorithms the results are constant at different number of measurement samples except for KLF algorithm. The KLF algorithm performs better if larger measurement samples are available. This is because the KLF algorithm decision depends on every single pair of RSS which sometimes does not give enough information about the location is small measurements are available. The results of KLF algorithm shows the accuracy of KLF is in decreasing order of Building C, B, and A which accordingly follows the average number of available AP signals as mentioned in column 5 of Table 6.2 that is directly related to the number of measurement samples available. The reduction of floor accuracy in Building C, B, and A from 100 to 4 samples is 8.6%, 16.4%, and 26.6% respectively. However, the performance of KLF is still better than Liu and Marques algorithms if 70 e.g. in Building A or more samples are used as seen in Figure 6.4. The results in all tested buildings also show that the proposed AKF algorithm is robust even using small amount of fingerprint samples to determine the floor location. This means the process of collecting measurement samples during the offline stage could be done in less time and less effort. Additionally, reduction of measurement samples also related to the reduction of cost to develop a localisation system.

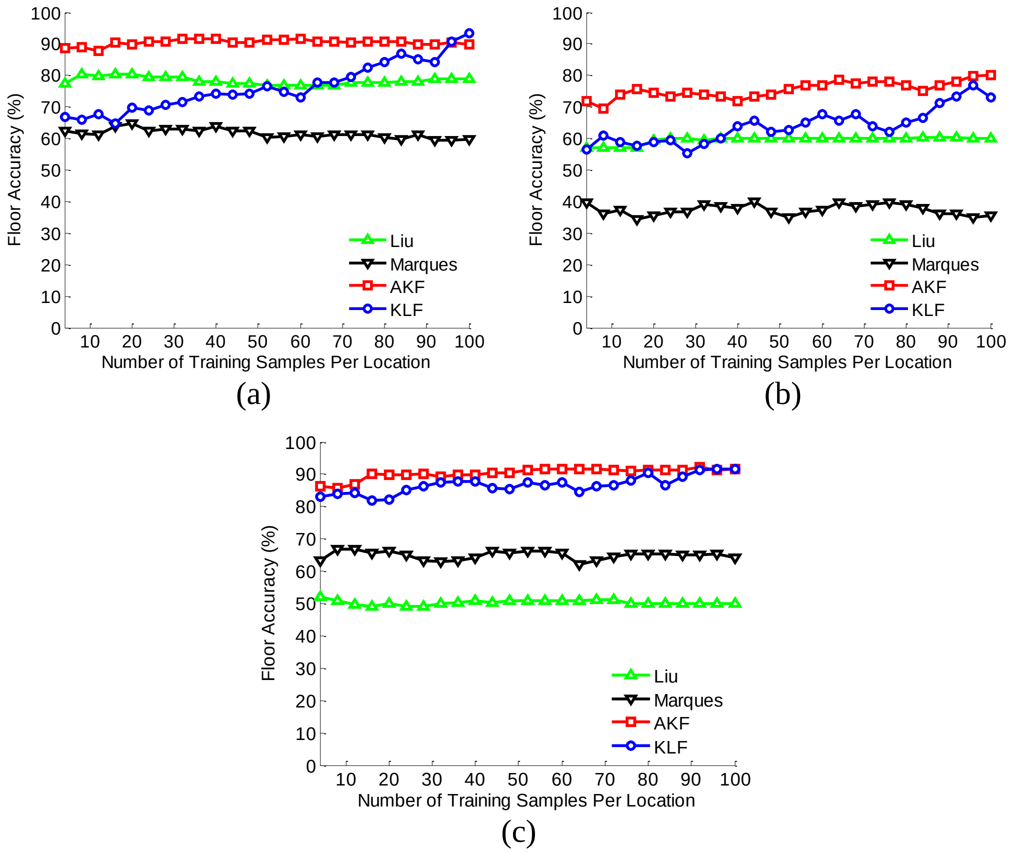


Figure 6.4. Effect of different number of measurement samples to the accuracy of floor detection by the floor localisation algorithms in (a) Building A, (b) Building B, and (c) Building C

6.8 Combining The Proposed Max Kernel AP selection with AKF and KLF Floor Localisation Algorithms

The performance of the proposed floor localisation algorithms is further studied by pre-selecting certain subset of APs for each cluster using the developed AP selection algorithm, Max Kernel. Figure 6.5 shows the floor accuracy for AKF and KLF algorithms according to different subset of selected APs in each cluster for Building A, B, and C. The comparison is not made with Liu and Marques algorithms as the algorithms will not be used for multi-floor location estimation. Also, the floor accuracy of both algorithms is lower compared to proposed AKF and KLF algorithm and it is meaningless to do pre-selection of APs for the algorithms. In agreement with the results of AP selection discussed in Chapter 5, the number of APs in the cluster could be reduced further to reduce the computation time of processing the algorithm. For AKF algorithm, optimal floor accuracy could be achieved by using only 18, 39, and 48 number of APs for each cluster in Building A, B, and C respectively. For KLF algorithm, highest floor estimation could be obtained by using only 19, 34, and 31 APs from each cluster respectively in Building A, B, and C. This means reduction of APs for both algorithms are 37.9%, 44.3%, and 15.8%, and 34.5%, 51.4%, and 45.6% in the corresponding building. The results prove that utilising database optimisation using AP selection could improve the computational complexity of floor localisation algorithm by optimising the required AP to perform the localisation estimation.

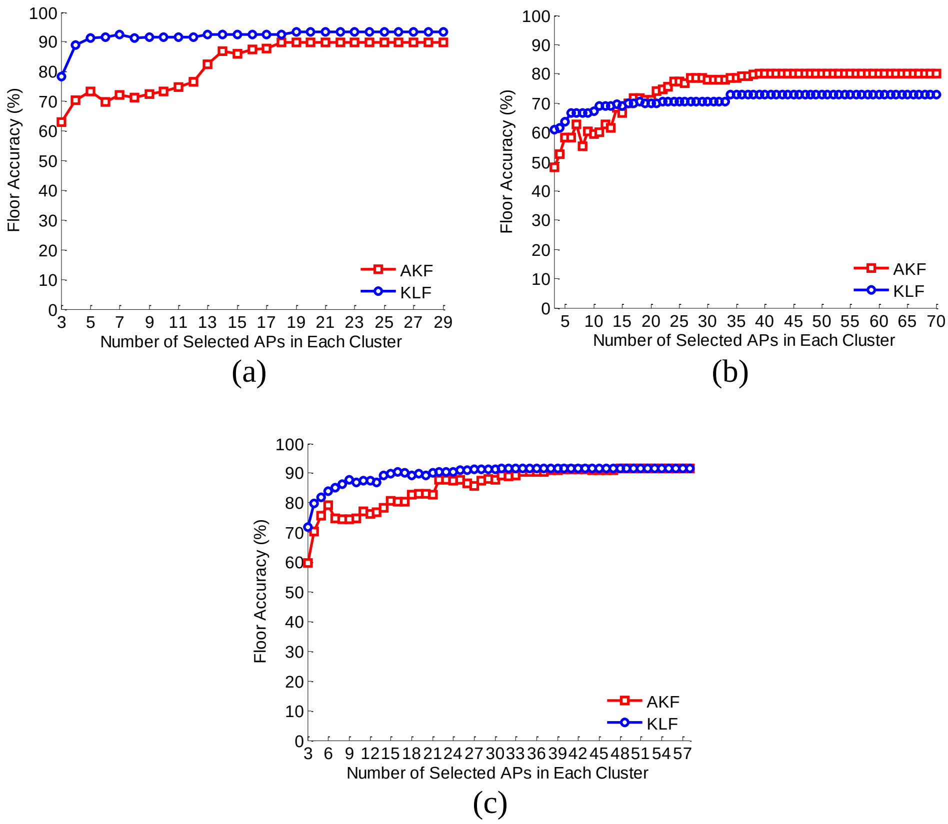


Figure 6.5. Effect of floor accuracy estimated by AKF and KLF algorithms when Max Kernel AP selection algorithm is applied to the cluster according to different number of AP in (a) Building A, (b) Building B, and (c) Building C

6.9 Summary

The novel floor localisation algorithms of AKF and KLF have shown excellent performance in improving the accuracy of the floor estimation at reduced computational time. The accuracy improvements are dependent on implementation of theory of kernel density estimate and kernel logistic regression classification to estimate the floor. It is demonstrated that high floor accuracy of 93.4% is achievable by the proposed algorithm in the tested environments. Lower processing time of up to 29 times compared to existing floor algorithm is also obtained due to the advantage of the signal strength clustering technique which reduces the amount of elements to be processed by the algorithm.

CHAPTER 7

RESULTS AND DISCUSSION III: PERFORMANCE OF NORMAL AND KERNEL MULTI-CLASS KNN CLASSIFIER HORIZONTAL LOCALISATION ALGORITHMS

7.1 Introduction

Previous section describes the performance of vertical or floor determination algorithms. Once the floor level is determined, the exact coordinate on horizontal floor plan of the building should be estimated. In this work, the normal and kernel multi-class kNN classifiers described in Chapter 4 are proposed for determination of the horizontal coordinate in any floor. The proposed kNN classifiers algorithms first requires tuning of the parameter k and ϖ and additional parameter σ for kernel version as given in Equation 3.34 (for k), 3.35 (for ϖ), and 3.15 (for σ), before localisation is performed. Section 7.2 discusses the effect of using different number k and ϖ of the proposed kNN classifier algorithm regarding the localisation performance. Section 7.3 explains the effect of using different σ in kernel kNN classifier algorithm. Then, the performance of the proposed normal and kernel kNN classifier horizontal localisation algorithms are compared with three classical localisation algorithms discussed in Chapter 3 which are kNN, univariate kernel, and multivariate kernel. Further sections describe the performance comparison of the kNN classifier algorithms with the three classical localisation algorithms. Section 7.4 investigates performance of the algorithms in terms of precision and accuracy. Section 7.5 studies the effect of reducing the number of APs to be used with the localisation algorithms. Section 7.6 gives details on the effect of using different number of offline training samples to the algorithms. In Section 7.7, the kNN classifier algorithms are paired with Max Kernel AP selection algorithm to examine appropriate subset of APs to be used for the multi-floor localisation system.

7.2 Effect of Number of Estimated locations k and Optimisation Parameter ϖ of the Multi-class kNN Classifier

The proposed multi-class kNN classifier algorithm performance depends on the appropriate selection of k and ϖ which are the number of nearest estimated locations and the number of selected data with the closest distance to the online test sample respectively. The proposed normal and kernel multi-class kNN classifiers require tuning these two parameters in order to obtain the best localisation accuracy. This is done by determine at what value of k and ϖ would give the lowest mean distance error (Equation 4.3). Figure 7.1 shows the horizontal mean error variation of by using different values of in all six tested environment. Tuning the appropriate value of parameter ϖ reduces the localisation errors in all environments. This shows that in all indoor environments exists noise of the signal that affects the determination of accurate position. Therefore if using only the closest distance for each class ($\varpi = 1$), the mean error of positioning increases. Antwerp dataset is a special case. It can be seen that the lowest mean error could be achieved for any value of ϖ e.g. 1 to 3 as in Figure 7.1(e). This is because in Antwerp environment, each fingerprint location already contains unique set of AP identification and therefore the error of positioning is unaffected by very small value of ϖ.

On the other hand, the number of k that gives the best positioning errors for multi-floor environments (Building A, B, C, and Antwerp) are in agreement with the classical kNN algorithm as discussed in Section 4.10 which is one. However for single-floor datasets (CRAWDAD and Dublin), the required k to achieve lowest mean positioning error is three. This indicates that the distribution of signal vector of each AP pair is less discriminating from each other and several positions may have overlapping signal vector.

7.3 Effect of σ in Kernel kNN Classifier

Kernel kNN classifier (Equation 3.37 and Equation 3.15) has one additional parameter compared to the normal kNN classifier which is σ. The adjustable parameter σ describes similarity between data of kernel RBF function in kernel space. To choose suitable value of σ for efficient positioning using kernel method, graph of different values of σ versus mean error of positioning is plotted. In this case the best k and w as in Section 7.2 is used, the value of sigma is investigated within range of 1 to 50. However, since the value of σ coverged to a constant value around 30 for all environments the only range of 1 to 30 is shown in Figure 7.2. It can be seen that for all environments except for Antwerp dataset the mean error is decreasing as σ increases. Once σ reaches a certain value, the accuracy of the positioning does not monotonically increase and in fact beyond certain value, the mean error increases. This could be explained by the nonlinearity of the signal vectors collected in those environments. If an environment consists of multiple non-linear signal vectors, the small change of σ could affect the accuracy of positioning.

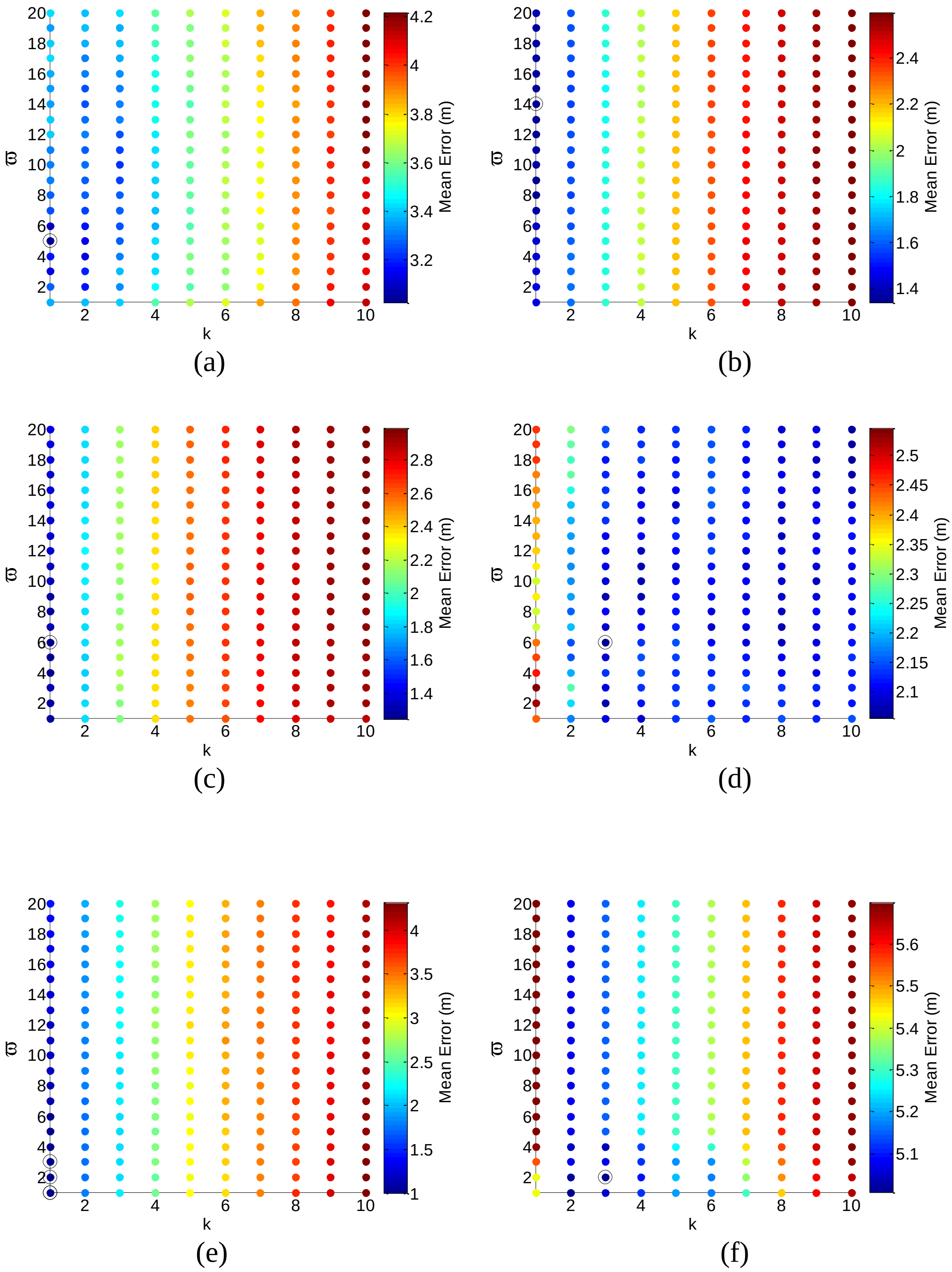


Figure 7.1. Mean error variation by using different *k* and ϖ in all six environments: (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin. The lowest mean error is marked with black unfilled circle

For example, it can be seen from the figure that the CRAWDAD environment consists more non-linear signal vectors compared to other environments because small change of σ value e.g. from 1 to 3 quickly reduces the mean positioning error. For Antwerp environment, changing σ does not change the positioning accuracy. This is because the group of APs that identifies the fingerprint location is distinguishable. This happens as the amount of APs observed in the environments is large. This is similarly explained as in Section 7.2. From the figure, the best sigma is 9, 5, 5, 11, 1, and 6 for Building A, B, C, CRAWDAD, Antwerp, and Dublin dataset respectively.

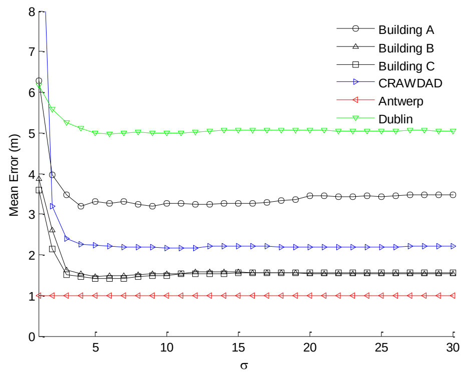


Figure 7.2. Different values of σ versus mean positioning error in all tested environments.

7.4 Precision and Accuracy of Multi-Class *k*NN Classifier and Comparison with Other Horizontal Localisation Algorithms

To evaluate the performance of the proposed multiclass *k*NN classifiers, the cumulative density function of the distance error is plotted using Equation 4.6 as in Figure 7.3. The CDF represents the precision of the localisation algorithm as explained in Chapter 4. The comparison is made between the proposed algorithms – the normal and kernel multi-class *k*NN classifiers, and three classical localisation algorithms (which is discussed in Chapter 3) – the *k*NN, and two probabilistic algorithms which are the univariate kernel, and multivariate kernel. The *k*NN, UK, MK, MkNN, and KMkNN refers to kNN, univariate kernel, multivariate kernel, multi-class kNN classifier, and kernel multi-class kNN classifier algorithms respectively. Additionally, Table 7.1 shows detail result of performance of the localisation algorithms numerically. In the table, five error measures are depicted which are mean, standard deviation, median, 90% and 95% of CDF.

