Dual-sensor fusion for seamless indoor user localisation and tracking

Milan Redžić, MScEE

A Dissertation

Presented to the Faculty of Engineering and Computing

of Dublin City University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance by the

School of Electronic Engineering

Supervisors

Professor Noel O'Connor

Dr. Conor Brennan

September, 2012.

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed (Candidate): Milan Redžić

ID No.: 58118594

Date: 01/09/2012

Acknowledgements

It would have been impossible for me to do this work without the help of several people. First of all I would like to thank my supervisors Dr. Conor Brennan and Professor Noel O'Connor for giving me opportunity for doing Ph.D. research in CLARITY: Centre for Sensor Web Technologies, DCU, Dublin. They gave me the full freedom to develop my own ideas, while arranging all the necessary conditions for doing good research. I would like also to mention all my friends and colleagues from CDVP, CLARITY and DCU for providing an inspiring and pleasant atmosphere. Also I need to thank all my friends from Serbia, Ireland and all over the world for nice times and get-togethers, thus making my research years more pleasant and enjoyable. The last but not the least I would like to thank my family: Dragan, Slobodanka, Milena, Vučeta, Mica, Steva (in memoriam), Mira (in memoriam) i Sneška, for being with me during all ups and downs and who have been enriching my everyday life. The work presented in this thesis is kindly supported by Science Foundation Ireland under grant 07/CE/I1147. Thank you all! Hvala vam svima!

In Dublin, 01.09.2012.

Abstract

Indoor localisation based on ubiquitous WLAN has exhibited the capability of being a cheap and relatively precise technology and has been verified by many successful examples. Its performance is subject to change due to multipath propagation and changes in the environment (people, building layouts, antenna characteristics etc.) which cannot be easily eliminated. This thesis addresses the automatic localisation of indoor user and proposes solutions for both positioning and seamlessly tracking a user using WLAN technology in addition to image sensing. By fusing these modalities we obtain better performance than using them individually. A fusion function designed to merge both analysis results into one semantic interpretation of user location is presented. Also a tracking approach based on an adaptive function that converts times between locations into probabilities and employs a Viterbi-based solution is proposed. An indoor localisation algorithm is described which is based on the creation of a database of WLAN signal strengths at pre-chosen calibration points (CPs). The need for fewer CPs than in standard methods is achieved due to the use of a novel interpolation algorithm, based on the specification of robust range and angle-dependent likelihood functions that describe the probability of a user being in the vicinity of each CP. The actual location of the user is estimated by solving a system of equations with two unknowns derived for a pair of CPs. Different pairs of CPs can be chosen to make several estimates which can then be combined to increase the accuracy of the estimate. The effectiveness of the fusion and the tracking approaches is evaluated on a very challenging dataset throughout a university building. Results that are presented demonstrate high accuracy that can be achieved. The methods are compared to several competing localisation

methods and are shown to give superior results. The potential usefulness of this work is envisaged in a range of ambient assisted living applications including lifelogging and as an assistive technology for the memory or the visually impaired.

Contents

List of Figures

List of publications

Journals under review

• M. Redžić, C. Brennan and N. O'Connor, *On energy consumption and time efficiency in dual-sensor indoor localisation*, submitted to: IEEE Transactions on Mobile Computing (under review).

• M. Redžić, C. Brennan and N. O'Connor, *SEAMLOC: seamless indoor localisation based on reduced number of calibration points*, submitted to: IEEE Transactions on Mobile Computing (under review).

Peer reviewed conferences and workshops

• M. Redžić, C. Brennan and N. O'Connor, *User tracking using a wearable camera*, In: IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), 13-15 November 2012. Sydney, Australia.

• M. Redžić, C. Brennan and N. O'Connor, *Dual-sensor fusion for indoor user localisation*, In: International Conference on Multimedia - ACM Multimedia, 28 Nov - 1 Dec 2011, Scottsdale, AZ.

• M. Redžić, C. Brennan and N. O'Connor, *Indoor localisation based on fusing WLAN and image data*, In: IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), 21-23 Sept 2011. Guimaraes, Portugal.

• M. Redžić, C. Brennan and N. O'Connor, *Using SenseCam images in a multimodal fusion framework for route detection and localisation*, In: The 2nd Annual SenseCam Symposium, 16-17 Sept 2010. Cambridge, UK.

• M. Redžić, C. O'Conaire, C. Brennan and N. O'Connor, *Multimodal identification of journeys*, In: IMPRESS 2010 - 1st International Workshop on Interactive Multimodal, Pattern Recognition in Embedded Systems in conjunction with DEXA 2010, September 2010. Bilbao, Spain.

• M. Redžić, C. Brennan and N. O'Connor, *A route identification algorithm for assisted living applications fusing WLAN, GPS and image matching data*, In: Research Colloquium on Wireless as an Enabling Technology, August 2010. Dublin, Ireland.

• M. Redžić, C. O'Conaire, C. Brennan and N. O'Connor, *A Hybrid Method for Indoor User Localisation*, In: The 4th European Conference on Smart Sensing and Context (EuroSSC), 2009. Guildford, UK.

Projects

• M. Redžić, C. Brennan and N. O'Connor, *Advances in indoor location based on signal strength and image data*, In: COSTProject: IC1004 TD(12) 03056, Barcelona, Spain.

Chapter 1

Introduction

1.1 Localisation

Location awareness is a key feature in the context-aware computing paradigm [168, 125]. Moreover, location information can provide context for additional mobile computing services [94]. The meaning and the relevance of data can be interpreted differently as the user's location changes with time. Therefore, location determination represents a key goal for mobile computing. Localisation refers to a process of obtaining the location information of a user. There are different types of location information: symbolic location, relative location, physical location and absolute location. Symbolic location matches physical location with a symbolic location name. Physical location is represented in coordinates, which is shown as a point on a 2D or 3D map. Absolute location uses a common reference system for all located objects. A relative location is determined in the context of predefined reference system. Thus, location is usually measured relative to known reference points. In the literature, localisation process is variously known as location sensing [78], geolocation [134], position location [138] or simply localisation [97]. In this work these terms will be used interchangeably. There are two types of localisation possible based on the characteristics of the physical space: indoor and outdoor localisation.

The success and uptake of outdoor positioning and localisation systems, GPS in the

first instance, has thrown down a challenging gauntlet for indoor localisation services and applications. Unfortunately, the GPS system cannot be used effectively inside buildings and in dense urban areas (so called urban canyons) due to its weak signal reception (i.e. when there is no line-of-sight to the satellites). Thus, indoor localisation relies on different means to precisely position a user without the help of signals from GPS satellites. Commercial indoor positioning systems typically use RF, infrared and ultra sound signals. Different types of sensors are required to detect electromagnetic signals that actually depend on location (e.g. a photo-diode detector is commonly used for infrared signal detection). A sensing process converts these signals into a measurable metric such as distance or angle in the localisation process [134]. Then, the measurable metrics are processed by a positioning algorithm to estimate the position of a user [134]. Unlike outdoor areas, indoor areas impose various challenges on user localisation due to the dense multipath and building material which affect propagation [170]. GPS specifies the latitude, longitude, and altitude coordinates and thus computes the absolute position of a user. On the contrary, some indoor localisation systems have to use some other points of reference such as offices, rooms and bigger indoor spaces, calibration points etc.

The most important performance metrics of an indoor localisation system are accuracy and precision. These are usually expressed as a percentage of correctly guessed locations (precision) and as an absolute distance between the estimated and the real location of a user (accuracy). Precision represents a confidence that the location detection process has been successful. Several authors mention coverage, delay, scalability and capacity of the positioning system as other metrics. The coverage metric defines the (maximal) area within users can be located [107]. The delay metric refers to the time taken between sensing of the signal and reporting the information. Scalability indicates system performance when the number of possible user locations increases and/or when the area of user localisation becomes larger [107]. The capacity metric measures the number of detected locations that a system can process in one second [107]. All of these metrics heavily depend on characteristics of the indoor environment, the sensing technologies, the bandwidth of the sensed

signal, infrastructure, positioning algorithm, and complexity of signal processing techniques employed to estimate the location information [107].

The cost of an indoor positioning system depends on the cost of extra infrastructure required, reliability desired, and characteristics of the technologies used [30]. It also includes installations during the deployment period. It is generally accepted that it is better to reuse infrastructure that already exists in the environment in order to make the setup easier. This can also save some equipment and infrastructure cost. For instance, in an in-band communication system, the existing (communication) signals can also be used for location sensing. After the system becomes operational, the extra power consumption can be considered as a cost for the positioning system [30]. The complexity of the approach used in a localisation system needs to be balanced and optimized with its performance. This is usually based on various trade-offs between a system's accuracy/precision and complexity.

The major application requirements for location information are the availability, the performance and the granularity. These requirements are different from one application to another. The availability discusses situations when the location information cannot be obtained and determines the obstacles that prevent a localisation system from obtaining the location information. The performance requirements can include any combination of performance metrics discussed above. There are two types of granularity: temporal and spatial. Temporal granularity determines the rate at which the location information is requested. The spatial granularity determines the level of detail of location information. We may distinguish two approaches based on where localisation is performed: self-positioning and remote-positioning [94]. In self-positioning a user determines his (her) location while in remote-positioning someone/something else determines position of a (distant) user.

Another important issue and concern in the world of localisation systems is security. This means that (some) users do not want to be revealed or tracked when their location is detected. This is also associated with the way in which devices detect a user's position. Moreover, it also depends on the type of application as the level of possible security is often determined by the sensing approach [119]. For example there is a location tracking system,

$E-911$, whose privacy can be violated [119] compared to GPS that can be used absolutely securely. Therefore, in case it is needed a location system should have a security protocol embedded within the system to protect the location information.

There are two types of integration of multiple modalities: fusion and hybrid. In case of fusion all features or results are integrated at all times while in case of hybrid they are integrated only when needed. These terms will be explained in greater detail in section 3.12. Due to complex indoor environments and given the modalities that are currently available, using more that one improves accuracy/precision. By using two sensing modalities, e.g. images and WLAN data, one can make a more robust localisation system. In this work, a localisation system is proposed that represents the fusion of two components: WLAN-based and image-based localisation. In the fusion, images and WLAN data are used at all times i.e. in every localisation test. In other case (a hybrid approach) images in addition to WLAN data (used in every test), are used only when necessary i.e. when WLAN breaks down and/or is unreliable. Moreover, an improved localisation approach based on WLAN is proposed as the subcomponent of the fusion algorithm. These options are represented in figure 1.1.

The limitations and challenges of existing approaches are closely related to the propagation of RF signals in indoor environments which poses a central challenge. Certain materials within the indoor environment affect the propagation of radio waves. For example materials such as wood or concrete attenuate RF signals, while materials such as metals or water cause reflections, scattering and diffraction of radio waves. These effects lead to multipath radio wave propagation, which encumbers accurate calculation of the distance between the transmitter and the receiver [66,36,89,32,84,97,85,168]. Several authors have proposed techniques to compensate for these inaccuracies by automatically generating radio maps which consider the structure of the building [110,96,58]. However, a comprehensive model of all the materials in a complex environment such as a health care facility or a patient's home is a non-trivial problem. The propagation of radio waves are adversely affected by changes to the physical environment such as the rearrangement of furniture, structural modifications

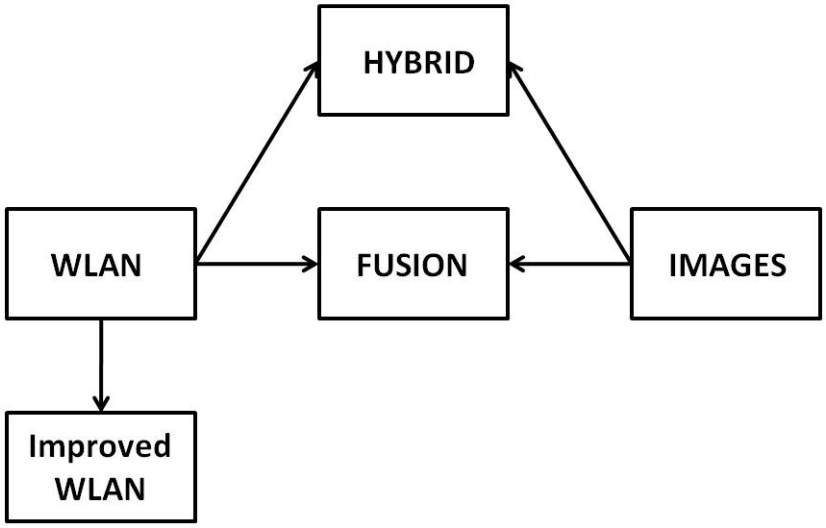


Figure 1.1: The diagram of the dual-sensor localisation system. An improved approach based on WLAN data only is used when localising a user between pre-selected locations

or movement of personnel within a building. In these environments, the radio properties are highly dynamic, and a radio map captured at a certain point in time cannot be used reliably for localization without accounting for these dynamic changes [62,38,114,152]. Interference and noise are often-mentioned challenges [48,135]. Localisation technologies based on RF technologies can be attractive due to the ubiquity of certain infrastructural technologies, such as wireless data networks, that may already be present in the facilities. Care must be taken, however, in evaluating the impact of the physical environment on the RF localization technologies, as the solution may be rendered non-operative in certain clinical settings. For image-based localisation challenges exist in case of significant occlusion or lightning changes. Also most of image-based localisation methods cannot easily distinguish very similar locations thus introducing localisation errors in case of locations physically far from each other.

Image and wireless based technologies have penetrated the world of consumer electronics and include public safety, industrial, medical, logistics and transport systems along with

many other applications. Some other related areas are context-based information services, tracking, guiding and touring applications and devices. The main breakthroughs in the indoor localisation area have been achieved in the last 20 years. Both industry and academia are doing their best to ensure developments and advancements of localisation systems. Numerous wireless and image-based technologies have been used for indoor localisation. Only a few works however have tried to fuse these two modalities to achieve better overall accuracy/precision as proposed in this thesis.

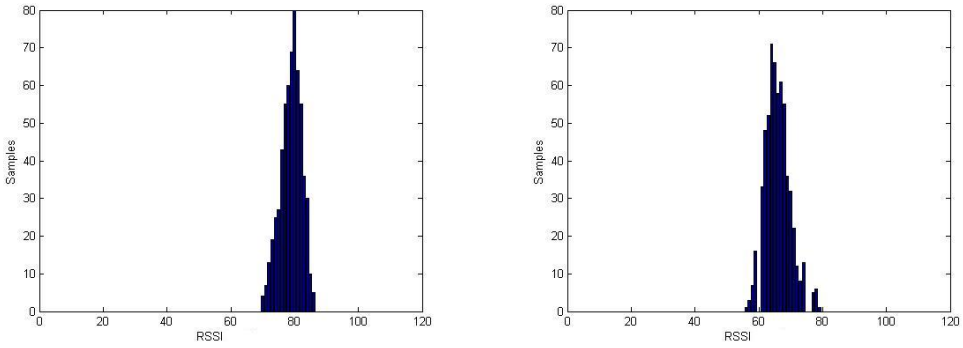
1.2 Mobile data sensing

Wearable camera technology has evolved to the point whereby small unobtrusive cameras are now readily available, e.g. the Vicon Revue (formerly known as Microsoft Research's SenseCam)[1] . This has allowed research effort to focus on analysis and interpretation of the data that such devices provide [88]. Even in the absence of bespoke platforms such as the Vicon Revue, any smart phone can be turned into a wearable camera. The Campaignr[2] configurable micropublishing platform has demonstrated the capability of mobile platforms to act as WLAN (and more general sensor) data gathering hubs. Researchers in Dublin City University are developing a device using an Android-based smart-phone worn on a lanyard around the neck that in addition to image capture also senses a variety of other modalities e.g. motion, GPS, Bluetooth, WLAN. The motivation for this is to use this platform in a variety of ambient assisted living applications as well as assistive technology for the memory and visually impaired. Although this platform does not exist commercially at the moment it is likely that such devices will start to appear in the near future. Novel technologies such as robust indoor localisation will help drive this. These platforms allow users to regularly collect data at many (indoor) locations. This large collection of data needs some means of structuring it to make it understandable and searchable.

In this work the problem of structuring the data is addressed by examining the automatic

[1] http://www.viconrevue.com
[2] http://www.campaignr.com

identification of indoor locations. The indoor localisation problem is complicated by a number of factors. Although GPS has become synonymous with user localisation, indoors its signals are weak or non-existent. Using WLAN as a solution has given promising results, but its performance is subject to change due to multipath propagation and changes in the environment, such as number of persons present in a given location, variable orientation, temporary changes to building layout, etc. [135]. Figure 1.2 displays an example of typical histogram data collected for WLAN-based localisation and illustrates the variation in signal characteristics which enables us to distinguish between locations. In an IEEE 802.11 system RSSI is the relative received signal strength (RSS) in a wireless environment, in arbitrary units. RSSI is an indication of the power level being received by the antenna. Therefore, the higher the RSSI number the stronger the signal. Vendors provide their own accuracy, granularity, and range for the actual RSS (measured in mW or dBm) and their range of RSSI values (from 0 to $RSSI_{Max}$).



(a) (b)

Figure 1.2: Two histograms of WLAN data (RSSI) each representing a different location

Variations in the environment, such as temporary changes to building layout or presence of foliage, can affect received signal strength (RSS) [93, 169]. Its performance also depends on the material the building is made from, size of spaces where measurements take place, antenna orientation, directionality, etc. [135]. Using images to determine location is an alternative technique to radio-frequency wireless signal based approaches. As we have access to the images on our capture platform it costs us nothing in terms of extra hardware.

Moreover, there are situations when WLAN breaks down, thus making the use of images the only option. Image-based localisation techniques have provided some promising results, as in [171, 24], but the limitations continue to be due to computational expensiveness, occlusion, changes in lighting, noise and blur [24]. The challenge investigated in this thesis is to find the best way to fuse or hybridize both sources of information. In this work, an approach that combines image and WLAN data is proposed to leverage the best of both of these complementary modalities. A novel fusion function designed to merge both analysis results into one semantic interpretation of a user location is presented. By combining the strengths of these two complementary approaches, we hope to achieve high accuracy and robustness to the problems that affect individual modalities. A method that addresses the automatic tracking of the user indoors using a fusion of WLAN and image data is also presented. Eventually an approach for localising a user anywhere within space defined by preselected locations is presented. As these methods can be orientated towards the needs and capabilities of the user based on context they are potentially useful for ambient assisted living applications.

1.3 Research contributions

In this thesis the problem of indoor user localisation and tracking using WLAN-based and image-based localisation techniques is investigated. First both modalities are investigated separately. A precise WLAN-based algorithm is used, along with a vocabulary tree method for image-based localisation. The vocabulary tree [131] was designed using 64-dimensional Speeded Up Robust Features (SURF) descriptors [24]. WLAN and image matching data are fused to improve localisation results. The user can be tracked and eventually the sequence of consecutively visited locations is obtained. Moreover, position anywhere in the space can be estimated using different algorithms. In this thesis all methods are verified on a large and very challenging experimental datasets thus proving the robustness of the proposed methods. The key research contributions are as follows:

- A novel image-based localisation method is proposed based on fine tuning the cluster centers of the hierarchical vocabulary tree of the SURF feature descriptors. Cluster centers were calculated recursively using the previously calculated cluster centers. This shows great robustness over the simpler approach where the centers are calculated only once.

- A novel fusion function is presented which takes localisation results from both sensing modalities simultaneously to create a new ranking of the locations. It uses weighted linear combination of confidences of both modalities and together with adaptively calculated thresholds obtains better accuracy/precision than when using any single modality. The proposed fusion approach is very general and thus potentially applicable to various sensing modalities.

- A novel tracking method is employed when using an image-based, WLAN-based or fusion-based approach *only*. The method represents a simple Viterbi-based multiplestate model using simple Hidden Markov Model (HMM) states. An approach for converting times (between the two consecutive locations) into probabilities in order to construct the most likely route traversed by the user is proposed.

- Finally this thesis proposes and verifies a novel approach for localising the user anywhere within space defined with a rectangular grid of known locations of size $3 \times 5m^2$. The novel interpolation algorithm is based on the specification of robust range and angle-dependent likelihood functions that describe the probability of the user being in the vicinity of known, pre-selected locations in space. This approximation showed the best trade-off between complexity and accuracy and moreover introduced flexibility into the system. Contribution is the ability to reduce the number of calibration points (CPs)[3] needed.

[3]Received signal strength data are collected at many pre-selected locations which will be referred to in this thesis as calibration points - CPs

1.4 Organization

An overview of some applications motivated by the use of a platform capable of sensing different sensors (in our case WLAN and image data gathering hubs are of main interest) is given in chapter 2. Wearable cameras and multiplatform devices more easily contribute to collecting and indexing a wearer's experiences by (unobtrusively) collecting various sensing data. Vicon Revue in particular is a successful example of such a camera and is discussed in detail. In this chapter potential application scenarios such as aid for memory-impaired, visually-impaired, tourist-orientated applications and applications in health and medicine are discussed.

In chapter 3 we focus on a review of indoor localisation and tracking systems, particularly on those that are of main interest in this work: wireless local area network (WLAN) based, image/video-based and localisation systems based on hybrid and fusion of these two modalities. It also discusses the most often used RF-based and image-based localisation metrics. An overview of RF and WLAN-based localisation techniques is given together with the image-based counterpart. For the image-based techniques the most famous feature detectors and descriptors used in the localisation process are also explained. Various different hybrid and fusion examples of these two modalities are presented including location-based services for ambient-assisted living scenarios, identification of frequent indoor trips, e-services and mobile services, etc.

In chapter 4 some necessary technical background for indoor localisation and tracking are presented. First we introduce Bayes classification method also known as Bayes localisation method. Using a probabilistic chain rule and a naive assumption of conditional probabilities this approach can be transformed into the more convenient, and for this work more important, Naive Bayes approach. For the image-based localisation, presented in section 5.4, we use a hierarchical vocabulary tree of Speeded Up Robust Features (SURF) descriptor vectors presented in section 5.4.2 of chapter 5. The SURF method itself is described in greater detail in section 4.3.1 of this chapter. Hierarchical k-means clustering is achieved using simple k-means clustering repeatedly. The ranking list of (possible) user

locations is achieved using K nearest neighbour classifier which is explained in detail in section 4.4. In this chapter we also discuss how two or more modalities could be fused in general using weights and/or grid search engine to achieve better performance than using them separately.

In chapter 5 methods for indoor user localisation and tracking inside a university building are presented. An approach that would enable user localisation to within an office is described. The WLAN-based solution, given in section 5.3, uses a extended Naive Bayes approach to find the user location. Image-based localisation is realized using a novel approach based on a hierarchical vocabulary tree of SURF descriptor vectors and is given in section 5.4. A novel fusion approach is proposed to overcome the shortcomings of each of the individual sensing modalities. We also propose a tracking method (in section 5.8.1) that can be employed when using image-based, WLAN-based or the fusion-based approach. Two very challenging experimental setups (ESs) are discussed in sections 5.6 and 5.8.2. Both localisation and tracking results are given in sections 5.7 and 5.8.3 respectively to demonstrate the effectiveness of the proposed methods.

Grid based fingerprinting methods can only localise users to one of the points at which data was collected. Consequently a fine grid of points is needed to ensure accuracy. Chapter 6 describes a method to localise the user anywhere in a space defined with a grid of calibration points. It uses fewer number of calibration points (CPs) than standard, well-known localisation approaches and still achieves good performance. Here we also give a short analysis as to why linear interpolation is needed and useful for finding a likelihood of the user being at specific CP. Moreover, we tried different rectangular grid spacing and tested how it affects the overall accuracy and performance. The experimental setup was performed in the case of a relatively open-plan space and in the case of walls presented along an observed line which would show how likelihoods change when this kind of obstacle is present. We compare results against well-known competing approaches showing the superiority of the proposed method.

The last chapter - chapter 7, discusses potential future work, reviews the thesis's con-

clusion and highlights the achievements of the work. Here an overview of Support Vector Machines (SVMs) classifier together with Neural networks classifier is presented. Their comparison to k-nearest neighbor is briefly mentioned. One section is dedicated to multi-modal fusion framework using adaptive-based weighting. It is a very general algorithm from which many particular ones can be derived and used separately or even fused. Research contributions are summarized together with the difficulties encountered when evaluating this work. Finally some concluding remarks are given.

Eventually, the Appendix discusses initial WLAN and image data analysis together with a description of the properties of both sensing modalities. Signal strength and image data processing can be difficult because of the data complexity of each modality. In this chapter the RSSI and image data collection processes are explained: InSSIDer as WLAN network scanner software for Microsoft Windows produced by MetaGeek, LLC [6] and image acquisition using a camera and FFmpeg software.

Chapter 2

Applications and Context

2.1 Introduction

Localisation of people is considered to be a key requirement for ambient assisted living technology platforms. In the previous chapter we mentioned that wearable cameras together with a variety of other sensors can be used in a range of ambient-assisted living applications. In this chapter we give an overview of these applications motivated by the use of a multi-sensor platform. This includes the huge area of lifelogging, described in section 2.2, enabled when using the well known Vicon Revue or more generally wearable cameras and other similar multiplatform devices. Wearable cameras and their applications are described in general in section 2.2.1. More specifically, Vicon Revue and its technical characteristics and applications are given in section 2.2.2. The potential applications as assistive technologies for the memory and visually impaired are presented in sections 2.3.1 and 2.3.2 respectively. Tourist-oriented applications that would greatly enhance user experience in museums, galleries and other similar institutions are described in section 2.3.3. Finally, various potential applications in health and daily life monitoring in general are given in section 2.3.4.

Various sensors have been used in indoor localisation, GPS, WLAN, images, video, audio, accelerometer to name just a few, and have been part of many successful commercial examples. Moreover, some authors tried to integrate two or even three sensors to improve

the recognition performance, accuracy or precision. Although GPS signals are stable indoors they are weak or non-existent. Audio represent reliable source and have been used to localise users indoors but its performance is worse than performance of other descriptive sensing modalities such as images. Also as this work describes the user context, using images is more informative than using audio or accelerometer data. Using video is more informative but computationally much more expensive than using images thus being worse option compared to images. Thus, in this thesis, WLAN and image sensors are chosen and used in the data collection and later in the fusion process. These sensing modalities are widely available nowadays on almost any smart device thus being cheap and easy to use.

2.2 Lifelogging

Memory represents the ability of an organism to store, retain, and recall information and experiences. Since all important information and events become part of oneself, it is very important to have access to them at all times. In the past, exaggeration and repeated storytelling were used to remember. Shared memories can help make stronger bonds between the people who share them (e.g. family members, group of close friends). Of course, it goes without saying that sight is one of the key ways in which we interpret information about our environment. Unfortunately, the number of memory and visually impaired people is not small, thus giving researchers an opportunity to develop assistive technologies [80].

Lifelogging is defined as digitally recording different aspects of one's daily life for one's own exclusive personal use. It is a form of reverse surveillance, often called *sousveillance*, referring to subjects performing the watching of themselves [52]. A diary as a form of memory of activities and aspects of an individual (and more recently a blog) is a well known example of lifelogging. One of the main goals of lifelogging is the effect of total memory/recall [52]. Users have tried to achieve this using automatic capturing of data from numerous sources such as e-mails sent and received, web pages visited, audio recordings of conversations, etc. An important area in this field is visual lifelogging, i.e. an individual capturing activities using the medium of images or video (camera). Visual lifelogging can

be associated with different devices such as the Vicon Revue and more generally wearable cameras, other multiplatform devices, etc. These will be explained in greater detail in the following subsections. Many research papers have concentrated on the area of visual lifelogging, more specifically in miniaturising the hardware devices and/or managing a large amount of data. Only relatively few works have tried to structure and organize these large amounts of image data. In this work we structure such data to find a user's position in an indoor environment.

2.2.1 Wearable cameras and multiplatform devices

The main attraction of wearable cameras is their advantage as a novel, pervasive, lightweight and wearable technology. The Vicon Revue, shown in figure 2.1, (previously known as SenseCam) has been used by a number of research institutions and companies and thus represents the most common example of this technology. Wearable cameras are usually fixed to some part of the wearer's body (neckworn or handworn). Although wearable cameras contain mainly image or video sensors, recently many cameras are emerging with various different sensors. They constitute an easier way of indexing and collecting a wearer's experiences by taking images and in the near future other sensed data such as motion, GPS, WLAN data, etc. Image capture, for example, can be done with user intervention or automatically. This is achieved whenever a change in movement, lighting or temperature trigger the internal sensors. The camera also contains accelerometers for improving the image quality such as removing/reducing blur and stabilizing the image. Also the camera can take the data in a timer mode. One can select the various options regarding these capturing methods.

Humans receive most of information through their eyes. Thus, the Vicon Revue images taken in a month or just a single day usually contain most of what the wearer can experience. Using a computer application the files can be automatically uploaded and analyzed. They can be observed at different speeds (e.g. images at $3 - 10$ frame/s using a technique called Rapid Serial Visual Presentation - RSVP). Another way of analyzing the data is to automatically segment data into events and then to use an event-detector tool which can

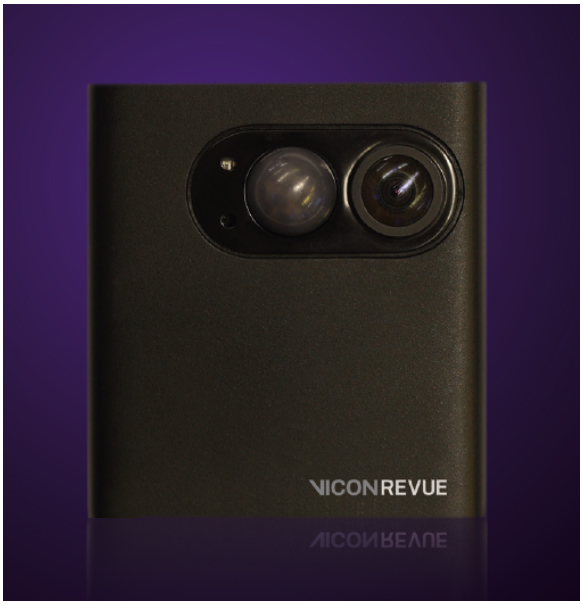


Figure 2.1: The Vicon Revue camera - an example of a wearable camera

detect an event using features which are representative of that event.

2.2.2 Vicon Revue

The Vicon Revue evolved from the SenseCam developed by Microsoft Research in Cambridge [80]. It is a small lightweight (94g) wearable device (of size: 6.5cm (w) $\times$7.0cm (h) $\times$1.7cm (d)), illustrated in figure 2.2, that is worn on a lanyard around the neck. The camera incorporates a 3 megapixel digital camera (giving images of resolution 2048×1536 pixels) and multiple other sensors including: a thermometer, sensors that detect different levels of light, 3-axis magnetometer (compass), accelerometer with a multiple axis to detect motion and a passive infrared sensor to detect whether a person is present. It includes a fish-eye lens that can provide a full field view of 130 degrees in total. The Vicon Revue device uses a very simple algorithm to trigger its camera: a photo will be taken every 30 seconds by default (this interval is configurable) or alternatively a change in sensor readings will trigger capture [81]. Some image examples are given in figures 3.1(a) and 3.1(b).

An image can be taken automatically if during a fixed time period (up to 50 seconds) no image is captured based on the activity of the sensors. It can store up to 8GB of

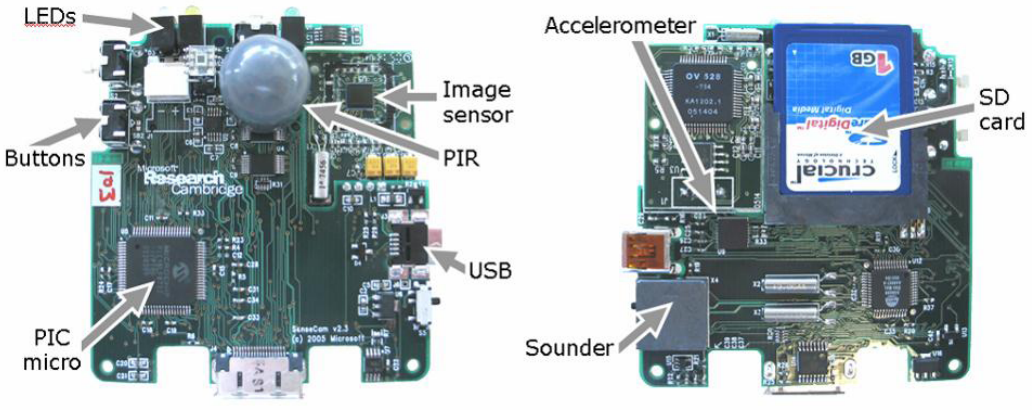


Figure 2.2: The front (left) and the back (right) of the Vicon Revue PCB [81]

data in its memory (around 18500 images). Around 2000 images on average can be stored in the camera, together with other sensor data. The Vicon Revue is capable of storing approximately 15 days worth of data before it is required to download the images to a standard PC via a USB cable. Beside images, Vicon Revue also senses and records data from other sensors. The log file also records the reason for capturing each image (e.g. change in sensor readings, timed capture or manual shutter press). The Vicon Revue has a built-in real-time clock that accurately timestamps all files.

(a) (b)

Figure 2.3: Vicon Revue image examples

Many researchers have discussed the capabilities of Vicon Revue or SenseCam in recent years. For example in [66] a novel application is developed to replay the images of the camera in specific order. Some others, such as [81], discuss the so called visual diary created using

a Vicon Revue. This particular work showed the great improvement in everyday life of a female patient with limbic encephalitis. She had to recall events that occurred during that day or week, and she managed better with the help of Vicon Revue. Throughout the world there have been investigations conducted and research works on the benefits of visual lifelogging using Vicon/SenseCam devices, not only in the lifelogging community, but also increasingly in the cognitive psychology community.

2.3 Potential application domains

2.3.1 Memory impaired

A key application of lifelogging is as a memory prosthesis. It is possible to automatically segment and cluster images into specific events during a day or activity (when he/she had dinner in the evening, when he/she went for a walk in the afternoon, when he/she was talking to colleague Milan, etc.). By inspecting images of this event this person can remember and recall talking to a friend Milena, also about how quickly her granddaughter is growing, etc. For certain events one may prompt the person about what one was doing. The person can be provided with a list of other potentially relevant events, which one quite enjoys looking at as they trigger some memories. Clearly, location information, indoor or outdoor can play an important role here [29, 37].

Wearable camera images can be collected and used as part of a larger dataset for developing algorithms that allow image content to be categorized automatically, thus quickly determining what the wearer did [36, 53]. This would be of great help to the memory impaired. Moreover, the categorization of recorded images can be used to alter the way in which images and sequences are presented to the user [51], which will become important as the size of datasets become bigger. Work has also been carried out to explore ways of automatically detecting unusual and pertinent images from large data sets [53]. The examples given in this section demonstrate the importance of indoor and outdoor localisation based on images as a potentially enabling technology for all these potential application domains.

2.3.2 Visually impaired

For most humans, the largest amount of information is received through eyesight. This is precious information about the outside world without which it is very difficult to perform everyday activities. There are many devices that can assist visually impaired people in localisation. From the 50's significant efforts have been made to provide everyday aid for visually impaired individuals [21]. They range from simple canes to very sophisticated computer-based aids and have proved to be very effective. However, adequate solutions for assistance in navigation for the visually impaired have not been developed yet. The development of the smart wearable camera could potentially provide navigation, context awareness and aid. Environmental camera-based sensing is a very attractive solution due to the information available through this sensor, its low cost, and low power consumption and its functional similarity to the human eye.

Visual recognition for the visually disabled is a very challenging task since the images represent complex and time-consuming data to process. The most sought after device is probably a wearable camera that can detect text regions anywhere in space around a user and translate and transform the text into a representation understandable to him such as sound or braille (an example is shown in the figure 2.4). Some camera devices can recognize and read out characters and the whole text information for the user [84, 101, 108]. A text capturing device equipped with a camera is presented in [157]. It can detect and track several multiple regions of text in the surrounding scene. A complete framework for text detection and tracking in real time is demonstrated and presented in [118]. This wearable camera automatically and successfully detects and tracks text regions in indoor scenes. Another example of a wearable camera employs edge detection and color correction together with the optical zoom to help people who suffer from gradual visual loss over the total view field. It is also used by color-blind people.

Most previously published wearable camera works use camera systems that can easily show tactile information to the user (e.g. one example is given in [18]). For recognizing gestures, standard camera systems perform well as the background which is stationary

Figure 2.4: Wearable Camera System

can be separated from the image, so that visually impaired can see clearly with help of simple computer vision processing. The background changes quickly when the camera moves fast and it is difficult to distinguish between background clutter and nearby objects. The VibraVest [116] is a specialized collision avoidance helper that provides 3D range information but requires a very small radar system which can be quite expensive hardware. The user gets the depth information (similarly to Microsoft Kinect) as a tactile feedback i.e. with an array of vibrotactile actuators shown in figure 2.5.

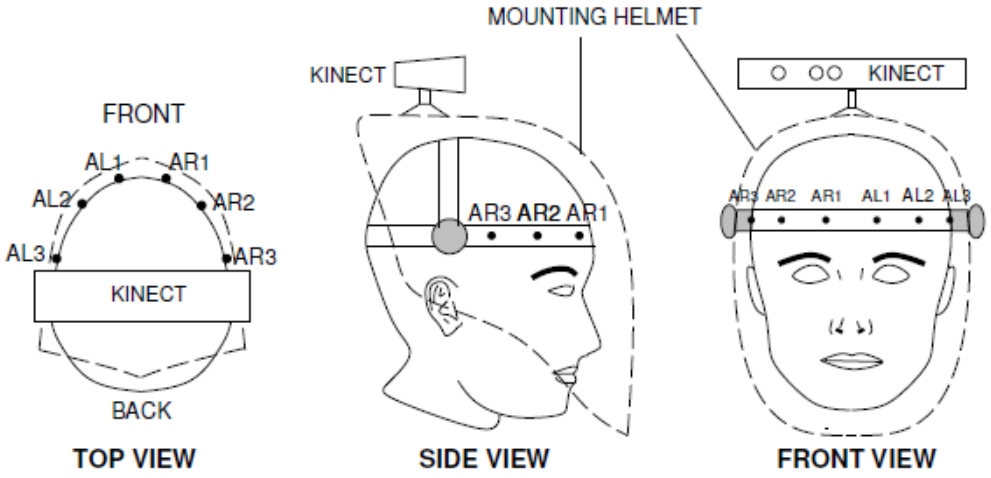


Figure 2.5: Wearable actuator and sensor together with a helmet; the motor actuators are placed around a user's forehead [116]

The Stereo Vision based Electronic Travel Aid (SVETA) system consists of a head-set

on which stereo earphones and stereo cameras are placed [22], as shown in figure 2.6(a). The stereo cameras are placed in the front of the headgear, located slightly above the position of the eyes as shown in the figure 2.6(b). The stereo camera takes video and images of the scenes that are in front of a visually impaired user. This data is then processed and the information is transfered to the blind user by voice commands and musical tones using earphones. All these fruitful examples show that robust user localisation, and its extensions to collision avoidance, scene recognition, etc, is a much sought after technology as a key component of eventually realising truly useful systems.

(a)

(b)

Figure 2.6: SVETA system: (a) Prototype system (b) SVETA system worn by a blind user [22]

2.3.3 Tourist-oriented

Tourist-oriented application of wearable cameras can help visitors inside galleries and museums while visiting exhibits of their preference. They could provide visitors with descriptions of exhibits in museums. With the help of some other built-in sensors such as WLAN or simply punching in an exhibit ID number cameras can become aware of specific location

and give more information regarding the specific exhibit. It can also deliver relevant audio content. Sometimes wearers are required to take a preselected and marked route. If the device is made to be aware of a user's location this would enable visitors to experience the museum freely, instead of using a predefined route. Images taken during the visit can give more precise information about the particular things related to the exhibits the user had seen while he was present inside the museum/gallery. The tour experience could be potentially enriched by providing extra information on the exhibits that are currently observed or that are of particular interest. Information on how visitors explore the museum could also be determined, thus showing the frequency of visits of the most popular exhibits. It can also show whether they are rarely visited or placed in sub-optimal positions in terms of user experience. Something similar was presented in [15]. Clearly, an indoor localisation system based on WLAN and image data has many potential uses as a tourist-orientated application in this context.

2.3.4 Other applications in health and medicine

Besides its benefits to memory and the visually impaired or museum visitors, many other applications have been proposed for wearable cameras many of which require the user location as a key enabler. For example, some works show how using the camera as a tool can assist people with physical and mental health problems such as learning disabilities, neuro-logical conditions and autism [61]. Examples include assessing how much (physical) activity a patient undertakes or monitoring the number of interactions in a typical week. Moreover, wearable cameras can provide ground truth data (usually difficult to obtain) and in the near future it is conceivable that manual data analysis will be replaced by automated approaches.

A wearable camera has been used in innovative ways to get a first-hand account of the lives of certain groups of people for whom this would otherwise be difficult (children with autism and people with learning disabilities wear the device during the course of a day). The images are reviewed by their carers to better understand how they experience

(indoor) daily life. It can be monitored how lighting conditions in indoor environments affect people present. Researchers have used wearable cameras as a tool for ethnography as it can provide recording behaviour for both wearer and people the wearer interacts with [34]. In education scenarios a wearable camera provides a method to enhance the reflective practice techniques sometimes used during on-the-job development [61]. In all scenarios the user location provides additional useful context for subsequent analyses.

2.4 Conclusion

From the many examples given in the previous sections it is clear that indoor localisation using wearable cameras or multiplatform devices has potential to be an important enabler in many application areas. Although some applications and devices are not developed yet or exist only as a proof of concept it is very likely that they will start to appear and be sought after in the near future. This also provides the possibility for integrating other sensing modalities such as audio and tactile information that would enrich user experience and give useful feedback. As this technology can be implemented on a small platform, e.g. a smart phone, and as its costs are low it is potentially accessible to everyone. Moreover, due to its size it can be worn on a daily basis thus proving to be useful for lifelogging and various health and medicine analyses. As such, it can be considered as a technology whose "time has come". Robust indoor localisation will be a key ingredient in its ultimate usefulness and successful take-up.

Chapter 3

Literature Overview

3.1 Introduction

Localisation is the process of determining the location of a person's (or object's) position with respect to other already defined reference position(s) or system (relative location). Many different technologies can provide location information. Nowadays there are many devices that in addition to WLAN capture also sense a variety of other modalities e.g. motion, Global Positioning System (GPS), Bluetooth. Whilst outdoor localisation is taken care of on this platform via GPS, indoor localisation is still an unsolved issue. Indoor localisation can be important enabling technology when using these platforms in a variety of ambient assisted living applications as well as assistive technology for the memory and visually impaired. In this chapter we concentrate on the two important technologies related to this field of research: RF-based and image-based localisation.

There are indoor and outdoor localisation techniques. Indoor localisation can be based on WLAN, RFID (passive tags are very cost-effective, but do not support any metrics), Ultrawide band (UWB) that give reduced interference with other devices, Infrared (IR) which is previously included in most mobile devices, Visible light communication (VLC) that can use existing lighting systems and Ultrasound waves that move very slowly which results in much higher accuracy. Also indoor localisation systems are based on audio,

image, accelerometer, magnetic, etc. data. Outdoor localisation systems are: GPS together with Differential GPS (DGPS), GLONASS, Galileo and Compass. Also some image-based approaches have been used outdoors as well.

First we give an overview of outdoor localisation systems in section 3.2. Although Global Positioning System (GPS) is an outdoor localisation technology, nowadays it represents a (standard) ubiquitous localisation tool, and is thus presented here briefly together with its localisation error simply for completeness (section 3.4) and for conceptual comparison against indoor localisation methods presented in this thesis. To employ RF indoor technologies various metrics are defined. The most well-known RF-based localisation metrics used in various scenarios are given in section 3.5. Section 3.7 presents received signal strength characteristics and discusses mapping between different devices used in RSS collection. An overview of RF-based (indoor) localisation techniques is given in section 3.8 while some specific RF technologies and applications used in indoor user localisation are presented in section 3.8.1. Section 3.8.2 discusses some prominent examples of RF-based localisation systems.

The well known image-based localisation methods employ feature detection algorithms as metrics in the localisation process. Section 3.9 gives an overview of the most used feature detectors. Image-based localisation techniques are presented in section 3.10 consisting of a short overview of prominent image-based localisation methods (section 3.10.1). We then discuss how RF and image-based approaches could be fused in order to obtain better results in the localisation process. A discussion that indicates why the fusion of image and WLAN data is meaningful is given in section 3.11. Section 3.12 discusses fusion of different modalities in general. Various detailed examples of hybrid and fusion of WLAN (RF) data and image data is presented in section 3.13. These examples show that fusion of different modalities has attracted the attention of researchers throughout the world and should be further investigated. Motivation for using fusion of WLAN and image data can be given in the context of:

• using more than one modality to improve accuracy and/or precision

• obtaining information about user context using more than one (or even more) sensing modality (in this case - images)

• having these sensing modalities on a single easy-to-use and widely available device (e.g. a smartphone)

• using modalities robust enough to environmental changes

• sensing modalities that need to be processed relatively fast (almost in real time)

3.2 Outdoor localisation systems

Localisation systems can be indoor and outdoor localisation systems based on where a user is localised. The most prominent examples of outdoor localisation systems are: Global Navigation Satellite System (GNSS), broadcast networks developed for some other purposes such as cellular phone networks, radio-frequency identification (RFID) tags and RAdio Detection And Ranging (RADAR). There are currently four main GNSS systems in different states of development. The United States NAVSTAR Global Positioning System (GPS) is the only fully operational GNSS. The Russian GLONASS is a GNSS in the process of being restored to full operation which should be reached by 2011. The European Union's Galileo positioning system is a next generation GNSS in the initial deployment phase. Full Orbit Constellation (FOC) should be reached in 2015. China is building up a global system called COMPASS but also referred to as Beidou-2. Beidou-1 is the a regional augmentation system. On these navigation system pages we give an overview of the details of these different systems: GPS, GLONASS, Galileo, and Compass 3.1.

The number of satellites notation $21+3$ for GPS and GLONASS mean that the minimal

Features	GPS	GLONASS	GALILEO	COMPASS
Number of Satellites	$21+3$	$21+3$	$27+3$	$30+5GEO$
Number of orbital planes	6	3	3	$notgiven$
Semi-major axis	$26600km$	$25440km$	$29600km$	$21500km$
Orbital revolution period	$11:58H$	$11:15H$	$14:07H$	$12:35H$
Inclination	$55deg$	$64deg$	$56deg$	$55deg$
Satellite Mass	$1100kg$	$1400kg$	$700kg$	$2200kg$
Solar panel area	$14m2$	$23m2$	$13m2$	$notgiven$

Table 3.1: The table shows a comparison of some of the key features of different GNSS systems

system constellation consist out of 21 satellites with 3 active in orbit spares. So there are a total of 24 satellites in orbit sending signals. Galileo will consist out of a minimum of 27 satellites with 3 active in orbit spares. So there will be 10 satellite in each orbital plane. One of these satellites is the active in orbit spare. The higher number of satellites in the case of Galileo is caused by the fact that the Galileo satellites fly above the GPS and GLONASS satellites.

3.3 Fingerprinting and Non-fingerprinting-based solutions

Fingerprinting-based localisation solutions are based on mobile station which extracts radio fingerprints, i.e., features from one or more metrics of the radio signal measured at predefined points in the environment. These radio fingerprints are proportional to the distance between the mobile receiver and the emitting station. Common metrics include the direction or angle of arrival (AOA), Received Signal Strength (RSS), or time of flight (TOF) of the incoming radio signal [52]. A radio map or database of fingerprints is created, storing signal feature values at each location along with the corresponding spatial coordinates. Localisation is commonly achieved by proximity techniques but more accurate localisation can be achieved using a triangulation-like process, in which several candidate locations (each with a fingerprint bearing some resemblance to that of the received signal) are geometrically combined to provide an estimate of the receiver location in space. Fingerprinting seems to

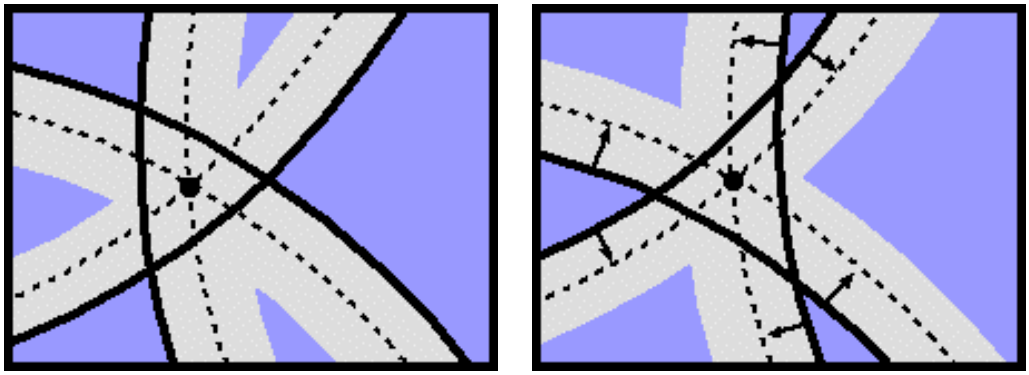
provide reasonable localisation accuracies without excessive hardware requirements. The most pressing challenge however is the non-stationarity of the radio map. This is reflected as differences in the measured signals during the on-line and off-line phases at the same exact location.

Non-fingerprinting-based solutions are achieved without a priori analysis of the radio properties of the environment. Four of these articles, all of them based on UWB radio signals, rely on signal triangulation as the sole localisation technique [28,88,23,41], while in [20] localisation is achieved by proximity and scene analysis. Indoor localisation based on triangulation of radio waves is a non-trivial problem because the transmitted signal can suffer obstructions and reflections. As a consequence, Non-Line-of-sight (NLOS) conditions emerge. In the presence of NLOS conditions, the radio signal can travel to the receiver through a non-direct path, giving rise to erroneous distance estimates. To overcome these problems, the use of UWB radio signals has become the most novel solution in radio frequency-based solutions. The properties of ultra-wide band, short duration pulses mitigate the propagation problems associated with multipath propagation.

3.4 The Global Positioning System (GPS)

The Global Positioning System (GPS) is the most prominent technology in positioning and navigation technology. GPS uses satellites orbiting around the Earth and ground devices to obtain a user's position anywhere on Earth. Using a small cheap receiver anyone can localise oneself. GPS has drastically advanced in terms of accuracy in recent years and has become an important device for everyday activities [50]. Global Positioning System satellites transmit signals to receivers which are on the Earth. These receivers receive these signals passively and they do not transmit any signals. GPS performs poorly when the receiver does not have line of sight to the satellites, due to obstacles. In addition the presence of significant multipath can cause that GPS signals are practically non-existent. The main problems outdoors are urban canyons and/or forested areas. Each GPS satellite transmits data containing its location and the current time measured by its atomic clocks.

They have all their operations synchronised in order to start their operation simultaneously. Transmitted signals reach receivers, traveling at the speed of light. These traveling times can be different due to different distances between transmitters and receivers and also due to different atmospheric content. The distance between them is sometimes difficult to calculate as receivers do not have atomic clocks. This can introduce significant errors while calculating the distance as discussed in [57].



(a) (b)

Figure 3.1: User localisation using GPS: the intersection point and the area of uncertainty (gray bands) [57]

A GPS receiver knows the location of the satellites, because that information is included in satellite transmissions. By estimating how far a satellite is, the receiver also can determine its location somewhere on the surface of an imaginary sphere centered at the satellite. Each satellite defines an imaginary sphere. The spheres will be used in obtaining position of the GPS receiver. The typical operation of GPS is shown in figure 3.1. The dashed lines in figure 3.1 on the left give the intersection point together with the area of uncertainty given with gray bands. Due to the various circumstances given above, the distances to the GPS satellites can only be estimated but the relative distance can be calculated accurately.

The radii of the spheres are known and there is one physically meaningful intersection point (two in total) of the three spheres. A GPS receiver gives the location of the spheres which is drawn with the solid lines. As they are given with errors they will typically not intersect at one single point. Thus the sizes of the spheres need to be adjusted until a single intersection point is obtained and its coordinates are calculated (to the right in figure

3.1). This is all done by the GPS receiver. Three spheres are needed to calculate a two-dimensional position and four spheres to find the three-dimensional position. At least 24 satellites are available at all times. The satellites, operated by the U.S. Air Force, orbit with a period of 12 hours. Ground stations are used to precisely track each satellite's orbit [57]. The accuracy of a GPS receiver depends on the actual receiver and its technical characteristics. Accuracy usually ranges from 2 to 8 meters. There are very precise GPS systems, so called Differential GPS (DGPS), that are highly accurate giving an accuracy of about 0.5 meters. DGPS requires that another receiver is fixed at a known position in the near vicinity. Measurements recorded by the stationary receiver are used to correct the positions obtained by the mobile unit.

3.5 RF-based localisation metrics

There are several different technical approaches employed in RF-based localisation nearly all of which utilize a variety of RF metrics. The most often used RF metrics are: Time of Arrival (TOA), Time Difference of Arrival (TDOA), Received Signal Strength (RSS) or Received Signal Strength Indication (RSSI) and Angle of Arrival (AOA).

3.5.1 Time of Arrival

In Time of Arrival (TOA)-based localisation, the arrival time of the First Detected Peak (FDP) of the received signal is measured to calculate the time of flight and eventually the distance between the transmitter and the receiver. There are two types of errors in TOA localisation: Undetected Direct Path (UDP) conditions and multipath effects. In UDP, the direct path (DP) is obstructed by objects and is below the level of noise, so the receiver receives an incorrect path as the direct path, thus obtaining errors when measuring position [13]. Multipath effects arise as a result of the fact that the direct path and the reflected path are received at the receiver which therefore leads to errors in localisation (due to the overlap of the signals at the receiver, as shown in figure 3.2) [82]. The need for synchronisation represents the key drawback of this approach.

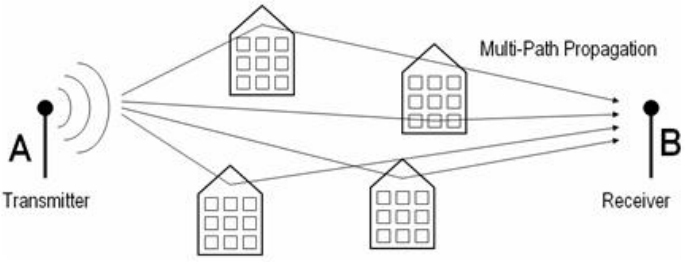


Figure 3.2: How presence of physical structures produce multi-path propagation [82]

3.5.2 Time Difference of Arrival (TDOA)

Using TDOA measurements is especially suited to localisation of high-bandwidth transmitters, e.g. radars. Knowing the time difference of arrival between the transmitter and two receivers localises the transmitter to the points of a hyperbola. Introducing the third sensor (second TDOA measurement), one can localise the transmitter at the intersection of two hyperbolae. Time Difference of Arrival (TDOA) measures the difference in arrival time $\Delta\tau$ from the transmitter to the receivers. A typical situation is presented in figure 3.3.

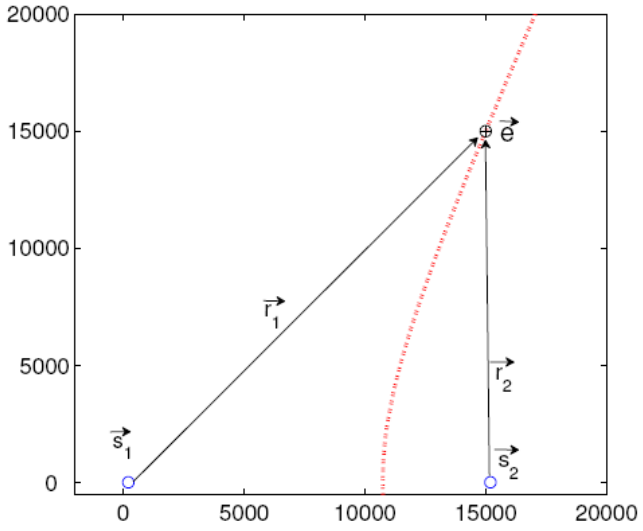


Figure 3.3: TDOA approach and constant TDOA curve (red dashed line) [128]

$$\Delta\tau = \Delta r/c = (r_1 - r_2)/c \tag{3.1}$$

True time difference of arrival is directly proportional to the difference in distances between the transmitter and the receivers (given in equation 3.1) where c denotes the speed of light; r_1 and r_2 denote the distances between the transmitter and the first receiver and between the transmitter and the second receiver respectively [128]. All points on the red dashed line in figure 3.3 have the same distance difference to the two receivers, and therefore the same true time difference of arrival. This constant TDOA curve can be constructed as a function of x and y transmitter coordinates, and solving it one can obtain the transmitter coordinates, i.e. its location.

3.5.3 Received Signal Strength (RSS)

The simplest localisation metric for a mobile host (MH) in a WLAN environment is Received Signal Strength (RSS). The received signal strength value generally decreases with distance from the transmitter. If one knows how the signal attenuates with distance as well as transmitted and received power one is able to estimate the distance between the transmitter and the receiver. A reasonable model for the received power P_r at distance d from the transmission source is given in equation 3.2:

$$P_r(d) = P_0 - 10n_p \log(d/d_0) + X \tag{3.2}$$

where P_0 is the reference power (dBm) at the distance d_0, n_p is a path loss exponent (PLE - covers deterministic parameters of the environment) and X (dBm) represents a random variable which follows a log-normal distribution [126]. There are several challenges when using an RSS-based approach to achieve the distance estimation. In different environments (urban, outdoor, indoor, free space, with obstacles, etc.) the performance and accuracy is affected by many factors e.g. transmitter-receiver distance, RSSI uncertainty, non-line-of-sight (obstructions), antenna radiation pattern, gain and height, reflections due to multi-path transmission, vegetation diffraction and scattering [126]. Nevertheless this is very often the first choice as the basis of a localisation technique.

3.5.4 Localisation using Angle of Arrival

Angle of Arrival (AOA) is defined as the angle between the propagation direction of an incident wave and a given reference direction (orientation). Orientation is measured in degrees in a clockwise direction from the North. When the orientation is 0 deg or pointing to the North, the AOA is absolute, otherwise, relative. An antenna array is typically used to obtain AOA measurements. In figure 3.4 (from [144]) angles θ_1 and θ_2 are measured at receiver u, are the relative AOAs of the signals sent from transmitters b_1 and b_2, respectively [144]. If the orientation of the receiver is denoted by $\Delta\theta$, the absolute AOAs from b_1 and b_2 can be calculated as $(\theta_i + \Delta\theta)$, $i = 1, 2$.