The mean error indicates the accuracy of the localisation algorithm as mentioned in Chapter 4. Figure 7.3 shows that proposed algorithms especially the normal multi-class kNN robustly performs in all buildings. In Building B, Building C, CRAWDAD and Antwerp, the performance of normal kNN classifier is better than other classical algorithms (UK, MK, and kNN). Also, even though the normal multi-class kNN classifier algorithm in Building A and Dublin is not the best, it is the second best performer in those buildings.

It can be seen that the best classical algorithms is the multivariate kernel, but it performs better than the normal multi-class kNN classifier only in Building A. This case is similar to classical kNN algorithm which only outperforms other algorithms in Dublin environment. Additionally, the performance of multivariate kernel is notable worse when applied in single-floor buildings (CRAWDAD and Dublin). Contrarily, the performance of classical kNN is good in single-floor buildings but worse in multi-floor buildings as can be seen in Figure 7.3. The performance of univariate kernel is generally lower compared to multivariate kernel. This shows that the classical localisation algorithms perform well only in certain buildings.

At 90% of localisation error, the multi-class kNN classifier could obtain as high as 69.7% compared to existing algorithms. It is noted that the number and placement of APs in those environments explains the result. The normal multi-class kNN classifier algorithm could achieve higher accuracy if many APs are installed within the floors and well-distributed such as in Antwerp setting as the signal vector of pairwise APs are more distinguishable between different locations. In single-floor CRAWDAD environment, the multi-class kNN classifier outperforms any other algorithms. Improvement of 4.6% (0.1 m of 2.2 m), 22.2% (0.6 m of 2.7 m), and 25% (0.7 m of 2.8 m) of mean distance error by normal kNN classifier respectively compared to kNN, univariate kernel and multivariate kernel is seen. The environment is rich of AP signals and the APs are well-distributed and therefore the result replicates the Antwerp case. On the other hand, the performance of the multi-class kNN classifier algorithm is outperformed by kNN algorithm in single-floor Dublin environment. The main reason is the information of RSS at each fingerprint location in Dublin environment is captured at only short period of time compared to other environments and therefore increases variance of the signals at every fingerprint location. This condition gives disadvantage to algorithms that considers each distribution of signal. However in Dublin environment, the multi-class kNN classifier still outperforms the other two probabilistic algorithms. Additionally, the normal kNN multiclass classifier generally performs better than the kernel version of the algorithm. This shows that transforming the dataset into higher dimension as kernel function does not improve the classification of dataset when using kNN classifier and in most cases using kernel degrades the performance of the algorithm.

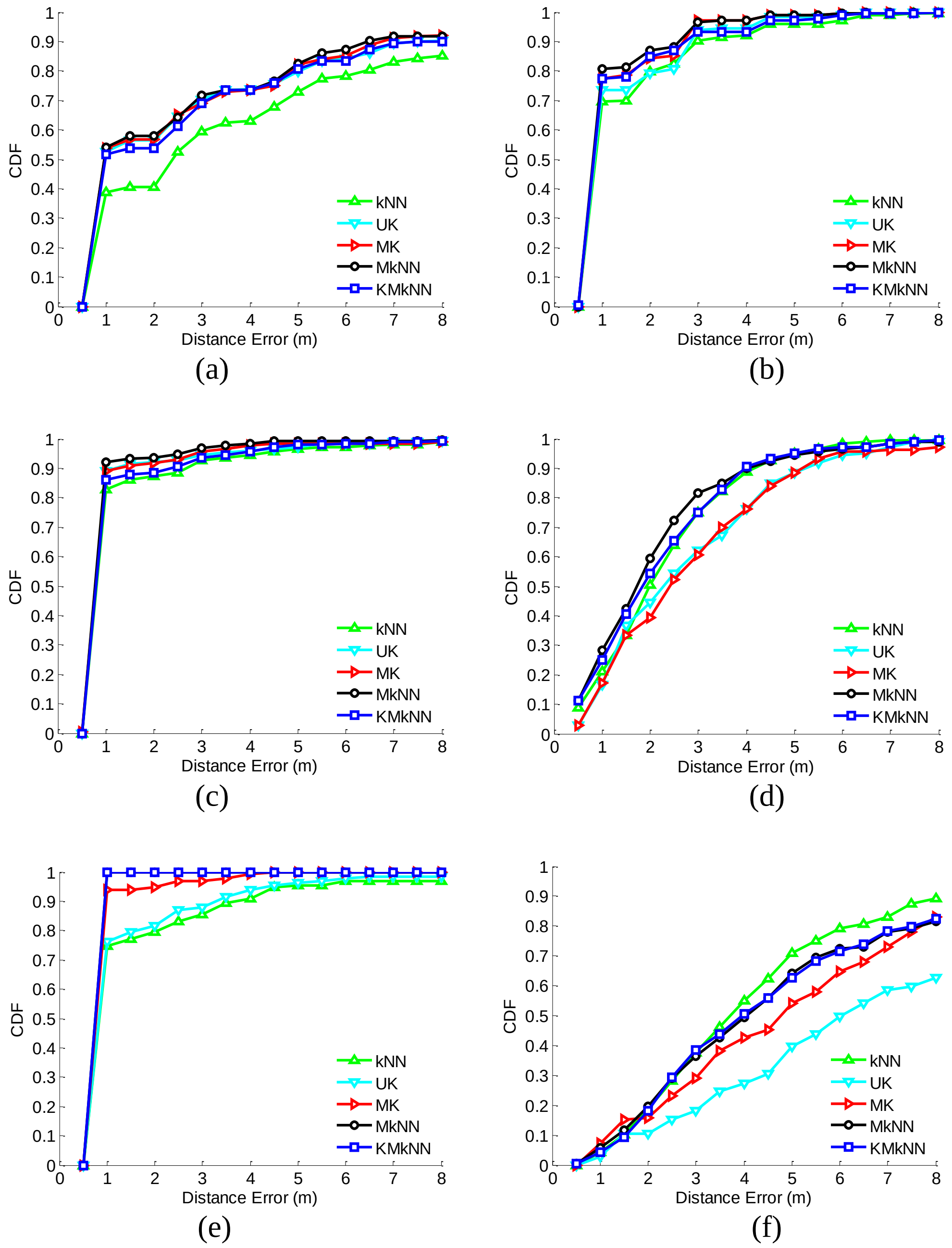


Figure 7.3. Cumulative density function of mean errors of tested horizontal positioning algorithms using six environment datasets (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin. In Antwerp environment, errors of multiclass *k*NN classifier and its kernel version are overlapped

The precision measure also reflects the accuracy of the algorithm. The mean error of normal multi-class *k*NN classifier is the lowest compared to other algorithms in all buildings except in Dublin which is in agreement with the 90% and 95% CDF measures as shown in Table 7.1. The results presented in Figure 7.3 and Table 7.1 indicates classification of signal vectors of group of pairwise APs works better to discriminate information of fingerprint datasets at different locations thus giving good and robust location estimates. .

Table 7.1. Error measures of different horizontal positioning algorithm in Building A, B, C, CRAWDAD, Antwerp, and Dublin

Dataset	Algorithm	Error measures				
		Mean	Standard Deviation	Median	90% of CDF	95% of CDF
A	k Nearest Neighbour	3.9	3.7	2.3	9.0	11.0
	Univariate Kernel	3.2	3.7	1.0	7.0	11.0
	Multivariate Kernel	3.0	3.3	1.0	6.9	9.1
	Multi-class kNN classifier	3.0	3.6	1.0	6.1	11.0
	Kernel multi-class kNN classifier	3.2	3.6	1.0	7.1	10.9
B	k Nearest Neighbour	1.6	1.4	1.0	2.7	4.2
	Univariate Kernel	1.5	1.0	1.0	2.7	4.0
	Multivariate Kernel	1.4	0.8	1.0	2.6	2.7
	Multi-class kNN classifier	1.3	0.8	1.0	2.6	2.7
	Kernel multi-class kNN classifier	1.4	1.1	1.0	2.6	4.1
C	k Nearest Neighbour	1.6	3.0	1.0	2.6	4.2
	Univariate Kernel	1.4	2.0	1.0	1.1	3.4
	Multivariate Kernel	1.3	1.9	1.0	1.1	2.7
	Multi-class kNN classifier	1.2	1.7	1.0	1.0	2.5
	Kernel multi-class kNN classifier	1.4	1.9	1.0	2.1	3.5
CRAWDAD	k Nearest Neighbour	2.2	1.5	2.0	4.1	4.9
	Univariate Kernel	2.7	1.9	2.3	5.1	6.3
	Multivariate Kernel	2.8	2.0	2.5	5.1	5.7
	Multi-class kNN classifier	2.1	1.6	1.7	4.0	4.9
	Kernel multi-class kNN classifier	2.2	1.6	1.8	3.9	4.9
Antwerp	k Nearest Neighbour	1.7	1.6	1.0	3.8	4.9
	Univariate Kernel	1.6	1.5	1.0	3.3	4.1
	Multivariate Kernel	1.1	0.5	1.0	1.0	2.0
	Multi-class kNN classifier	1.0	0.0	1.0	1.0	1.0
	Kernel multi-class kNN classifier	1.0	0.0	1.0	1.0	1.0
Dublin	k Nearest Neighbour	4.3	2.8	3.7	8.1	9.6
	Univariate Kernel	6.7	3.9	6.0	11.8	13.1
	Multivariate Kernel	5.4	3.9	4.9	9.7	13.5
	Multi-class kNN classifier	5.0	3.8	4.1	9.9	12.6
	Kernel multi-class kNN classifier	5.0	3.7	3.9	9.8	12.2

Moreover, the main advantage of the multiclass kNN classifier algorithm could be seen in terms of localisation accuracy within spacing of fingerprint location (Equation 4.7) as plotted in Figure 7.4. The accuracy within fingerprint spacing is defined as percentage of estimated location errors within each fingerprint distance which is 1 m for CRAWDAD and 2 m for other environments. This means the higher the percentage, the better algorithm performs. In all environments, the performance of multiclass kNN classifier is better that other tested algorithms. It is noted that the accuracy of the localisation within fingerprint spacing increases 1.3%, 2.8%, 1.6 %, 8.9%, 19.3%, and 0.8% in Building A, Building B, Building C, CRAWDAD, Antwerp, and Dublin respectively compared to the best version of classical algorithms when multiclass kNN classifier is used. This shows that most of the classification works correctly using the proposed classifier to identify the nearest available location in the fingerprint dataset.

7.5 Effect on Number of Fingerprint Measurement Samples, *T* to Multi-Class *k*NN Classifiers

The performance of the normal and kernel kNN classifier algorithms is further analysed by varying the number of training sample in the radio map database. The aims of varying the number of trainings samples are twofold. First, to investigate if similar or better performance of the algorithm could be achieved by using different number of training samples. Second, to study at what range of number of training samples the performance of the proposed algorithms is better than existing algorithms. Figure 7.5 presents the mean error of every horizontal localisation algorithms at different number of training samples. The result is discussed based on dataset of Building A, Building B, and Building C as these are the datasets that are administered in this work. The number of samples at each fingerprint location in the dataset is chosen between 4 to 100 samples at 4 sample interval. The implementation

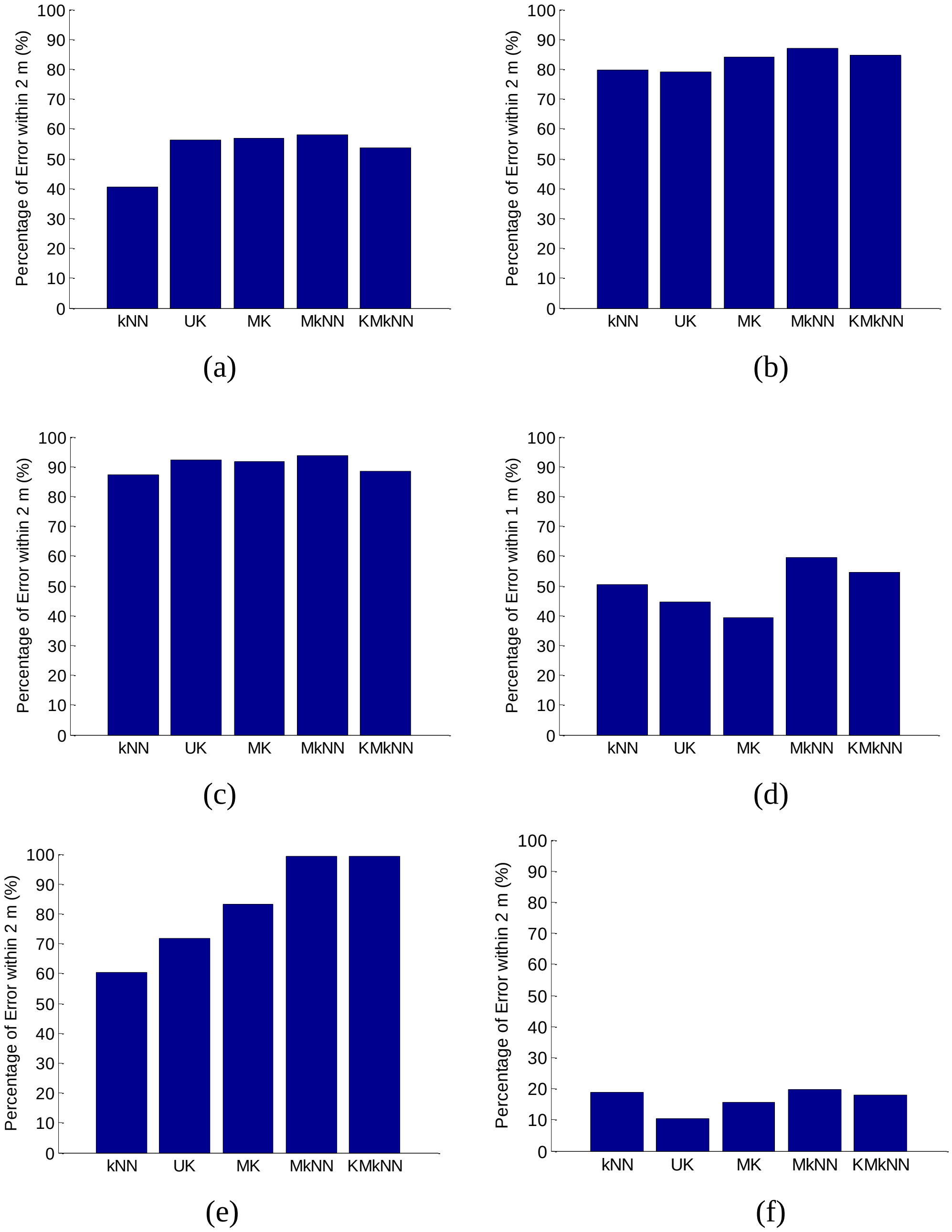
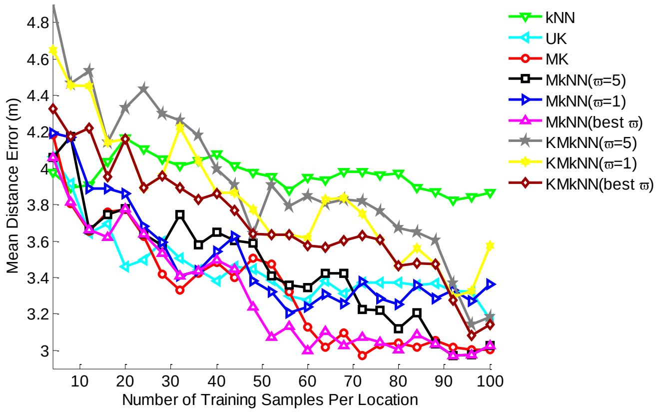


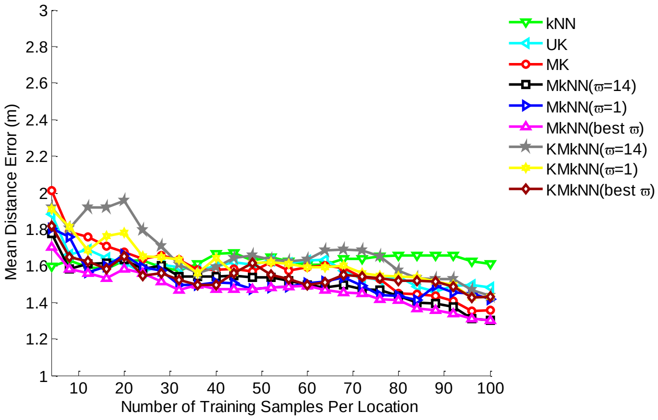
Figure 7.4. Localisation accuracy within fingerprint spacing of each tested horizontal localisation algorithm in (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

of the multi-class kNN classifiers as plotted in Figure 7.5 are observed in three different versions which are optimised with the best value of ϖ (pink and brown markers), with value of ϖ optimised using 100 samples which are 5, 14, and 6 for Building A, B, and C respectively (black and grey markers), and without optimised

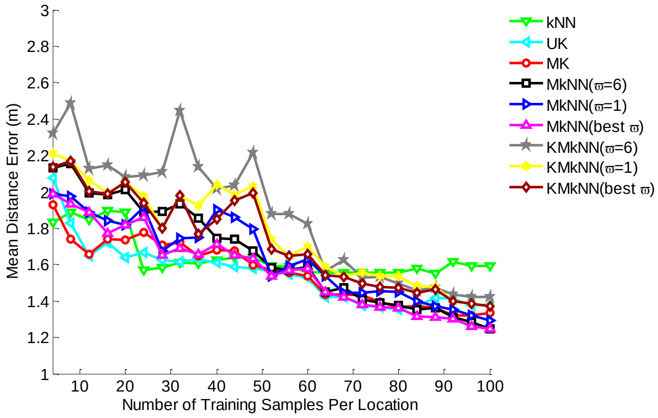
ϖ where the value of ϖ is equals to 1 (blue and yellow markers). Details of the optimised ϖ values are given in Appendix E1. The figure shows that in all buildings optimisation of ϖ value is required at different number of samples to obtain lowest mean error. This indicates that the change of sample number significantly affects the performance of both normal and kernel kNN classifier algorithms. Additionally, in all environments it is better to tune the ϖ value to obtain higher accuracy compared to using only nearest signal distance ($\varpi = 1$) for both algorithms. Also, using constant ϖ optimised at 100 training samples for different number of samples increases the mean error of the algorithms. It is also noted that the mean error of kernel multi-class kNN classifier is higher compared to normal multi-class kNN classifier if compared at similar ϖ. This indicates that it is better to use normal kNN classifier compared to using its kernel version in every case. Using the optimised ϖ of normal multi-class kNN algorithm (MkNN(best ϖ)), lowest mean error could be obtained at 60 training samples in Building A. The performance of the normal algorithm in Building A however is comparable to multivariate kernel algorithm at certain number of training samples. The kernel version of the algorithm with best ϖ performs better than univariate kernel and classical kNN algorithms at large trainings samples of 96 to 100 samples. In Building B, the normal multi-class kNN classifier algorithm with best ϖ gives the lowest mean error within 8 to 100 training samples which outperforms other localisation algorithms. In Building C, the algorithm performs better than other algorithms at training samples of 80 to 100. The results for the kernel version in Building B and C are almost identical to Building A where it outperforms univariate kernel and classical kNN algorithm from 92 to 100 samples. In both buildings however lower mean errors could not be achieved by normal kNN classifier algorithm if smaller than 100 samples are used.



(a)



(b)

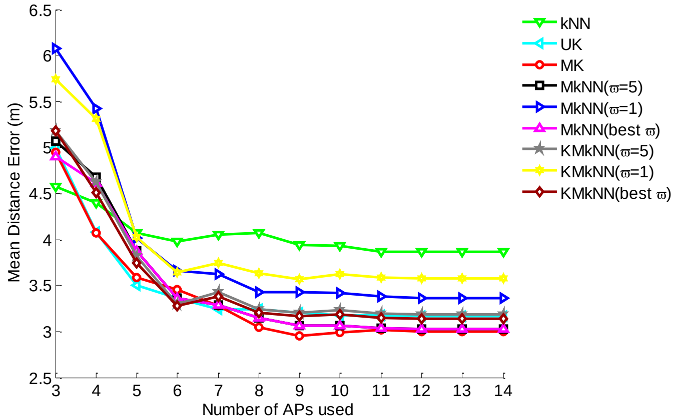


(c)

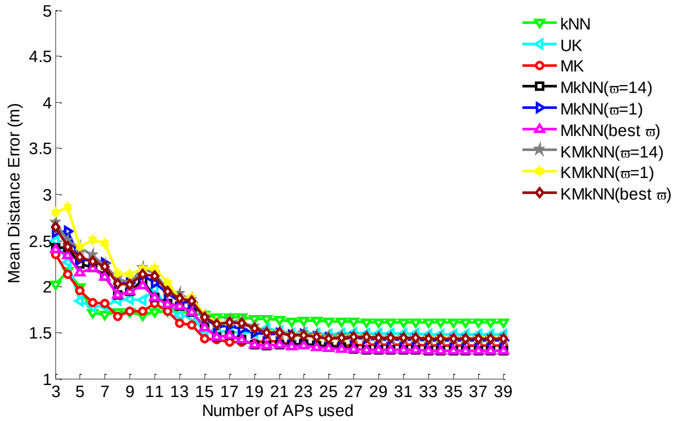
Figure 7.5. Effect of reducing the number of training samples for all localisation algorithms. The normal and kernel *k*NN classifier algorithms are tested in three different versions using different value of ϖ. The plot of M*k*NN(best ϖ) and MK*k*NN(best ϖ) are overlapped

7.6 Effect of the Number of APs Per Fingerprint on the Multi-Class kNN Classfiers

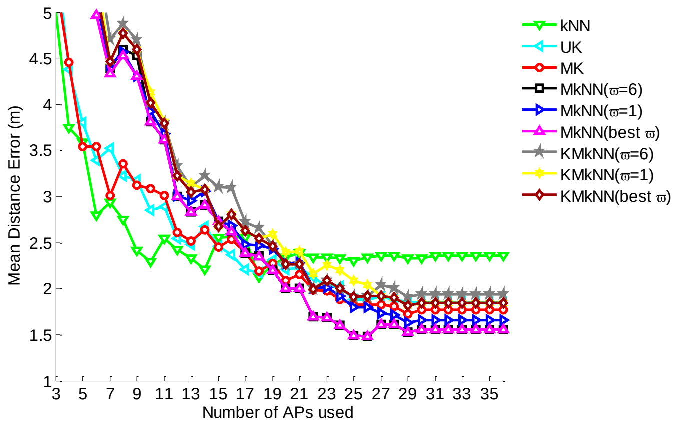
The performance of the proposed algorithms is further analysed by reducing the number of APs of every fingerprint element in the radio map database. The result of this analysis is depicted in Figure 7.6 for Building A, B, C, CRAWDAD, Antwerp and Dublin. The effect of reducing the APs indicates that whether the algorithm could provide similar accuracy at smaller subset of APs as using all APs available at each fingerprint location. This can be evaluated by the change of mean errors at different number of APs used in the database. Lower or similar mean errors at lower number of APs indicate that small AP numbers is enough for the algorithm to give similar estimation compared to using all AP information in the database. The selection of APs is made by considering the decreasing average signal strength value of the AP signal at each fingerprint location. For example, for 3 selected APs, the 3 APs at each fingerprint elements with highest average RSS is chosen for testing. In all environments except Building B, large variation of mean errors obtained at small number of selected APs (3 to 4 in Building A, 3 to 11 in Building C, 3 to 8 in CRAWDAD, 3 to 4 in Antwerp, and 3 to 9 in Dublin) by different algorithms because the selection of APs is not the optimally configured to be used by all algorithms in all environments. The classical kNN algorithm however produces low mean error at small number of selected APs due to advantage of selecting the APs using average signal strength value. This is because the algorithm's principle also use average signal strength value (Equation 3.3) to estimate the location. The better performance of classical kNN is obviously seen in CRAWDAD building (Figure 7.6(d)). However, as the number of APs increased, the mean errors of all algorithms converge. In Antwerp environment, convergence is not seen in any algorithms because the removal of one AP from the fingerprint dataset affects the location estimation accuracy. Both proposed multiclass kNN classifiers give almost similar error pattern in all environments. Using the optimised normal multi-class kNN classifier (as the parameters given in Appendix E1), the number of APs could be reduced up to 21.4% (14 to 11), 15.4% (39 to 33), 30.6% (36 to 25), 15.4% (13 to 11), and 77.8% (36 to 8) in Building A, Building B, Building C, CRAWDAD, and Dublin respectively by using normal multiclass kNN classifier. If kernel multiclass kNN classifier is implemented, the APs could be reduced up to 35.7% (14 to 9), 15.4 % (39 to 33), 30.6% (36 to 25), 7.7% (13 to 12), and 77.8% (36 to 8) in related order of buildings. It is also interesting to see that the errors produced by both kNN classifiers of using optimised ϖ are almost similar. In some environments, it is also noted that similar accuracy as another classical algorithm could be achieved by using lower subset of APs. Normal multiclass kNN classifier similarly performs as multivariate kernel by using 22 APs compared to 29 APs in Building C and 19 APs compared to 20 APs in Antwerp. Lower mean error compared to using all APs could also be observed in Building C by using multiclass kNN classifier i.e. 1.5 m compared to 1.6 m. Another interesting thing observed is that at certain subset of APs, the error increases even though the error is already converges at lower subset of APs e.g. the mean error increases at 25 to 27 APs in Dublin dataset (Figure 7.6(f)). This is due to the algorithm finds that additional APs provide duplicate information and at the time such duplicate information influence the algorithm to provide wrong location estimation. This case is also happened to classical algorithms of kNN, univariate kernel, and multivariate kernel. Additionally, it can also be seen that in all environments kNN algorithm quickly converges at low number of APs compared to other algorithms. This is because additional APs information does not help the algorithm to discriminate different locations and consequently provide poorer estimation compared to other algorithms in most tested environments.



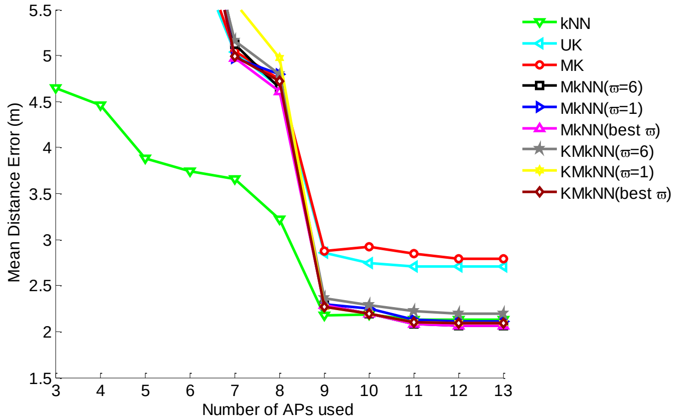
(a)



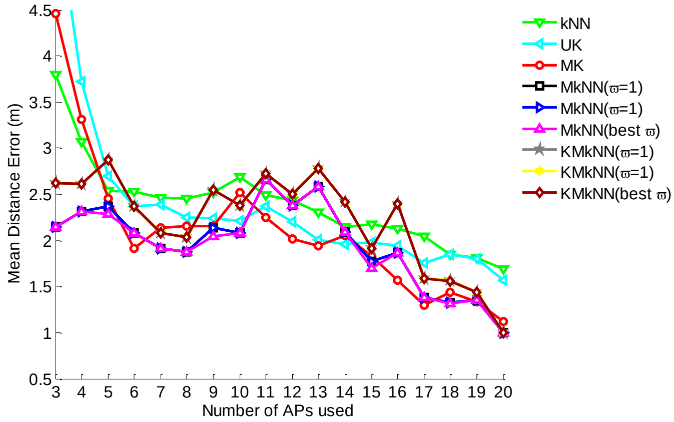
(b)



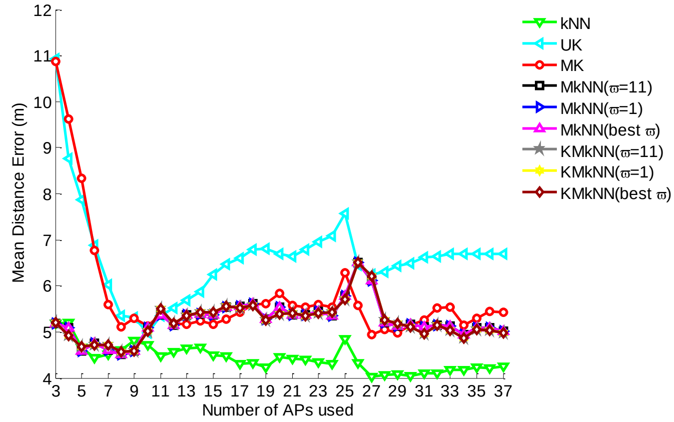
(c)



(d)



(e)



(f)

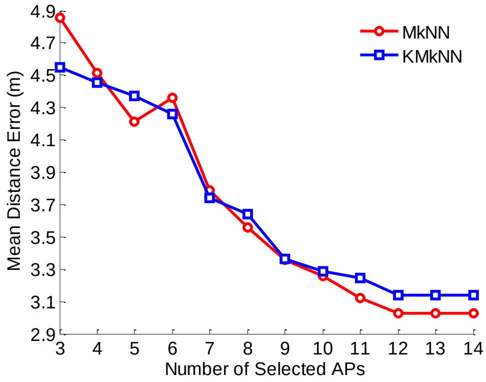
Figure 7.6. Effect of reducing number of APs based on average signal strength value at each fingerprint location for all tested algorithms in all environments of (a) Building A, (b) Building B, (c) Building C, (d) CRAWDAD, (e) Antwerp, and (f) Dublin

7.7 Combining the Proposed Max Kernel AP Selection with Multi-Class kNN Classifier Algorithms

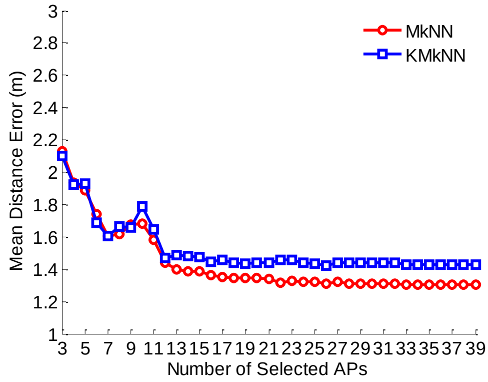
The proposed Max Kernel AP selection algorithm with multi-class kNN classifier algorithms were used to investigate the optimal subset of APs for each algorithm for localisation. Similarly as explained in Section 6.8, other algorithms are not compared in this analysis because they are not to be implemented for multi-floor localisation algorithm. Figure 7.7 plots the mean distance error against variation of subset of APs for both algorithms after applying AP selection in Building A, B, and C. It is shown that the MkNN algorithm is outperforms its kernel version in all of three tested buildings. The Max Kernel algorithm is effective to further reduce the number of AP required to obtain lowest mean error of both algorithms. Both algorithms saturate at 12 out of 14 APs, 12 out of 39 APs, and 20 out of 39 APs in Building A, B, and C respectively. This means reduction of 14.3%, 69.2%, and 44.4% of APs in each environment. Thus, computational complexity to estimate the location using kNN classifiers could be lowered as lower subset of APs could be used.

7.8 Summary

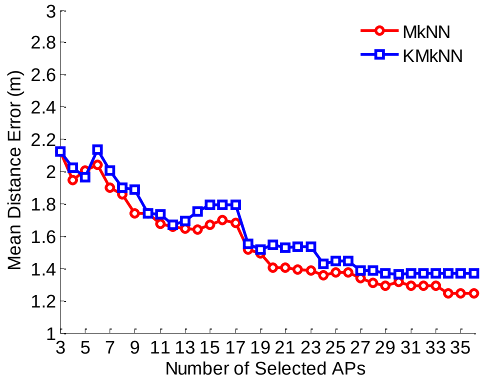
The proposed multi-class kNN classifier horizontal localisation algorithm performs better than classical horizontal localisation in multiple buildings. Unlike classical algorithms which perform better at specific environments, the proposed multi-class kNN especially the normal classifier could provide good localisation estimation in multiple tested buildings. At 2 m distance, the accuracy of estimation increased as high as 19.3% compared to classical horizontal localisation algorithms.



(a)



(b)



(c)

Figure 7.7. Mean error of both *k*NN classifier algorithms at different number of selected APs when combined with Max Kernel AP selection algorithm in (a) Building A, (b) Building B, and (c) Building C

CHAPTER 8

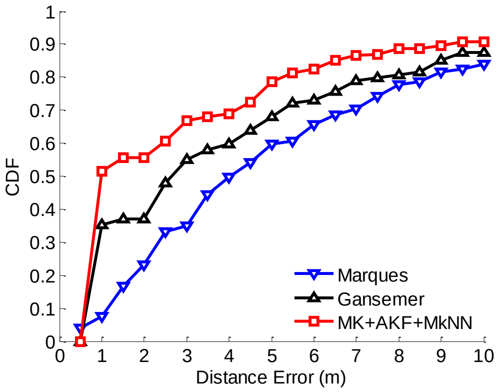
RESULTS AND DISCUSSION IV: MULTI-FLOOR LOCALISATION ALGORITHM

8.1 Introduction

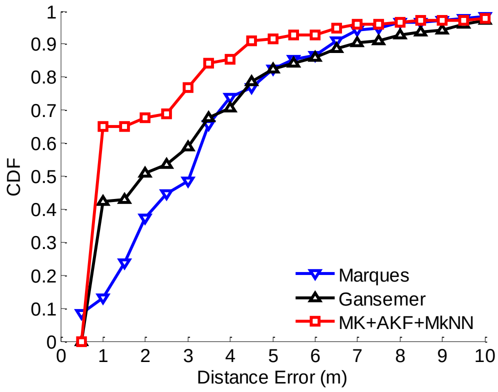
Performance of proposed AP selection, floor localisation, and horizontal localisation algorithms has been described in Chapter 5 to 7. This chapter integrates the proposed algorithms as a multi-floor algorithm. The best proposed algorithm of each respective section is chosen to investigate the performance in multi-floor building namely, Max Kernel AP selection, Averaged Kernel Floor (AKF), and normal multi-class kNN classifier. The results are compared using existing multi-floor localisation algorithms proposed by Gansemer and Marques of which implementation details was discussed in Appendix F1. Here, the parameters of Gansemer's and Marques' algorithms were chosen based on the lowest accuracy in respective scenarios. The detailed of the parameters are given in the Appendix F2. Additionally, for certain online samples processed by Gansemer and Marques algorithms could not provide estimation because the samples do not meet the criteria according to the algorithms and for fair evaluations the results for both algorithms using those samples is processed using classical kNN algorithm. The performance is studied in terms of precision and accuracy as described in Section 8.2 and processing-time of the algorithms as given in Section 8.3.

8.2 Precision and Accuracy of Proposed Multi-Floor Localisation Algorithm

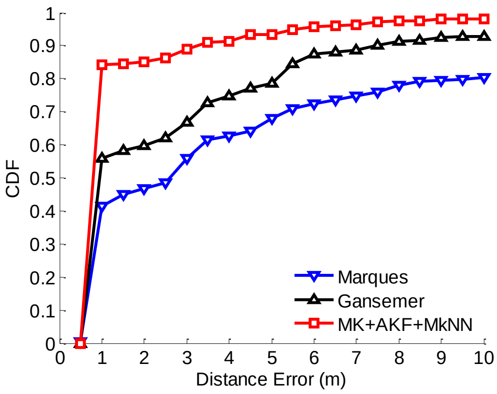
The precision of the multi-floor algorithms are investigated by plotting the cumulative density function graph (Equation 4.6) in terms of multi-floor distance error as depicted in Figure 8.1. The figure shows that the proposed multi-floor localisation algorithm which combines the Max Kernel, AKF, and MkNN algorithms is superior compared to other multi-floor localisation algorithms. This shows that the proposed algorithm could provide improvement to existing approach of multi-floor localisation. The implementation of kernel density theory for floor localisation and multi-class machine learning classification for horizontal localisation has been shown effective to improve the performance of multi-floor localisation algorithm.



(a)



(b)



(c)

Figure 8.1. Multi-floor mean errors of multi-floor algorithms tested in (a) Building A, (b) Building B, (c) Building C

The detailed error measure of the CDF is tabulated in Table 8.1. At 90% of CDF, the distance error is reduced up to 15.4% (11.0 m to 9.3 m), 33.3% (6.3 m to 4.2 m), and 54.7% (7.5 m to 3.4 m) respectively in Building A, B, and C by employing the proposed multi-floor localisation algorithm. The mean distance error (Equation 4.4) of the proposed algorithm is the lowest in all tested environments. The reduction of mean distance error is up to 21.3% (4.7 m to 3.7 m), 29.0% (3.1 m to 2.2 m), and 48.5% (3.3 m to 1.7m) for Building A, B, and C.

Table 8.1. Error measure of Gansemer, Marques, and proposed combined Max Kernel, AKF, and MkNN algorithms

Dataset	Algorithm	Error measures				
		Mean	Standard Deviation	Median	90% of CDF	95% of CDF
A	Marques	6.2	7.0	4.0	14.2	18.1
	Gansemer	4.7	5.4	2.9	11.0	14.3
	MK+AKF+MkNN	3.7	5.0	1.0	9.3	13.3
B	Marques	3.3	2.4	3.2	6.3	7.5
	Gansemer	3.1	3.0	1.7	6.7	9.3
	MK+AKF+MkNN	2.2	2.5	1.0	4.2	6.5
C	Marques	5.7	7.7	2.7	16.7	20.2
	Gansemer	3.3	4.9	1.0	7.5	11.5
	MK+AKF+MkNN	1.7	2.6	1.0	3.4	5.5

From CDF plot of Figure 8.1, the result of accuracy within fingerprint spacing i.e. at 2 m is extracted as illustrated in Figure 8.2 using Equation 4.7. In agreement with the CDF plot, the proposed algorithm shows excellent performance compared the other two multi-floor localisation algorithms. Improved accuracy of 49.6%, 36.8% and 43.5% within 2 m respectively in Building A, B, and C can be achieved by the proposed algorithm when compared to other multi-floor localisation algorithm.