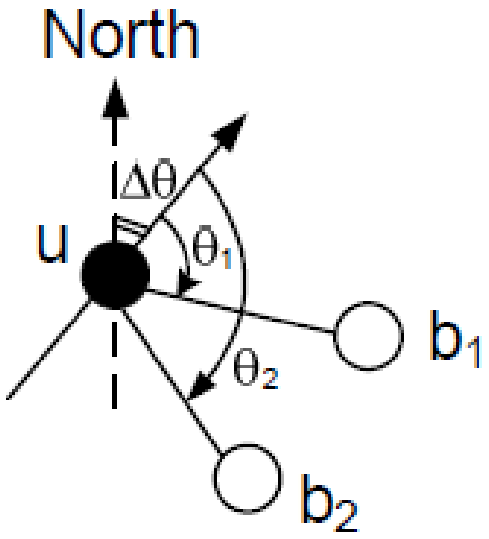


Figure 3.4: Localisation using AOA [144]

Every AOA measurement that corresponds to a transmitter constrains the location of the receiver to lie along the line starting at the transmitter in the direction given by the angle. The location of the receiver u is given by intersection of all lines when two or more non-collinear transmitters are operating. This technique is well known and commonly used in cell-phone positioning.

3.6 Fingerprinting

A location fingerprint based on WLAN such as RSSI is the key component in a user location representation. The fingerprint is labelled with a location information. The fingerprinting technique requires a training phase (or off-line phase) to collect location fingerprints for all CPs in the operating area, before the actual deployment - a testing phase (or on-line phase). The measurement dataset collected during the off-line and on-line phase are called a training and a testing set respectively. The location fingerprints and their labels are maintained in database and used during the testing phase to estimate the user's location. The label and fingerprint are usually denoted as a tuple. The tuple of real coordinates can vary from one dimension to 5 dimensions which includes the 3 spatial dimensions and 2 orientation variables expressed in spherical coordinates. A location information of a two-dimensional system with an orientation is usually expressed as a triplet (x, y, d) where x and y represent CP coordinates and d represents one of the 4 orientations (North,East,South,West). It is commonly acknowledged that the RSS is the simplest and most effective signature for location fingerprints because it is readily available in all WLAN interface cards. The RSSI is found to be more location-dependent than the signal-to-noise ratio (SNR) because the noise component is rather random in nature. However, the RSSI itself can fluctuate over time for each access point and location. Each RSSI element can be considered as a random variable. The location fingerprint is usually denoted as an array or vector of signal strength received at any position in the location-based area. The size of the vector is determined by the number of access points that can be heard. In this thesis all non-confident access points were not taken into consideration and thus removed from the localisation process: weak access points (with RSSI values bigger than 80), non-stable access points (appear very irregularly and/or disappear quickly) and access points whose RSSI values oscillate a lot (e.g. going too quickly from 66 to 49 and rising again back to 66 too quickly). Also the RSSI measurements in the training phase were collected using laptop and using all four orientations per calibration point. Each orientation had the same number of RSSI measurements consisting of RSSI values taken at regular time intervals (time stamps) from all confident access points - a RSSI

observation. In the testing phase only one RSSI observation and arbitrary user orientation was used.

3.7 RSSI characteristics and mapping

It is desirable that the same device is used in the training and in the testing phase. This is to prevent the database to be filled with data values that are too high or too low, in which the amount of over or underestimation can be diminished. Second, localisation algorithms can provide false information about a user's location, as they are dependent on the building structure, and thus highly influent on multipath effects. Fingerprinting requires a lot of pre-processed work, in the form of site surveying [2]. Also, the fact that it is capable of handling multipath makes it an advantage at the same time. As soon as large objects, including people, are moved, the fingerprinting process will be affected. The same applies for the fixation of MAC addresses [155]. Since the RSSIs are dependent on the MAC address, patterns might not be found, as soon as an access point (AP) has been replaced by another. This needs to be constantly updated, and from time to time a new survey has to take place.

The papers [183] and [23] have mentioned that fingerprinting and thus localisation result can be different when using different WLAN card and devices (laptop, mobile phone, iPAD, antennae), sometimes substantially. It is also well known [23] that different WLAN card distributors choose to measure WLAN RSSI data differently [23]. Some cards, for example IEEE 802.11 Cisco's wireless card, have 100 different values, while e.g. Atheros wireless card has 65 levels. Device drivers and WLAN card use these values internally and every distributor of WLAN cards has its own specific set of RSSI values going from 0 to some $RSSI_{Max}$ [183]. Moreover, the distributor defines granularity and range for the power (in dBms) [183]. A WLAN card with good granularity or bigger set of RSSI values gives better localisation accuracy since it better differentiates between two locations. A receiver with smaller standard deviation is better for localisation as it shows fewer different values of the RSSI at that location. If RSSIs have a small standard deviation, the probability to choose a nearby position instead of the real one is smaller. A wireless card with the widest RSSI

range can obtain a higher resolution signal. Thus, some RSSI measurements of similar value (close to each other) can be in fact represented as one RSSI value! Since collection of the training data depends on wireless card distributor it is crucial that the same card is used during the training and testing dataset collection as the localisation process in case when different cards are used in the training and testing processes might differ significantly [183] unless a special mapping function is developed for localising a user when using different devices in the training and the testing phase. In the indoor localisation area a WLAN card that has a bigger mean RSSI value at the same position is better but for positioning purposes, the standard deviation and the range of RSSI values are more important [23]. An essential feature of WLAN cards used in indoor positioning is the capability to scan the nearby APs actively, passively, or using both. Experiments with different wireless cards showed that the scanning information from only SmartMedia card (SMC) can be acquired correctly [183]. For indoor positioning purpose, the card that has the widest range and the lowest standard deviation of RSS should be employed. The software tools provided by these cards are different in quality. The cards that have a good standard deviation and RSSI measurable range should be used for most measurement experiments. The distance between APs also differentiates various locations. In case of moving these APs the system will have a different region where position location can be performed. The distance in signal space is very different to the distance in physical space. In-depth analysis to estimate the resolution according to the quantization is beyond the scope of this work due to limited information of actual quantization step of each wireless card.

3.8 Overview of RF-based (indoor) localisation

RF-based localisation technologies face difficulties such as obstructed wireless propagation in complex environments and unreliable performance in indoor areas. RF propagation is usually multipath and low probability Line Of Sight (LOS) propagation of the signal between a transmitter and a receiver. Thus, accurate RF-based positioning is a difficult and challenging task.

3.8.1 RF technologies used in indoor user localisation

Analyzing and measuring different properties of waves generated by a transmitter and received by a receiver, present opportunities for estimating the location of a mobile user. These estimates can be divided into several categories based on the RF technology used such as IEEE 802.11, Ultra-Wideband (UWB), ZigBee, or Bluetooth, whether RF is the sole localisation technology or part of a hybrid solution [77, 159, 9]. A mobile station extracts features from one or more metrics of the radio signal at the selected points in the space. These features represent radio fingerprints and they give a good basis for fingerprinting-based localisation. Some of them (not AOA for example) depend on, and are proportional to, the distance between transmitter and receiver. A radio map or database of fingerprints is created by storing signal feature values together with corresponding spatial coordinates. Localisation is commonly achieved using RF localisation metrics and proximity techniques (i.e., finding the closest match between the features of the received radio signal and those stored in the radio map). More accurate results can be obtained using a triangulation process. It is a process of determining the location of a point by measuring angles to it from known points. For a selected location, fingerprinting measurements are taken for each possible orientation (for each orientation the fingerprint will be different). Fingerprinting is a low cost localisation method without significant hardware requirements and yields good accuracy [159]. Non-fingerprinting-based solutions (UWB) are achieved without a priori analysis of the radio properties of the environment [9]. UWB, for example, relies on signal triangulation as the localisation technique (with degraded signals this becomes a complex and difficult task).

Indoor localisation based on personal and local area networks

WLAN is aimed to provide local wireless access to fixed network architectures. The IEEE 802.11 working group published 802.11b in 1999, and 802.11g. WLAN is becoming increasingly popular today, especially in indoor and public areas. Most of the WLAN products are based on 802.11b, and work in the 2.4 GHz band which is unlicensed and can be used for data

transmission if a number of rules are followed. Many signal strength (RSSI) [165, 41, 8, 77], angle of arrival (AOA) [46, 77], time of arrival (TOA) [176, 177, 77] and time difference of arrival (TDOA) [103, 9] based techniques have been employed for location estimation in WLAN indoor environments.

ZigBee is a low-cost and low-power wireless network standard which is widely deployed in wireless control and monitoring applications. It uses three bands: 868 MHz, 915 MHz and 2.4 GHz. The localisation techniques based on this standard usually use fingerprinting RSSI values which are pre-stored in a database and retrieved to locate a user's position [154]. It means that a "blind" node is placed at pre-defined anchor positions in advance. The blind node continuously sends requests to its surrounding reference nodes and receives responses from these reference nodes. The system continuously record these responses to analyze them until they become stable [154]. The mobile object is located by comparing the current RSSI values with the pre-stored maps of RSSI values.

Some other RF-based localisation systems used mainly for "near to user" application scenarios employ Bluetooth. This is an ad-hoc network whose inquiry signals make inquiries about near-by Bluetooth stations. This process is achieved using different RSSI power levels sequence. Thus, low power levels will detect devices in close proximity and higher power levels will locate devices further away [63]. This approach requires a fixed or "anchor" node which establishes the position of nearby mobile nodes. The nodes that were localised can be used for the position estimation of the other (undetected) nodes, creating an ad-hoc network. The Bluetooth protocol operates at $2.402 - 2.480$ GHz and error estimation is around 1.88 meters [63].

Ultra-Wideband (UWB) technology can be used to track and/or detect the position of a (moving) object with even centimeter level accuracy thanks to its cooperative symmetric two-way metering technique [13]. The reliable detection of the direct path from the transmitter to the receiver enables accurate positioning calculations. Not only can interfering reflections be mitigated, but also the direct path can be principally captured even if it is attenuated by an object to some extent. A very important feature of UWB radio

technology is the possibility to calculate time of arrival (TOA) of the direct path of the radio transmission between the transmitter and receiver at different frequencies. Important characteristics of pulse-based UWB are very short pulses, less than 23 cm for a 1.3 GHz and less than 60 cm for a 500 MHz bandwidth pulse, so most signal reflections do not interfere with the direct signal, thus ensuring that the usual multipath fading of narrow signals does not occur. Some reported works use AOA, TDOA and RSSI for localisation purposes as well [89].

Radio-frequency Identification (RFID) tags

RFID tag solutions are based on one or more reading devices that are capable of wireless detection of ID tags present in the neighborhood (using TOA, AOA, TDOA and RSSI metrics) [160, 150, 85]. The tags present in the environment reflect the signal that was previously transmitted by the reader. They modulate it by adding a unique identification code. The tags can be active or passive. Active tags are usually powered by a battery while the passive tags draw energy from the incoming radio signal [150, 85]. The detection range of the latter is therefore more limited. Reference tags and the reader are located in known fixed positions in the environment. To locate a mobile tag, the reader scans through different power levels for tags in the vicinity. When a mobile tag is detected, the receiver compares the power returned by the reference tags and the mobile tag, determining the closest reference tags by using a nearest neighbor algorithm. The position of the mobile tag is determined by triangulating the position of the nearest reference tags [160].

3.8.2 RF-based localisation systems

Many localisation technologies and applications have been proposed so far. The Campaignr[1] configurable micropublishing platform has shown the capability of mobile platforms to act as WLAN (and more general sensor) data gathering hubs and thus can be used in the localisation process [139]. Although Active Badge system [74] and the systems presented

[1] http://www.campaignr.com

in [167] and in [55] do not belong to RF-based localisation systems, they are the first, well-known examples of indoor localisation systems and thus described below.

The Active Badge system [74] was among the first systems (1992) and thus an important contribution in the field of localisation systems. This badge is an infra-red (IR) system and is worn by a user. Every 10 seconds a unique IR signal is emitted. Sensors are placed at known positions inside a building. They receive the unique identifiers and send them to the location manager software. However, the system cannot exactly position a user, only the room he/she is currently in. This system requires significant maintenance and installation costs and it performs poorly in the presence of direct sunlight. Another method based on IR technology consists of IR beacons that are placed on the ceiling at known positions [167]. An optical sensor on a headmounted unit receives the IR beacons, which allows the system to obtain the user's position.

Another work reported in [55] discusses a combination of ultrasound (US) and IR sensors. A user that needs to be localised is, in this case, a mobile robot that emits an active US chirp. Beacons, scattered throughout environment can detect this signal and, after a predefined waiting time, the beacon replies to the chirp with an IR burst containing its location. The distance (between the active beacon and the robot) is determined by the elapsed time interval. Using a database of distance measurements a position can be calculated. The paper reported an accuracy of less than 10 cm.

RADAR [19] records and processes RSSI data from multiple base stations positioned such that their signals overlap in the area in which a user is to be localised. This paper shows the first fingerprinting system localisation system obtaining localisation accuracies of $2-3$ meters using around 70 calibration points (CPs) that are placed non-uniformly at least every 2.5 meters. Empirical measurements are combined with signal propagation models to obtain user position. The signal propagation model approach makes the localisation process easier but the empirical method gives much better accuracy.

The HORUS system [180] models the signal strength distributions using parametric and non-parametric distributions thus reducing the effect of temporal variations. The experi-

mental setup and the results show that under the independence assumption, as the number of RSSI values increases, the performance decreases. Therefore, the authors introduced the correlation modeler and handling modules that use an autoregressive model for handling the correlation between RSSI values from the same AP. HORUS estimates the location of a user from the discrete set of CPs. The distance between two consecutive CPs on average was around 1.52 meters. 110 CPs in total were used. The authors stated they had achieved very high accuracy (the average accuracy reported was less than 0.8 meters) which was questionable when replicated in this thesis and some other works [58, 10].

Another localisation system is described in [91]. The COMPASS system utilizes both WLAN 802.11 and digital compasses. The authors have shown that incorporating information about the user orientation could significantly improve the accuracy of the localisation system. Moreover a probabilistic algorithm was used to calculate probability distribution over CPs and a simple averaging algorithm to further improve the performance of the system was presented. 166 CPs were uniformly distributed 1 meter apart throughout a university building in order to set up the positioning system and to evaluate it in a real-world environment. An average error distance of less than 1.65 meters was achieved.

The DAEDALUS project [79] represents a system for coarse-grained user localisation consisting of the base stations that transmit beacons augmented with their physical coordinates. A mobile host estimates its location using the location of the base station to which it is attached. It was reported that the accuracy of the system was limited by the (possibly large) cell size.

The EKAHAU real time location system (RTLS) [1] is a WLAN-based indoor localisation solution which achieves sub-room, room-, floor- and building-level accuracy (in general $1.4 - 5$ meters). The system uses patented software-based algorithms to compute the location of tracked objects and can easily scale to support more than $30,000$ tags on a single server. Several global companies (McKesson, Nortel, HP, 3M, Siemens, etc.) have benchmarked EKAHAU and stated that it offers the best combination of performance and cost [1].

The NIBBLE location system, from UCLA, uses a Bayesian network to infer a user location [39]. Their Bayesian network model includes nodes for location, noise, and access points (sensors). The signal to noise ratio (SNR) observed from an AP at a given location is taken as an indication of that location. The SNR is distinguished into four levels: high, medium, low, and none. The system stores the joint distribution between all the random variables of the system. There are 9 room locations that were used for accuracy testing: 3 conference rooms, 2 other offices (cubicles), two private offices, a bigger office (lounge), and a patio outside the building. A non-uniform distribution of CPs was used within these spaces. Reported accuracy was $1.7 - 2.5$ meters.

Another two systems [96, 145] use Bayesian inversion to return the location that maximizes the probability of the received signal strength (RSS) vector. The first system [96] stores the signal strength histograms and uses them in the testing phase to obtain the location of a user. Average estimation of the error is around 2 meters and 155 CPs are used. The spacing between two consecutive CPs was around 2 meters.

One of the latest WLAN-based indoor positioning systems is presented in [58]. The proposed technique is based on principal component analysis and offers a more efficient mechanism to utilize information from all APs. These algorithms replace sets and subsets of available APs by a subset of principal components. The reported average accuracy is around 2.2 meters using 45 CPs, positioned approximately every 1.5 meters.

A novel Time of Arrival-based (TOA) approach is presented in [68]. It needs a small modification to the WLAN physical layer to achieve localisation error far less than using RSSI-based measurements (RMSE is between 0.4 and 5.5 meters). Location Estimation using Model Trees (LEMT) [178] is based on radio map reconstruction in real time. This approach takes RSSI values and creates dependency between CPs and the estimated locations.

Another recent WLAN-based indoor positioning systems is presented in [59]. The paper discusses the reduction of severe fluctuations of RSS and proposes a scheme that efficiently extracts the signal for user localisation. The authors state that their system achieves high

accuracy with the average distance error around 0.65 meters. 86 CPs, separated by 1 meter, are used.

The most similar paper to the work presented in this thesis in terms of reduction of calibration effort is [40]. By reducing both the number of calibration data and the number of CPs, the radio map can be successfully rebuilt using an interpolation approach. A learning algorithm can employ unlabeled trace data to further improve localisation performance.

Thus, in conclusion we believe that the work represented in this thesis extends the state of the art in terms of accuracy (under 2.2 meters on average as compared to around 3 meters in [40]) and due to the fact that an area can be covered with fewer CPs. Furthermore, the work presented in this thesis takes into account user orientation in the localisation process.

3.9 Image-based localisation metrics

3.9.1 Introduction

A digital image is a numeric representation (normally binary) of a two-dimensional image. Depending on whether the image resolution is fixed, it may be of vector or raster type. The term digital image usually refers to raster images also called bitmap images. Raster images have a finite set of digital values, called picture elements or pixels. The digital image contains a fixed number of rows and columns of pixels. Pixels are the smallest individual element in an image, holding quantized values that represent the brightness of a given color at any specific point. In the RGB color space, brightness can be thought of as the arithmetic mean μ of the red, green, and blue color coordinates. In image processing, the histogram of an image refers to a histogram of the pixel intensity values. This histogram shows the number of pixels in an image at each different intensity value found in that image. For an 8-bit grayscale image there are 256 different possible intensities, and so the histogram will show 256 numbers showing the distribution of pixels amongst those grayscale values. Histograms can also be taken of color images, either individual histograms of red, green and blue channels can be taken, or a 3D histogram can be produced, with the three axes

representing the red, blue and green channels, and brightness at each point representing the pixel count. Intensity range of an image represents the range of its pixel intensity values. Image normalisation is a process that changes the range of pixel intensity values. Image features represent point correspondences between images of different scenes and objects [24]. They are analyzed in many computer vision and image processing applications, including object recognition, image retrieval, image registration and camera calibration to name just a few. Image features are usually identified in several steps including detection and descriptor building phases. Both phases should be robust to changes and show reliability in detecting the same feature points under different (lighting, viewing, etc.) conditions. In the following subsections the most well-known detector and descriptor algorithms that are used as metrics in image-based localisation are presented.

3.9.2 Hessian and other detectors

The Hessian detector uses the determinant of the Hessian matrix ($I_{xx}I_{yy} - I_{xy}^2$) where I_{xx} is the second partial derivative of the intensity function I, i.e. the luminance component of an image. The Hessian detector uses the second moment matrix as the basis of its corner decisions. The matrix, denoted by A, has also been called the autocorrelation matrix and has values closely related to the derivatives of image intensity

$$A = \sum_{x,y} w(x,y) \begin{bmatrix} I_{xx}(\mathbf{x}) & I_{xy}(\mathbf{x}) \\ I_{xy}(\mathbf{x}) & I_{yy}(\mathbf{x}) \end{bmatrix} \tag{3.3}$$

where I_{xx} and I_{yy} are the respective derivatives (of pixel intensity) in the x and y direction at point $\mathbf{x}$. The weighting function $w(x,y)$ can be uniform, but is more typically an isotropic, circular Gaussian function. By analyzing the eigenvalues of A, this characterization can be expressed in the following way: A should have two large eigenvalues for an interest point. Upon calculating the magnitudes of the eigenvalues, if λ_1 and λ_2 have large positive values, then a corner is found [120]. There is an extended detector called Hessian-Laplace which is capable of finding rotation and scale invariant points (local maxima of the Laplacian-of-Gaussian). Using the Laplacian operator the scale selection is determined;

the intensity gradient of the elliptical regions is estimated using the second moment matrix eigenvalues. The Harris-affine detector [27] is reliable to determine scale and localisation while the second moment matrix of the intensity gradient determines the affine neighborhood. The MSER (Maximally Stable Extremal Region) detector extracts regions closed under monotonic transformation of the image intensities and under continual transformation of the image coordinates [95]. Another detection scheme called the Salient Regions detector measures the entropy of the pixels' intensity histograms in order to detect regions [122]. In the EBR (Edge-Based Region) detector [151] regions are extracted combining image edges (extracted with a Canny operator) and interest points (detected with the Harris operator). The IBR (Intensity extrema-Based Region) detector detects affine-invariant regions by analyzing the image intensity function and its local extrema [122].

3.9.3 MPEG 7

Despite of the fact that it is mainly used for providing audio-visual content description, MPEG 7 can be used for image description as well. The MPEG-7 standard consists of a set of descriptors and each of them defines semantics and syntax of visual low-level features e.g. color, shape [12]. Usually the problem of image localisation, similarity and matching is solved using three visual MPEG-7 descriptors extracted from an image. Several descriptors can be analyzed. Color Layout Descriptor (CLD) is a resolution-stable and compact MPEG-7 visual descriptor. It is defined in the YCbCr color space (Y is luminance, Cb and Cr are the blue-difference and red-difference chroma components respectively), and developed to capture the spatial color distribution of an image or a region of arbitrary shape. The Scalable Color Descriptor (SCD) is based on a Haar-transform encoding scheme that calculates color distribution over an entire image in the HSV color space that is uniformly quantized to 256 bins. The spatial distribution of edges is detected using the Edge Histogram Descriptor (EHD).

3.9.4 Scale Invariant Feature Transform

The Scale Invariant Feature Transform (SIFT) is a method which detects interest points and extracts these features together with their feature vector. A feature vector represents the region around that interest point. This can be used to perform reliable matching between different views of an object or scene [49]. The method starts with the extraction of interest point features from images. The SIFT features are invariant to scale and orientation of the image. They are also robust to occlusion. Features are generated based on local geometric processing by analyzing gradients in the image at various scales and in numerous orientation planes. Two images are matched in the following way. Each feature from the 1^{st} image is compared to a feature in the 2^{nd} image. The features are matched using the distance ratio test [111]. To check if a feature point from the 1^{st} image has a match in the 2^{nd}, its two most similar descriptors are analyzed. If the ratio of the nearest distance to the second nearest distance is less than some fixed threshold (0.7 in this case), a match is found (see figure 3.5). It is also possible to achieve robust matching across a wide range of transformations such as noise, illumination changes, various affine distortions, 3D viewpoint changes, etc. The features exhibit other properties such as matching that is easy to perform against a big database (mismatch is of low probability), being easy to extract and very distinctive. Its application domains can also be found in 3D scene reconstruction, tracking and image similarity [111]. Recognition can be achieved in near-to-real time for medium to small databases.

Figure 3.5: SIFT matching between two images of the same scene but at different scales. Matches are represented with yellow lines

3.9.5 Speeded Up Robust Features

Speeded Up Robust Features (SURF) is a fast, scale and rotation-invariant interest point detector and descriptor. It was developed and inspired using the SIFT method by Bay et al. [24]. A SURF matching example is given in figure 3.6. The extraction and description of interest points is speeded up and state-of-the-art performance is achieved in the feature matching. This important speed gain is achieved by using integral images. An integral image, denoted by $ii(x, y)$, at pixel (x, y) contains the sum of the pixel values below and to the left of (x, y) (see equation 3.4).

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{3.4}$$

where $i(x, y)$ is the input pixel. They drastically decrease the number of operations needed and are also independent of the chosen scale. SURF is mainly used in object recognition, image retrieval and image similarity. Because of all of these reasons SURF is chosen as the main processing tool in the image-based part of this work. It is described in greater detail in section 4.3.1.

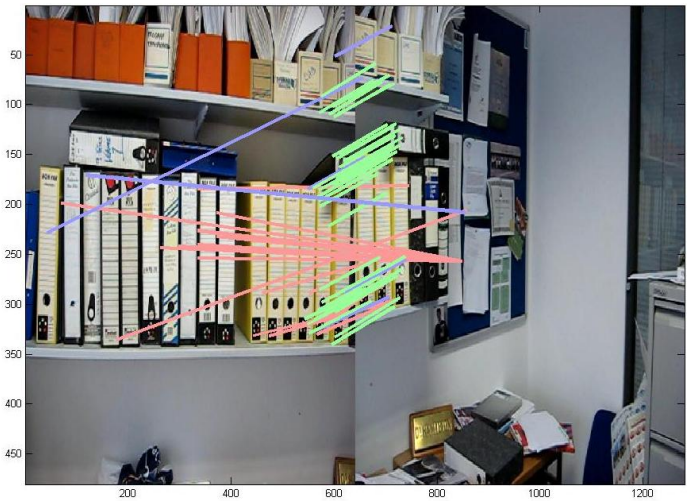


Figure 3.6: SURF matching between two similar images. Unidirectional matches in the right image that correspond to the left image are represented with red lines (vice versa for blue lines) and bidirectional matches with the green lines

3.10 Overview of image-based localisation

Image-based localisation has been an important part of user localisation in the last 20 years. It has many applications including robot and indoor navigation [143, 184], augmented reality [16, 69] or 3D browsing and visualization of photo collections [146, 104].

Some localisation systems only provide positioning information, while view registration of a camera with respect to a 3D scene gives full six degree of freedom pose information and moreover works indoors and in urban canyons. For Augmented Reality (AR) applications [17] fast and accurate image alignment to a given scene is particularly useful for registration of 3D content to the live view of the world as captured from a camera. Image-based localisation is non-intrusive since only image and video data are needed to compute accurate 3D pose and orientation information. A prerequisite of an accurate and fast image-based localisation system is to have an exact 3D model of the environment. The ideal case would be a precomputed visual map of the environment that encodes original illumination and viewing conditions from any desired viewpoint.

3.10.1 Image-based localisation systems

In the world of computer vision, the problem of location recognition has been analyzed in several different contexts [143, 143, 185]. Generally, self-localisation in indoor and outdoor environments using image or video data is explained in the visual SLAM (simultaneous localisation and mapping) literature. The most prominent methods rely on wide baseline matching techniques that use image features such as the local image descriptors and scale invariant interest points presented in the previous section. The idea behind these methods is to find the most similar image searching from a database of registered labeled images [147], planar surfaces [143] or 3D models [129]. For a static scene, a geometrical solution can be used to determine the actual camera pose with respect to the given database. Various illumination or viewpoint changes are successfully solved by robust SIFT [111] and SURF [24] features i.e. using descriptors of local image correspondences. The work presented in [153] describes one of the first works for 3D scene modelling, tracking and recognition with

invariant local image features. First, the authors reconstructed a sparse 3D model from the object using multiview video coding and vision methods. Then, hierarchical clustering into a *kd*-tree structure is performed using SIFT descriptors connected with the sparse 3D points, and a nearest neighbour search engine is employed to find the position most similar to a user position. For geometric verification a robust pose estimation method is employed and it gives the accurate pose of the query image with respect to the 3D model (illustrated in figure 3.7). In [147] a city scale location recognition approach is presented based on video streams that are geotagged and vocabulary trees appropriately trained using SIFT features. The vocabulary tree method that allows a sub-linear search of large descriptor databases and inverted file scoring was first presented in [131]. Nister et al. in [130] build dense 3D models out of incoming data based on multiview linking, which is computationally and memory demanding and also not appropriate for planar objects.

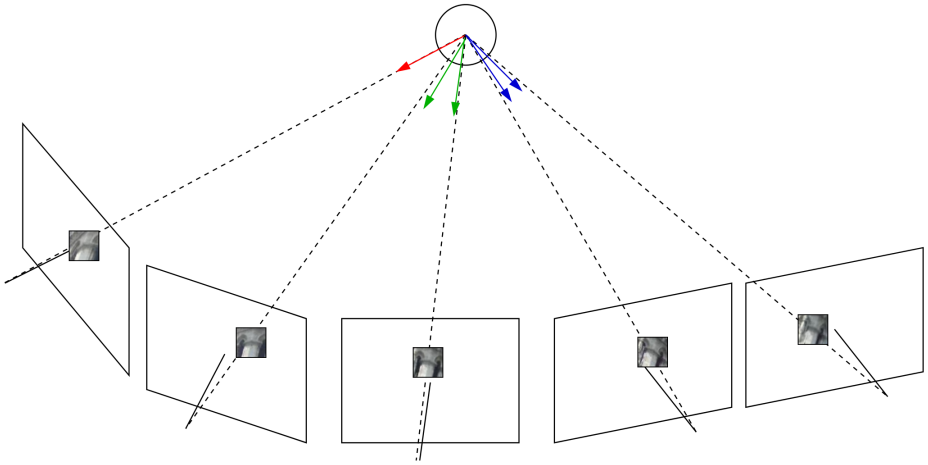


Figure 3.7: 3D point reconstructed from multiple views. Each measurement corresponds to an image patch associated to the region from which the feature point was extracted [129]

In contrast, in [100] a classification problem called recast matching employs a decision tree and makes a trade off between the increase of the memory space and expensive descriptor computations. In [56] the authors present a vocabulary tree-based approach, based on a reduced SIFT-like descriptor. In the augmented reality area, the related work is found in [69, 141, 92]. Recently, Li et al. [105] presented a location recognition approach based on priority-based feature matching using features of the stable scenes. Another bag-of-features

approach combined with geometric verification for object recognition was proposed in [174]. In [14] visibility prediction of known 3D locations with respect to a query image was investigated. Se et al. constructed the world only by special features called visual landmarks [148]. It is still not available for large datasets. In [45], Goedeme et al. developed a complete autonomous mobile robot system using omni-images. It is useful for both indoor and outdoor environments. The authors used a topological map in which the world is represented as collection of connected nodes.