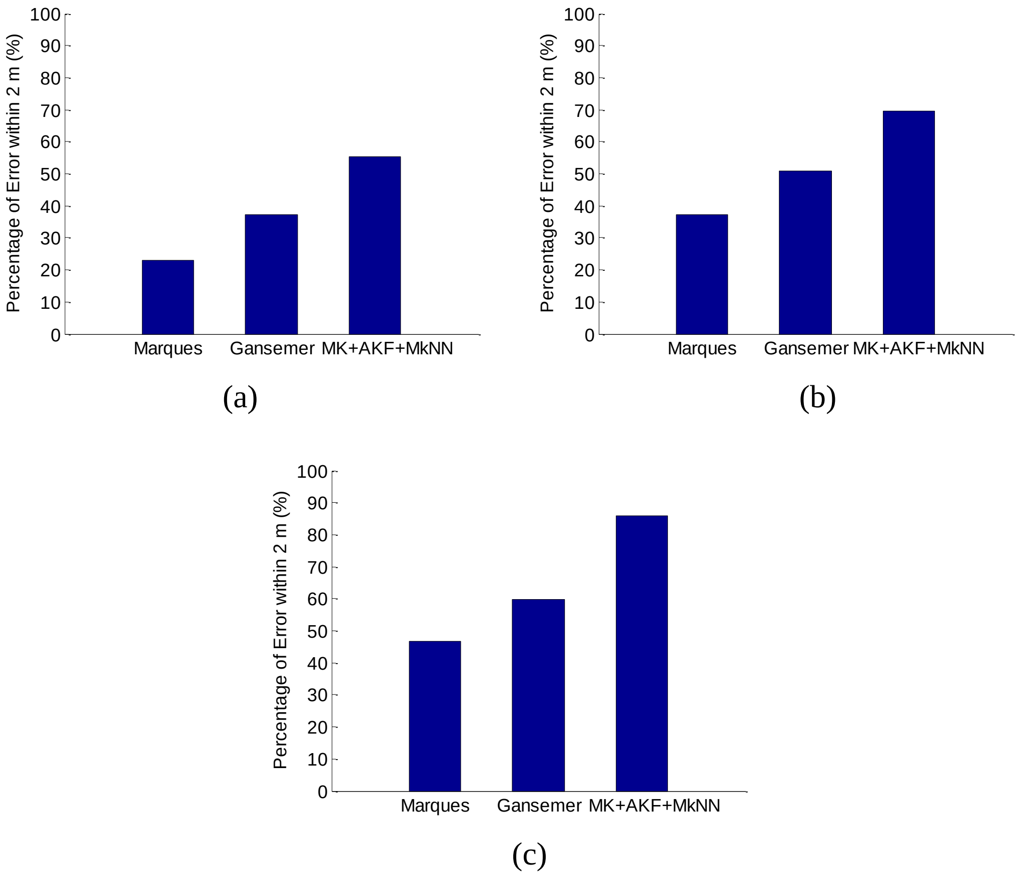


Figure 8.2. Localisation accuracy within fingerprint spacing of Marques, Gansemer, and proposed multi-floor location algorithm in (a) Building A, (b) Building B, and (c) Building C

8.3 Computational Complexity of the Proposed Multi-Floor Localisation Algorithm

The computational complexity or processing time of the algorithm is investigated using an Intel-i5 3.2 GHz desktop computer as described in Chapter 4. The result of computational complexity of the multi-floor localisation algorithms are compared in Figure 8.3. The results are based on averaging the computation time of the algorithm after 100 times of processing. The proposed multi-floor algorithm shows very good computation time in all buildings. The computational complexity of the proposed multi-floor localisation algorithm is tremendously reduced by the because of application of Max Kernel AP selection algorithm for both floor and horizontal localisation algorithm and also the signal strength clustering for floor localisation algorithm. The proposed multi-floor localisation algorithm could obtain 24.4, 7.2, and 15.2 times of reduced computation time in Building A, B, and C respectively compared to Gansemer algorithm. However, the time of Marques algorithm is not far to the proposed algorithm in Building B and C. This is due to the filtering and restructuring of the fingerprint elements which reduce the number of elements to be processed by the localisation algorithm. The number of elements in those buildings is smaller compared to the number of elements produced by Max Kernel and signal strength clustering thus resulting lower processing time. However, the filtering degrades the performance of Marques algorithm as previously discussed in Section 8.2. On the other hand, Gansemer algorithm attains high processing time because the each of the fingerprint element is extracted to multiple elements according to the direction of measuring device which increase the elements to be processed by the algorithm.

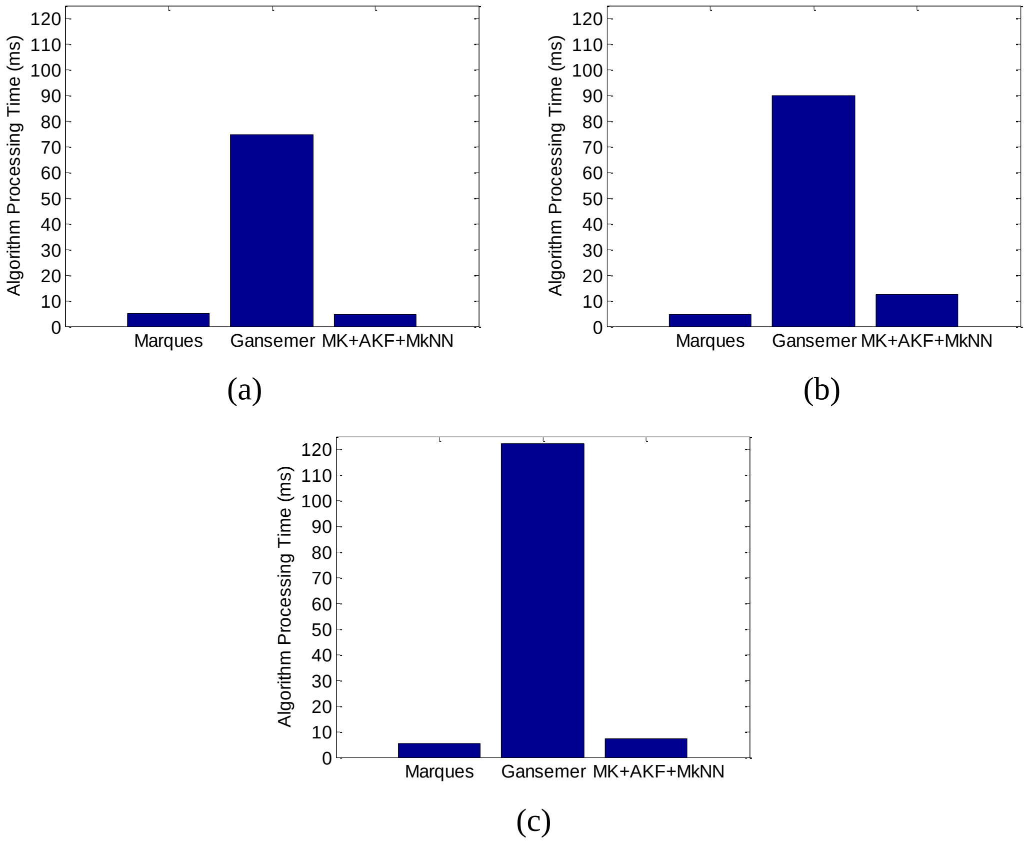


Figure 8.3. Processing time of each multi-floor localisation algorithm in (a) Building A, (b) Building (b), and (c) Building C

8.4 Summary

The combination of Max Kernel, AKF and MkNN algorithm has shown advantage in improving the accuracy and precision in estimating the location in multi-floor setting. The multi-floor mean distance error could be reduced up to 48.5% compared to existing multi-floor localisation algorithm. The accuracy of the estimation within 2 m distance is also could be achieved up to 85.8%. At the same time, the algorithm acquires lower processing time of up to 24.4 times compared to existing multi-floor algorithm reduces the computation cost of developing the multi-floor localisation system.

CHAPTER 9

CONCLUSIONS AND FUTURE WORKS

9.1 Summary of The Thesis

The main objective of this thesis is to develop a time-efficient and accurate multi-floor WLAN localisation system. The objective is achieved by developing novel algorithms based on theory of kernel density estimation and multi-class machine learning classification for the multi-floor localisation algorithms.

Existing WLAN multi-floor localisation system is either computationally complex or inaccurate. This is because of non-optimised fingerprint database is applied in the localisation. Using original database without optimisation burdens the computational process as every single fingerprint elements needs to be computed during online localisation phase especially for large database of multi-floor environments. Additionally, some AP signals could give redundant information which sometimes degrades the localisation accuracy. On the other hand, the localisation algorithms for detecting the floor and localising on horizontal locations mainly has been worked around similar type of algorithm i.e. similarity function. Majority of multi-floor localisation system based on WLAN has acceptable localisation accuracy but still requires improvement.

The theory of kernel density estimate has only been used for horizontal localisation algorithm e.g. univariate and multivariate kernel algorithm but has not been applied for floor localisation and AP selection. In this work, kernel density is applied for floor localisation algorithm by averaging the kernel density estimate of clustered fingerprint elements according to floor level of the building. , the *Maximum A*

Posteriori (MAP) calculation of kernel density is used to determine APs either weak or strong which are then selected for positioning. Machine learning classification technique using Support Vector Machine (SVM) has been shown promising for indoor localisation application for horizontal AP selection and localisation algorithm. In this work, another machine learning classification approach based on Kernel Logistic Regression (KLR) is investigated for application in multi-floor localisation. KLR method is applied for AP selection and floor localisation. The AP selection problem has been addressed by taking advantage of the Negative Likelihood (NLL) function of multinomial KLR. The APs are sorted and selected according to increasing value of NLL for pairwise AP signals found at all fingerprint locations. The floor localisation is solved by computing the sigmoid function of the KLR according to the value of parameter w of the function which is trained based on clustered fingerprints. Another type of classifier by kNN is also studied to improve the horizontal localisation performance. Both kernel and linear kNN classifiers are applied. The modified multi-class kNN classifier using optimised parameters has successfully been implemented to address horizontal localisation problem.

Performance of the algorithms has been tested in multiple WLAN environment scenarios. The measurement of dataset has been carried out in three multi-story buildings of different heights and structures. Additionally, the algorithms are also tested using three single-floor dataset. The WLAN scenarios of the datasets are different from each other with different placement of APs and floor plan. The measured signals are also verified using statistical path loss model to validate feasibility of the signal used in the development the localisation algorithm.

The develop algorithms have shown excellent performance in reducing the computational complexity and increase localisation accuracy for both floor and horizontal location. The performance of each algorithm has been independently tested and compared with related algorithms for AP selection, floor localisation and horizontal localisation cases. The performance of both floor and horizontal localisation algorithms has been tested in terms of accuracy, precision, varying the number of training samples, and reducing the number of APs. For AP selection, reduction of 77.8% of AP is seen using proposed Max Kernel algorithm. In some cases, the accuracy could be slightly improved further at lower subset of AP. Also, The proposed KLPD AP selection algorithm which utilise variant AP selection reduce the computational time of up to 21.6% of classical AP selection technique. The proposed KLF floor localisation could achieve high floor accuracy of 93.4% in one of the tested building which is at least 14.4% more accurate compared to existing floor localisation algorithms. As the proposed floor localisation utilising signal strength clustering technique, the computational complexity of the algorithm could be reduced further where AKF algorithm demonstrated 14 times lower processing time compared to existing approach. The horizontal localisation algorithm based on an enhanced multi-class kNN classifier also improved estimation of horizontal position where the distance error at 2 m increased 19.3% compared to other existing algorithms. The best developed algorithms of AP selection, floor detection, and horizontal localisation namely Max Kernel, Averaged Kernel Floor (AKF), and multi-class k-nearest neighbour (MkNN) classifier algorithms are combined to perform location estimation in multi-floor setting. The performance of the combined algorithms is tested both in terms of accuracy and time-efficiency and the results are compared with existing multi-floor localisation algorithms. The multi-floor accuracy within 2 m in one of the tested building could be achieved at 88.5% which shows improvement of 43.5% compared to existing multi-floor algorithm. At the same time, the processing time is 15.2 times lower than the existing algorithm.

In the next section, main contributions arise from the work are highlighted and explained. Finally, recommendations for further improvement of the multi-floor WLAN localisation system are described.

9.2 Contributions

The development of the proposed multi-floor positioning system has shown superiority in terms of accuracy and processing time by application of new theory and knowledge to improve the system. Application of the theory and knowledge implies areas of novelty in this thesis and the following research contributions are made and highlighted:

1. Multiple multi-floor signal fingerprint datasets are analysed to provide statistics and the path loss parameters are extracted using Seidel's model to validate the signal's propagation and also to confirm reliability of the dataset for indoor localisation.

2. Three popular classical positioning algorithms which are k-nearest neighbour, univariate kernel and multivariate kernel in are evaluated in different WLAN environments including multi-floor environments. The positioning errors of the algorithms are compared and analysed. The algorithms are also used for comparison with the develop algorithms in AP selection and horizontal localisation.

3. Theory of kernel density estimates is proposed and analysed for AP selection and floor detection. For AP selection, APs are sorted according to group of strong and weak which are calculated based on density estimate. For floor detection, the fingerprint elements are clustered according to AP with strongest signal strength and the kernel density estimates of grouped fingerprints are calculated and averaged. The application of the kernel density in floor localisation decreases the computation time of the developed algorithm compared to existing algorithms.

4. Logistic regression technique is introduced for the first time for WLAN localisation problem. The logistic regression technique is applied for AP selection problem and floor localisation. For AP selection, the information of negative likelihood function is exploited to choose the AP. For floor localisation, the classification of the floor level is determined based on calculating the sigmoid function with trained value of w vector.

5. The proposed AP selection performance has been tested and analysed not only in multi-floor environment but also in single-floor environment. The floor localisation algorithm is tested in three different multi-floor buildings. The comparison is made with existing AP selection and floor localisation algorithms. It is shown that the proposed algorithms work well in the environments.

6. New normal and kernel kNN classifier for horizontal positioning is investigated and the performance is analysed. The technique is developed based on machine learning classification theory and the proposed algorithms are optimised with tuneable parameters. The performance of both techniques is compared with the

classical algorithms. The proposed algorithms in general show improvements in terms of localisation accuracy.

7. The three proposed algorithms i.e. the AP selection, floor localisation, and horizontal localisation are combined together for multi-floor indoor positioning. The performance of the proposed multi-floor positioning technique is compared with existing multi-floor positioning algorithms.

9.3 Recommendations for Future Works

Further directions of research could be described as follows:

1. The tracking algorithm to locate device in motion state could be developed and analysed. Existing tracking algorithms has mainly been applied for single-floor location but using other sensors e.g. inertial sensor to perform the tracking. The tracking algorithm using solely WLAN signals is challenging and available algorithms are limited.

2. The measurements and tests made in this work are based on single device. It has been shown that different devices in WLAN environment give different readings of signal strength value. Introduction of multiple devices in the environments requires different type of algorithm to mitigate variation of the signal strength before localisation algorithm is implemented.

3. The APs installed within the investigated building is not optimised for localisation purpose. Enhancing the placement and location configuration of

the APs could boost the performance of localisation algorithm. Also, effect of signal interference of APs operating within the same and adjacent channel could also be considered.

4. Introduction of multiple additional devices requesting for location requires recurring computation of the algorithm which may burden the computational processing core. A new approach by co-locating the device based on inferring the location of the additional device according to existing device in the environment may speed up the computation and using less computation resource. A novel co-locating technique by computing the signal data similarity between devices could be investigated.

5. The algorithms developed in this thesis are based on WLAN signal strength data. The accuracy of the localisation could be further improved if the channel impulse response of the WLAN signal is considered. However, the measurement and testing of the localisation performance should be made using special measuring equipment such as vector network analyser.

6. The developed multi-floor localisation algorithm requires active participation of the user by requesting the location. However, for surveillance purpose the administrator of the building could not depend on user's participation. A passive device-free localisation technique could be developed without requiring direct communication with the device. The technique could be studied by installing monitoring APs to measure the signal transmitted from any device inside the buildings.

7. The computation of the algorithms in this thesis has been made using MATLAB software. However to improve the computational speed and for easier development of the real system, lower level language such as C, C++, or Java could be applied

REFERENCES

Ahmad, U., A. Gavrilov, Lee Sungyoung, and Lee Young-Koo. 2006. "Modular Multilayer Perceptron for WLAN Based Localization." Neural Networks, 2006. IJCNN '06. International Joint Conference on, 0-0 0.

Ahmad, U., B. J. D'Auriol, Y.-K. Lee, and S. Lee. 2008. "Multi-Floor Semantically Meaningful Localization Using IEEE 802.11 Network Beacons." *IEICE Transactions on Communications* E91-B:3450-3460. doi: 10.1093/ietcom/e91-b.11.3450.

Akl, Robert G, Dinesh Tummala, and Xinrong Li. 2006. "Indoor propagation modeling at 2.4 GHz for IEEE 802.11 networks." The Sixth IASTED International Multi-Conference on Wireless and Optical Communications, July 3-5.

Al-Ahmadi, Abdullah Saad Mohammed, Abdusamea. I. A. Omer, Muhammad Ramlee Bin Kamarudin, and Tharek Abdul bin Abd Rahman. 2010. "MULTI-FLOOR INDOOR POSITIONING SYSTEM USING BAYESIAN GRAPHICAL MODELS." *Progress In Electromagnetics Research B* 25:241-259. doi: 10.2528/PIERB10081202.

Alsehly, Firas, Tughrul Arslan, and Zankar Sevak. 2011. "Indoor positioning with floor determination in multi story buildings." *2011 International Conference on Indoor Positioning and Indoor Navigation*:1-7. doi: 10.1109/IPIN.2011.6071945.

Arya, Azin, Philippe Godlewski, Marine Campedel, and Ghislain Du Chéne. 2013. "Radio database compression for accurate energy-efficient localization in fingerprinting systems." *IEEE Transactions on Knowledge and Data Engineering* 25:1368-1379. doi: 10.1109/TKDE.2011.241.

Bahl, P., and V.N. Padmanabhan. 2000. "RADAR: an in-building RF-based user location and tracking system." *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)* 2. doi: 10.1109/INFCOM.2000.832252.

Bhargava, Preeti, Shivsubramani Krishnamoorthy, and Aditya Karkada. 2013. "Locus: An Indoor Localization, Tracking and Navigation System for Multi-story Buildings Using Heuristics Derived from Wi-Fi Signal Strength." 120:212-223. doi: 10.1007/978-3-642-40238-8.

Bolliger, Philipp Lukas. 2011. "Robust indoor positioning through adaptive collaborative labeling of location fingerprints." Doctor of Sciences PhD Dissertation, ETH ZURICH (20109).

Campos, Rafael Saraiva, Lisandro Lovisolo, and Marcello Luiz R. de Campos. 2014. "Wi-Fi multi-floor indoor positioning considering architectural aspects and controlled computational complexity." *Expert Systems with Applications* 41:6211-6223. doi: 10.1016/j.eswa.2014.04.011.

Chen, Qiuyun, and Bang Wang. 2015. "FinCCM: Fingerprint Crowdsourcing, Clustering and Matching for Indoor Subarea Localization." *IEEE Wireless Communications Letters* 4:1-1. doi: 10.1109/LWC.2015.2482971.

Chen, Yiqiang, Qiang Yan, Jie Yin, and Xiaoyong Chai. 2006. "Power-efficient Access-Point selection for indoor location estimation." *IEEE Transactions on Knowledge and Data Engineering* 18:877-888. doi: 10.1109/TKDE.2006.112.

Cheng, Jiujun, Yueqiao Cai, Qingyang Zhang, Junlu Cheng, and Chendan Yan. 2014. "A New Three-Dimensional Indoor Positioning Mechanism Based on Wireless LAN." *Mathematical Problems in Engineering* 2014:1-7. doi: 10.1155/2014/862347.

Chintalapudi, Krishna, Anand Padmanabha Iyer, and Venkata N. Padmanabhan. 2010. "Indoor localization without the pain." *Proceedings of the sixteenth annual international conference on Mobile computing and networking - MobiCom '10*:173. doi: 10.1145/1859995.1860016.

Fang, Shih-Hau Fang Shih-Hau, and Tsung-Nan Lin Tsung-Nan Lin. 2010a. "Accurate Indoor Location Estimation by Incorporating the Importance of Access Points in Wireless Local Area Networks." *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*:1-5. doi: 10.1109/GLOCOM.2010.5684309.

Fang, Shih-hau, and Tsung-nan Lin. 2009. "Projection-Based Location System via Multiple Discriminant Analysis in Wireless Local Area Networks." *IEEE Transactions on Vehicular Technology* 58:5009-5019. doi: 10.1109/TVT.2009.2025134.

Fang, Shih-Hau, and Tsung-Nan Lin. 2010b. "A dynamic system approach for radio location fingerprinting in wireless local area networks." *IEEE Transactions on Communications* 58:1020-1025. doi: 10.1109/TCOMM.2010.04.090080.

Fang, Shih Hau, and Tsung-nan Lin. 2012. "Principal component localization in indoor wlan environments." *IEEE Transactions on Mobile Computing* 11:100-110. doi: 10.1109/TMC.2011.30.

Fang, Shih Hau, and Tsung Nan Lin. 2010c. "Cooperative multi-radio localization in heterogeneous wireless networks." *IEEE Transactions on Wireless Communications* 9:1547-1551. doi: 10.1109/TWC.2010.05.090061.

Farid, Zahid, Rosdiadee Nordin, and Mahamod Ismail. 2013. "Recent Advances in Wireless Indoor Localization Techniques and System." *Journal of Computer Networks and Communications* 2013:1-12. doi: 10.1155/2013/185138.

Gansemer, S., S. Hakobyan, S. Püschel, and U. Großmann. 2009. "3D WLAN indoor positioning in multi-storey buildings." *Proceedings of the 5th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2009*:669-672. doi: 10.1109/IDAACS.2009.5342893.

Gansemer, Sebastian, Uwe Grossmann, and Syuzanna Hakobyan. 2010. *RSSI-based Euclidean Distance algorithm for indoor positioning adapted for the use in dynamically changing WLAN environments and multi-level buildings*: IEEE.

Genming, Ding, Tan Zhenhui, Zhang Jinbao, and Zhang Lingwen. 2013. "Fingerprinting localization based on affinity propagation clustering and artificial neural networks." Wireless Communications and Networking Conference (WCNC), 2013 IEEE, 7-10 April 2013.

Gu, Yu, and Fuji Ren. 2015. "Energy-efficient Indoor Localization of Smart Hand-held Devices Using Bluetooth." *IEEE Access* PP:1-1. doi: 10.1109/ACCESS.2015.2441694.

Gupta, Pankaj, Sachin Bharadwaj, Sthanunathan Ramakrishnan, and Jaiganesh Balakrishnan. 2014. "Robust floor determination for indoor positioning." *2014 Twentieth National Conference on Communications (NCC)*:1-6. doi: 10.1109/NCC.2014.6811285.

He, Suining, and S.-H. Gary Chan. 2015. "Wi-Fi Fingerprint-based Indoor Positioning: Recent Advances and Comparisons." *IEEE Communications Surveys & Tutorials*:1-1. doi: 10.1109/COMST.2015.2464084.

Honkavirta, Ville, Tommi Perala, Simo Ali-Loytty, and Robert Piche. 2009. *A comparative survey of WLAN location fingerprinting methods*: IEEE.

Hossain, A K M Mahtab, and Wee-seng Soh. 2007. *A Comprehensive Study of Bluetooth Signal Parameters for Localization*: IEEE.

Jianyong, Zhu, Luo Haiyong, Chen Zili, and Li Zhaohui. 2014. "RSSI based Bluetooth low energy indoor positioning."526-533. doi: 10.1109/IPIN.2014.7275525.

Khaoampai, Kornkanok, Kulit Na Nakorn, and Kultida Rojviboonchai. 2014. "Low complexity floor localization algorithm for mobile phone." *2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*:1-6. doi: 10.1109/ECTICon.2014.6839765.

King, Thomas, Stephan Kopf, Thomas Haenselmann, Christian Lubberger, and Wolfgang Effelsberg. 2006. "COMPASS: A probabilistic indoor positioning system based on 802.11 and digital compasses." *Online*:34-40. doi: 10.1145/1160987.1160995.

Kung, Sun Yuan. 2014. *Kernel methods and machine learning*: Cambridge University Press.

Kushki, Azadeh, Konstantinos N Plataniotis, A N Venetsanopoulos, and Senior Student Member. 2007. "Kernel-Based Positioning in Wireless Local Area Networks." *Mobile Computing, IEEE Transactions on* 6:689-705. doi: 10.1109/TMC.2007.1017.

Laoudias, C., P. Kemppi, and C. G. Panayiotou. 2009. "Localization Using Radial Basis Function Networks and Signal Strength Fingerprints in WLAN." Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE, Nov. 30 2009-Dec. 4 2009.

Laoudias, Christos, George Constantinou, Marios Constantinides, Silouanos Nicolaou, Demetrios Zeinalipour-Yazti, and Christos G Panayiotou. 2012. "The airplace indoor positioning platform for android smartphones." Mobile Data Management (MDM), 2012 IEEE 13th International Conference on.

Letchner, Julie, Dieter Fox, and Anthony LaMarca. 2005. "Large-scale localization from wireless signal strength." Proceedings of the national conference on artificial intelligence.

Liang, Dong, Zhaojing Zhang, and Mugen Peng. 2015. "Access Point Reselection and Adaptive Cluster Splitting-Based Indoor Localization in Wireless Local Area Networks." *IEEE Internet of Things Journal* 2:573-585. doi: 10.1109/JIOT.2015.2453419.

Lin, Tsung-nan, Senior Member, Shih-hau Fang, Wei-han Tseng, Chung-wei Lee, and Jeng-wei Hsieh. 2014. "A Group-Discrimination-Based Access Point Selection for WLAN Fingerprinting Localization." 63:3967-3976.

Liu, Hui, Houshang Darabi, Pat Banerjee, and Jing Liu. 2007. "Survey of wireless indoor positioning techniques and systems." *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 37:1067-1080. doi: 10.1109/TSMCC.2007.905750.

Liu, Hung-Huan, Chung-Wei Liao, and Wei-Hsiang Lo. 2015. "WiFi-based Floor Estimation for Multi-Storey Building: A Case Study for a Crowded AP Environment." Intelligent Systems and Applications: Proceedings of the International Computer Symposium (ICS) Held at Taichung, Taiwan, December 12–14, 2014.

Liu, Hung-Huan, and Yu-Non Yang. 2011. *WiFi-based indoor positioning for multi-floor Environment*.

Lohan, Elena Simona, Jukka Talvitie, Pedro Figueiredo e Silva, Henri Nurminen, Simo Ali-Loytty, and Robert Piche. 2015. *Received signal strength models for WLAN and BLE-based indoor positioning in multi-floor buildings*: IEEE.

Maneerat, Kriangkrai, and Chutima Prommak. 2014. "An Enhanced Floor Estimation Algorithm for Indoor Wireless Localization Systems Using Confidence Interval Approach."1058-1062.

Maneerat, Kriangkrai, Chutima Prommak, and Kamol Kaemarungsi. 2014. *Floor estimation algorithm for wireless indoor multi-story positioning systems*: IEEE.

Marques, Nelson, Filipe Meneses, and Adriano Moreira. 2012. *Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning*: IEEE.

Marsland, Stephen. 2014. *Machine learning: an algorithmic perspective*: CRC press.

Martinez, Wendy L, and Angel R Martinez. 2007. *Computational statistics handbook with MATLAB*: CRC press.

Mirowski, Piotr, Dimitrios Milioris, Philip Whiting, and Tin Kam Ho. 2014. "Probabilistic Radio-Frequency Fingerprinting and Localization on the Run." *Bell Labs Technical Journal* 18:111-133. doi: 10.1002/bltj.21649.

Mu, Zhou, Zhang Qiao, Tian Zengshan, Qiu Feng, and Wu Qi. 2014. "Integrated location fingerprinting and physical neighborhood for WLAN probabilistic localization." Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on, 11-13 July 2014.

Murphy, Kevin P. 2012. *Machine learning: a probabilistic perspective*: MIT press.

Park, Jun-geun, Ben Charrow, Dorothy Curtis, Jonathan Battat, Einat Minkov, Jamey Hicks, Seth Teller, and Jonathan Ledlie. 2010. "Growing an organic indoor location system." *Proceedings of the 8th international conference on Mobile systems applications and services MobiSys 10*:271. doi: 10.1145/1814433.1814461.

Petrellis, N, N Konofaos, and G P Alexiou. 2006. "Target localization utilizing the success rate in infrared pattern recognition." *Ieee Sensors Journal* 6:1355-1364. doi: Doi 10.1109/Jsen.2006.881342.

Prieto, J, S Mazuelas, A Bahillo, P Fernandez, R M Lorenzo, and E J Abril. 2012. "Adaptive Data Fusion for Wireless Localization in Harsh Environments." *IEEE Transactions on Signal Processing* 60:1585-1596. doi: 10.1109/TSP.2012.2183126.

Rappaport, Theodore S. 1996. *Wireless communications: principles and practice*. Vol. 2: prentice hall PTR New Jersey.

Razavi, Alireza, Mikko Valkama, and Elena-Simona Lohan. 2015. "K-Means Fingerprint Clustering for Low-Complexity Floor Estimation in Indoor Mobile Localization."

Redzic, Milan D., Conor Brennan, and Noel E. O'Connor. 2014. "SEAMLOC: Seamless indoor localization based on reduced number of calibration points." *IEEE Transactions on Mobile Computing* 13:1326-1337. doi: 10.1109/TMC.2013.107.

Roos, Teemu, Petri Myllyma, Henry Tirri, Pauli Misikangas, and Juha Sieva. 2002. "A Probabilistic Approach to WLAN User Location Estimation." 9.

Scott, David W. 2015. *Multivariate density estimation: theory, practice, and visualization*: John Wiley & Sons.

Seidel, Scott Y, and Theodore S Rappaport. 1992. "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings." *Antennas and Propagation, IEEE Transactions on* 40 (2):207-217.

Shih-Hau, Fang, and Lin Tsung-Nan. 2008. "Indoor Location System Based on Discriminant-Adaptive Neural Network in IEEE 802.11 Environments." *Neural Networks, IEEE Transactions on* 19 (11):1973-1978. doi: 10.1109/TNN.2008.2005494.

Snedecor, George W, and Witiiam G Cochran. 1989. "Statistical methods 8. ed." *Iowa State Univ*.

Sorour, Sameh, Yves Lostanlen, Shahrokh Valaee, and Khaqan Majeed. 2015. "Joint Indoor Localization and Radio Map Construction with Limited Deployment Load." *IEEE Transactions on Mobile Computing* 14:1031-1043. doi: 10.1109/TMC.2014.2343636.

Sun, Lin, Zengwei Zheng, Tao He, and Fei Li. 2015. "Multifloor Wi-Fi Localization System with Floor Identification." 2015.

Tao, Shuai, Mineichi Kudo, Bing-Nan Pei, Hidetoshi Nonaka, and Jun Toyama. 2014. "Multiperson Locating and Their Soft Tracking in a Binary Infrared Sensor Network." *IEEE Transactions on Human-Machine Systems* 45:1-12. doi: 10.1109/THMS.2014.2365466.

Thomas, King, Kopf Stephan, Haenselmann Thomas, Lubberger Christian, and Effelsberg Wolfgang. 2008. CRAWDAD dataset mannheim/compass (v. 2008-04-11).

Tran, D. A., and Nguyen Thinh. 2008. "Localization In Wireless Sensor Networks Based on Support Vector Machines." *Parallel and Distributed Systems, IEEE Transactions on* 19 (7):981-994. doi: 10.1109/TPDS.2007.70800.

Wang, B, S Zhou, W Liu, and Y Mo. 2015. "Indoor Localization Based on Curve Fitting and Location Search Using Received Signal Strength." *Industrial Electronics, IEEE Transactions on* 62:572-582. doi: 10.1109/TIE.2014.2327595.

Wang, Hua-Yan, V.W. Zheng, Junhui Zhao, and Qiang Yang. 2010. *Indoor localization in multi-floor environments with reduced effort*: IEEE.

Wu, Chenshu, Zheng Yang, Yunhao Liu, and Wei Xi. 2013. "WILL: Wireless indoor localization without site survey." *IEEE Transactions on Parallel and Distributed Systems* 24:839-848. doi: 10.1109/TPDS.2012.179.

Yijun, Mo, Cao Zuo, and Wang Bang. 2012. "Occurrence-Based Fingerprint Clustering for Fast Pattern-Matching Location Determination." *Communications Letters, IEEE* 16 (12):2012-2015. doi: 10.1109/LCOMM.2012.111412.121909.

Youssef, M.a., a. Agrawala, and a. Udaya Shankar. 2003. "WLAN location determination via clustering and probability distributions." *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).*143-150. doi: 10.1109/PERCOM.2003.1192736.

Youssef, Moustafa. 2004. "HORUS: A WLAN-based indoor location determination system." Doctor of Philosophy, Department of Computer Science, University of Maryland.

Youssef, Moustafa A, Ashok Agrawala, A Udaya Shankar, and Sam H Noh. 2002. "A Probabilistic Clustering-Based Indoor Location Determination System Location Determination System Architecture The Joint Clustering Technique."

Zhu, Ji, and Trevor Hastie. 2005. "Kernel logistic regression and the import vector machine." *Journal of Computational and Graphical Statistics* 14 (1).

APPENDIX A1

ITERATIVE REWEIGHTED LEAST SQUARE OPTIMISATION ALGORITHM FOR NEGATIVE LIKELIHOOD FUNCTION OF LOGISTIC REGRESSION

The working principle of IRLS algorithm is to find the convergence value of $\boldsymbol{w}$ by iteratively updating the gradient and Hessian of NLL function given as follows (Zhu and Hastie 2005):

$$\boldsymbol{w}^{(\mathbb{k}+1)} = \boldsymbol{w}^{(\mathbb{k})} - \mathbf{H}^{-1}\mathrm{g}^{(\mathbb{k})} \tag{A1.1}$$

where $\mathbf{H}$ is the second-derivatives of the NLL function or Hessian and g is the gradient or first derivative of NLL function, and $\mathbb{k}$ is number of iteration. First derivation of NLL function gives the gradient of the function given (Zhu and Hastie 2005):

$$\mathrm{g} = \frac{d}{d\boldsymbol{w}}\mathrm{NLL}(\boldsymbol{w}) = \boldsymbol{K}^{\mathrm{T}}\mathbf{W}\boldsymbol{z} \tag{A1.2}$$

where

$$\boldsymbol{z} = \boldsymbol{K}\boldsymbol{w}^{(\mathbb{k})} + \boldsymbol{W}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \tag{A1.3}$$

and

$$\boldsymbol{W} = diag\left[\mu_\rho\left(1 - \mu_\rho\right)\right]_{\Phi\times\Phi} \tag{A1.4}$$

with

$$\mu_\rho = \mathrm{p}(y_\rho|\boldsymbol{K}_\rho, \boldsymbol{w}_\rho) \tag{A1.5}$$

where T is the transpose function, $\boldsymbol{K}$ is the kernel function, ρ is each of the number of class label for total of Φ labels or vector signal data existed in a pairwise classes, and $\boldsymbol{\mu} = \mu_{1:\Phi}$. The hessian is derived from gradient as follows:

$$\mathrm{H} = \frac{d}{d\boldsymbol{w}}\mathrm{g}(\boldsymbol{w})^{\mathrm{T}} = \boldsymbol{K}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{K} + \lambda\boldsymbol{K} \tag{A1.6}$$

where λ is the regularisation parameter of a small value ($\lambda = 0.001$).