Most works published so far on image-based location recognition target small-area settings [130, 147]. An exception is the work of Schindler et al. [147] who proposed information theoretical criteria for using informative features to build vocabulary trees for wide outdoor location recognition in a $300,000$ image database. The typical flow of stages of operations in image-based localisation based on image features and some sort of a vocabulary tree is given in figure 3.8. The main contribution of these previous works is in their ability to cope with extremely large image databases. Hays and Efros downloaded about 20 million images from which they excluded all photos containing text-labels (6.5 million images in total).

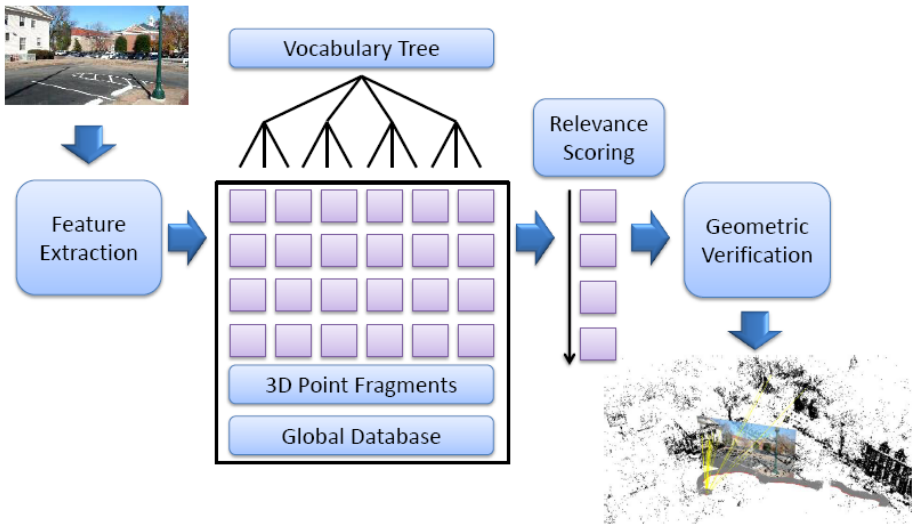


Figure 3.8: Typical flow of operations in feature-based image localisation [174]

The authors in [98] present an approach to reduce the matching time, complexity and the size of SIFT features for indoor use in robot localisation and image retrieval. The complexity

and the number of SIFT features needed to describe and define a scene are reduced by structural analysis of indoor spaces. While there is a significant reduction in matching time and image descriptor size, a minimal loss of accuracy in feature retrieval is achieved. The scale value of SIFT features improves the accuracy of filtering and increases localisation performance. A context vision system for object and place recognition is presented in [163]. Its goal is to categorize new environments, identify locations and to use this information to provide context for object recognition. A global low-dimensional representation of an image gives useful information for place recognition. The algorithm has become a part of mobile system which provides feedback in real time to the user. Sattler et al. [146] achieved the best performance of 79.64% while [105], [131], [148] and [56] achieved 76.2%, 71.42%, 63.13% and 61.21% respectively. Average distance error for the above mentioned approaches is 0.72, 0.91, 1.642, 1.3535 and 1.496 meters respectively.

To summarize, it is clear that image-based localisation is an active research field. In this thesis, we adopt an approach based on hierarchical vocabulary tree of SURF descriptor vectors as this has shown to be an effective approach to date.

3.11 Fusion of WLAN and image data: justification

Localisation based on WLAN technologies can be attractive due to the ubiquity of the infrastructural requirements, such as wireless data networks, that may have already been implemented within the facilities. The propagation of WLAN signals can be affected and significantly changed due to many reasons. For example movements, structural modifications, rearrangements or settings that are under constant change (such as metallic vehicles) can affect signal strengths distribution in space [38]. Interference and noise are often-mentioned challenges. Most of the research done in the area of WLAN-based localisation deals with open spectrum bands although there are some works related to the reserved radio band. This means these solutions risk receiving interference due to other systems sharing the same frequency bands of the radio spectrum. WLAN continues to evolve giving additional standards regarding quality of service and localisation employment. Images as a complementary

modality can be used when WLAN breaks down and/or is unreliable. It is unlikely that two locations with similar RSSI values have similar images. Also images give extra contextual information about the user activities and thus represent the core information about his/her behavior.

In the work reported in this thesis a Naive Bayesian approach [86] is used to calculate the most probable user location. It models the relationship between the frequency of appearance of access points (APs) and their respective RSSI values. Speeded Up Robust Features (SURF) are chosen as the metric in image-based localisation because they provide the best combination of precision, speed and complexity and achieve good results in indoor scenarios. A hierarchical vocabulary tree for image-based localisation approach is used as it has a good trade-off between accuracy and complexity. Both WLAN as ubiquitous technology and images as technology used for contextual description of the user behavior are part of the fusion process. This fusion process overcomes difficulties when using each sensing modality.

3.12 Data integration and fusion

In the proposed approach, the problem of structuring data into indoor locations cannot be solved without some way of integrating data from multiple sources (namely WLAN signal strengths and image matching data). Multimodal systems need to acquire and process data from various different sources which are recognized by different modules and merge them to form one single representation of data [172]. This representation should be then interpreted semantically. One can differentiate several levels of integration of multiple modalities: early fusion (integration at the feature level), late fusion (integration at a semantic level) and hybrid fusion. Hybridization and fusion techniques and systems are introduced to find solutions when one modality is not sufficient and when the fusion system should deliver higher accuracy/precision, speed and/or efficiency.

Early fusion integrates the signals based on their features. From different modalities there are signals at different sampling rates. Thus they need to be modified to have the

same sampling rates. At every time stamp of the signal the N_i features from M different modalities must be merged into one feature vector of length $\sum_{i=1}^{M} N_i$ [173]. Early fusion processes high-dimensional vectors and is thus computationally very expensive. Usually large training datasets are also needed, which sometimes cannot be obtained or at least it can be very costly to do so. The data from different modalities need to be accurately synchronized otherwise the results can be very misleading and result in very low precision/accuracy. If the signals are highly correlated and synchronous, integrating them leads to acceptable results as the correlation structure of the modes can be obtained using training learning [173]. Early fusion has been successfully applied in audio-visual [113] and mitigation systems [158], coupled Hidden Markov Model (HMM), the multistream HMM recognition [123], and interest recognition [71].

In late fusion the signals are combined on a semantic level. Signals are processed separately and merged later, in the decoding phase. Since each modality is trained separately, learning of the joint probability of the modalities is not done explicitly [172]. Trained unimodal data are used in a late fusion process. Furthermore, as no re-training is necessary late fusion systems can use larger datasets more easily. Late fusion has some advantages over other types of fusion. These are better combination and separation of modalities and easier manipulation with the modalities that are missing at specific time stamps [11].

Hybrid fusion has been created to take advantages of both early and late fusion. A hybrid system has freedom and flexibility of a late fusion system (its database can be extended easily and is able to process data which are asynchronous in time) and during the recognition process it can exploit mutual information from other modalities. A classification process can be obtained on the first modality and then the classification process using the second modality can be applied on the feature space reduced using the first classification process. The Asynchronous Hidden Markov Model [28] is a well-known example of hybrid fusion and represents the main method for comparison to other different approaches [172].

When user context is needed at all times and when high accuracy/precision is priority it is must to use both modalities. On the other hand when one needs to faster processing and

an energy efficient process during the evaluation process the hybrid solution is a preferable choice.

3.13 Fusion and hybrid solutions for WLAN-based and image-based localisation

In this section we present an overview of some research work on fusion and hybrid solutions for RF-based and image/video-based indoor localisation.

3.13.1 Examples based on fusion of RF and image/video data

Most existing localisation methods are based on a single modality. In fact, to our best knowledge, even in other application domains there are only a few techniques based on fusion of RF and image sensing methods. One previous work [47], showed a proof of concept of how WLAN received signal strengths (RSSs) and image matching data could be fused to do coarse localisation for a small number of locations. Histogram similarity for RF and a hierarchical vocabulary tree for image-based localisation were used [47]. A simple fusion function was derived to take into account the strong points of both approaches. RSSI values are not able to easily differentiate nearby locations. Image data is thus applied on the remaining locations, or if this data fails, the users motion priors restricts the location detection. Average mean precision for image-based, RF-based and fused localisation approach were 82.09%, 77.42% and 88.26% respectively. The distance error rate was 3.24 and 2.02 meters for the image-based and WLAN-based localisation methods respectively.

As the sequel to that work, and as the basis of this thesis, a more precise WLAN-based algorithm [140], a different vocabulary tree concept for image-based localisation and a novel more complex and effective function in the fusion process were developed. The approach is verified on a much larger and more challenging dataset.

Multiple modalities have been used in the complementary but related challenge of tracking specific objects. In [67] the authors discuss an approach for actively tracking humans in

crowds using robots. It consists of 360° RFID system and video camera placed on remote directional and zoom control unit. The authors have developed a multisensor control strategy based algorithm for tracking using RFID data. A particle filtering method is used to fuse heterogeneous data to make the tracking more robust. Tracker outputs and RF data are used as a basis for the multisensor control platform (the RF tracking system is shown in figure 3.9). With the fixed dataset the average accuracy error in robot tracking was 0.8 meters.

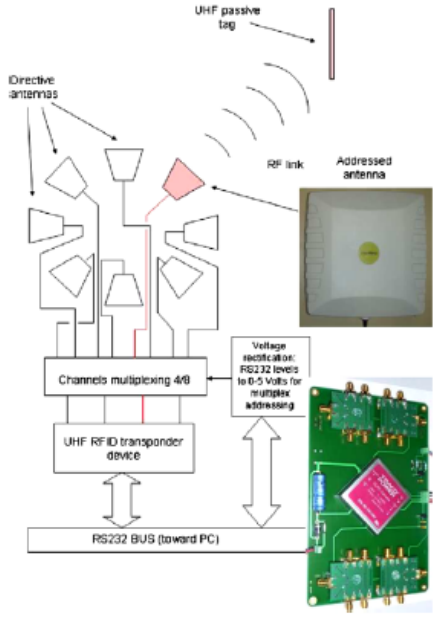


Figure 3.9: Eight antennae addressed by a RF multiplexing prototype [67]

Other related work describes an approach for object tracking using a different particle filtering model [124]. It consists of a camera recording method based on color features of the target and a WiFi-based localisation system. Sensor fusion consists of video and WiFi data merged together for obtaining position and tracking. Due to WiFi performance and its RSSI characteristics the method can be utilized in both outdoor and indoor spaces. To track the targets seamlessly a particle filtering method that merges the two sensors is used. A WiFi observation model is involved in a video particle filtering approach to find the particular weights for every particle. It is proved that the fusion outperforms any of

the modalities separately and is useful when any of the modalities fails. In this system, the particle filtering observation model consists of two parts: one is a video based model that uses color features of the target, and the other is an approximated location system based on WiFi RSSI which access points transmit to PDAs. The method was compared with the ground-truth data showing maximal error distance of 18 meters. The reported precision was 68.63%.

There does exist work which uses fusion of three different sensors for three independent complementary modalities [181]. This system consists of an inertial sensor, positional sensor and visual sensor. Visual information is given by a video camera, acceleration is acquired using an accelerometer and the information about the position is obtained using an 802.11g receiver. These sensors are low cost and widely available. This ubiquitous platform represents a typical example of a context-aware device that is able to give the real feel of a user environment through information sent to a user. To obtain reliable position and achieve least error distance three sensors are fused using a Discrete Kalman filter. In the case of a correct initial estimate of the system's position, the average accuracy error does not go beyond 8.26 meters.

Another paper proposes an algorithm that fuses WiFi and video camera data for indoor localisation [164]. The algorithms differ to other solutions, by fusing the sensor data in the measurement model before calculating an estimated position based on the individual technologies. The purpose of fusing WiFi and video data is to have a smaller localisation error in the rooms where there is a camera, in contrast to only WiFi, but still offer room level localisation where there are no cameras. Data measured by the sensors are sent to a data aggregator (it stores the incoming sensor data). The aggregator selects which measurement models to use, a WiFi or image measurement model or both. The sensor data is then sent to a fusion engine where the particle filter algorithm is applied. The fused approach achieves error distance less or equal than 2 meters 67% of the time when a user walks around the test area without interference and less or equal than 4.3 meters 87% of the time with interference.

In the work presented in [133] a unified approach for a camera tracking system based on an error-state Kalman filter algorithm is presented. The filter uses relative (local) measurements obtained from image-based moving sensors to estimate change in position over time, as well as global measurements produced by landmark matching through a built visual database and range measurements obtained from RF ranging radios. The results of the work are shown by using the camera poses output by the system to render views from a 3D graphical model built upon the same coordinate system as the landmark database. The localisation distance error did not go below 2.46 meters with precision of 73.364%.

3.13.2 Hybrid localisation and tracking solutions based on RF and image/video data

In work preceding this thesis, the authors in [47] discuss a museum simulation where the main objective is to identify exhibits that a visitor is observing . A simulated museum with multiple exhibits is constructed, with two devices which capture images and wireless signal strength readings. Image-based localisation and RF-based localisation are used together to overcome the difficulties of each approach. In image-based localisation, it was examined whether it is possible to locate a user accurately enough by capturing an image at its current location. In the RF-based part RSSI values from wireless access points are collected. These two sources of data are complementary and using different fusion approaches good localisation accuracy was achieved (the precision was around 74%).

Another work describes an indoor localisation solution which relies on relevant infrastructure of aware mobile devices using WLAN and video data. While many previous works assume users can be positioned anywhere, this approach assumes that movement is constrained to a network and that robots can only move along a specific grid. All locations of interest are also reachable via the network. In the given settings, the method also finds routes nearest to user position in indoor environment [35].

Mobile users often take the same route to a particular destination. During these trips important contextual data for various services can be provided. The system proposed in

[31] stores all the relevant contextual data received from a receiver for a user. It uses image and WLAN data. The detailed architecture of the system together with its design is given in [31].

An electronic service that gives localisation and contextual information about mobile objects (robots) using wireless communications technologies in addition to other services such as video, audio, etc is presented in [149]. k-nearest neighbor approach is employed to obtain the location. The model that can be easily implemented is presented in [149]. An outline of this prototype system which provides information about the moving objects is shown in figure 3.10. Its complexity is directly proportional to its efficiency.

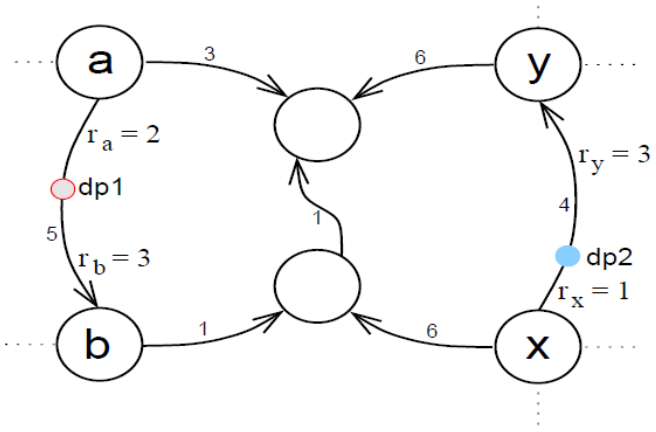


Figure 3.10: Two mobile objects on a route grid (network) [149]

A new method, described in [179], proposes a similarity model which finds paths most similar to the current one. Using the selected path, a user location can be determined. The user's adaptive services are obtained using RF and video data coming from multiple streams [179].

The authors in [45] propose a method of robot navigation and tracking in indoor environments based upon fixed camera view and WLAN information. The system is equipped with a WLAN receiver and a camera. The route scene can be described by three-dimensional objects extracted as landmarks from camera views. For an environment having limited routes, a two-dimensional map can be made based upon indoor routes scenes, assuming

that the topological relation of routes at intersections is known. By using this information a coarse method is used to generate an indoor environment map and locate a mobile robot [45]. First, a robot finds its approximate position based on the WLAN information. Then, it identifies its location from the image information. Promising experimental results in an indoor environments are given.

3.14 Conclusion

Hybrid and fusion solutions discussed in the sections above show huge potential for the further expansion and development of similar localisation systems. This applies not only to the fusion of the two specific sensing modalities used in this thesis but also for the use of other modalities and fusion techniques. The fusion examples show a variety of different sensors used and show how inertial and positional sensors together with accelerometers could be successfully integrated into a system together with RF and image/video data. The examples presented above range from simple proof of concept methods to more complex fusion approaches (the discrete and error Kalman filters). The stability of these systems is directly proportional to the complexity, and the accuracy is comparable to the state of the art localisation systems. In many hybrid examples, indoor route detection, i.e. finding the nearest indoor route for a query, has been a popular topic in recent years. This has been achieved by using different tracking algorithms or by finding similarity functions that would find the nearest routes. The nearest routes can also deliver information related to the user context or different services available on these routes. The main motivation for the fusion of two different modalities is to potentially overcome difficulties when using them separately. Moreover, this approach can have another benefit such as acting as a descriptor of the user's activities. Based on this review it can be concluded that the fusion of different sensing modalities can be exploited cheaply in a short amount of time and can be enhanced using other sensing modalities. These hybrid and fusion techniques provide an opportunity to integrate WLAN and image data beforehand and to deliver it as a (standard) smart phone application useful for various ambient-assisted living scenarios. Thus, motivated and

inspired by these prior works, we propose a hybrid/fusion approach to localisation.

Chapter 4

Technical background

4.1 Introduction

In this chapter, some necessary technical background useful for indoor localisation and tracking are presented. We believe this is useful for understanding the proposed algorithms described in chapter 5 and chapter 6. First we introduce a basic approach in WLAN-based localisation: Bayes classification method or Bayes localisation method. Using a probabilistic chain rule and a naive assumption for conditional probabilities this approach can be transformed into the more convenient, and for this work more important, Naive Bayes approach. It is presented in section 4.2.1. In this thesis this algorithm is transformed and optimized to better suit the specific environmental challenges encountered. It is used to determine the most likely location of a user and will be presented in section 5.3 of chapter 5.

For the image-based localisation, presented in section 5.4, we use a hierarchical vocabulary tree of Speeded Up Robust Features (SURF) descriptor vectors which are taken from all available images. The tree will be presented in section 5.4.2 of chapter 5. The SURF method itself is described in greater detail in section 4.3.1 of this chapter. The detection, the description and the matching phase are explained in each subsection. A hierarchical vocabulary tree is formed using hierarchical k means clustering of SURF descriptor vectors.

Hierarchical k means clustering is achieved using simple k-means clustering repeatedly until the final clusters contain less than k SURF descriptor vectors. The simple k-means clustering divides the dataset of all SURF descriptor vectors into k subsets using k cluster centers. It is described in section 4.3.2. Descriptor vectors are extracted from all images and from a query image as well. After each descriptor of the query image has voted for a location, a ranked list of locations, from the most probable to the least probable location is obtained.

For each localisation method we obtain a ranking of possible user locations. The ranking list of (possible) user locations is achieved using K nearest neighbour classifier which is explained in detail in section 4.4. All nearest neighbours are found using simple Euclidean distance. Also the K nearest neighbour approach is used when determining a user's position between calibration points using just WLAN which is discussed in chapter 6. Moreover, in section 4.5, we discuss how two or more modalities could be fused in general using weights and/or grid search engine to achieve better performance than using them separately. Eventually, we decided to use the simplest method where the two sensing modalities are treated equally (with the weights equal to one) in the fusion (section 5.5 of chapter 5) and in the tracking process (section 5.8 of chapter 5).

4.2 WLAN-based localisation

There are three methods/approaches that can be used in WLAN-based localisation. These are: Bayes, Naive Bayes and Hidden Naive Bayes approach. Bayes takes into account simple received signal strength value from the corresponding access point. Naive Bayes approach takes into account not only signal strength value but also the frequency of appearance of these access points as well. Eventually Hidden Naive Bayes discusses correlation between access points of the network and establishes a mathematical law that they follow. Since the correlation is negligible and Hidden Naive Bayes is much more complex it is decided to choose Naive Bayes as it is more sophisticated and accurate than the simple Bayes approach. Here we give an overview of two well-known classification (localisation) methods.

4.2.1 Bayes and Naive Bayes classifiers

The Bayes method shows the connection between two events A and B represented with their probabilistic values $P(A)$ and $P(B)$ respectively and the conditional probabilities of A given B and B given A, represented as $P(A|B)$ and $P(B|A)$ in the following form:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4.1}$$

In its extended form when event A consists of or is partitioned into several events A_i with corresponding probabilities $P(A_i)$ and $P(B|A_i)$, then with the help of the total probability law the value of $P(B)$ can be eliminated as shown in equation 4.2.

$$P(B) = \sum_j P(B|A_j)P(A_j) \tag{4.2}$$

Equation 4.3 gives the relationship between the probabilities.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \tag{4.3}$$

In the case of a Naive Bayes classifier, a probabilistic model is defined as a conditional model for a class variable C that depends on n feature variables $F_1, F_2, ..., F_n$. Using Bayes the conditional probability

$$p(C|F_1, \ldots, F_n) \tag{4.4}$$

can be expressed as

$$p(C|F_1, \ldots, F_n) = \frac{p(C)p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)} \tag{4.5}$$

The numerator in equation 4.5 does not depend on C and $F_1, F_2, ..., F_n$ are given and moreover the same for all variables and so do not affect the overall conclusions and can be omitted. Using a probabilistic chain rule [86] we have:

$$
\begin{aligned}
& p(C,\, F_1, \ldots, F_n) \\
& \quad \propto p(C)p(F_1, \ldots, F_n|C) \\
& \quad \propto p(C)p(F_1|C)p(F_2, \ldots, F_n|C, F_1) \\
& \quad \propto p(C)p(F_1|C)p(F_2|C, F_1)p(F_3, \ldots, F_n|C, F_1, F_2) \\
& \quad \propto p(C)p(F_1|C)p(F_2|C, F_1)p(F_3|C, F_1, F_2)p(F_4, \ldots, F_n|C, F_1, F_2, F_3) \\
& \quad \propto p(C)\ p(F_1|C)\ p(F_2|C, F_1)p(F_3|C, F_1, F_2) \ldots p(F_n|C, F_1, F_2, F_3, \ldots, F_{n-1})
\end{aligned}
$$

(4.6)

If we can assume that for each i and for each $j \neq i$, F_i is independent of F_j, equation 4.2.1 can be written:

$$p(F_i|C, F_j) = p(F_i|C) \tag{4.7}$$

So equation 4.2.1 given above can be written in a simpler way:

$$p(C, F_1, \ldots, F_n) \propto p(C)\ p(F_1|C)\ p(F_2|C) \cdots = p(C) \prod_{i=1}^{n} p(F_i|C) \tag{4.8}$$

This is the so-called naive assumption. Under the assumptions given above, the conditional distribution over the class variable can be formulated as:

$$p(C|F_1, \ldots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i|C) \tag{4.9}$$

where Z represents $P(F_1, F_2, \ldots, F_n)$ and is usually omitted in the analysis as it is same for all instances and thus not needed in an evaluation process. This formula shows how the probability of C under conditions of $F_1, F_2, \ldots, F_n$ depends on specific, particular and conditional probabilities $p(F_i|C)$. It is of a huge importance for the work represented in the following chapters and will be transformed and optimized in section 5.3 of chapter 5. It will be also used in chapter 6.

4.3 Image-based localisation

As it was discussed in previous chapter there are several choices for an image-based localisation. These are various bag of words, bag-of-features, image retrieval, robot pose estimation, city scale location recognition approaches, kd-tree based localisation structures, etc. These localisation methods were using different localisation metrics such as Hessian, Harris, SIFT, MPEG, SURF detectors and descriptors. Many efficient image retrieval and image-based localisation methods have been proposed so far [42, 143, 184, 171, 147, 140, 16, 69, 146, 104]. Some use discrete image point correspondences (features) found in the image with its neighbourhood represented by a feature vector (robust to geometric and photometric deformations, noise, detection displacements etc). [147, 130, 56, 174, 98]. Nister and Stewnius proposed an efficient and precise algorithm based on hierarchical vocabulary tree of descriptors based on SIFT [171].

Hierarchical vocabulary tree based on SURF is chosen as image-based localisation technique as it suits best indoor localisation environment (SURF is robust to noise and lighting changes to some extent) and is fastest to process and obtain user's position. Also it is the least computationally expensive approach from the set of ones given here. Here a similar method based on Nister's work is given to effectively find similar database images to the query image. Instead of SIFT we use 64-dimensional Speeded Up Robust Features (SURF) descriptors. Speeded Up Robust Features (SURF) are based on the SIFT approach, but achieve faster extraction and description of feature points and achieve very good performance in the matching process [24]. The important speed gain is achieved by using integral images, which can drastically decrease the total amount of processing for basic box convolutions, regardless of the scale [24]. During the detection stage one can use the previously calculated trace of the Hessian matrix, which is in fact the sign of the Laplacian. The sign of the Laplacian is able to differentiate light blobs on dark backgrounds from the reverse. When we compare features we only compare those that have the same type of contrast.

4.3.1 Speeded Up Robust Features algorithm

SURF (Speeded Up Robust Features) is a well known robust image detector and descriptor that can be used in computer vision tasks [24]. It is inspired by, but several times faster and more robust against different image transformations than, the SIFT descriptor [49]. It is based on sums of 2D Haar wavelet responses and effectively employs integral images. It builds on the strengths of the best existing detectors and descriptors and gives novel state-of-the art detection, description, and matching steps. SURF's usefulness is in a wide range of topics especially for image registration, image-based localisation, 3D reconstruction, object recognition and detection, image similarity, camera calibration, etc. Image features represent the most important image characteristic. First, interest points must be selected (detected) at distinctive locations (T-junctions, corners, blobs), an example is given in figure 4.1.

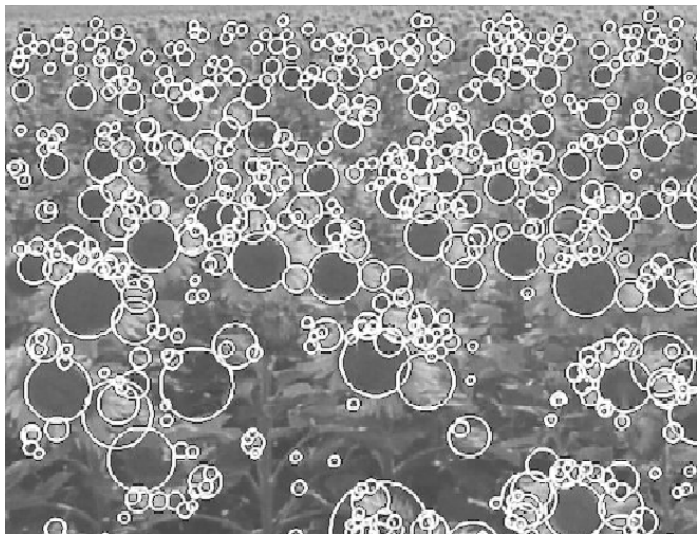


Figure 4.1: An example of feature point detection in an image. Interest points are detected using Hessian-based detector [24]

The main property of the detector is its robustness to changes related to noise, scale, rotation and invariation [60]. This means that it is reliable in finding the same physical interest points under various viewing conditions. Then every interest point's neighbourhood is represented by a feature vector (descriptor). This descriptor needs to be distinctive

and also robust to detection displacements, scale, noise, photometric and geometric deformations. Eventually, the descriptor vectors that belong to different images are matched using e.g. the nearest neighbour rule. Here, the matching (figure 4.2) is achieved using an Euclidean distance between the image feature descriptor vectors.

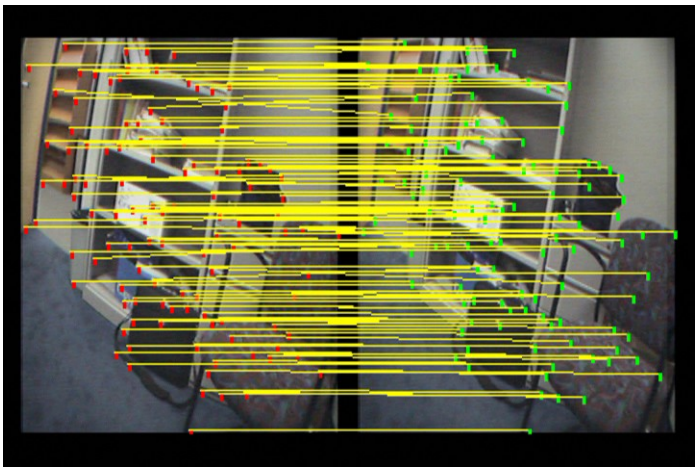


Figure 4.2: Example of matching process between two images

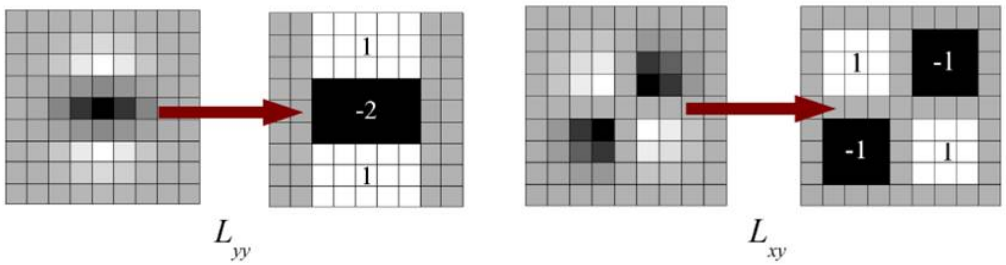
The dimension of the descriptor is proportional to the time it takes to be built, and to achieve fast processing and matching smaller number of dimensions are encouraged. Nevertheless, feature vectors with smaller number of dimensions are more difficult to distinguish. The method for robust interest point detection does not depend on various image scales, i.e. the approach is scale invariant. Another version of descriptor shows invariance with respect to rotation and scale thus making it even more robust to changes.

SURF Detector

The level of invariance, when processing local features, is clearly influenced by different viewing conditions which consequently produce various geometric and image deformations. Our main focus is on detectors and descriptors which are rotation and scale invariant. Other characteristics such as anisotropic scaling, skew, and blur are of less importance and they are included in the robustness of the descriptor [87]. The detector is an approximation on the Hessian matrix:

$$\mathcal{H}(\mathbf{x},\sigma) = \begin{bmatrix} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{xy}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{bmatrix}, \tag{4.10}$$

where $L_{xx}(\mathbf{x},\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}g(\sigma)$ of the image I at point $\mathbf{x}$, and similarly for $L_{xy}(\mathbf{x},\sigma)$ and $L_{yy}(\mathbf{x},\sigma)$. Figure 4.3 shows Gaussian second order partial derivative and its approximation in y-direction and xy-direction.



(a) (b)

Figure 4.3: (a) Gaussian second order partial derivative in y-direction (L_{yy}) and its approximation (D_{yy}) (b) Gaussian second order partial derivative in xy-direction (L_{xy}) and its approximation (D_{xy}). The grey regions are equal to zero. [24]

It is based on the Hessian matrix and gives good accuracy with fast computation time, just like Difference of Gaussians (DoG) [24] which is also a Laplacian-based detector. The determinant approximation is given in equation 4.11.

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx}D_{yy} - (wD_{xy})^2 \tag{4.11}$$

The relative weight $w \approx 0.92$ of the filter responses is used to balance the expression for the Hessian's determinant. This is needed for the energy conservation between the Gaussian kernels and the approximated Gaussian kernels. The Fast-Hessian detector reduces the time of the computation as inside a feature point a distribution of Haar-wavelet responses will affect the actual detector. At a specific location $\mathbf{x} = (x, y)$ an integral image is defined as the sum of all pixels in the input image I within a rectangular formed by the $(0, 0)$ origin and the point $\mathbf{x}$. A sum of all the intensity values over random upright rectangular loop of arbitrary size is then calculated. The analysis of how the scale affects the performance is achieved by increasing the scale of the filter size instead of reducing the size of the image

iteratively. This can be achieved because of the use of box filters and integral images [106]. Gaussian second order derivatives ($\sigma = 1.2$) are approximated with the 9×9 box filters. They represent the lowest scale. They are denoted by D_{xx}, D_{yy}, and D_{xy}. The next layers are obtained by applying bigger and bigger filters on the image (see figure 4.4.

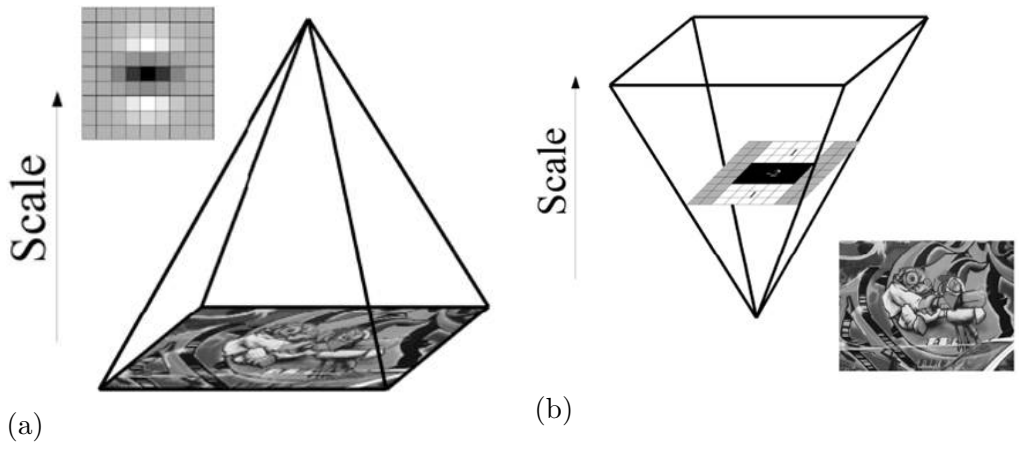


Figure 4.4: Instead of iteratively reducing the image size (left), the use of integral images allows the up-scaling of the filter at constant cost (right). [24]

Taking the specific structure of the filters and the discrete nature of integral images one can apply filters of different sizes: 15×15, 21×21, 27×27, etc. Gaussian derivatives scale accordingly. The normalised responses of filters are obtained using calculated scale spaces [106]. The images are repeatedly sub-sampled and smoothed in order to get to higher pyramid level. At the output of a previously filtered layer one applies the next filter. These filters process the original image at the very same speed or in parallel if needed. Parallel processes are not discussed in this thesis. To better localise feature points on different scales a $3 \times 3 \times 3$ neighbourhood non-maximum suppression is applied. Then the results are interpolated in scale and space as the error between the first $2 - 3$ layers of every octave is significantly large.

SURF Descriptor Components

In order to build the descriptor vector one has to know the vector orientation and each dimension value for a specific interest point. Orientation of an interest point is calculated

using Haar wavelet responses [90]. These responses are calculated in the x and y direction within a circle of radius $6s$ around the same interest point, where s represents the scale at which the interest point was detected.

The wavelet responses are obtained and weighted using Gaussian centering at the feature point. The responses are represented as points with the horizontal response strength along the x and the vertical response strength along the y axis. The sum of all responses within a smoothly moving orientation window is then calculated. The horizontal and vertical responses within the window are summed and yield a local orientation vector [112]. The longest such vector among all windows gives the orientation of the interest point [121]. The descriptor vector is calculated by splitting up the square region around the interest point (the square size is $20s$) into 16 small sub-squares (4×4 within one square). One has to compute Haar wavelet responses at 5×5 regularly spaced sample points. There are four sums: $\sum d_x$ (the sum of Haar wavelet responses in horizontal direction) and $\sum d_y$ (the sum of Haar wavelet responses in vertical direction) and two sums of absolute values $\sum |d_x|$ and $\sum |d_y|$. Filter size is equal to $2s$.

To represent its intensity structure every subregion is represented with a 4-dimensional descriptor vector $\mathbf{v}$. As there are 4×4 sub-square regions in total this means that the resulting descriptor vector is 64 dimensional. Invariance to contrast is obtained by turning this descriptor into a unit vector. Building of descriptor is briefly presented in figure 4.5. In case of a descriptor vector of length 128, $d_x \geq 0$ and $d_x < 0$ are calculated for each sub-square sum $\sum d_y$ and $\sum |d_y|$ separately and similarly for $\sum d_x$ and $\sum |d_x|$ for $d_y \geq 0$ and $d_y < 0$. Eventually every interest point is represented with its 64 or 128 dimensional descriptor vector.

Matching

Matching is based on comparing different interest points. Every interest point in the test image is compared to every interest point in the reference image by calculating the Euclidean distance between their descriptor vectors. The nearest neighbour ratio matching strategy

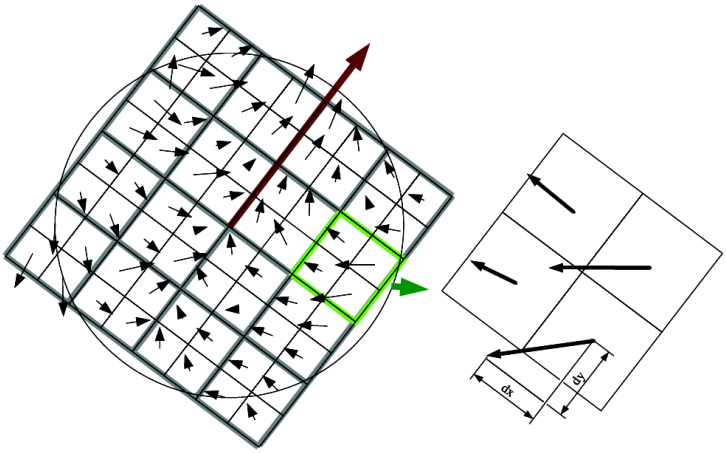


Figure 4.5: Building the descriptor: an oriented squared-shaped grid with 4×4 smaller regions around its feature point [24]

gives a feature pair detected, if the nearest neighbour match (calculated using Euclidean distance) is closer than 0.7 times the second nearest neighbour match distance [166, 111]. If that condition is satisfied then we can say that the match has been established. This asymmetrical measure can be also used in the other direction (from the reference to the test image) and the matches that exist in the both directions can be counted (so called *bidirectional* matches). These matches are more stable thus confirming a good match [111].

4.3.2 k-means clustering

The k-means clustering algorithm tries to partition n features, given with a dataset $\mathbf{x}_i \in \Re^d$, $i = 1...n$, where each feature is given as a d-dimensional vector. k-means clustering algorithm groups the n features into k clusters where $k \in N^+$. Here, in case of a hierarchical vocabulary tree, a feature refers to a SURF descriptor vector. The goal is to find an assignment of features to clusters and also the centroids $\mathbf{c}_j \in \Re^d$ such that the sum of squares of the distances of each data point to its closest vector $\mathbf{c}_j$, is a minimum. Thus, the optimal set of k centroids (C) can be formed by minimizing equation 4.12

$$\phi = \sum_{i=1}^{n} \sum_{j=1}^{k} \delta_{ij} \|\mathbf{x}_i - \mathbf{c}_j\|^2 \tag{4.12}$$

where δ_{ij} represents binary values 0 and 1 describing to which cluster (k in total) the data point $\mathbf{x}_i$ belongs. This is an NP hard problem [54] but there is a non-optimal approximation, i.e. the so called the Lloyd algorithm [109]. For high-dimensional data the vector direction is crucial. A unit vector representation, $\|\mathbf{x}\| = 1$, additionally takes into account gradient variations. This is achieved using a spherical (a unit hypersphere) k-means method. Only a small error is obtained using the approach and convergence is achieved. As $\mathbf{x}$ and $\mathbf{y}$ are both unit vectors, the cosine similarity is equivalent to the Euclidean distance (equation 4.13):

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}\mathbf{y} = 2 - 2\mathbf{x}\mathbf{y} \tag{4.13}$$

In our specific case, if it occurs that the number of SURF descriptors increases, k-means clustering approach slows down for clustering large descriptor sets. Computational time greatly depends on the real nearest neighbours calculation between the features and centers of clusters. These can be speeded up using an k-means approximation shown in [43], thus reducing the complexity from $O(nk)$ to $O(nlog(k))$.

4.4 K Nearest Neighbour Classifier

Supervised classification algorithms aim at producing a learning model from a labeled training set. Various successful techniques have been proposed to solve the problem in the binary classification case. The multiclass classification case is more delicate, as many of the algorithms were introduced basically to solve binary classification problems. In this short survey we investigate the various techniques for solving the multiclass classification problem. Several algorithms have been proposed to solve this problem in the two class case, some of which can be naturally extended to the multiclass case, and some that need special formulations to be able to solve the latter case. The first category of algorithms include decision trees [5, 16], neural networks [3], k-Nearest Neighbor [2], Naive Bayes classifiers [19], and Support Vector Machines [8]. The second category include approaches for converting the multiclass classification problem into a set of binary classification problems that

are efficiently solved using binary classifiers e.g. Support Vector Machines [8, 6]. Another approach tries to pose a hierarchy on the output space, the available classes, and performs a series of tests to detect the class label of new patterns. K-NN classifier is chosen as it is a simple and easy to use classifier. It is fast and usually gives transparent classification results.

The K nearest neighbour (K-NN) classifier is the simplest and the most straightforward classifier in the set that was considered [65]. The nearest neighbours for each query were found and used in the decision process for calculating the class of the query (figure 4.6).

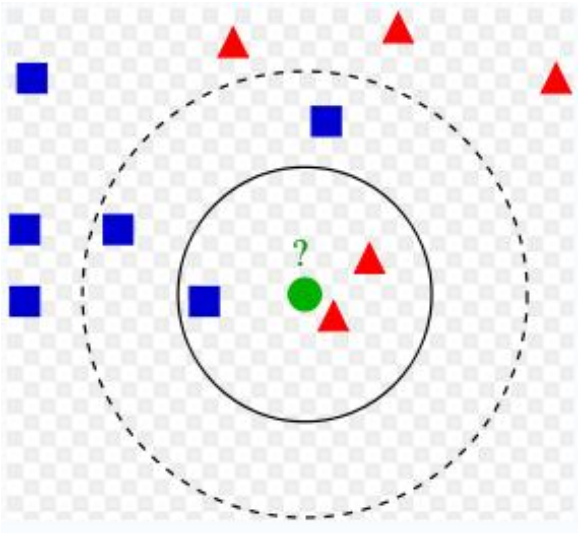


Figure 4.6: K-NN classification example. The classification of a query (represented as a green circle) depend on the value of K: e.g. for $K = 1..3$ the query is classified as red triangle. For $K = 5$ it is classified as blue square. For $K = 4$ one has to include the weights to determine the class [65]

Today K-NN classifier is much more popular thanks to the computational power that is available. The main characteristic for this type of classifier is that the data are classified according to the class of their nearest neighbours, where K represents the number of nearest neighbours we are observing in the class determination. There are three types of classification: the training examples must be in memory during the run (Memory-Based Classification), examples are processed during the run (Lazy Learning technique) and training example based classification (Example-Based Classification) [142].

The determination can be calculated using a distance weighted voting system or by majority. In the case that it can not be decided to which class the query belongs (e.g in 4-NN classifier the first and the second nearest neighbour belong to one and the third and fourth to some other class) it is said that K-NN classifier is undecided (not definable). Let the set D consists of $\mathbf{x}_i$ $(i \in N)$ elements and let the elements be described using a set of features F. If every training example is labelled with a class label $y_j \in Y$ and if one wants to classify an unknown feature $\mathbf{q}$ one should calculate the difference $\mathbf{q} - \mathbf{x}_i$ for each $\mathbf{x}_i \in D$ as [70]:

$$d(\mathbf{q}, \mathbf{x}_i) = \sum_{f \in F} w_f \delta(\mathbf{q}_f, \mathbf{x}_{if}) \tag{4.14}$$

Although there are many options for such distance metric, for continuous/discrete attributes the most general version would be:

$$\delta(\mathbf{q}_f, \mathbf{x}_{if}) = \left\{ \begin{array}{ll} 0 & f \text{ discrete and } \mathbf{q}_f = \mathbf{x}_{if} \\ 1 & f \text{ discrete and } \mathbf{q}_f \neq \mathbf{x}_{if} \\ \|\mathbf{q}_f - \mathbf{x}_{if}\| & f \text{ continuous} \end{array} \right\} \tag{4.15}$$

One can select the K nearest neighbours using this distance metric. To obtain the class of $\mathbf{q}$, one has various different options. The easiest would be to attach the majority class of the nearest neighbours to the given query. Sometimes assigning greater weight to the nearer neighbours decides the class of the query. In order to find a method to achieve this we can employ a voting scheme based on weighted distance where the neighbours vote (noted as V) on the class of the query with vote weights equal to the reciprocical value of their distance to the query.

$$V(y_j) = \sum_{c=1}^{K} \frac{1}{d(\mathbf{q}, \mathbf{x}_c)^n} I(y_j, y_c) \tag{4.16}$$

Therefore the vote given to class y_j by $\mathbf{x}_c$ is 1 over the distance to that neighbour, i.e. $I(y_j, y_c)$ returns 1 if the classes match. Otherwise it is equal to 0. If we take higher values

for n, $n \geq 2$, it would decrease the effect of further neighbours. Also the voting scheme can be achieved using equation 4.17.

$$V(y_j) = \sum_{c=1}^{K} e^{-\frac{d(\mathbf{q},\mathbf{x}_c)}{h}} I(y_j, y_c) \tag{4.17}$$

where h represents a vote exponential coefficient (fixed value). The K-NN classifier can be enhanced to better reflect a particular scenario. For example, speed-up methods can significantly decrease the processing time and handle large datasets. Computation problems can be solved using dimension reduction. With a simple Minkowski distance (the general formula is given in equation 4.18) complexity is equal to $O(S_D S_F)$ where S_D denotes the size of the training dataset and S_F is the size of the set of features that represent the data. The comparison process is directly proportional with the amount of data [161].

$$MD_p(\mathbf{q}, \mathbf{x}_i) = \left(\sum_{f \in F} \|\mathbf{q}_f - \mathbf{x}_{if}\|^p \right)^{\frac{1}{p}} \tag{4.18}$$

Since the complexity of the compression metrics is more difficult to characterise, a K-NN classifier is likely to be $O(nS_D logn)$ where n is the number of clusters [161]. Much research has been conducted regarding alternatives to the time consuming search process of the K-NN classifier approach. Also research focus was directed towards reducing the size of the training data and using less features to define the data.

In summary K-NN classifier is a simple, easy to implement classifier and can yield a solution to many classification tasks. Its main advantages can be deduced from its interpretability and simplicity and should not be underestimated. K-NN classifier is a very simple classifier, easy to debug and implement with a transparent process. Sometimes the classifier output is very useful so it is essential in situations where analysis of the neighbours gives needed information [64]. Noise reduction methods can improve its precision and efficiency. An extension of a memory-based classifier is a case-retrieval network that can drastically improve computational performance on big datasets [161]. If the training set is very large it can have poor performance. As all features contribute to the classification, K-NN classification is not effective in case of irrelevant or redundant data features because all

features are part of the classification process. This can be attenuated by thorough feature weighting or selection. Some other techniques such as Support Vector Machines or Neural Networks can outperform K-NN classification on more challenging classification tasks. However, in the work reported in this thesis, K-NN classification with Euclidean distance is used to form a ranked list of locations, from the most probable to the least probable location, for both sensing modalities as it is shown to perform well. Also it is used when proposing an algorithm for localising the user between calibration points (chapter 6).

4.5 Weighting and grid search engine

Two or more data sources can be fused in various different ways as described in section 3.12. Sometimes weighted combination of the sources in a multimodal fusion should be deployed to achieve best performance. This means that these sources should be suitably weighted and then added. Using a training dataset (which is separate from the testing set) a set of optimal weights for each combination of sources can be identified using an exhaustive grid-search [162]. A grid-search algorithm (illustrated in figure 4.7) can be generally described in the 3D case where it successively achieves more thorough volume searches within Euclidean space, to achieve an optimal estimate of probability density function (PDF) for the so called misfit function [26]. Comparing to linear and stochastic micro techniques it is time consuming and consecutive grid searches might become either very large or very small [26]. Therefore it needs very precise selection of grid size and node spacing.

The grid search implementation is often formed using a nested approach grid search consisting of using one or several grids. The first grid used is strictly defined with its location, size, number of nodes, etc. Furthermore, the following grids which are nested are defined in terms of node number and size using the first grid. The position of the nested grids are given automatically along one or all three axes x,y,z [62]. For each location grid and for every node the quality PDF or misfit value is calculated. For a subsequent nested grid, its location for each node is set to the center of the maximal PDF or the minimum misfit node of the observed grid. The initial grid must be within every next grid. Subsequent grids

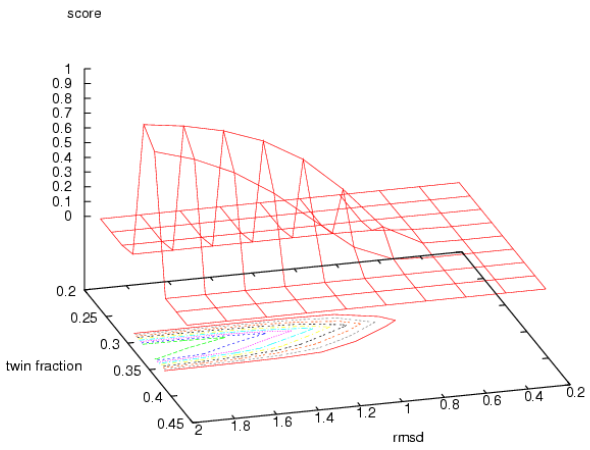


Figure 4.7: Grid search engine example [162]

must be totally contained inside the initial grid (also if their position is set automatically). Grid translation that intersects with a boundary of the first grid should be done as if it is placed inside the initial grid [156]. For every observation and for each node of the position grid, the grid-search approach has to perform systematically throughout each location with the x or y or z index varying the last. Processing times for every iteration may be very large so the algorithm can be extremely time consuming. In much simplified usage employed in this thesis all possible combinations of w_1 and w_2 from the $[0, 1]$ domain (grid size in one dimension) are reduced to the simple case in which the weights are equal to $w_1 = w_2 = 1$. In this way sensing modalities are of equal importance in the fusion and the tracking process.

4.6 Conclusions

In this chapter, an overview is presented of the important state-of-the-art algorithms which will be used as a basis for the proposed algorithms presented in chapters 5 and 6. The algorithms were selected as they showed the best trade-off between complexity and performance and are thus popular in this and many other data processing tasks. Moreover, the

algorithms showed flexibility over some more recent but less attractive and/or less-suitable approaches for indoor localisation tasks.

Chapter 5

WLAN and image-based localisation and tracking

5.1 Introduction

In this chapter, methods for indoor user localisation and tracking inside a university building are presented. An approach that would enable user localisation to within an office is described. The WLAN-based solution, given in section 5.3, uses a extended Naive Bayes approach to find the user location. A novel image-based localisation is realized using a novel approach based on a hierarchical vocabulary tree of SURF descriptor vectors and is described in section 5.4. The method introduces fine tuning of cluster centers iteratively and finding the centroid for each cluster center thus improving the basic hierarchical vocabulary tree approach. A novel fusion approach is proposed to overcome difficulties of each of the individual sensing modalities. Thus, the fusion function is designed to take both localisation results into account and to successfully merge them to achieve better performance (especially when WLAN breaks down and/or is unreliable). It is described in section 5.5. We propose a tracking method (in section 5.8.1) that can be employed when using image-based, WLAN-based or the fusion-based approach. It introduces novel transitional probability function which converts times taken for traversing (between) locations

into probabilities. The effectiveness of combining the strengths of these two complementary modalities is demonstrated for a very challenging dataset. Two experimental setups (ESs) are discussed in sections 5.6 and 5.8.2. The first one discusses the user's location based on fusion of WLAN and image data. This setup consisted of several adjacent offices with a fixed number of CPs placed inside every office. The main goal is to localise a user to a specific CP and to that specific office as well. The second setup is based on the first one and is used in the tracking process. Both localisation and tracking results are given in sections 5.7 and 5.8.3 respectively to demonstrate the effectiveness of the proposed methods.

5.2 Overview

Naive Bayes approach is used to localise user using WLAN. This algorithm is transformed and optimized to better suit the specific environmental challenges encountered. It is used to determine the most likely location of a user. These probabilities were rescaled to sum up to one and denoted as WLAN-based confidences for a location.

For the image-based localisation a hierarchical vocabulary tree of Speeded Up Robust Features (SURF) descriptor vectors is used. SURF vectors are taken from all available images. A hierarchical vocabulary tree is formed using hierarchical k means clustering of SURF descriptor vectors. Hierarchical k means clustering is achieved using simple k-means clustering repeatedly until the final clusters contain less than k SURF descriptor vectors. The simple k-means clustering divides the dataset of all SURF descriptor vectors into k subsets using k cluster centers. A novel fine tuning of cluster centers of the hierarchical vocabulary tree is employed to fix the cluster centers and to achieve better precision and accuracy compared to standard hierarchical vocabulary tree approach. Descriptor vectors are extracted from all images and from a query image as well. After each descriptor of the query image has voted for a location, a ranked list of locations, from the most probable to the least probable location is obtained. These values were denoted as image-based confidences for a location.

For each localisation method we obtain a ranking of possible user locations. The ranking

list of (possible) user locations is achieved using K nearest neighbour classifier. All nearest neighbours are found using simple Euclidean distance. Also the K nearest neighbour approach is used when determining a user's position between calibration points using just WLAN. Moreover, these two modalities are fused in general using weights and grid search engine to achieve better performance than using them separately. Localisation precision based on fusion outperforms image and WLAN-based localisation precision when localising to specific location and to within an office.

Eventually a tracking approach is proposed to find the most likely sequence of locations visited by the user. This approach proposes a function that converts times for visiting consecutive locations into probabilities and finds the total probability of visiting a sequence of locations using WLAN-based, image-based or fusion method separately. The one with the highest value gives the order of visited locations.

5.3 Naive Bayes localisation

Probabilistic WLAN-based localisation techniques based on fingerprinting start with the acquisition of training observations consisting of signal strength information at calibration points distributed along a dense grid throughout the building [86, 73, 140]. To calculate the probability of a user being at a particular CP when the user is positioned at some point in space observed only by signal strength values at that specific point, it was decided to employ a Naive Bayes method. The approach represents an extension of the Bayes and Naive Bayes classifier presented in section 4.2.1. This algorithm takes into account the access points' (APs) signal strength values (RSSI) and also the frequency of the appearance of these APs.

A signature for each CP is defined as a set of W distributions of signal strengths of W APs and a distribution representing the number of appearances of W APs observed at this CP. $C \in \{1, 2, ..., K\}$ denotes the CP random variable where K is the number of CPs, $X_m \in \{1, 2, ..., W\}$ represents the m^{th} AP random variable, $Y_m \in \{s_1, ..., s_V\}$ is the signal strength (assumed to take on discrete values) received from the m^{th} AP, where W is the number of APs, M is the number of APs present in an observation and V is the number of dis-

crete values of signal strength. It is not necessary that each AP produce receivable signals at each CP, and indeed whether or not an AP signal can be obtained at a CP can vary with time depending on the state of the radio channel. $D = \{\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_N\}$ is a set of N training observations where the n^{th} training observation is defined as $\mathbf{o}_n = (c^{(n)}, x_1^{(n)}, y_1^{(n)}, ..., x_M^{(n)}, y_M^{(n)})$, for $n = 1, .., N$. The joint distribution $P(C, X_1, Y_1, ..., X_M, Y_M)$ is given by

$$P(C) \prod_{m=1}^{M} P(X_m|C)P(Y_m|C, X_m) \tag{5.1}$$

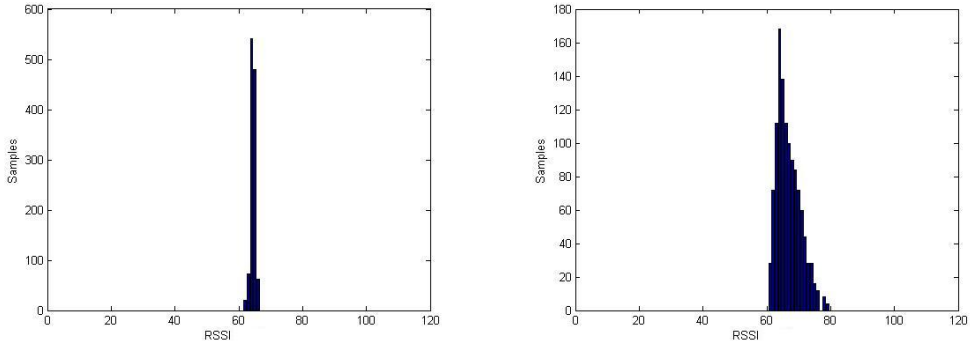
Using the Naive Bayes approach described in section 4.2.1 and one testing observation $\mathbf{o}$ the likelihood that the user is at location c can be written as

$$P(c|\mathbf{o}) \propto P(c) \prod_{m=1}^{M} P(x_m|c)P(y_m|c, x_m) \tag{5.2}$$

The user location is estimated as the value c^* which maximises $P(c|\mathbf{o})$, that is

$$c^* = \arg\max_c P(c) \prod_{m=1}^{M} P(x_m|c)P(y_m|c, x_m) \tag{5.3}$$

In order to estimate the user location, information about the various probability distributions on the right hand side of equation 5.3 must be obtained. In the absence of any other information the *a priori* probability distribution of the user location, $P(C = c)$, is presumed to be uniform. The distribution of AP x given a location c, $P(X_m = x|C = c)$ is multinomial (W-size parameter π_c), the probability of signal strength y given location c and AP x, $P(Y_m = y, C = c, X_m = x)$, can be estimated from the normalised histogram (V - size vector parameter $\gamma_{c,x}$). The majority of the histograms have shape of slightly left-skewed, almost symmetric and slightly right-skewed distributions. Thus, they can be approximated by a lognormal distribution. The RSSI values are usually concentrated around 1-3 dominant modes. On average the histograms varied little with time but in case of significant changes in time, such as a number of people present in an office (compared to an empty office), they varied a lot. An example that illustrates this phenomenon is presented in figure 5.1.
Using the identity function



(a) (b)

Figure 5.1: Effect of users' presence on WLAN RSSI histogram in an office space: no users present (left), users present and moving (right). RSSI measurements are collected at a CP in all four orientations

$$I(s,t) = \begin{cases} 1 & \text{for} \quad s = t \\ 0 & \text{for} \quad s \neq t \end{cases} \tag{5.4}$$

In a maximum likelihood estimation framework the sufficient statistics are

$$n_c = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) \tag{5.5}$$

$$n_c^{(x)} = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) I(x_m^{(n)}, x) \tag{5.6}$$

$$n_{c,x}^{(y)} = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) I(x_m^{(n)}, x) I(y_m^{(n)}, y) \tag{5.7}$$

The probability of AP x given location c, $P(X_m = x|C = c)$ is given by

$$P(X_m = x|C = c) = \frac{n_c^{(x)} + 1}{n_c + W} \tag{5.8}$$

while the probability of signal strength y given location c and AP x, $P(Y_m = y|C = c, X_m = x)$ is given in equation 5.9. These are estimates of the signature parameters.

$$P(Y_m = y|C = c, X_m = x) = \frac{n_{c,x}^{(y)} + 1}{n_c^{(x)} + V} \tag{5.9}$$

The algorithm chooses the location which maximises equation 5.2 as being the user location. We rescaled these probabilities to sum to one and denoted their new values as the CP confidences, p_i.