APPENDIX B1

PHOTOS OF ACTUAL BUILDINGS OF MEASUREMENT LOCATIONS

Figure B1.1. Building A

Figure B1.2. Building B

Figure B1.3. Building C

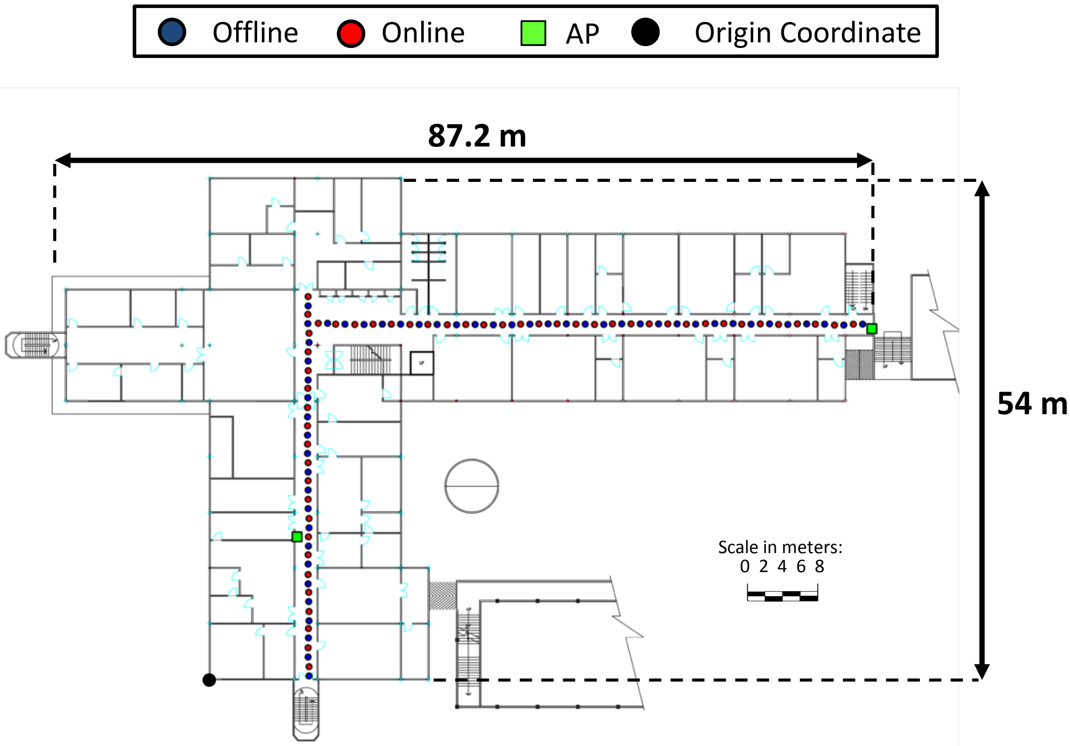
Figure B1.4. Antwerp. The measurements are made at the seventh floor level of the building (top most floor) of the building as marked in red colour.

Figure B1.5. Dublin

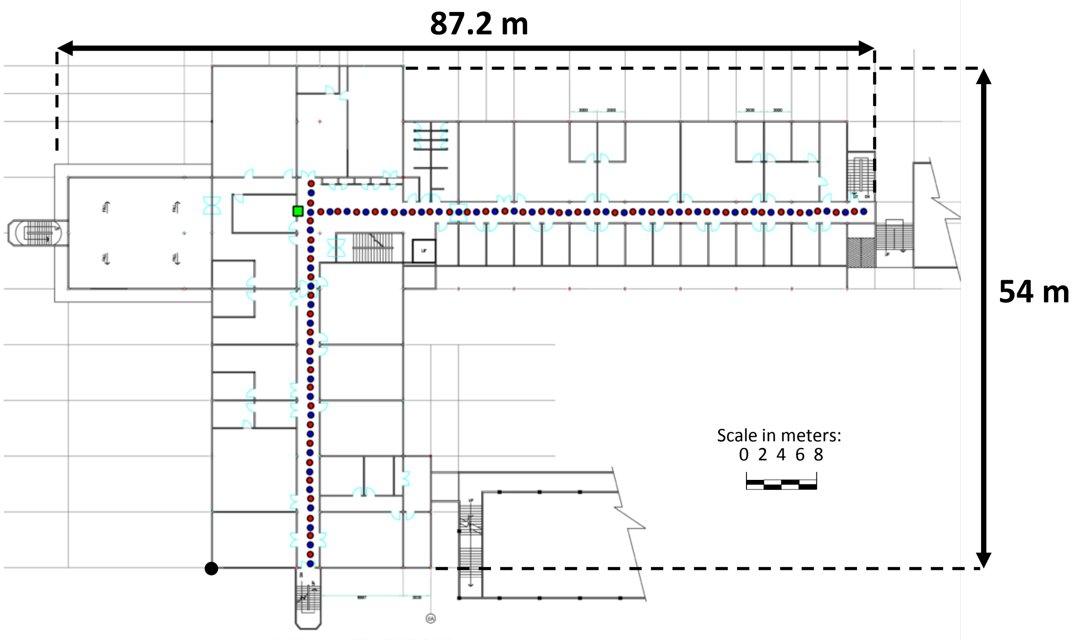
APPENDIX B2

FLOOR PLAN OF FIRST FLOOR AND ABOVE OF MEASUREMENT

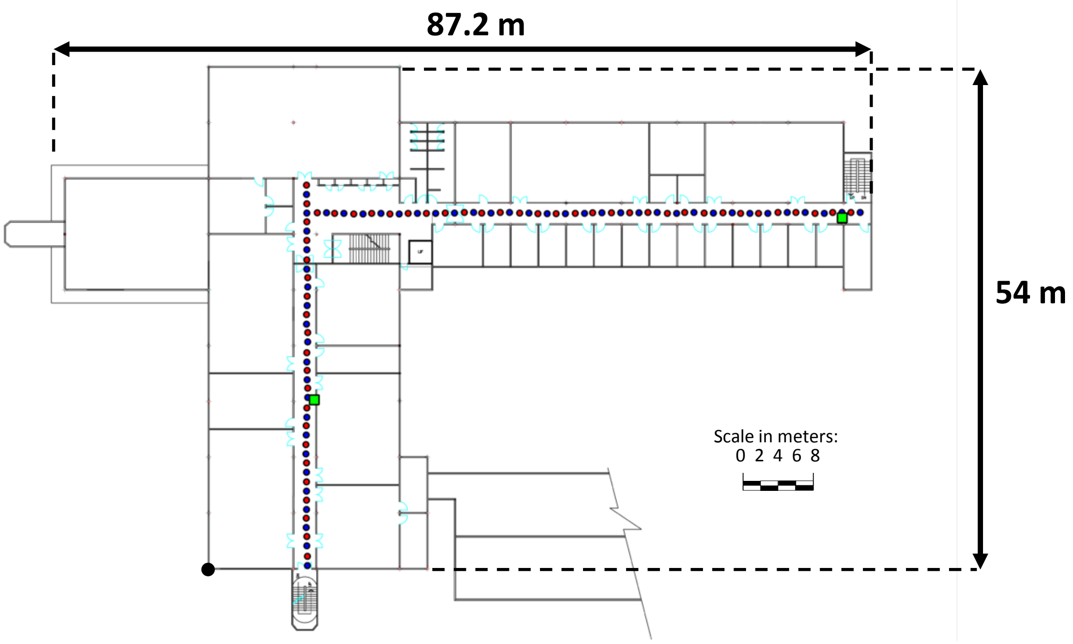
LOCATIONS IN BUILDING A, B, AND C



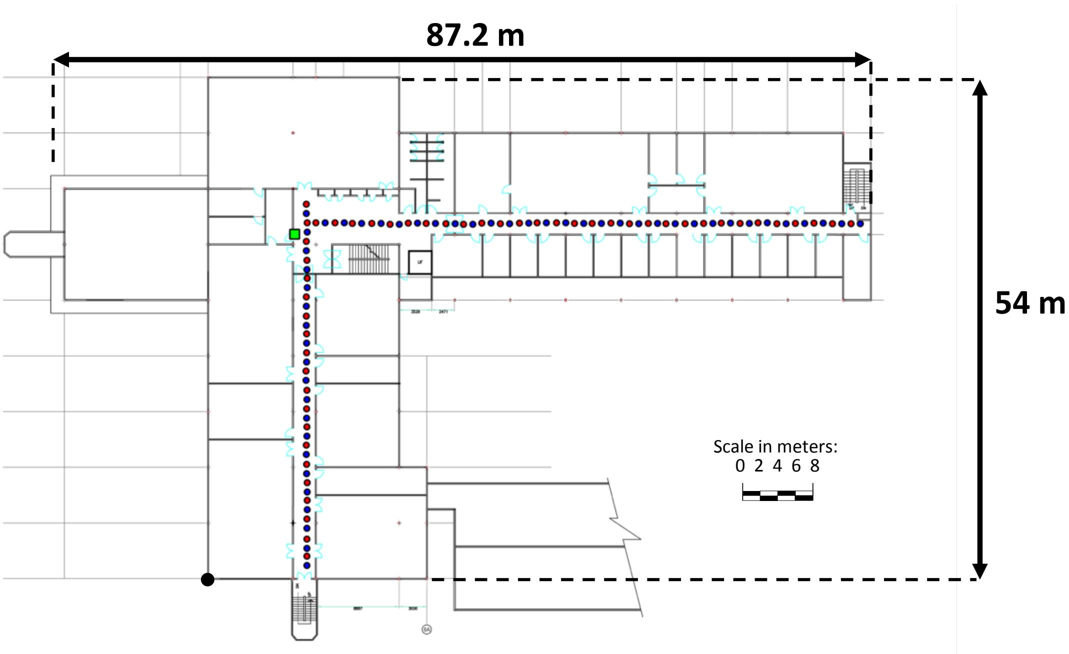
(a)


87.2 m
54 m
Scale in meters:
0 2 4 6 8

(b)

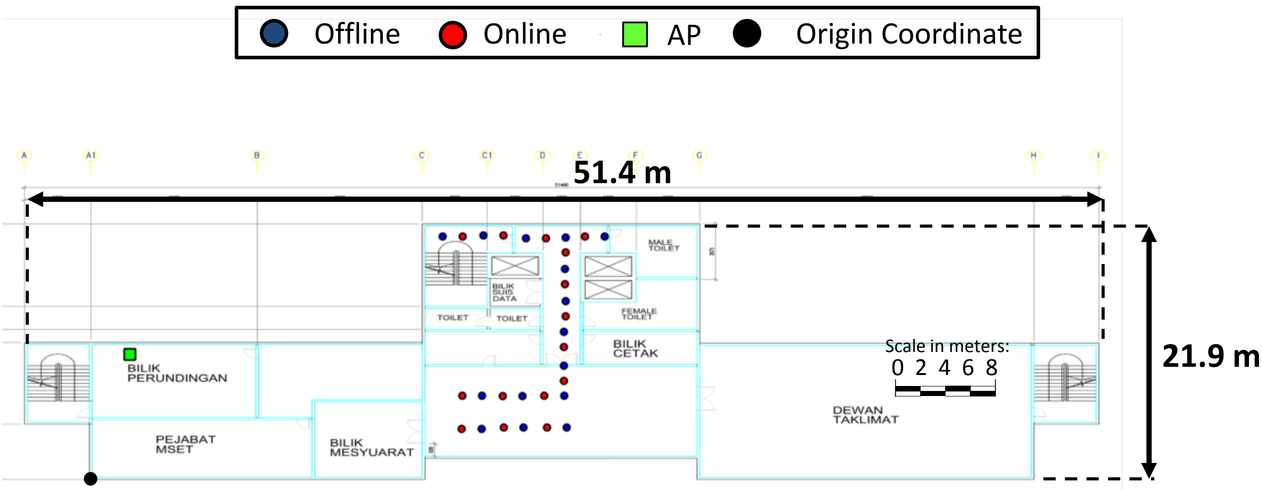

87.2 m
54 m
Scale in meters:
0 2 4 6 8

(c)

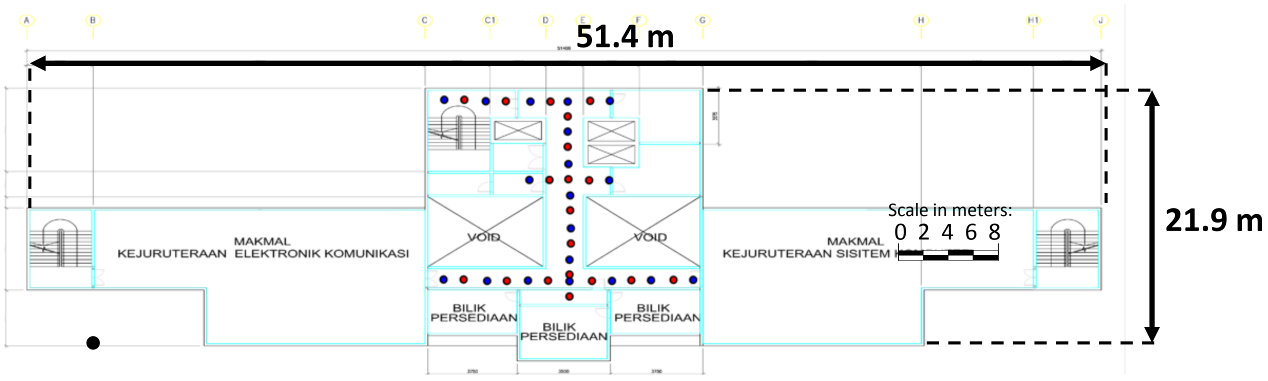


(d)

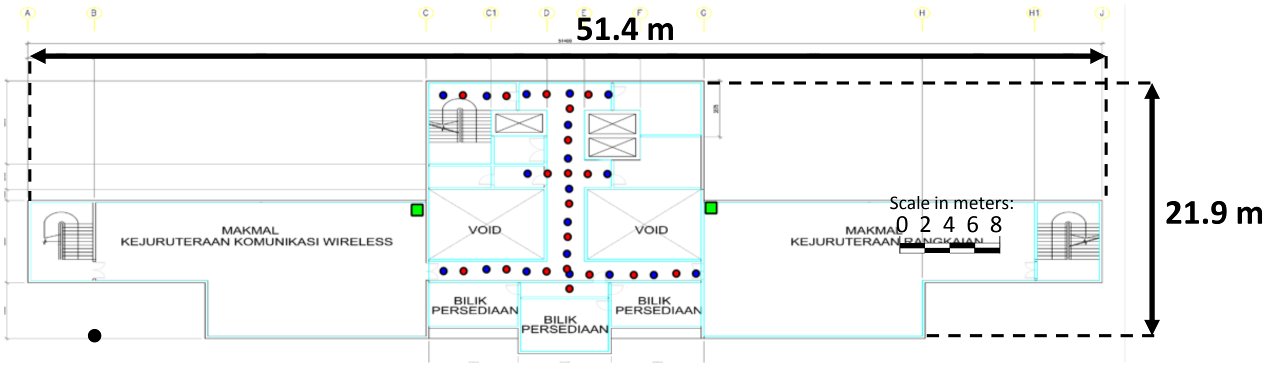
Figure B2.1. Floor plan of Building A where WLAN signals are measured. Location of fingerprint, measured online samples, AP locations, and origin location coordinate of the measurement are marked as shown in the legend. The location according to the floor levels are (a) first, (b) second, (c) third, and (d) fourth



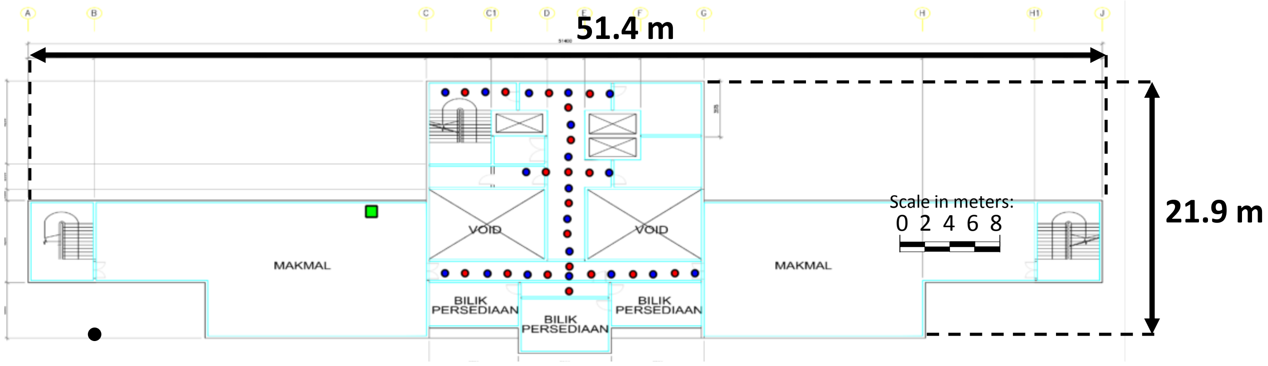
(a)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
KEJURUTERAAN ELEKTRONIK KOMUNIKASI
VOID
VOID
MAKMAL
KEJURUTERAAN SISITEM
BILIK PERSEDIAAN
BILIK PERSEDIAAN
BILIK PERSEDIAAN

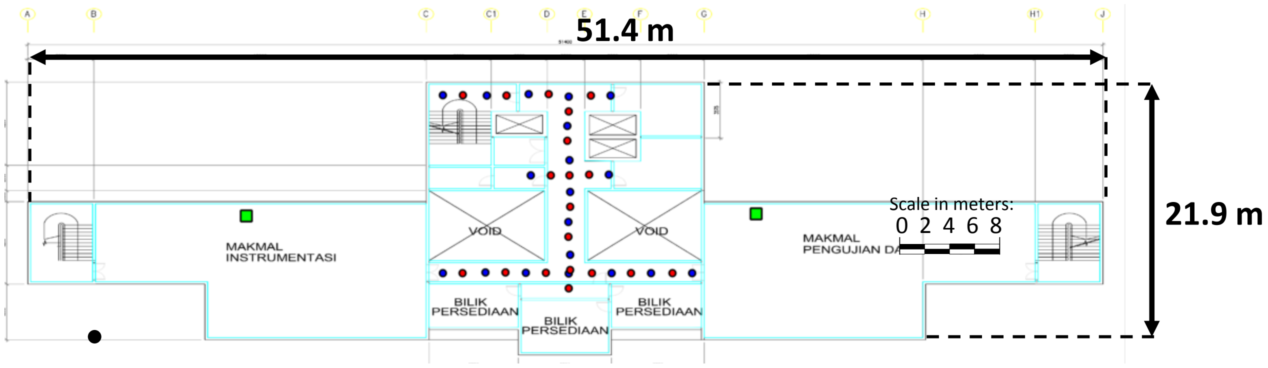
(b)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
KEJURUTERAAN KOMUNIKASI WIRELESS
VOID
VOID
MAKMAL
KEJURUTERAAN RANGKAIAN
BILIK PERSEDIAAN
BILIK PERSEDIAAN
BILIK PERSEDIAAN

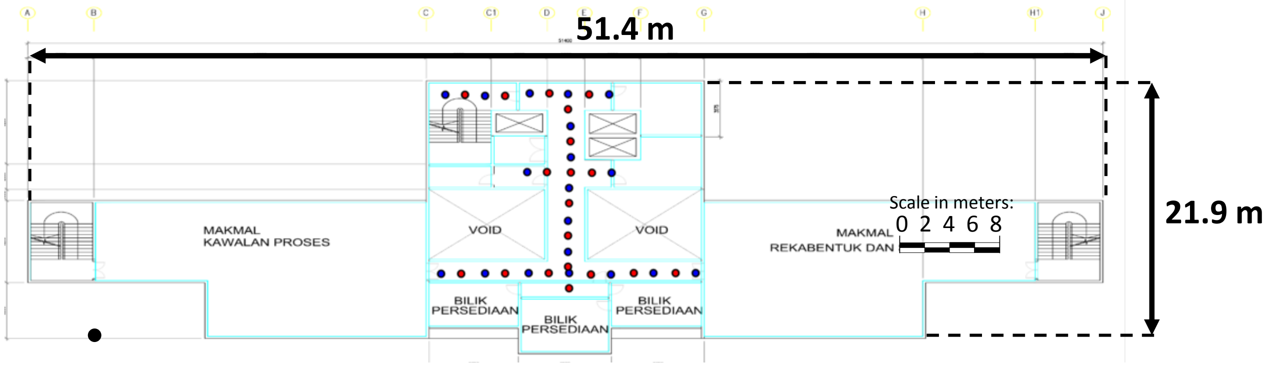
(c)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
VOID
VOID
MAKMAL
BILIK PERSEDIAAN
BILIK PERSEDIAAN
BILIK PERSEDIAAN