5.4 Image-based localisation using hierarchical vocabulary trees

5.4.1 Hierarchical vocabulary tree: introduction

Many research works on efficient image retrieval are based on a hierarchical vocabulary tree [132, 54]. This enables users to find an image-to-an image similarity and to establish an appropriate similarity score in the case of large image datasets [132]. Similar works in the image retrieval and similarity field include [45, 44]. They take into account an image representation as a bag of visual words and find the number of similar words between images thus defining a similarity score between them. This approach is taken from recent research in text retrieval [32]. Consequently, only images with similarity score high enough are considered to be good candidates for image matching. These and some other approaches such as [83, 7] are a standard tool for large scale reconstruction. These methods reduce matching effort as in big databases an image rarely matches with the whole database due to occlusion and missing overlap. The bag-of-words representation of an image [152] based on SIFT features [110] has been one of the main large scale image retrieval methods of choice in recent years.

5.4.2 Hierarchical vocabulary tree

In most general terms an image is represented and described using a histogram of quantized feature occurrences based on a codebook of predefined cluster centers (so called visual words). This is usually organized as a structure that represents the entire image database. An efficient approach for approximated nearest neighbour search on the codebook can be done using hierarchical quantization of descriptor vectors, also denoted as a hierarchical

vocabulary tree [132]. SURF descriptor vectors are compared using the standard Euclidean distance, as it is explained in section 4.3.1. The features were split into two groups based on the sign of the Laplacian which enables us to search faster. Descriptor clustering for each group is achieved applying the k-means algorithm, described in section 4.3.2, recursively. Initially we created k clusters, then within each cluster, k more clusters, and so on until the last cluster contains less than k descriptor elements. Eventually, two hierarchical vocabulary trees are created. The maximum number of levels of the tree is denoted by L and each node is divided into k children (an example is shown in figure 5.2).

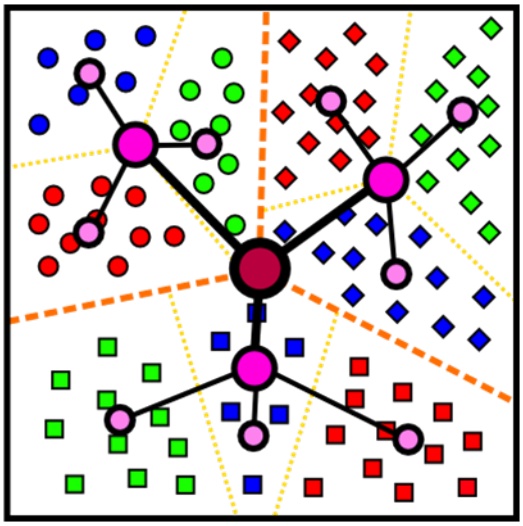


Figure 5.2: An example of hierarchical vocabulary tree of SURF descriptor vectors. In this example $k = 3$ and $L = 3$

The vocabulary tree concept uses the following rule: if the similarity between two features f_i and f_j is high, then it is highly likely that the two features are in fact the same visual word $w(f_i) \equiv w(f_j)$, i.e. the features represent the tree's same leaf node. Based on the quantized features from a query image Q and each database image D a scoring of relevance is derived. Scoring values can be found for a query and every database image. Scoring functions are usually based on a vector $tf - idf$ model (term frequency - inverse document frequency) which gives a location/data ranking based on the level of similarity between query and database images. In this thesis we make the following approximations:

• there is an 1 : 1 mapping between visual words and descriptor vectors meaning that we compare only (SURF) descriptor vectors.
• we do not use $tf - idf$ model but a simple model that calculates the number of similar descriptor vectors.

Creating a vocabulary tree is not an easy task because of the feature vector size and their number. In order to build a vocabulary tree of M leaf nodes, $n >> M$ data points are required. Moreover k-means clustering needs to be performed repeatedly to eventually cluster all SURF descriptor vectors and to form the tree. Initially we created k clusters, then within each cluster, k more clusters, and so on until the last cluster contains less than k descriptor elements. k-means clustering is feasible since it only requires linear memory $O(k+n)$ in the number of cluster centers k and data points n. Since in our SURF approach, k-means clustering and K nearest neighbour classifier are crucial for building hierarchical vocabulary tree they will be explained in greater detail in the following sections.

5.4.3 Propagation in a hierarchical vocabulary tree

A hierarchical vocabulary tree structure enables an efficient quantization of feature descriptors. Also, the hierarchical tree can employ a fast search using a Best Bin First (BBF) strategy [25]. Feature quantization for a vocabulary tree requires $O(kL)$ (k branch factor and L levels) dot products and it is observed that a broader tree yields superior performance since more descriptors are considered.

5.4.4 Fast localisation based on hierarchical vocabulary tree

SURF features are extracted from all R database images. Eventually we had F feature descriptors. Every feature is connected with the image from which it was extracted. Then two hierarchical vocabulary trees are built using the sign of Laplacian explained in section 4.3.1. There is a main characteristic in building hierarchical vocabulary tree which explains how the cluster centers are formed. For the first two and rarely three levels of the hier-

archical tree the following procedure was applied. k cluster centers for the first level were found calculating the mean value of several previously calculated cluster centers. In other words for I iterations there are I cluster centers vectors, each of length k. In the work presented in this thesis $I = 4$. Then the mean value for each dimension was calculated and the final vector of length V represents k cluster centers. The process is repeated for the second and eventually the third level of the tree. For the higher tree levels the process is not necessary as cluster centers are already properly placed (positioned). In general let $T = \{t_j \,|j = 1, ..., n\}$ be attributes of n-dimensional vector and $W = \{x_j \,|j = 1, ..., r\}$ be each data of T. The pseudocode is given as:

1. Set $W = \{w_j \,|j = 1, ..., r\}$ as each data of T, where $T = \{t_j \,|j = 1, ..., n\}$ is attribute of n-dimensional vector.
2. Set K as the predefined number of clusters.
3. Determine l as numbers of computation
4. Set $j = 1$ as initial counter
5. Perform K-means algorithm.
6. Record the centroids of clustering results as $C_j = \{c_{ju} \,|u = 1, , K\}$
7. Increment $j = j + 1$
8. Repeat from step 5 while $j < l$.
9. Assume $C = \{C_j \,|j = 1, , l\}$ as new data set, with K as predefined number of clusters
10. Apply hierarchical algorithm
11. Record the centroids of clustering result as $D = \{d_j \,|j = 1, , K\}$
12. One uses $D = \{d_j \,|j = 1, , K\}$ as initial cluster centers for K-means clustering.

The performance improves as the branch factor increases (slowly) and the number of nodes increases (dramatically). The branch factor and number of nodes of the tree were

chosen to best match tree performance and time of the traversals. This approach makes the localisation process efficient because the features can be matched precisely. Descriptor vectors are extracted from all images and from a query image as well. For every descriptor a match is found using $+1$ or -1 hierarchical tree, based on 1 nearest neighbour classifier (i.e. the nearest match is found) explained in section 4.4. Every match is connected via a label to the image it is extracted. If matched, it gives one vote for the location to which the image belongs. After each descriptor has voted for a location, a ranked list of locations, from the most probable to the least probable, is obtained. Similarly to the WLAN case, we assigned a confidence for each CP (q_i) as the ratio of the number of votes associated with that CP and the total number of votes. This improved hierarchical vocabulary tree achieved 14.82% better precision than standard hierarchical tree. Moreover, fixing cluster centres gives more stable results (precise results are three times more repeatable) than using standard vocabulary tree. Image resolution can be also analyzed in the context of localisation results. Higher resolution images have more SURF feature vectors, give more precise results but are also more computationally expensive. The trade-off between precision and real-time processing should be analyzed in the context of application.

5.5 Data fusion

To fuse information from the two modalities, we take confidences p_i and q_i from both sensing modalities P and Q into account, where in our case P and Q were WLAN and image sensing methods respectively. Here, i refers to a given CP. If we sort these confidences we can denote the first ranked, the second ranked, the third ranked, etc. confidence by p_{max1}, p_{max2}, p_{max3}, etc. respectively (or by q_{max1}, q_{max2}, etc. for the Q modality). It was decided to use a large passive training dataset of confidences of different CPs. This would help in building a robust fusion function which would be reliably used on (unknown) testing data. First, let us define for modality P

$$P_{gh} = p_{maxg} - p_{maxh} \tag{5.10}$$

and similarly for modality Q

$$Q_{gh} = q_{maxg} - q_{maxh} \tag{5.11}$$

We made a very large training dataset of more than 600 ranked confidence pairs for both the modalities. Measurements were taken at all the CPs and at different times of the day to make the fusion process more robust and self-contained. Observing P_{12} and Q_{12} in these training confidence pairs we concluded that for values P_{12} and/or Q_{12} beyond some reliably large thresholds, we were sure that the correct CP (location) was the 1^{st} ranked one, based either on P or Q (or both). All P_{12} and Q_{12} values were calculated and sorted and eventually the minimum of all P_{12} and the minimum of all Q_{12} values were found. They are found to be very solid boundary for each modality. We denoted them by T_1 and T_2 for P and Q modality respectively. Thus $T_1 = min\{P_{12r}\}$ and $T_2 = min\{Q_{12s}\}$. The thresholds were robust and made the localisation process stable for all available CPs. Also the process was ensured when condition $P_{12} \geq Q_{12}$ was met. So far only two thresholds were used.

There were confidences which did not satisfy the requirements given above. We deduced that introducing multiplication and/or addition functions under some conditions can improve precision (or at least average precision). Confidence pairs that could improve localisation precision (or average precision) were transformed to improve new ranking compared to the rankings of the single modalities. Actually, one could use different (and/or even more complex) functions but the process of finding the thresholds would be more challenging. It is important to note the improvement using addition and multiplication functions is only incremental, and the main increment comes when using only two thresholds. This is important in cases when training and testing sets are significantly different form each other which is not a common situation in practice.

Eventually, if the 1^{st} ranked confidence belongs to the correct location, the algorithm would discard it if P_{12} (or Q_{12} or both) is below this (these) threshold(s). Also we found that the ranking of the correct location did not fall below some positions in both sets of rankings. In general, these are the m^{th} position for P and the n^{th} position for Q modality.

Thr.	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Val.	0.019	0.015	0.0047	0.0069	0.0041	0.0062	0.008	0.0093	0.0078	0.0089

Table 5.1: Threshold values (Val.) used in the fusion process

The fusion function is thus as follows (equation 5.12), where f_i represents fusion confidence and k_i confidence of the method to which $min(n, m)$ corresponds. The location output by the algorithm is the one with the maximum value of the fusion confidence.

$$f_i = \begin{cases} p_i, & P_{12} \geq Q_{12} \wedge P_{12} \geq T_1 \wedge Q_{12} \geq T_2 \\ q_i, & Q_{12} \geq P_{12} \wedge P_{12} \geq T_1 \wedge Q_{12} \geq T_2 \\ p_i, & P_{12} \geq T_1 \wedge Q_{12} < T_2 \\ q_i, & Q_{12} \geq T_2 \wedge P_{12} < T_1 \\ p_i q_i, & T_3 \leq P_{12} \leq T_4 \wedge T_5 \leq Q_{24} \leq T_6 \\ p_i + q_i, & T_7 \leq P_{12} \leq T_8 \wedge T_9 \leq Q_{24} \leq T_{10} \\ k_i, & \text{else} \end{cases} \tag{5.12}$$

The fusion process goes from the top to the bottom. The fusion function is logically consistent thus ensuring that there are no conflicting situations. The thresholds values used in the fusion function and in the localisation process are successfully tested at all the CPs. They are given in table 5.1.

Some other integrative approaches were also tried to merge these two different modalities. Early fusion was difficult to apply as we were processing two non-compatible sources of information: one in the form of matrix (images) and the other in the form of a vector (WLAN). Moreover hybrid of these two modalities gave worse results than when using the late fusion method described here.

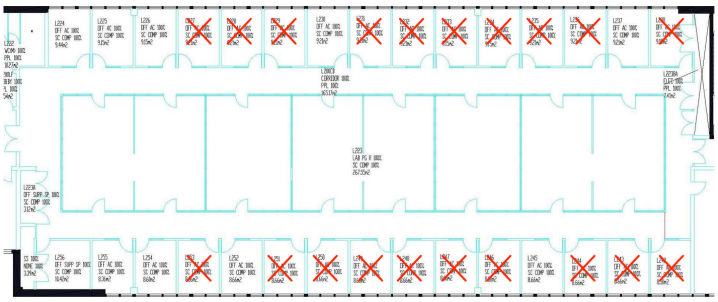
5.6 Experimental setup when localising to within an office

For this experimental test bed 20 offices on the second floor of a building (see figure 5.3) are used, where the average size of an office is $8.9m^2$. Within each office we use 5 calibrations

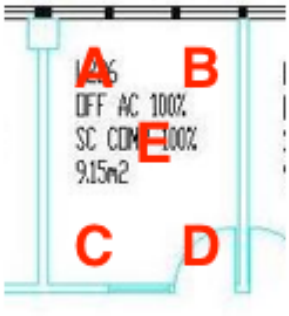
points (CP), A, B, C, D & E. Each orientation of a CP (North, South, West and East) is represented with 8 (640×480 pixels) images taken with a camera Canon PowerShot A560 (see figure A.4(b) for examples), and 300 RSSI (received signal strength indication) observations taken with a Dell Inspiron laptop with Intel Core 2 Duo Processor T5250 (2.0 GHz, 2 MB L2 cache, 667 MHz FSB), memory of 2×2048 MB, 667 MHz Dual Channel DDR2 SDRAM, SATA Hard Drive with 450 GB (5.400rpm) and Intel PRO/Wireless 3945ABG card using InSSIDer software[1] . Every CP is represented using data from all four orientations together. For one set we had $5,000$ images, of which $3,200$ were used for training (20 offices X 5 CPs X 4 orientations X 8 images) and $1,800$ for testing, and $125,000$ signal strengths observations of which $120,000$ were used for training (20 offices X 5 CPs X 4 orientations X 300 RSSI) and $5,000$ for testing. Each orientation of every CP is used at least four times in the testing phase. WLAN and image data collection processes are described in section A.2 and section A.3 of the Appendix respectively. One observation consists of received signal strengths from all confident APs: 14 in our case. In the best case it is the total number of APs. Offices are chosen to be next to each other and moreover, look very similar inside, thus resulting in very challenging data for both WLAN and image-based localisation methods.

In a test we used one image and one signal strength observation per CP and tested how precisely we could localise to a given CP. Clearly, if we can localise to a CP, we can localise to within the office that contains that CP. However, we wanted to understand how many (and which) CPs are necessary as this has an impact on the manual data collection effort required to perform accurate localisation. We also present results for localizing to a given office whereby the office selected is based on the 1^{st} ranked results corresponding to one of the CPs for that office, even if the top ranked CP is not the actual location CP. We examined localisation precision for 5 different combinations of 1, 2 and 3 CPs per office (giving 5 different sets of 20, 40 and 60 locations respectively in total). Precision (P) and average precision (AVP) were used as performance measures. The precision is calculated as the ratio between the total number of the first ranked correct locations in N_t tests and N_t.

[1] http://www.metageek.net/products/inssider/



(a)



(b)

Figure 5.3: (a) Map of office locations – red crosses indicate offices used; (b) Calibration points $ABCDE$ within an office

The average precision is computed as $AVP = \frac{\sum_{k=1}^{N_t} \frac{1}{P_k}}{N_t}$ where P_k represents the position of the correct location in the k^{th} test.

5.7 Localisation results

An example of the benefits of fusion, when 2 CPs are observed as individual locations (BE), is shown in figure 5.4. It shows the behaviour of precision considering the top N ranked results, thus illustrating how often each modality returned the correct location as the top ranked result, 2^{nd} ranked results, and so on (bars in the graph) and also how precision increases if the top N ranked results are considered (lines in the graph). In the top N, for $N = 1 \ldots 5$, the fusion approach outperforms both WLAN and image-based methods reaching precision of 91.82%. Also it can be seen that correct location rank doesn't drop below 8^{th} for WLAN, and 12^{th} for the image-based method. In this example we have

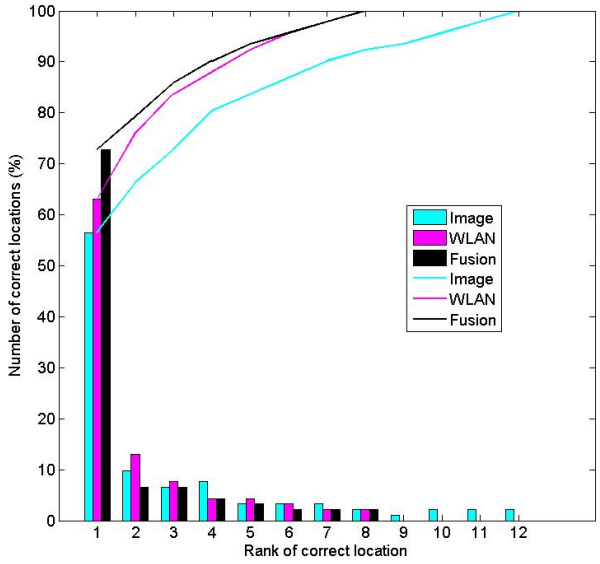


Figure 5.4: Number of correct locations (in %) found on the N^{th} rank (bars); Number of correct locations (in %) found in the top N ranks (lines)

$AVP_W = 75.18\%$, $AVP_I = 68.14\%$ and $AVP_F = 80.94\%$ for the WLAN, image-based and the fusion approach respectively.

Similar examples demonstrating the effectiveness of the approach are given in figure 5.5 and in figure 5.6. In these examples different combinations of 3 and 2 CPs per office are used. In all cases the fusion outperforms each modality separately.

The left hand side of table 5.3 shows results on average when using 1, 2 and 3 CPs per office (every CP represents a different location), using WLAN data only (P_W), image data only (P_I) and the fusion of both modalities (P_F). For 2 and 3 CPs we show a selection of results, corresponding to the best performing ones, rather than all possible combinations. The right hand side of the table shows results when we take into account the 1^{st} ranked result that is not the correct one but that belongs to a CP within that particular office. This gives the precision to a particular office, denoted by P_{WO}, P_{IO} and P_{FO} obtained using WLAN-based, image-based and the fusion method respectively. From the table it is clear that fusion of WLAN and images significantly improves the performance of using either approach on its own. Moreover, on average, P_W, P_I and P_F decrease while P_{WO},

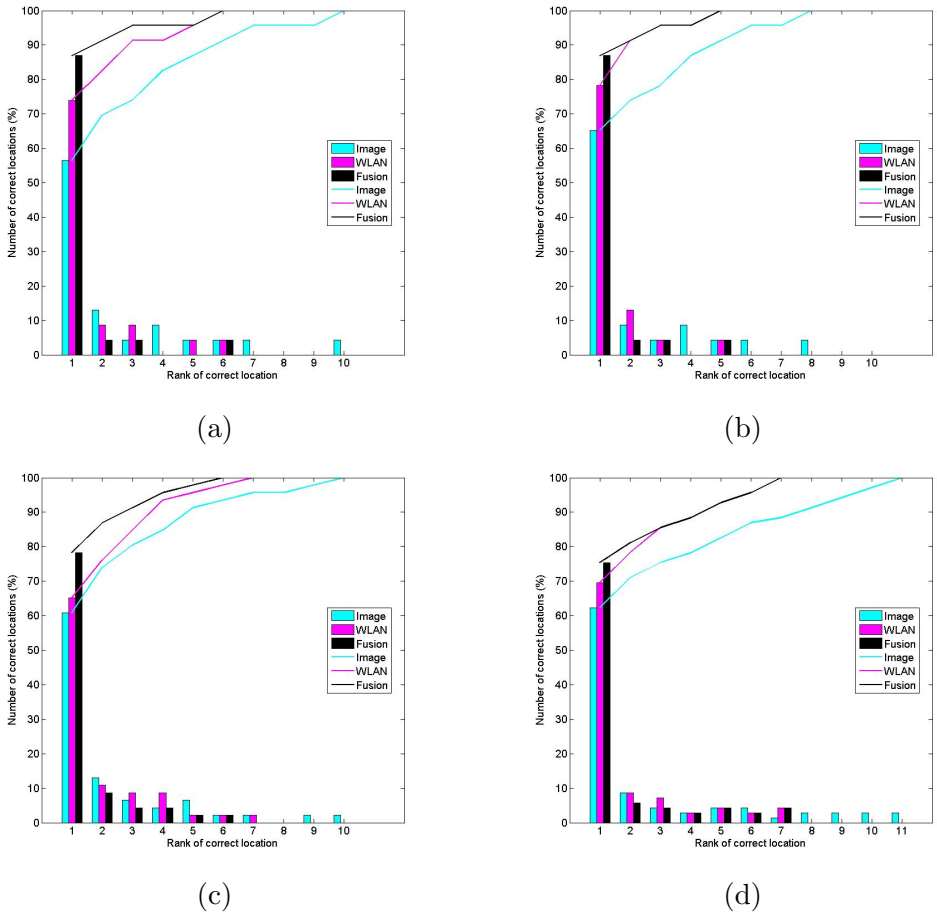


Figure 5.5: Number of correct locations (in %) found on the N^{th} rank (bars); Number of correct locations (in %) found in the top N ranks (lines)

P_{IO} and P_{FO} increase when the number of CPs per office increases. This is expected since the data within an office are very similar, thus sometimes making the algorithms *choose* the nearby CP instead of the correct one. For images we have an even more complex situation as locations that are not physically close by can look similar as well. When we consider localisation to an office many incorrect 1^{st} guesses become correct especially when the number of CPs in an office increases. Thus, in the case of 3 CPs, one can notice a large increase in precision, where on average it increased by 15.52%, 18.72% and 13.04% for WLAN-based, image-based and the fusion method respectively.

The performance variation for the localisation to within an office obtained by using a

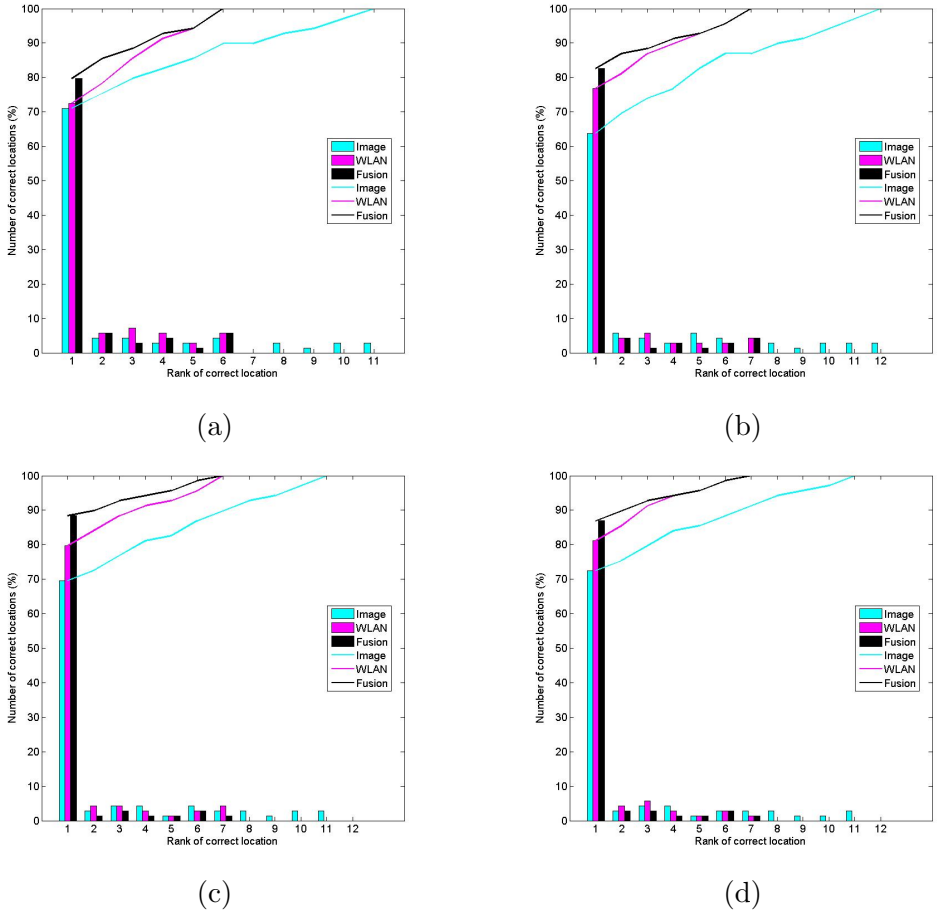


Figure 5.6: Number of correct locations (in %) found on the N^{th} rank (bars); Number of correct locations (in %) found in the top N ranks (lines)

variable number of CPs also gives an interesting conclusion. Whilst the best results are naturally always obtained by using all 5 CPs for each office, we can see that using only one CP produces reasonably good performance: 69.57% precision for the worst result (calibr. point A), 76.09% on average. This is important as it means that the manual data collection stage for model creation outlined in section 5.6 is viable as it only needs to be performed once (i.e. at one CP) per office in order to obtain reasonably accurate performance.

CP	P_W	P_I	P_F	P_{WO}	P_{IO}	P_{FO}
A	65.22	50.00	69.57	65.22	50.00	69.57
E	69.57	60.87	73.91	69.57	60.87	73.91
C	69.57	56.52	78.26	69.57	56.52	78.26
B	73.91	58.70	82.61	73.91	58.70	82.61
D	71.74	63.04	76.09	71.74	63.04	76.09
AB	61.96	47.83	69.57	65.22	60.87	71.74
BE	63.04	56.52	72.83	71.74	70.65	76.09
ED	66.30	54.35	73.91	73.91	60.87	78.26
AC	64.13	46.74	71.74	75.00	53.26	78.26
BC	68.48	53.26	73.91	77.17	58.70	80.43
ABE	55.07	46.38	63.77	69.57	62.32	75.36
AEC	58.70	45.65	66.67	72.46	65.22	79.71
EBD	58.70	49.28	70.29	76.81	63.77	84.06
ABD	61.59	47.83	69.57	79.71	69.57	83.33
ACD	63.04	44.20	71.74	81.16	65.94	84.78

Table 5.2: Localisation results: P_W, P_I, P_F are precision results for considering each CP as a separate location using WLAN, image and fusion respectively; P_{WO}, P_{IO}, P_{FO} are precision results for localising to a specific office

5.8 WLAN and image-based tracking

WLAN and image-based tracking is examined using experimental setup described in section 5.6. Calibration points are very close to each other within an office thus making them very difficult for algorithms to distinguish as different locations. Thus, we decided to have only one CP per office and 5 different experimental setups in total.

5.8.1 Proposed tracking method

This section addresses the automatic tracking of a user indoors using fusion of WLAN and image data. A tracking method is proposed based on a simple Viterbi multiplestate model [102] using simple Hidden Markov Model states [137]. In the tracking scenario we only used one CP per office using experimental setup given in figure 5.3: either A, B, C, D or E. Thus we have 5 different experimental scenarios. Let us denote by $t_{i,j}$ and $t^*_{i,j}$ time intervals measured in the training and the testing phase respectively between any two consecutively

visited locations i and j. Here we refer to i and j as the location output by any of three possible methods used (WLAN-based, image-based and fusion-based). Also let us denote by t^k, the k^{th} the nearest time interval to $t^*_{i,j}$ in the training phase, such that it refers to locations i and j which are output by any of the three methods. If i or j is not obtained by the algorithm output we discard that t^k and do not include it (and its corresponding i and j) in the tracking process. Transitional probability, $T^k_{i,j}$, which models how likely the user passes by the pair of locations i and j, $i \neq j$, is derived and given in equation 5.13.

$$T^k_{i,j} = 1 - \frac{\left|t^*_{i,j} - t^k\right|}{max_k \left\{\left|t^*_{i,j} - t^k\right|, k \geq 1\right\}}, \quad (1 \leq i, j \leq n) \tag{5.13}$$

At every location the user can estimate position using either WLAN-based (p_i), image-based (q_i) or fusion-based approach (f_i) and obtain the ranking of possible locations from the most probable to the least probable. For a path consisting of several locations, e.g. $I - J - K - L - M - P$ where $1 \leq I, .., P \leq p$ represent different locations with length equal to n, the total probability consists of the sum of probabilities of being at these locations and the sum of the transitional probabilities of visiting every two consecutive locations ($I - J$, $J - K$, $K - L$, $L - M$ and $M - P$). In our case $p = 20$. Equation 5.14 calculates the probability of visiting several locations.

$$P_{l_n} + \sum_{L_i=l_1}^{l_{n-1}} P_{L_i} + T^k_{L_i,L_{i+1}} \tag{5.14}$$

where P_{L_i} refers to the probability of being at location L_i and $T^k_{L_i,L_{i+1}}$ refers to the transitional probability $T^k_{i,j}$ as explained and given in equation 5.13. For each location we obtain a ranked list of possible locations from the most to the least probable. For a testing time stamp between two consecutively visited locations i and j we can find a ranked list of location pairs whose times (from the training phase) are very similar to the testing time-stamp (they are also ranked from the most to the least probable). The top k ranked time stamps are chosen (as explained before), denoted by t^k where $k \in N^+$, and since it is known which location pair this particular time stamp belongs to, it can be connected to the *same* two

location outputs given by any of the modalities used. Probabilities of being at specific locations and the corresponding transitional probabilities are normalised to $[0, 1]$ interval to reliably represent the influence of each of $(n-1)$ sections. Then these probabilities are added and the process is repeated (as given by equation 5.14) for all other locations until the last visited location is reached. Thus we have k different sequences each consisting of n locations. The one with the highest probability value gives the order of visited locations.

5.8.2 User tracking: experimental setup

Times measured between consecutively visited locations were collected using a standard stopwatch. In this thesis a user average speed of walking is approximately 1.1m/s and the user is able to pass a three meter distance in approximately 2.73 seconds. This was experimentally proved. Using this approximation and the university building map together with its scale one is able to reconstruct real distances from the map and calculate all the times. The times are also checked in real world scenario thus showing robustness of this approach. The walking path is chosen to be fixed and along a line that halves the corridor next to the offices. In the testing phase one image and/or one signal strength observation per CP together with time interval ($t^*_{i,j}$) measured between two consecutively visited locations denoted by i and j ($i \neq j$, $1 \leq i, j \leq n$, $n \in N^+$) are used to track the user. If $i = j$ the user is stationary and $T^k_{i,i} = 0$. n denotes the number of *visited* locations. Thus in total for each ES we had 100 CPs. The data were collected and the experiments were performed on Dell Inspiron laptop with Intel Core 2 Duo Processor T5250 (2.0 GHz, 2 MB L2 cache, 667 MHz FSB), memory of 2×2048 MB, 667 MHz Dual Channel DDR2 SDRAM, SATA Hard Drive with 450 GB (5.400rpm) and Intel PRO/Wireless 3945ABG card.

5.8.3 Tracking results

Table 5.3 shows the results comparing tracking performance when using either data source and the combination of both sources using analyses given in section 5.8.1. Here the precision is calculated as ratio of correctly guessed locations and total number of locations in a

Dataset ID	P_W	P_I	P_F
A	73.41	61.66	82.66
B	76.18	62.74	84.71
C	66.39	67.19	76.25
D	71.83	57.24	79.02
E	65.42	59.82	82.83
Avg.	70.65	61.73	81.11

Table 5.3: Results: P_W, P_I, P_F represent precision (in %) when using WLAN, image and fusion method respectively. Also the last line of the table shows the results on average thus demonstrating the effectiveness of both fusion and tracking approaches

tracking process. For each of 5 datasets 9 tracking process were performed. Each tracking process consisted of $n = 10$ locations. Then the average precision for each dataset is calculated. Not only does the combination of both sources increase the performance, the difference between them is notably reduced.

5.9 Conclusions

In this chapter, results of combining two complementary sources of data for classifying locations and finding the deemed position of the user are presented. Moreover a tracking approach based on using these two modalities separately and later the fusion approach is given as well. In both cases, in all experimental setups, by fusing wireless signal strength readings and image-based matching, we achieve better performance than any individual/combined modality. Thus, this demonstrates the need and usefulness for employing more (than one) sensing modalities. Also, in both cases the need for fusion is justified by using both sensing modalities at all times to achieve better accuracy/precision in localisation and tracking and to get information about the user's context. One can see that using only one CP produces reasonably good performance thus meaning that the manual data collection stage for model creation outlined before is viable as it only needs to be performed once (i.e. at one CP) in order to obtain reasonably accurate performance. Possible extensions of the work can be other fusion methods and more sophisticated classifiers in order to achieve higher accuracy

and more efficient performance.

Chapter 6

Localisation between calibration points

6.1 Introduction

In previous chapters we have introduced fusion algorithms that use information from two complementary modalities to locate a user at one of a finite set of locations which we call calibration points. Clearly an improvement to the performance of either modality can potentially improve the performance of the overall fusion process. This chapter addresses this, and specifically outlines a potential improvement to WLAN-based localisation only. The improvement allows us to locate users at general positions in space, rather than the finite set of calibration points. This has the potential to improve the location accuracy, or to achieve the same accuracy with a reduced set of calibration points, with an attendant reduction in system set-up costs. Consequently a dense grid of CPs were needed in order to keep the error small.

In this chapter a novel WLAN-based method used to localise the user between calibration points defined with a grid of calibration points is proposed. It uses fewer calibration points (CPs) than standard, well-known localisation approaches and still achieves good performance. The need for fewer CPs is due to the use of robust, range and angle-dependent

likelihood functions that describe the probability of a user being in the vicinity of each CP. The actual location of the user is estimated by solving a system of two non-linear equations with two unknowns derived for a pair of CPs. Different pairs of CPs can be chosen to make multiple estimates which can then be combined to increase the accuracy of the estimate. We also give short but important analyses as to why a linear approximation to likelihood function is appropriate. Moreover, we tried different CP spacing and tested how it affects the overall accuracy and performance. Two separate experiments were performed: in the case of a relatively open-plan space and in the case where CPs were separated by walls which would show how likelihoods change as one passes through an obstacle. We compare results against well-known competing approaches showing the superiority of the proposed method.

6.2 Unconstrained user localisation

6.2.1 Generalization of the Naive Bayes method

The model outlined in section 5.3 assumes that a user is located at one of the CPs. Any deviation in signal strength is deemed to be due to the natural variation in signal strength expected at these locations and explicitly accounted for in the model. The model outputs the CP that most closely matches the signal strength pattern seen by the user. In reality the user is likely to be positioned *anywhere* in space and the deviation in signal strengths is due to their different position relative to the APs compared to that of the CPs, as well as natural variation due to fading etc. However the algorithm, described in section 5.3, is only capable of locating to the nearest CP, leading to the necessity of using a fine grid of CPs in order to keep the location error under control. In this section a technique that can interpolate between CPs and locate users at positions other than the calibrated ones is presented.

Let $\mathcal{L}_{CP_i}(\vec{r})$ be the likelihood that, applying the methodology of section 5.3, a user at location $\vec{r}$ will be identified as being at CP $\overrightarrow{CP}_i$. While during the training phase there

is no direct information about this quantity for general points $\vec{r}$ the training observation data allow estimation of this quantity for a discrete set of locations, namely the CPs $\overrightarrow{CP}_j$ for $j = 1, ..., K$. To see how this might be done let $\mathbf{o} = (x_1, y_1, ..., x_M, y_M)$ be an item of training data collected at CP $\overrightarrow{CP}_j$. It can be used to compute

$$\mathcal{L}_{CP_i}(\vec{r}) \equiv P(\overrightarrow{CP}_i) \prod_{m=1}^{M} P(x_m|\overrightarrow{CP}_i)P(y_m|\overrightarrow{CP}_i, x_m) \tag{6.1}$$

for $\vec{r} = \overrightarrow{CP}_j$ but also for *any* $\vec{r}$ if the corresponding observation data are obtained. Note that various probability distributions on the right-hand side of equation 6.1 are $\vec{r}$-dependent. Consider now figure 6.1 where $\overrightarrow{CP}_i$ and $\overrightarrow{CP}_j$ are separated by distance r_{ij}.

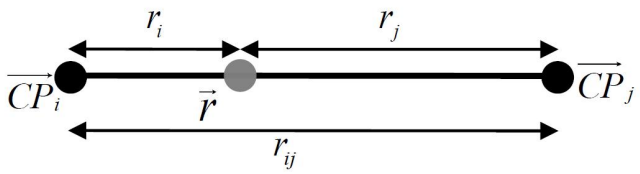


Figure 6.1: Linear interpolation

$\mathcal{L}_{CP_i}\left(\overrightarrow{CP}_i\right)$ and $\mathcal{L}_{CP_i}\left(\overrightarrow{CP}_j\right)$ can be computed as above and used to calculate $\mathcal{L}_{CP_i}(\vec{r})$ for a point $\vec{r}$ along the line between $\overrightarrow{CP}_i$ and $\overrightarrow{CP}_j$ using simple linear interpolation

$$\mathcal{L}_{CP_i}(\vec{r}) \simeq \frac{r_j \mathcal{L}_{CP_i}(\overrightarrow{CP}_i) + r_i \mathcal{L}_{CP_i}(\overrightarrow{CP}_j)}{r_{ij}} \tag{6.2}$$

This approach can be generalised to points not lying directly on a line between CPs. Consider figure (6.2). Here three CPs are given, denoted by $\overrightarrow{CP}_i$, $\overrightarrow{CP}_j$ and $\overrightarrow{CP}_k$

The following equations 6.3 and 6.4 for $\mathcal{L}_{CP_i}(\vec{r})$ apply only when the user is positioned along $\overrightarrow{CP}_i \rightarrow \overrightarrow{CP}_j$ and $\overrightarrow{CP}_i \rightarrow \overrightarrow{CP}_k$ lines respectively. Using one extra training observation $\mathbf{o}$ at each position $\overrightarrow{CP}_i$ and $\overrightarrow{CP}_j$, $\mathcal{L}_{CP_i}(\overrightarrow{CP}_j)$ and $\mathcal{L}_{CP_i}(\overrightarrow{CP}_i)$ can be calculated using equation 6.1. r_{ij} and r_{ik} are known distances. Thus, both equations 6.3 and 6.4 can be written in $-kr_i + n$ form where k and n are constant values and derived easily.

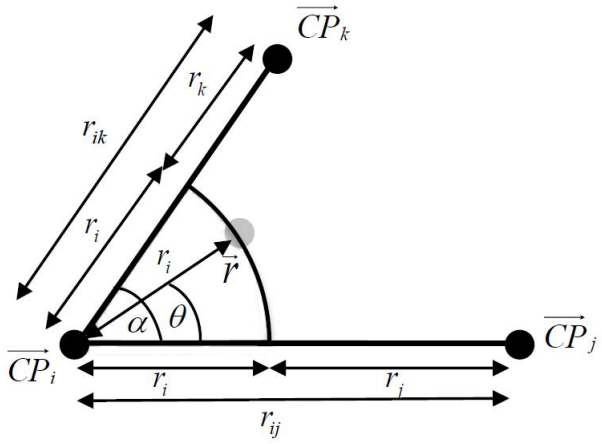


Figure 6.2: Bilinear interpolation

$$\mathcal{L}_{CP_{i1}}(\vec{r}) = \frac{r_j \mathcal{L}_{CP_i}(\overrightarrow{CP}_i) + r_i \mathcal{L}_{CP_i}(\overrightarrow{CP}_j)}{r_{ij}} \tag{6.3}$$

$$\mathcal{L}_{CP_{i2}}(\vec{r}) = \frac{r_k \mathcal{L}_{CP_i}(\overrightarrow{CP}_i) + r_i \mathcal{L}_{CP_i}(\overrightarrow{CP}_k)}{r_{ik}} \tag{6.4}$$

In the polar coordinate system with the origin at $\overrightarrow{CP}_i$, and the user at (r_i, θ), equation 6.5 can be used to calculate $\mathcal{L}_{CP_i}(\vec{r}) \equiv \mathcal{L}_{CP_i}(r_i, \theta)$, for $0 \leq \theta \leq \alpha$ degrees. For $\theta = 0$ and $\theta = \alpha$ degrees equation 6.5 reduces to equation 6.3 and 6.4 respectively.

$$\mathcal{L}_{CP_i}(r_i, \theta) = \frac{(\alpha - \theta)\mathcal{L}_{CP_{i1}}(r_i, 0) + \theta \mathcal{L}_{CPi2}(r_i, \alpha)}{\alpha} \tag{6.5}$$

Note that $\mathcal{L}_{CP_i}$, by virtue of being a function of $\vec{r}$ is a function of both r_i and θ as depicted in figure 6.2.

6.2.2 Proposed localisation algorithm

The improved localisation technique will be referred to as SEAMLOC or SEAM. SEAM means the same as SEAMLOC and is used due to lack of space. The method develops as follows. An observation $\mathbf{o} = (x_1, y_1, ..., x_M, y_M)$ is made by the user and the procedure

described in section 5.3 is followed. Thus $\mathcal{L}^*_{CP_i} = \mathcal{L}_{CP_i}(\vec{r})$, where $\vec{r}$ is the actual (to be determined) user position, are computed from equation 6.6 for $i = 1, ..., K$

$$\mathcal{L}^*_{CP_i} = P(\overrightarrow{CP}_i) \prod_{m=1}^{M} P(x_m|\overrightarrow{CP}_i) P(y_m|\overrightarrow{CP}_i, x_m) \tag{6.6}$$

but, rather than choosing c^* as prescribed by equation 5.3 as the estimate for the location, the top three-ranked locations are identified, that is the three CPs with the highest values of $\mathcal{L}^*_{CP_i}$. These CPs are denoted, from the highest rank downwards, as $\overrightarrow{CP}_1$, $\overrightarrow{CP}_2$ and $\overrightarrow{CP}_3$. Based on the observed measurements from the user they have likelihoods $\mathcal{L}^*_{CP_1} \geq \mathcal{L}^*_{CP_2} \geq \mathcal{L}^*_{CP_3}$.

The top three most probable CPs can either form a triangle or lie on a straight line (horizontal, vertical or diagonal). In the case the top three CPs are placed along a straight line the centroid of the two furthest CPs represents the user estimation. Alternatively these three CPs can form a triangle in which no angle is greater than $90\,^\circ$. In figure 6.3, one can see the triangle formed using the top three most probable CPs, denoted by $\overrightarrow{CP}_1$, $\overrightarrow{CP}_2$ and $\overrightarrow{CP}_3$ respectively. The sides of the triangle that connect $\overrightarrow{CP}_1$ and $\overrightarrow{CP}_2$, $\overrightarrow{CP}_1$ and $\overrightarrow{CP}_3$ and $\overrightarrow{CP}_2$ and $\overrightarrow{CP}_3$ are denoted by a, b and z respectively. Triangle angles are denoted by α, β and γ. It can be seen that $0 \leq \theta \leq \alpha$ degrees and $0 \leq \phi \leq \beta$ degrees. These are known values as the positions of the top three most probable CPs are known.

Based on equations 6.3 and 6.4 and using the notation given in figure 6.3 similar equations for $\mathcal{L}_{CP_{11}}(r_1, 0)$, $\mathcal{L}_{CP_{12}}(r_1, \alpha)$ and $\mathcal{L}_{CP_{21}}(r_2, 0)$, $\mathcal{L}_{CP_{22}}(r_2, \beta)$, for $\overrightarrow{CP}_1$ and $\overrightarrow{CP}_2$ respectively, can be written. For simplicity let these four be denoted by $f_1(r_1)$, $f_2(r_1)$, $g_1(r_2)$ and $g_2(r_2)$ respectively. Let them be written in $-k_i r_1 + n_i$ $(i = 1, 2)$ and $-k_j r_2 + n_j$, $(j = 3, 4)$ form respectively. It should be noted that various other interpolations including polynomial, sinusoidal, logarithmic, exponential functions, etc. were tested and their values compared to the real values of the probabilistic functions, i.e. the values of the functions at specific points inside the triangle. The linear approximation gave the best approximation, was better suited to solving the system of equations and moreover it is faster to compute. Using equation 6.5, $\mathcal{L}_{CP_1}(r_1, \theta)$ and $\mathcal{L}_{CP_2}(r_2, \phi)$ can be written following equations 6.7 and

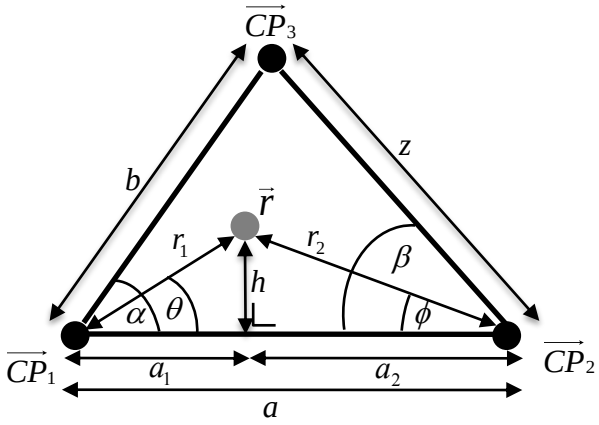


Figure 6.3: Triangle formed using the three nearest CPs with all the angles. r_1 and r_2 denote distances from the user ($\vec{r}$) to the CP_1 and CP_2 respectively. The horizontal side of the triangle is denoted by a and the line that is normal to a is denoted by h

6.8.

$$\mathcal{L}_{CP_1}(r_1, \theta) = f_1(r_1) + (f_2(r_1) - f_1(r_1))\frac{\theta}{\alpha} \tag{6.7}$$

$$\mathcal{L}_{CP_2}(r_2, \phi) = g_1(r_2) + (g_2(r_2) - g_1(r_2))\frac{\phi}{\beta} \tag{6.8}$$

$\mathcal{L}_{CP_1}(r_1, \theta)$ and $\mathcal{L}_{CP_2}(r_2, \phi)$ show how likelihood depends on distance and angle. Angle θ is measured starting from the $\overrightarrow{CP_1} \rightarrow \overrightarrow{CP_2}$ direction and it rises in the anticlockwise direction and angle ϕ is measured from the $\overrightarrow{CP_2} \rightarrow \overrightarrow{CP_1}$ direction and it rises in the clockwise direction. Some different approaches were tried as well. One was to use a two-dimensional signal strength function inside the triangle and then to find user location based on the multiple nearest neighbour approach (finding the closest locations in the signal space and calculating the centroid). Another approach was based on the intersections of the lines of constant probability values inside the triangle and finding the centroid of these points as the location of a user. Both of these methods were very difficult to employ as they needed

additional data collection and/or exhaustive computations. Also both of them gave worse results, were much more computationally expensive and required heavy data collection.

In the same manner, similar equations can be derived for the other two pairs of CPs: $\overrightarrow{CP}_1$ and $\overrightarrow{CP}_3$ and $\overrightarrow{CP}_2$ and $\overrightarrow{CP}_3$. To find the position of the user, a solution that satisfies both equations 6.7 and 6.8 needs to be found. Based on one RSSI observation, $\mathbf{o}$, $\mathcal{L}_{CP_1}(r_1, \theta)$ and $\mathcal{L}_{CP_2}(r_2, \phi)$ can be calculated for $\overrightarrow{CP}_1$ and $\overrightarrow{CP}_2$ respectively. This system will be transformed until two nonlinear equations with two unknowns are obtained. Based on the geometry of the triangle shown in figure 6.3, the following equations for the distances r_1 and r_2 can be derived.

$$\tan\theta = \frac{h}{a_1} \tag{6.9}$$

$$\tan\phi = \frac{h}{a_2} \tag{6.10}$$

$$a = a_1 + a_2 \tag{6.11}$$

$$\frac{\tan\theta}{\tan\phi} = \frac{a_2}{a_1} \tag{6.12}$$

Also one can derive the following:

$$a_1 = \frac{a}{(1 + \frac{\tan\theta}{\tan\phi})} \tag{6.13}$$

and similarly for a_2:

$$a_2 = \frac{a\tan\theta}{\tan\phi(1 + \frac{\tan\theta}{\tan\phi})} \tag{6.14}$$

Also it can be put

$$a_1 = r_1\cos\theta \tag{6.15}$$

and

$$a_2 = r_2 \cos\phi \tag{6.16}$$

Eventually r_1 and r_2 can be expressed as functions of θ and ϕ only:

$$r_1 = \frac{a}{\cos\theta(1 + \frac{\tan\theta}{\tan\phi})} = \frac{a \sin\phi}{(\sin\phi\cos\theta + \sin\theta\cos\phi)} \tag{6.17}$$

and similarly for the r_2 in equation 6.18

$$r_2 = \frac{a \sin\theta}{(\sin\phi\cos\theta + \sin\theta\cos\phi)} \tag{6.18}$$

Functions shown in equations 6.19 and 6.20 depend only on r_1 and θ for the $\overrightarrow{CP}_1$ and on r_2 and ϕ for the $\overrightarrow{CP}_2$.

$$\mathcal{L}_{CP_1}(r_1, \theta) = (-k_1 r_1 + n_1) + (-k_3 r_1 + n_3 + k_1 r_1 - n_1)\frac{\theta}{\alpha} \tag{6.19}$$

$$\mathcal{L}_{CP_2}(r_2, \phi) = (-k_2 r_2 + n_2) + (-k_4 r_2 + n_4 + k_2 r_2 - n_2)\frac{\phi}{\beta} \tag{6.20}$$

where α and β are fixed given angle values of the triangle. Let r_1 and r_2 in equations 6.19 and 6.20 be replaced with equations 6.17 and 6.18 respectively. Now they depend only on the two angles as given in equations 6.21 and 6.22.

$$\mathcal{L}_{CP_1}(\theta, \phi) = \frac{a \sin\phi((k_1 - k_3)\theta - k_1\alpha)}{(\sin\phi\cos\theta + \sin\theta\cos\phi)\alpha} + \frac{(n_3 - n_1)\theta}{\alpha} + n_1 \tag{6.21}$$

$$\mathcal{L}_{CP_2}(\theta, \phi) = \frac{a \sin\theta((k_2 - k_4)\phi - k_2\beta)}{(\sin\phi\cos\theta + \sin\theta\cos\phi)\beta} + \frac{(n_4 - n_2)\phi}{\beta} + n_2 \tag{6.22}$$

At a specific location inside the triangle these two functions, $\mathcal{L}_{CP_1}(r_1, \theta)$ and $\mathcal{L}_{CP_2}(r_2, \phi)$, become two positive real numbers less than one, $\mathcal{L}^*_{CP_1}$ and $\mathcal{L}^*_{CP_2}$, respectively and thus equations 6.21 and 6.22 become a system of two equations with two unknowns.

This system of two non-linear equations with two unknown angles θ and ϕ is solved iteratively using (N)solve commands in Mathematica 7.0 software. One can notice that the

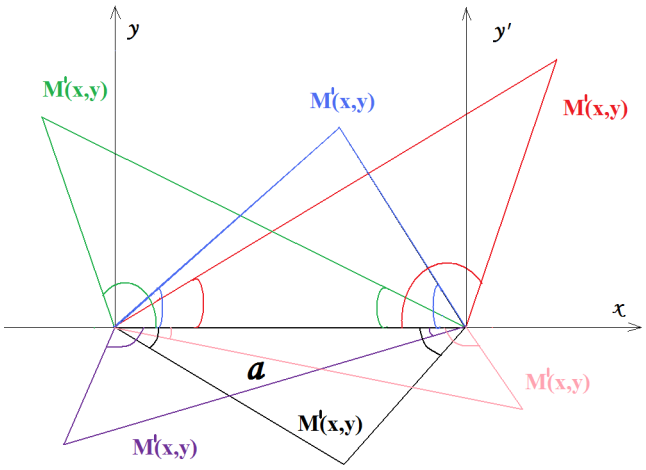


Figure 6.4: Six different cases of a user estimate each painted in different color

functions are not defined for $\theta = 0\,^\circ$ and $\phi = 0\,^\circ$ and no location output can be obtained. However, in practice, RSSI values don't change as a user moves slowly in the near vicinity, so values of the likelihood function don't change either thus making localisation error so small that it eventually does not matter whether a user is positioned somewhere on the line $\overrightarrow{CP}_1 \rightarrow \overrightarrow{CP}_2$ or in its immediate surroundings. Only the pair of solutions with real physical meaning was considered (both θ and ϕ have the same sign and intersect at a point i.e. for $\theta \neq 0\,^\circ$, $\phi \neq 0\,^\circ$, $180\,^\circ - |\theta| > |\phi|$ needs to be satisfied). In case there were two meaningful pairs, their centroid was taken to represent the location of the user. Output values outside of the triangle are also observed as valid estimates of the user position. This happens more often near the sides of the triangle, as algorithms are not precise enough to locate the user within the triangle. Generally, 6 different scenarios could be identified. These scenarios are shown in figure 6.4 each illustrated in a different color.

The estimated location of a user that was output by the algorithm is denoted by $M'(x, y)$, the edge of the triangle is denoted by a. The left angle is denoted by θ and the right one is denoted by ϕ (for all scenarios). In the first scenario (shown in blue) both angles are between $0\,^\circ$ and $90\,^\circ$; the opposite situation is shown in black and depicts the case when both angles are negative (between $0\,^\circ$ and $-90\,^\circ$). The situation when $0\,^\circ \leq \theta \leq 90\,^\circ$ and $90\,^\circ \leq \phi \leq 180\,^\circ$ is shown in red and similarly the opposite situation ($-90\,^\circ \leq \theta \leq 0\,^\circ$

and $-180^\circ \leq \phi \leq -90^\circ$) is presented in pink. Eventually one can see that the situation represented in green is when $90^\circ \leq \theta \leq 180^\circ$ and $0^\circ \leq \phi \leq 90^\circ$ while the opposite one ($-180^\circ \leq \theta \leq -90^\circ$ and $-90^\circ \leq \phi \leq 0^\circ$) is shown in purple. The vector approach is introduced to uniquely describe user position taking all these different scenarios into account. Based on the notations in figure 6.3, let $\vec{e}_1$, $\vec{e}_2$, $m, n \in \Re$, denote the unit vector of $\vec{r}_1$, the unit vector of $\vec{r}_2$, and the corresponding real coefficients respectively. The following equations can be written in the Cartesian coordinate system with the origin at $\overrightarrow{CP}_1$.

$$\vec{r}_1 = m\vec{e}_1 \tag{6.23}$$

$$\vec{r}_1 = \vec{a} + n\vec{e}_2 \tag{6.24}$$

If two vectors are equal then their corresponding scalar components, x and y, are equal as well (equation 6.25 and equation 6.26).

$$me_{1x} = a_x + ne_{2x} \tag{6.25}$$

$$me_{1y} = a_y + ne_{2y} \tag{6.26}$$

Solving this system one can find the values of m and n from equations 6.27 and 6.28:

$$n = \frac{a_x e_{1y}}{(e_{2y}e_{1x} - e_{2x}e_{1y})} \tag{6.27}$$

$$m = \frac{(a_x + ne_{2x})}{e_{1x}} \tag{6.28}$$

Based on the values of θ, ϕ, m and n one can determine the user position. The same process can be applied for the other two sides of the triangle separately and their location estimates obtained. This would give the same equations for the corresponding CPs ($\overrightarrow{CP}_1$ and $\overrightarrow{CP}_3$ & $\overrightarrow{CP}_2$ and $\overrightarrow{CP}_3$). Then the location of the user is found as a centroid of all three, or any two or just one side of the triangle. In case there is no meaningful solution of

the system of equations 6.21 and 6.21 for one side, the location of the user is found as the centroid of meaningful solutions obtained when using the other two sides or even only one side. In the worst case when there are no meaningful solutions, the user location was found using the weighted centroid approach described in equation 6.29.

$$\frac{\sum_{i=1}^{3} \mathcal{L}_{CP_i}(\vec{r})\overrightarrow{CP}_i}{\sum_{i=1}^{3} \mathcal{L}_{CP_i}(\vec{r})} \tag{6.29}$$

The last case describes the situation when the top 3 CPs formed the triangle in which one of the angles was greater than $90\,^{\circ}$. The user position was determined based on one, out of two methods, that gives the minimal distance error: the one given above applied on the side directly across from the angle greater than $90\,^{\circ}$ **only** and the weighted centroid approach described in equation 6.29.

6.3 Linearity model for likelihood function

The proposed localisation algorithm outlined in section 6.2.2 is predicated on the assumption that the likelihood functions $\mathcal{L}_{CP_i}(\vec{r})$ can be approximated as varying *linearly* along a path from one CP to another. In this section a series of experiments which examine to what extent this is true is described. A calibration point $\overrightarrow{CP}_i$ was chosen and a number of test points identified at equal intervals ($1.8m$) along a straight line emanating from the $\overrightarrow{CP}_i$. At each test point $\vec{r}$ a signal strength observation with an arbitrary orientation was collected, denoted by $\mathbf{o}$, and these data were used to compute a value for $\mathcal{L}_{CP_i}(\vec{r})$ as per equation (6.1).

This experiment was performed a number of times for test points on lines in different directions from the $\overrightarrow{CP}_i$ and for 15 different CPs as well. Data were collected at different times of the day to ensure a challenging data-set for the algorithms. These likelihood values for each testing point are plotted with blue asterisks in figures 6.5 and 6.6 as well as various lines of the best fit through the values. The best optimal curve with the least error was chosen to fit the data. In general, for a given set of values (x_i, y_i), for $i = 1..n$, we would like to find an optimal curve (equation 6.30) that would give minimal square error.

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... + a_j x^j = a_0 + \sum_{k=1}^{j} a_k x^k \tag{6.30}$$

This square error (err) can be calculated as in equation 6.31:

$$err = \sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right)^2 \tag{6.31}$$

and its minimal value can be found solving the following system of equations 6.32:

$$\begin{aligned}
&\tfrac{\partial err}{\partial a_0} = -2 \textstyle\sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right) = 0 \\
&\tfrac{\partial err}{\partial a_1} = -2 \textstyle\sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right) x_i = 0 \\
&\tfrac{\partial err}{\partial a_2} = -2 \textstyle\sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right) x_i^2 = 0 \\
&\vdots \\
&\tfrac{\partial err}{\partial a_j} = -2 \textstyle\sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right) x_i^j = 0
\end{aligned} \tag{6.32}$$

We find coefficients of the polynomial: a_0, a_1,... and a_j using matrix inversion $X = A^{-1}B$ of the equation $AX = B$ where matrix A is given in equation 6.33,

$$A = \begin{bmatrix} n & \sum x_i & \sum x_i^2 & ... & \sum x_i^j \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & ... & \sum x_i^{j+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^j & \sum x_i^{j+1} & \sum x_i^{j+2} & ... & \sum x_i^{j+j} \end{bmatrix} \tag{6.33}$$

matrix X in equation 6.34,

$$X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_j \end{bmatrix} \tag{6.34}$$

and matrix B in equation 6.35

$$B = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \\ \sum \left(x_i^2 y_i\right) \\ \vdots \\ \sum \left(x_i^j y_i\right) \end{bmatrix} \tag{6.35}$$

The linear fit is illustrated in green, the quadratic in red, the cubic in black and the quartic line in yellow. Note that the values of $\mathcal{L}_{CP_i}(\vec{r})$ are very small in practice and have been suitably normalised. In particular, figure 6.5 shows results for two CPs, $\overrightarrow{CP}_i$ and $\overrightarrow{CP}_j$, placed in two different rooms, where all the test points were in the same room as their corresponding CP. One can see that a linear fit through the data achieves reasonable accuracy while having the advantage of being easy to deal with. The same conclusions were reached for data taken at different CPs, and/or in different directions and/or at different times.