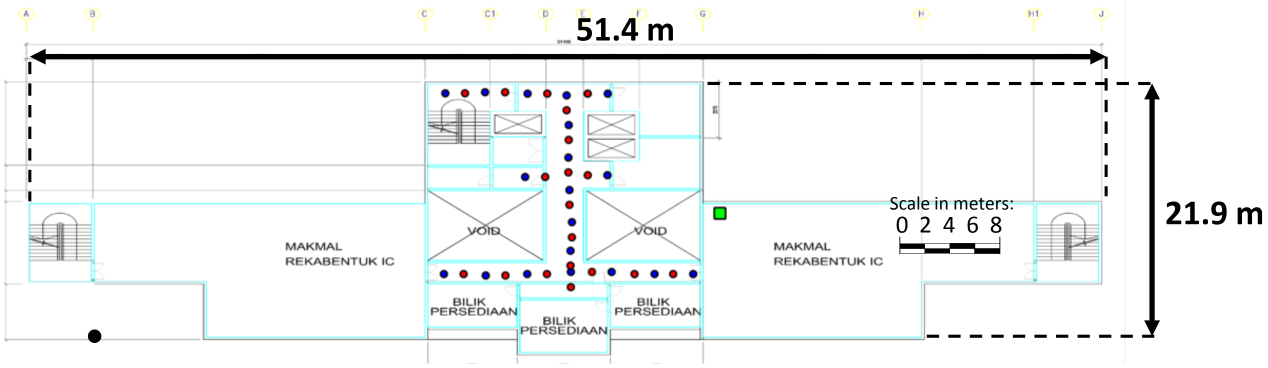
(d)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
INSTRUMENTASI
VOID
VOID
MAKMAL
PENGUJIAN DA
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN

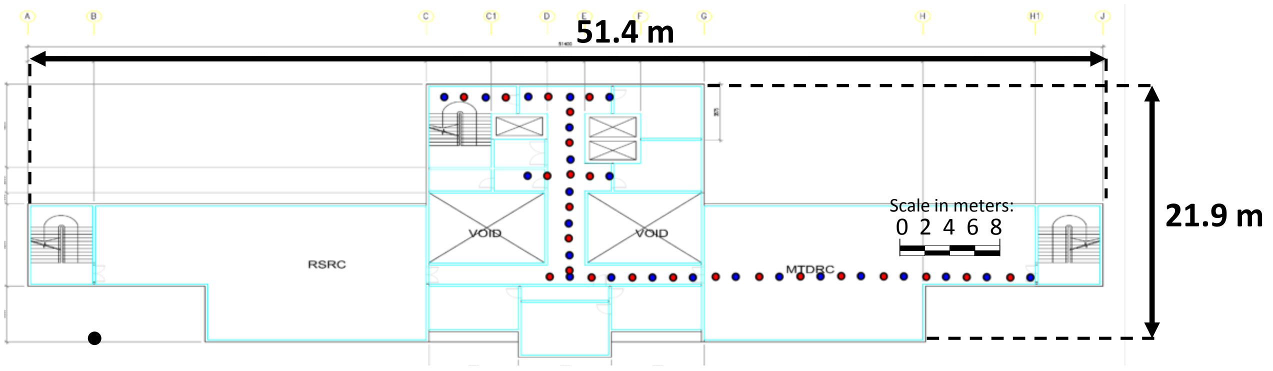
(e)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
KAWALAN PROSES
VOID
VOID
MAKMAL
REKABENTUK DAN
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN

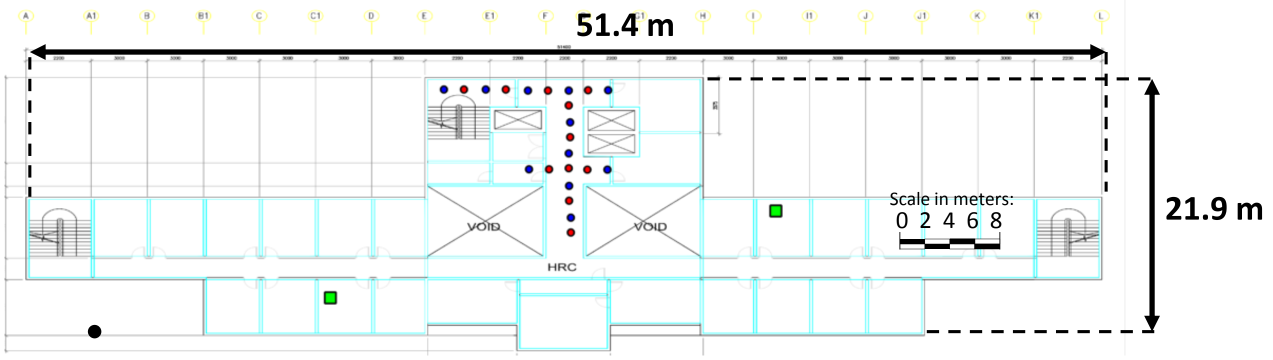
(f)


51.4 m
21.9 m
Scale in meters:
0 2 4 6 8
MAKMAL
REKABENTUK IC
VOID
VOID
MAKMAL
REKABENTUK IC
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN
BILIK
PERSEDIAAN

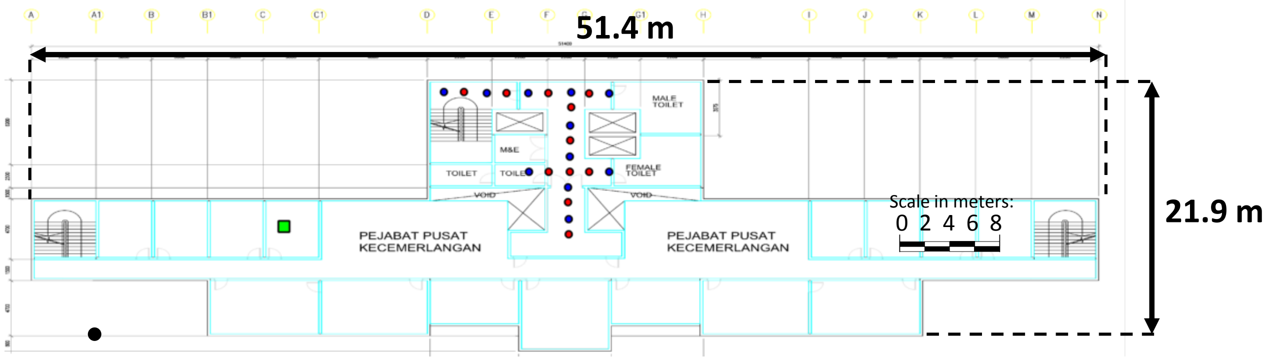
(g)



(h)

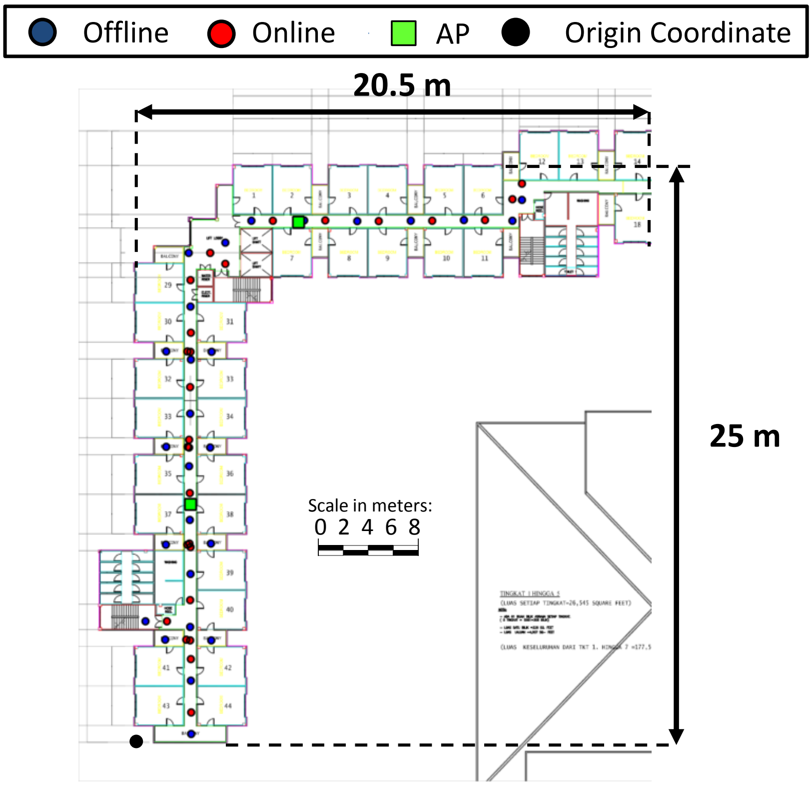


(i)



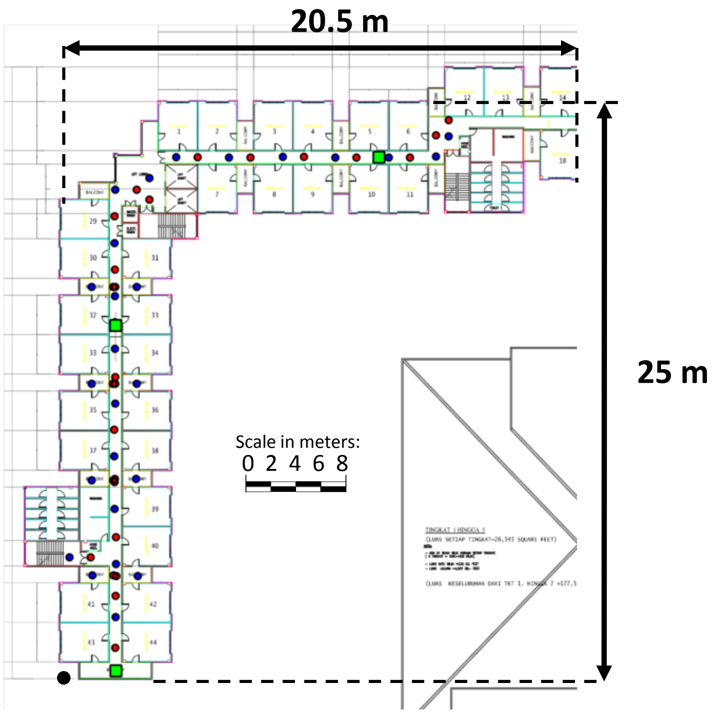
(j)

Figure B2.2. Floor plan of Building B where WLAN signals are measured. Location of fingerprint, measured online samples, AP locations, and origin measurement location coordinate are marked as shown in the legend. The location according to the floor levels are (a) first, (b) second, (c) third, (d) fourth, (e) fifth, (f) sixth, (g) seventh, (h) eighth, (i) ninth, and (j) tenth. The area marked with VOID is the open spaces between floors



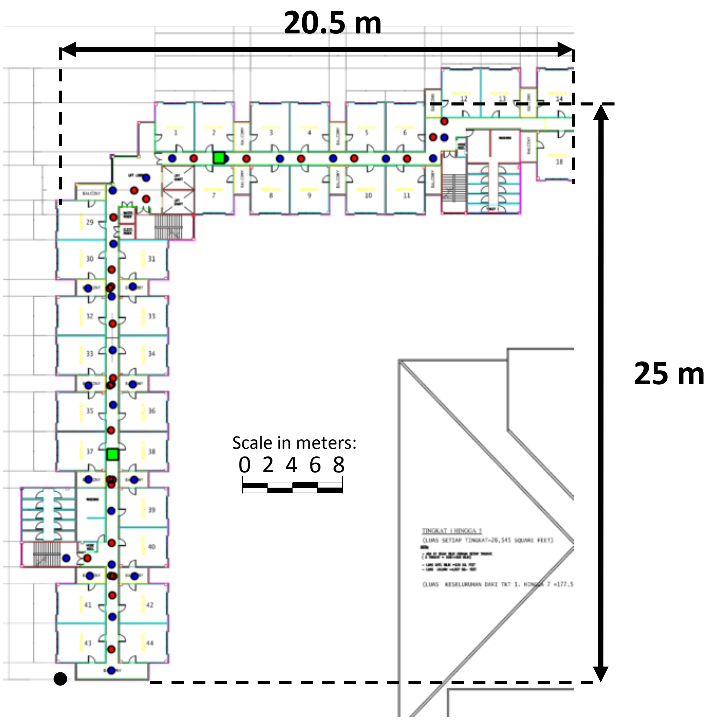
Offline
Online
AP
Origin Coordinate
20.5 m
25 m
Scale in meters:
0 2 4 6 8

(a)

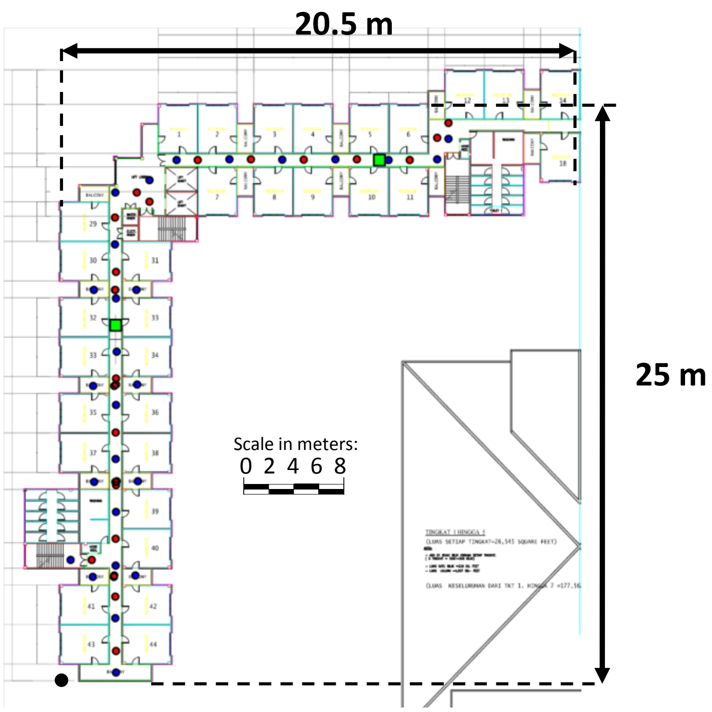


20.5 m
25 m
Scale in meters:
0 2 4 6 8

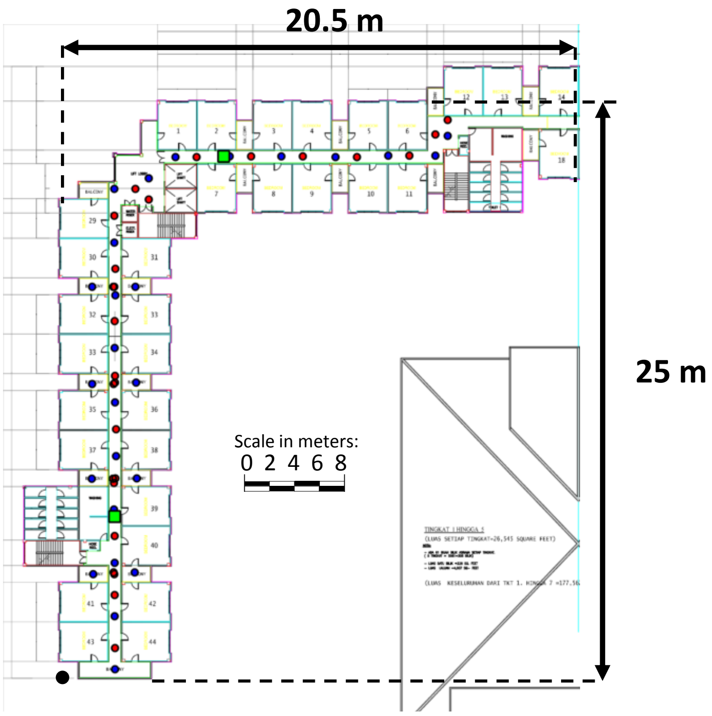
(b)


20.5 m
25 m
Scale in meters:
0 2 4 6 8

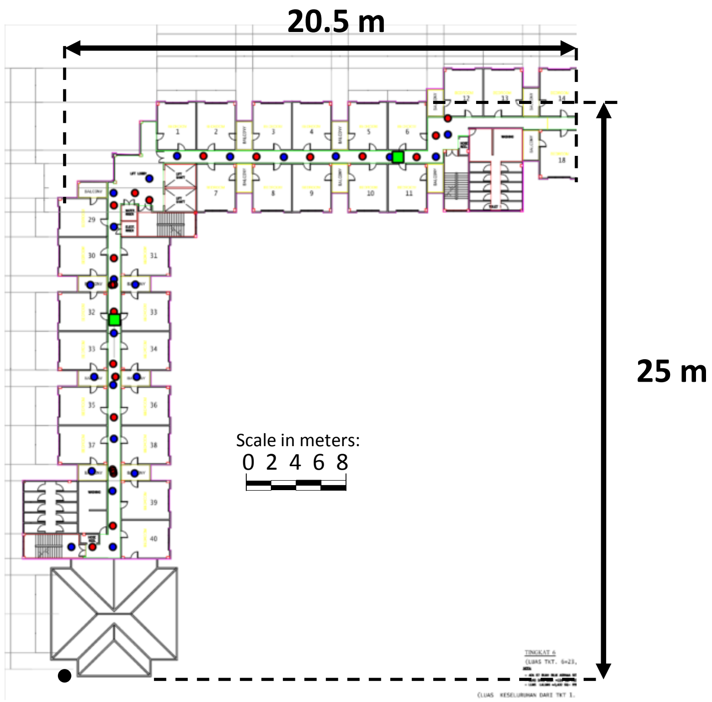
(c)


20.5 m
25 m
Scale in meters:
0 2 4 6 8

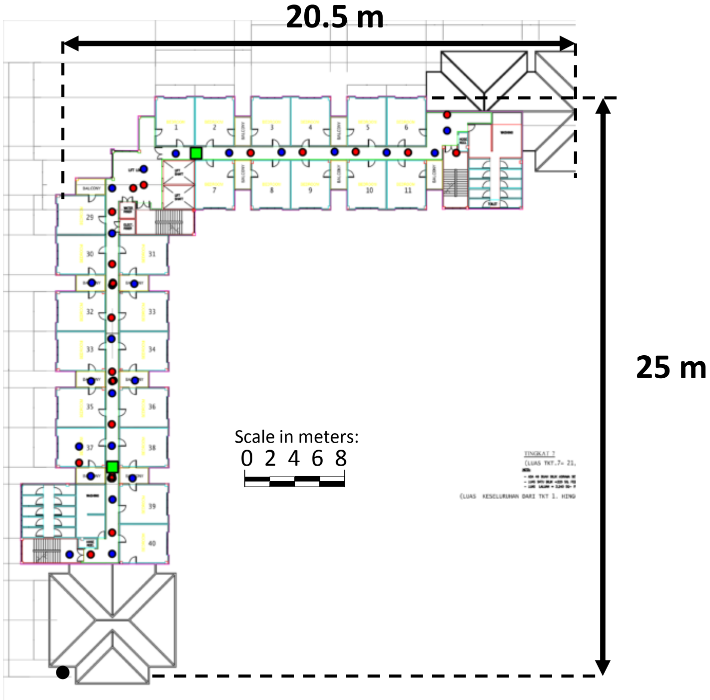
(d)


20.5 m
25 m
Scale in meters:
0 2 4 6 8

(e)


20.5 m
25 m
Scale in meters:
0 2 4 6 8

(f)



(g)

Figure B2.3. Floor plan of Building C where WLAN signals are measured. Location of fingerprint, measured online samples and AP locations are marked as shown in the legend. The location according to the floor levels are (a) first, (b) second, (c) third, (d) fourth, (e) fifth, (f) sixth, and (g) seventh

APPENDIX C1

COMPARISON OF MAX KERNEL AND KLPD WITH MAX MEAN AND INFOGAIN AP SELECTION ALGORITHMS

The proposed Max Kernel and KLPD algorithms are compared with Max Mean and InfoGain algorithms which are commonly used in existing research. Brief descriptions of Max Mean and InfoGain AP selection algorithms are given as follows:

a) Max Mean

Max Mean (Youssef et al. 2002)uses simple sorting of each AP listed at every fingerprint locations according to the APs average signal strength values. The top subset of the sorted APs is used for positioning. For example, at one fingerprint the listed AP and its average signal strength value in dBm is as follows: AP1 = -80, AP2 = -55, AP3 = -77 and AP4 = -66, then if the selection is for 3 APs, the selected APs in ranked order are AP2, AP4 and AP3.

b) InfoGain

InfoGain (Chen et al. 2006) implements Information Gain based AP selection technique. For every AP available in the fingerprint database, the Information Gain value is calculated as:

$$InfoGain\left(AP_{l_V}\right) = H(FP) - H(FP|AP_{l_V}) \quad \text{(C1.1)}$$

given

$$H(FP) = -\sum_{i=1}^{I} \mathrm{p}(FP_i) \log \mathrm{p}(FP_i) \quad \text{(C1.2)}$$

and

$$H(G|AP_{l_V}) = -\sum_{l_V}\sum_{i=1}^{I} \mathrm{p}(FP_i, AP_{l_V} = \forall) \log \mathrm{p}(FP_i|AP_{l_V} = \forall) \qquad \text{(C1.3)}$$

$H(G)$ is called the entropy of the fingerprint locations when AP_i's signal strength value unknown. The prior probability $\mathrm{p}(FP_i)$ of every fingerprint location FP_i is based on uniform distribution of the AP_i's signal strength. Since the probability of $\mathrm{p}(FP_i)$ is uniform for all locations, $H(FP)$ is constant. The term $H(FP|AP_i)$ computes the conditional entropy of grids given AP_i's value and $\forall$ is one possible value of signal strength from AP_{l_V} and the summation is taken over all possible values of AP. Computation of $\forall$ considers average signal strength value of AP_{l_V} at every fingerprint location. At every fingerprint location, the APs are ranked according to decreasing value of InfoGain. The subset of highly ranked APs is used in the positioning. The performance of AP selection algorithms are accessed by combining the algorithms with localisation algorithm. The detailed comparison in terms of performance of the all of the algorithms is discussed in Chapter 6.

Table C1.1 compares the Max Mean and InfoGain algorithms with proposed algorithms: Max Kernel and KLPD.

Table C1.1. Comparison of different AP selection techniques

AP Selection Technique	How subset of APs is selected?	How calculation is performed?
Max Mean [Youssef]	Strongest RSS p APs in each fingerprint	RSS of AP at each fingerprint is sorted according to highest value (strong signal)
InfoGain [Chen]	p ranked APs at each fingerprint selected according to maximum the AP Information Gain values in whole fingerprint database	Difference of entropy of location and entropy of location conditioned on each AP available in database and sorted in decreasing order
Max Kernel	p number of APs according to maximum value of maximum kernel density estimate at each fingerprint	Sorted AP in decreasing order according to maximum kernel density estimate of RSS distribution of each AP at each fingerprint
Kernel Logistic Pairwise Discriminant (KLPD)	p number of APs from sorted combination of pairwise APs according to minimum score function	Combination of APs that is sorted according to increasing value of summation of NLL at all pairwise location combination in the database

APPENDIX D1

COMPARISON OF AKF AND KLF ALGORITHMS WITH LIU AND MARQUES FLOOR LOCALISATION ALGORITHM

The proposed AKF and KLF algorithms are compared with existing floor localisation algorithms which are described as follows:

a) Liu

Liu and Yang (2011) account every fingerprint data to determine the floor. To determine the floor, the RSS difference based on Euclidean distance is calculated per floor and the total difference is normalized according to the number of fingerprint location on each floor. The least RSS difference of related floor decides the floor given as deciding equation as follows:

$$\hat{\mathrm{p}}_z = \min_F \left\{ \frac{1}{N_F} \left\{ \sum_{\varkappa=1}^{N_F} |\boldsymbol{\varphi} - \boldsymbol{\varphi}_{\varkappa F}|^2 \right\}^{\frac{1}{2}} \right\} \tag{D1.1}$$

where $\boldsymbol{\varphi}$ is the online signal vector and $\boldsymbol{\varphi}_{\varkappa F}$ is the $\varkappa$ out of N_F offline signal vector on floor F.

b) Marques

A two-stage method to reduce the computational load of radio map approach is proposed in Marques, Meneses, and Moreira (2012) by filtering the size of the database. In the first stage, the algorithm finds the fingerprints that match the strongest AP or two strongest APs of online sample. In the second stage, the related fingerprints of which their signal distances are sorted according to a proposed similarity function algorithm which is defined as follows:

$$S_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_i) = D(\boldsymbol{\varphi}, \boldsymbol{\varphi}_i) - \alpha \times n_{CM} + \beta \times n_{NCM} \quad \text{(D1.2)}$$

given

$$D_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_i) = \sum_{l=1}^{n_{CM}} |\boldsymbol{\varphi} - \boldsymbol{\varphi}_{il}| \quad \text{(D1.3)}$$

where n_{CM} and n_{NCM} are the number of matched and unmatched APs respectively between online sample and the related offline fingerprint data, and α and β are the weighting parameters. In the third stage, a second filtering process is done reduce the computational complexity of the algorithm by taking only the most similar fingerprint elements to the online sample. The most counted floor from the chosen most similar fingerprint elements is the estimated floor location. This work takes the best values of $\alpha = 6$ and $\beta = 0$ for the first step filtering and 50 most similar fingerprints for second step filtering are used to compute the floor where the decided floor is:

$$\hat{\mathrm{p}}_z = \max\{count\ of\ similar\ floor\ location\ of\ chosen\ S_i\} \quad \text{(D1.4)}$$

The summary of the Liu and Marques floor algorithms and the proposed Averaged Kernel Floor (AKF) and Kernel Logistic Floor (KLF) algorithms is given in Table D1.1.

Table D1.1. Comparison of floor localisation algorithm techniques

Algorithm	Summary of the Technique	How calculation is performed?
Liu	Floor Similarity Measure	The RSS distance of online samples with every offline sample on a floor level is calculated and averaged. The calculation is repeated to all floor level exists in the building. The floor with lowest averaged RSS distance is the determined floor.
Marques	Similarity Measure with Filtering and Majority Rule	All offline samples are chosen according to strongest signal of AP in the online sample. The signal distance of chosen offline samples with the online sample are calculated and sorted in increasing order. The signal distance at the bottom of the list is filtered according to a threshold value. The majority estimated floor location from the chosen signal distance is the determined floor.
AKF	Clustered Fingerprint with Averaged Kernel Density	The offline samples are first clustered according to strongest signal strength of the samples. Kernel density estimate of each AP signal in each offline sample within similar cluster is calculated and averaged. MAP estimate of the online sample signal and every cluster is computed. The floor location of cluster with maximum value of MAP estimate is the estimated floor.
KLF	Clustered Fingerprint with Logistic regression classification	The offline samples are clustered based on strongest signal of each sample. The clusters are trained using kernel logistic regression classification. The online sample that is classified based on sigmoid function to determine the most matched cluster. The floor location of the chosen cluster is the estimated floor.

APPENDIX E1

VALUE OF OPTIMISED ϖ AT DIFFERENT NUMBER OF TRAINING SAMPLES AND DIFFERENT NUMBER OF APS

Table E1.1. Value of optimised ϖ for linear and kernel kNN classifier algorithms in Building A, B, and C at different number of training samples

Number of Training Samples	Optimised ϖ (Left Column: MkNN, Right Column: MKkNN) Building A		Building B		Building C	
4	4	2	2	3	1	2
8	8	2	8	4	2	1
12	5	7	2	2	1	2
16	6	2	3	2	2	2
20	5	1	3	2	1	1
24	5	1	6	3	5	2
28	6	1	2	2	2	1
32	1	3	18	2	2	1
36	1	3	6	2	2	4
40	2	3	11	3	5	4
44	4	3	3	1	3	2
48	2	5	1	1	7	4
52	2	1	1	2	1	2
56	2	1	1	2	6	1
60	2	8	13	2	7	2
64	3	8	11	2	6	2
68	3	8	3	2	11	4
72	3	9	4	2	4	2
76	3	1	2	3	4	4
80	3	1	2	2	4	3
84	2	10	7	7	4	4
88	5	1	10	1	5	3
92	5	6	8	4	4	1
96	3	3	9	1	5	2
100	5	4	14	1	6	1

Table E1.2. Value of optimised ϖ for linear and kernel kNN classifier algorithms in Building A, B, and C, CRAWDAD, Antwerp, and Dublin at different number of selected APs

Number of Selected APs	Optimised ϖ (Left Column: MkNN, Right Column: MKkNN) Building A		Building B		Building C		CRAWDAD		Antwerp		Dublin	
3	8	10	7	15	2	10	9	9	1	1	1	2
4	9	4	7	16	7	10	6	9	1	1	5	2
5	5	3	7	9	3	2	7	8	2	1	3	1
6	5	4	12	9	4	2	9	10	1	1	4	1
7	5	4	12	10	10	1	1	8	1	1	4	4
8	5	4	13	4	2	1	10	10	1	1	3	4
9	5	4	13	10	1	1	6	1	3	1	1	1
10	5	4	8	6	8	6	6	1	1	1	1	1
11	5	4	12	5	5	6	6	1	1	1	1	4
12	5	4	8	8	5	1	6	1	1	1	5	5
13	5	4	14	4	5	2	6	1	1	1	1	1
14	5	4	14	4	5	1	-	-	1	1	5	5
15	-	-	19	12	4	1	-	-	4	1	5	4
16	-	-	19	12	5	1	-	-	1	1	4	4
17	-	-	19	1	4	1	-	-	1	1	5	4
18	-	-	14	1	4	1	-	-	2	2	5	4
19	-	-	14	1	5	4	-	-	1	1	5	5
20	-	-	14	1	5	6	-	-	1	1	5	5
21	-	-	14	1	5	6	-	-	-	-	3	5
22	-	-	11	1	5	5	-	-	-	-	5	3
23	-	-	11	1	5	6	-	-	-	-	2	2
24	-	-	11	1	5	6	-	-	-	-	2	5
25	-	-	11	1	5	6	-	-	-	-	1	5
26	-	-	11	1	6	5	-	-	-	-	1	5
27	-	-	14	1	6	1	-	-	-	-	4	3
28	-	-	14	1	6	1	-	-	-	-	5	3
29	-	-	14	1	6	1	-	-	-	-	5	4
30	-	-	14	1	6	1	-	-	-	-	5	5
31	-	-	14	1	6	1	-	-	-	-	3	2
32	-	-	14	1	6	1	-	-	-	-	4	3
33	-	-	14	1	6	1	-	-	-	-	1	1
34	-	-	14	1	6	1	-	-	-	-	2	3
35	-	-	14	1	6	1	-	-	-	-	2	3
36	-	-	14	1	6	1	-	-	-	-	2	3
37	-	-	14	1	-	-	-	-	-	-	2	1
38	-	-	14	1	-	-	-	-	-	-	-	-
39	-	-	14	1	-	-	-	-	-	-	-	-

Note: Dash(-) means the number of selected APs for the environment has reached the maximum value.

APPENDIX F1

GANSEMER AND MARQUES AND MULTI-FLOOR LOCALISATION ALGORITHMS

The description of Gansemer and Marques multi-floor localisation algorithm are given as follows:

a) Gansemer

Gansemer employed two-stage multi-floor localisation algorithm which computes every fingerprint elements using modified nearest neighbour (NN) approach. Each of the fingerprint elements are divided further according to direction of measurement of the signal for processing. For example, in this work each fingerprint element consists of 100 samples and since there are 4 directions at each fingerprint location, the elements are expanded to 400 samples (100*4). Additionally, some rules are applied compared to original NN by setting the threshold of the APs to be filtered or kept in the signal distance calculation. The rule is defined based on pre-defined signal threshold of TP1, TP2, and TP3 as follows:

1. If the φ_l (RSS of an online signal of AP_l) is higher than TP3 and if $\varphi_{\mu i}^{l}$ (averaged RSS of offline signal of AP_l) does not exist in the radio map database, the i-th fingerprint element is discarded from the ranking to compute signal distance.
2. If both the φ_l and $\varphi_{\mu i}^{l}$ are lower than TP2, the φ_l is removed from calculation of signal distance.

3. If either φ_l or $\varphi_{\mu i}^{l}$ is lower than TP1, the φ_l is discared from signal distance calculation.

The calculation signal distance of chosen fingerprints (after filtering) is according to:

$$D_{Gansemer} = \sqrt{\frac{1}{L_m}\sum_{l=1}^{L_m}\left(\varphi_l - \varphi_{\mu i}^{l}\right)} \quad \text{(F1.1)}$$

And the estimated floor location is determined based on:

$$\hat{\mathrm{p}}_z = \min_{\mathbf{p}_i} D_{Gansemer} \quad \text{(F1.2)}$$

And the horizontal location on the floor is calculated as:

$$(\hat{\mathrm{p}}_{x\xi}, \hat{\mathrm{p}}_{y\xi}) = \min_{\mathbf{p}_{i(z)=\hat{\mathrm{p}}_z}} D_{Gansemer} \quad \text{(F1.3)}$$

b) Marques

Marques also proposed two-stage localisation. In the technique, the original radio map is first restructured by normalising the database by including all APs observed in the environments for uniform computation for all fingerprint elements. The execution of the algorithm is started with sorting the online sample according to AP with strongest signal. Next, the strongest signal AP of the online sample is searched in the radio map to choose the fingerprint that contains the AP. A similarity function defined as in Equation D1.2 is used to calculate the signal distance of online and offline sample. The signal distance is sorted according to lowest value and a threshold value is determined to choose only certain top signal distance of list for estimating the floor. The floor is estimated according to majority of vote or majority rule of similar floor in the list as given in Equation D1.4.

From the chosen signal distance list to estimate the floor, another threshold is applied to the list to calculate the horizontal location. The horizontal location is computed by average position of the chosen lowest signal distance of the list where:

$$(\hat{\mathrm{p}}_{x\xi}, \hat{\mathrm{p}}_{y\xi}) = \left(\frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \mathrm{p}_{x_{S_m}}, \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \mathrm{p}_{y_{S_m}}\right) \tag{F1.4}$$

where m is each index of S_i in the total $\mathcal{M}$ number of S_i in the majority rule. The optimal values of majority rule are experimentally defined according to the lowest mean distance error in the tested environment.

APPENDIX F2

PARAMETERS OF MULTI-FLOOR AP SELECTION ALGORITHMS FOR MULTI-FLOOR LOCALISATION IN CHAPTER 8

Table F2.1. Number of selected APs for proposed multi-floor localisation algorithm in different buildings

Number of Selected APs	Building A	Building B	Building C
Floor Cluster	18	39	48
Horizontal Fingerprint Elements	12	12	20

Table F2.2. Majority rules for Marques' multi-floor localisation algorithm in different buildings

Majority Rules	Building A	Building B	Building C
Floor Determination	2	2	1
Horizontal Estimation	2	4	1

Table F2.3. Parameters of Gansemer's multi-floor localisation algorithm in all tested buildings

Parameters	Values
nMin	4
TP1	-85 dBm
TP2	-80 dBm
TP3	-75 dBm

BIODATA OF STUDENT

Mohd Amiruddin Bin Abd Rahman received Bachelor of Science degree in Electrical Engineering from Purdue University, United States of America in 2006. He has been working at Physics Department, Faculty of Science, Universiti Putra Malaysia (UPM) as a tutor since that. In 2011, he obtained Master of Science at in the field of Sensor and Instrumentation from Universiti Putra Malaysia, Malaysia. He is currently working towards his PhD degree of which will be jointly awarded by Universiti Putra Malaysia and Universiy of Sheffield, United Kingdom. During his PhD study, he has joint Bell Labs, Ireland to as a post-graduate research intern. He has also served as an invited researcher at Bell Labs, Belgium. His current research interest is in indoor localisation and signal processing.

LIST OF PUBLICATIONS

Rahman, Abd, Mohd Amiruddin, Mina Dashti, and Jie Zhang. "Localization of unknown indoor wireless transmitter." In *Localization and GNSS (ICL-GNSS), 2013 International Conference on*, pp. 1-6. IEEE, 2013.

Rahman, Abd, Mohd Amiruddin, Mina Dashti, and Jie Zhang. "Floor determination for positioning in multi-story building." In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pp. 2540-2545. IEEE, 2014.

Liu, Yang, Mina Dashti, Abd Rahman, Mohd Amiruddin, and Jie Zhang. "Indoor localization using smartphone inertial sensors." In *Positioning, Navigation and Communication (WPNC), 2014 11th Workshop on*, pp. 1-6. IEEE, 2014.

Dashti, Marzieh, Mohd Amiruddin Abd Rahman, Hamed Mahmoudi, and Holger Claussen. "Detecting co-located mobile users." In *Communications (ICC), 2015 IEEE International Conference on*, pp. 1565-1570. IEEE, 2015.

Mohd Amiruddin Abd Rahman, Zulkifly Abbas, and Alyani Ismail. “Path Loss Extraction of 2.4GHz WLAN Radio Waves in Various Multi-Floor Buildings According to Seidel’s Model.” In *Journal of Solid State Science and Technology*, 2016 – Accepted.