Figure 6.6 shows data from a more complicated setup with data collected and processed in the same way. In this case not all the test points were in the same room with the CP $\overrightarrow{CP}_k$ and $\overrightarrow{CP}_l$. In figure 6.6(a) there was a wall (of width 15 cm) between the first and the second test point as well as between the third and the fourth points. In figure 6.6(b) there were walls between the first and the second test points and between the the fourth and the fifth. It is evident from the data that the values of $\mathcal{L}_{CP_k}(\vec{r})$ and $\mathcal{L}_{CP_l}(\vec{r})$ drop sharply as one passes through a wall. The same experiments when walls were positioned between other test points were also carried out. A linear fit still captures the general trend of the data. The same conclusions were reached for data taken at different CPs, and/or in different directions and/or at different times.

In total, 150 radial tests (in different directions) for the first and 150 for the second (more complicated) setup were performed. These analyses show that the linear interpolation model between two CPs (equation 6.2) approximates well the linear fit model for the distances between a specific CP and test points which would correspond to the distances between two CPs in equation 6.2. For each analyzed CP, it was noticed that the difference between $\mathcal{L}_{CP_i}(\vec{r})$ function values calculated using equation 6.1 and the linear interpolation model did

not exceed 27.81% of $\mathcal{L}_{CP_i}(\vec{r})$ function value for the first 7.5 meters from that CP (26.06% and 15.78% on average in total in case of with and without walls respectively) thus making this approximation useful for further analyses. The possibility of combining information from several CPs to generate higher order models for $\mathcal{L}_{CP_i}(\vec{r})$, and to create piece-wise models which account for the sharp drops experienced as one passes through a wall is an interesting one and will be pursued in future work.

6.4 Grid spacing analysis

In this subsection localisation results obtained using SEAMLOC as described in sections 5.3 and 6.2 are presented. An $I \times Jm^2$ rectangular grid (henceforth just grid) is defined as the spacing between CPs equal to I meters in the x direction and J meters in the y-direction. Tests were performed inside five big indoor spaces, where 6 different grids were used ($3 \times 4m^2$, $3 \times 5m^2$, $5 \times 5m^2$, $4 \times 6m^2$, $4 \times 8m^2$ and $5 \times 10m^2$). Each orientation of a CP (North, South, West and East) is represented with 150 RSSI observations (600 RSSI measurements per CP in total). Each grid was tested using 600 RSSI measurements and the results are presented in figure 6.7.

The accuracy is shown on the x-axis and the cumulative distribution function (CDF) for the accuracy on the y-axis. $CDF(a)$ for a given value of accuracy a is defined as the percentage of occurrences whose accuracy does not exceed the given value. In this work $0 \leq CDF(a) \leq 1$. Accuracy is defined as the absolute difference between the actual location of the user and location output obtained by SEAMLOC. One can notice that for smaller grids CDF increases steadily and reaches 1 more quickly. Also, for each grid, the accuracy is comparable with the size of that grid. As accuracy of a typical WLAN-based localisation system is around $2-3$ meters, one should use either $3 \times 4m^2$ or $3 \times 5m^2$ grid to obtain similar performance. Moreover the difference between performances for $3 \times 4m^2$ and $3 \times 5m^2$ grids is very small but a larger area can be covered by $3 \times 5m^2$ grid thus making it preferable in further analyses. In table 6.1 one can see how the average accuracy depended on the grid size. A description of the competing approaches is given in section 6.6.2.

Grid (m^2)	3×4	3×5	5×5	4×6	4×8	5×10
$SEAM$	1.7524	1.8901	3.2267	3.4942	5.1131	6.4055
HOR	2.1326	2.2851	3.7295	3.6255	5.9525	7.5149
COM	2.4956	2.6125	3.6932	3.8922	6.8473	8.4154
RAD	2.8425	2.9522	3.3843	4.1093	7.2473	8.3554

Table 6.1: Average accuracy (in meters) vs grid size. SEAMLOC ($SEAM$) was compared with 3 well-known methods: HORUS (HOR) [180], COMPASS (COM) [91] and RADAR (RAD) [19]

Grid (m^2)	3×4	3×5	5×5	4×6	4×8	5×10
$SEAM$	0.1741	0.1784	0.2313	0.2911	0.4815	0.7311
HOR	0.2015	0.2111	0.3001	0.3131	0.6134	0.9114
COM	0.2113	0.2515	0.2913	0.3317	0.7312	1.1058
RAD	0.2225	0.2712	0.3131	0.3662	0.8411	1.1041

Table 6.2: Standard deviation of the average accuracy (in meters) vs grid size. SEAMLOC ($SEAM$) was compared with 3 well-known methods: HORUS (HOR) [180], COMPASS (COM) [91] and RADAR (RAD) [19]

Also table 6.2 shows how the standard deviation depended on the grid size.

6.5 Experimental setup when localising the user between calibration points

Two distinct experimental setups (ESs) were considered in order to validate the ideas presented in section (6.2). The first setup, ES_1, was the less challenging of the two. A relatively open-plan indoor environment was chosen with few walls and a lot of objects (figure 6.8). 60 CPs were distributed along an even grid covering $834.08m^2$. All offices and corridors were used except the two small ones on the very left side of the figure.

The second experimental setup , ES_2 was designed to be more challenging. It was carried out on the second floor of the Dublin City University Computer Applications building. Personal offices, of average size $8.9m^2$ were available for use (see figure 6.9). Every second office was chosen to host two CPs (placed in the left-most corners), yielding a total of 32 CPs, with one office left empty (containing no CPs) between each two that were sampled.

Thus in total there were 92 CPs. The wall thickness between offices was 15 cm. To test the algorithm a user was allowed to move freely through all three experimental setups (including the offices with no CP information), collecting data which were used to locate him. WLAN data collection processes is described in section A.2 of the Appendix.

Each CP is represented using 600 RSSI observations from all four orientations taken with a laptop using InSSIDer-based software[1] . The application records RSSIs from all confident APs. One observation consists of received signal strengths from all confident APs (in the best case the total number of APs: 17 and 14 in ES_1 and ES_2 respectively). In the testing phase, using one signal strength observation, $\mathbf{o}$, and arbitrary user orientation, the user can be located between CPs. A total of 500 testing observations were examined for each ES. Also $3 \times 5m^2$ grid was used in ES_1 and ES_2. The experiments were conducted at times of considerable levels of human presence throughout the environment. A number of separate investigations were performed in respect of the proposed algorithm:

- Number of sides of the triangle used in the localisation process.
- Comparison with other methods.
- Number of APs that can be sensed throughout the environment.
- Number of training and testing data.

6.6 Localisation between calibration points: results

6.6.1 Effect of number of sides of the triangle used

The SEAMLOC approach (denoted by $SEAM$ as well) of section (6.2) takes the three top-ranked CPs. These will usually form a triangle but can be collinear. Using one side of the triangle only allows us to interpolate and locate the user at some intermediate point (within or outside the triangle). There are three choices one could make to do this. Intuitively,

[1] http://www.metageek.net/products/inssider/

Method	$SEAM$	$SEAM_{ab}$	$SEAM_{az}$	$SEAM_{bz}$	$SEAM_a$	$SEAM_b$	$SEAM_z$
ES_1	1.7942	2.2180	2.3059	2.3936	2.5491	2.7851	2.8822
ES_2	2.4081	2.8401	2.9726	3.1282	3.3972	3.4711	3.5612

Table 6.3: Average accuracy (in meters) in ES_1 and ES_2 respectively using: all 3 sides, different combinations of the 2 sides and 1 single side of the triangle. **Method** denotes which combination of sides of the triangle was used.

using all three sides of the triangle and then choosing the centroid of the resultant three estimates would offer the best accuracy and this is examined in this section.

In what follows the term $SEAM_a$ denotes an estimate based on using side a of the triangle only, $SEAM_{ab}$ denotes an estimate based on taking the average of that given by $SEAM_a$ and $SEAM_b$ while $SEAM$ denotes the centroid of the three estimates $SEAM_a, SEAM_b$ and $SEAM_z$. The average accuracy achieved using these approaches, for both experimental setups is shown in table 6.3. The data suggest that the accuracy increases as one combines more interpolated estimates together. In order to give a fuller picture of the experimental statistics figure 6.10 shows cumulative distribution functions for the accuracy for the various scenarios. The first plot shows the data for ES_1 while the second shows the data for the more challenging ES_2.

6.6.2 Comparison with other methods

SEAMLOC is compared with 3 well-known methods: HORUS (HOR) [180], COMPASS (COM) [91] and RADAR (RAD) [19] (see figure 6.11). For RADAR the Multiple Nearest Neighbours method i.e. the average of the k-nearest neighbours approach is used to obtain the user position. Various values of k are considered; the k that gives the best accuracy ($k = 3$) is chosen.

HORUS is implemented to maximise the system of equations given by equation 6.1. There is a degree of correlation between signal strengths from the same AP and to obtain the user position an autoregressive model is used. It can capture the correlation between different RSSI values from the same AP. As in [180], if the distribution of RSSI values is

Approach	$SEAM$	HOR	COM	RAD
ES_1	1.7942	2.3048	2.6133	3.3030
ES_2	2.4081	2.9542	3.2717	3.3844

Table 6.4: Average accuracy (in meters) in ES_1 and ES_2 using SEAMLOC, HORUS, COMPASS and RADAR

Gaussian with mean value μ and the variance σ^2, then for the average of n correlated RSSI values the equivalent variance is given by equation 6.36. The value of α is estimated to give a realistic value of the autocorrelation between the values. Two techniques are used to estimate the position of a user: the center of mass technique (weighted k-means average approach) and the time-averaging technique. The following parameters (the meaning of the parameters is given in [180]) are used in ES_1 and ES_2 respectively: $\alpha = 0.76$, $N = 4$, $W = 6$ and $\alpha = 0.66$, $N = 4$, $W = 6$. The time-averaging technique gives higher accuracy.

$$\frac{(1+\alpha)\sigma^2}{(1-\alpha)} \tag{6.36}$$

The COMPASS method is implemented with 4 orientations. A weighted k-means approach is employed to find the user position. To compare SEAMLOC with HORUS, COMPASS and RADAR, 500 RSSI observations are used for each ES.

These competing approaches are implemented using on-line implementations of RADAR [19] , HORUS [180] and COMPASS [91]. Their original implementations are kept the same regardless of the distance between the nearest CPs. Also all the methods including SEAMLOC were tested in the same way, using the same training and testing dataset available. In the testing phase, in all four approaches, one signal strength observation was used and localisation accuracy was obtained.

From figure 6.11 it is clear that SEAMLOC outperforms all other methods. In ES_2 the difference between these accuracies is smaller as there exists more complex environment which degrades the performance of SEAMLOC (i.e. making it more equal to the other three methods). In table 6.4 one can see achieved results for both ESs.

6.6.3 Effect of number of APs on accuracy

In figure 6.12 one can see how the average accuracy changes when the number of available APs decreases. Let the maximum number of available APs be denoted by N_a. For a small number of available APs k, $(k < N_a)$, p different k-combinations are found (here $p = 7$). Each combination has k APs and for each combination the accuracy is calculated. Eventually, for that specific k the average accuracy is obtained. SEAMLOC clearly outperforms all other methods. Already three APs, for all localisation methods, in both ESs, give much better localisation results. It is also interesting to notice, in both ESs, that the average accuracy does not drop significantly when using more than 8 APs. This result is useful meaning that with lower budget and quicker installments reasonably accurate performance can be achieved.

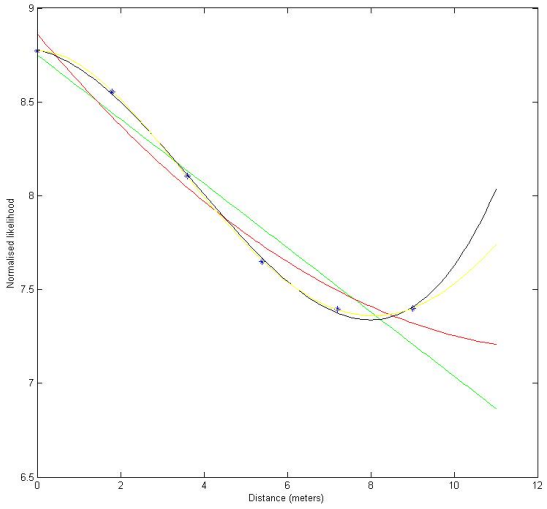
6.6.4 Effect of amount of training and testing data on accuracy

Results that show how the average accuracy changes if the amount of training and testing data is changed are presented in figures 6.13 and 6.14 respectively. The same number of RSSI observations is used in both ESs. For different training/testing datasets that have the same number of RSSI observations the accuracies were calculated, and the average accuracy for that specific number of RSSI observations is plotted.

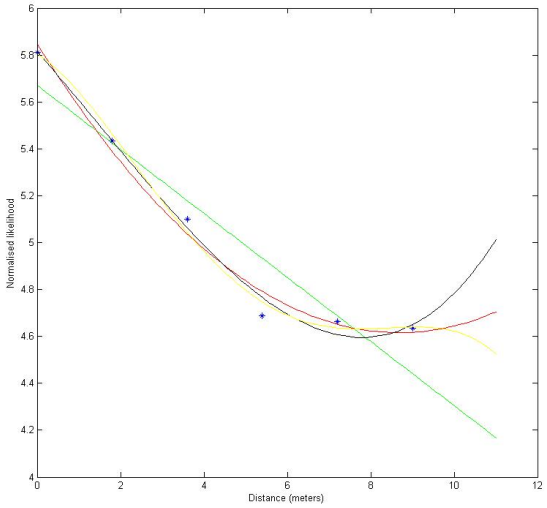
In both ESs the average accuracy becomes almost constant when the amount of the training data reaches 500 RSSI observations. Clearly, SEAMLOC again shows better performance over the other three methods. Thus, the calibration phase can be done more quickly to obtain good performance. In the case when the amount of testing data is changed, the average accuracy remains almost constant (with small variations), in both ESs, thus proving the reliability of SEAMLOC with a smaller number of tests.

6.7 Conclusions

In this work, a novel approach to WLAN-based indoor localisation is described and results are presented for two different environmental settings. Results of comparisons are also presented between this and other localisation methods demonstrating its robustness and improved performance over others in terms of accuracy. Moreover, fewer CPs than in other methods are used thus making it practical and easy to deploy. Also compared to results of RADAR, HORUS and COMPASS, SEAMLOC shows greater robustness to noise outperforming other three competitive methods significantly. As it can be seen form the results SEAMLOC shows good accuracy when using bigger grid size thus becoming useful in case of using fewer CPs and large coverage. The algorithms are stable when using more than 500 training observations, robust to the change of testing data and can be used in case of smaller number of APs available based on a trade-off between hardware requirements and needed accuracy. Also it is interesting to notice that accuracy obtained when using two sides of the triangle in SEAMLOC can be also useful as it also can outperform other three methods and moreover is slightly faster to compute. This reduces the calibration time while the processing time is very similar to the other approaches. The work given here can be extended to possibly obtain better accuracy and greater robustness. The possibility of seamlessly tracking a user indoors using Kalman filters, using different device(s) in training and testing phase and eventually employing more sophisticated classifiers such as SVM or neural networks, are all clear targets for future work in this area.

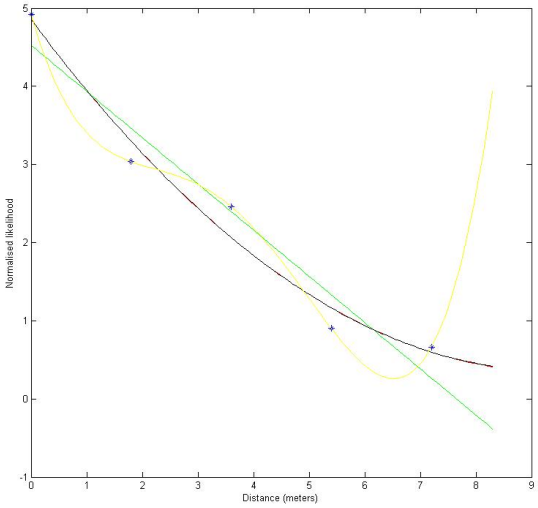


(a)

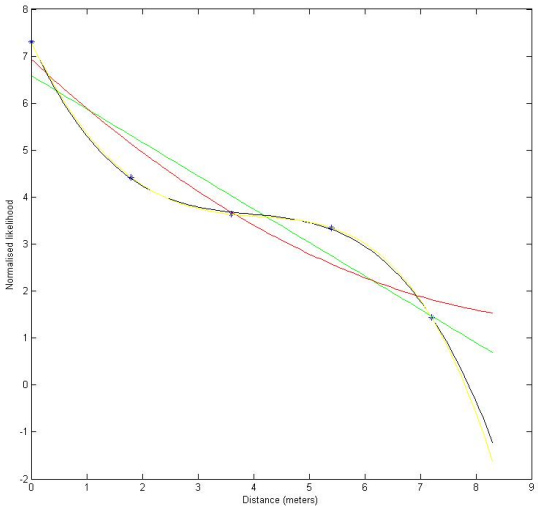


(b)

Figure 6.5: (a) Likelihood function $\mathcal{L}_{CP_i}(\vec{r})$. On the x-axis: distance from $\overrightarrow{CP}_i$ given in meters. On the y-axis: normalised $\mathcal{L}_{CP_i}(\vec{r})$ values (b) The same graph for $\mathcal{L}_{CP_j}(\vec{r})$ function



(a)



(b)

Figure 6.6: (a) Normalised $\mathcal{L}_{CP_k}(\vec{r})$ function values in the case of walls between test points; (b) The same graph for the $\mathcal{L}_{CP_l}(\vec{r})$ function. In both examples the likelihood function drops rapidly

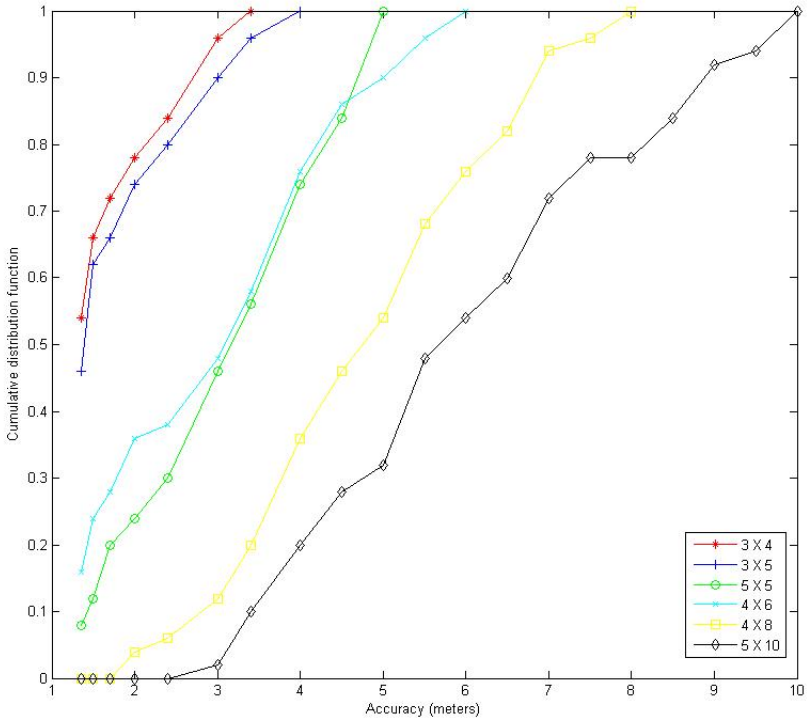


Figure 6.7: Cumulative distribution functions for the accuracy using SEAMLOC

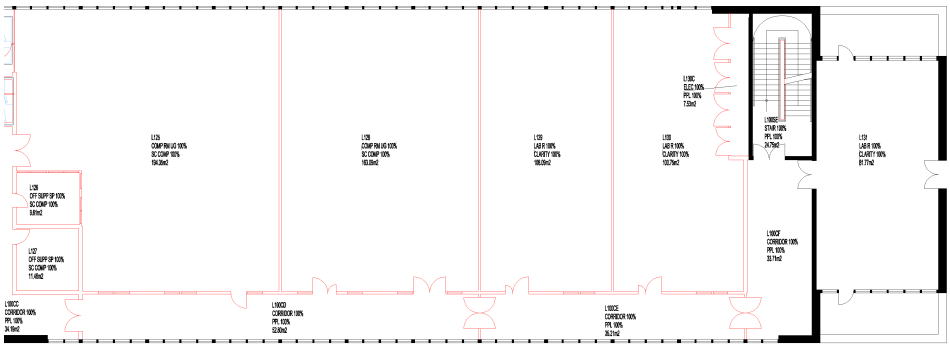
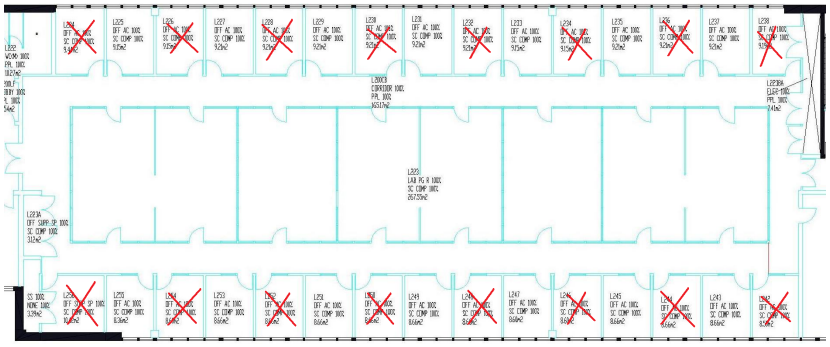
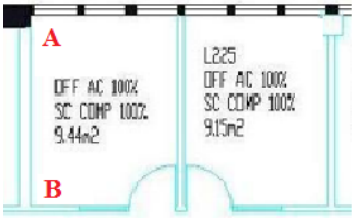


Figure 6.8: The offices where the measurements were taken for ES_1.

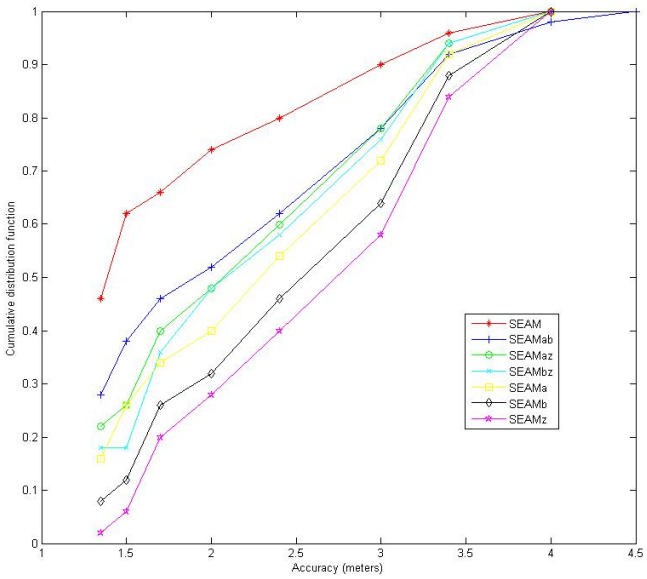


(a)

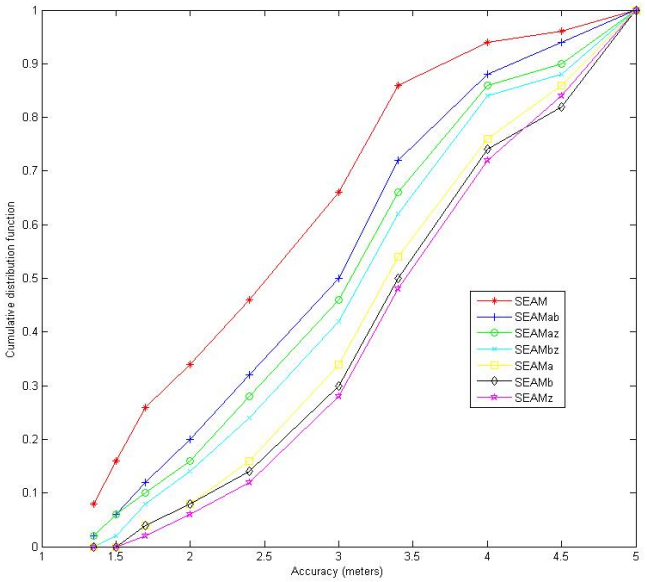


(b)

Figure 6.9: (a) Map of office locations used in ES_2: red crosses indicate offices used; (b) Examples of CPs A and B placed in a used office. Next to this office there is an empty office

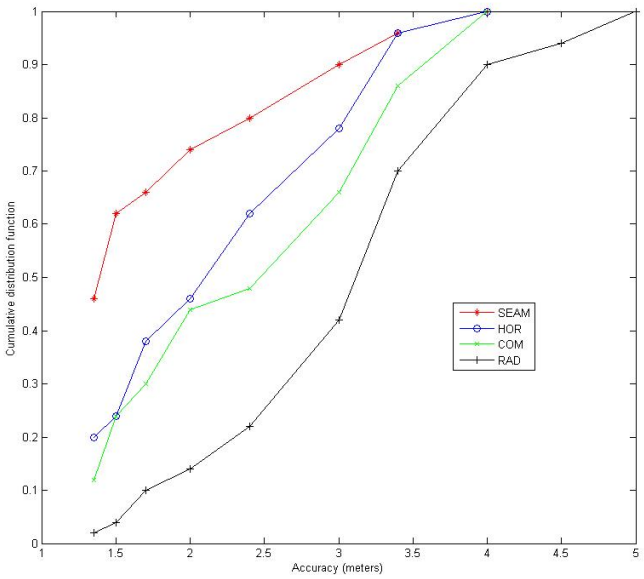


(a)

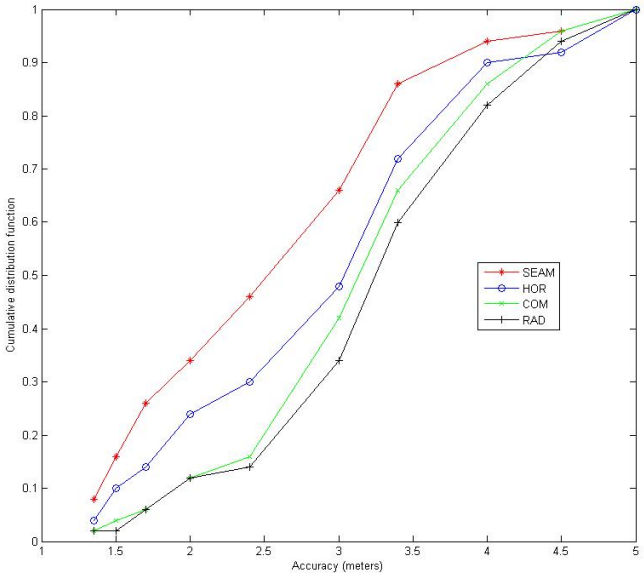


(b)

Figure 6.10: (a) Cumulative distribution functions for the accuracy in ES_1; (b) Cumulative distribution functions for the accuracy in ES_2

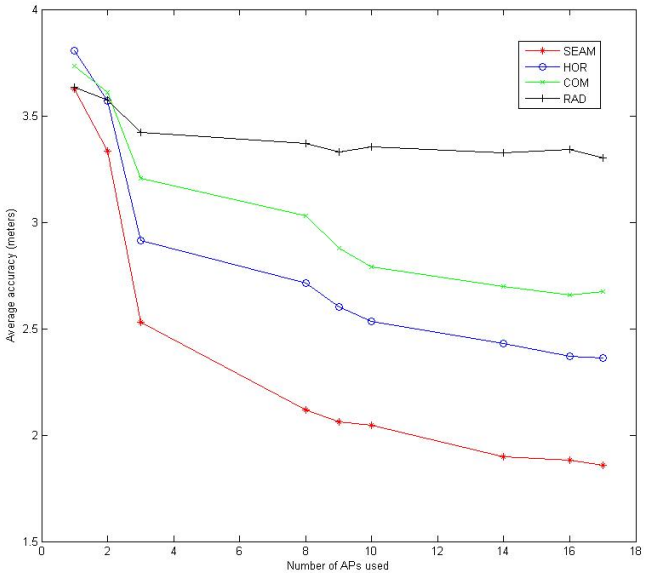


(a)

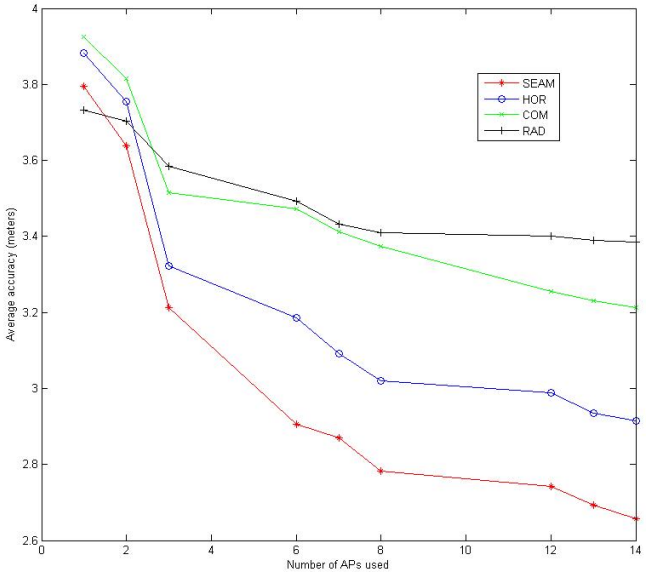


(b)

Figure 6.11: (a) Cumulative distribution functions for the accuracy in ES_1; (b) Cumulative distribution functions for the accuracy in ES_2

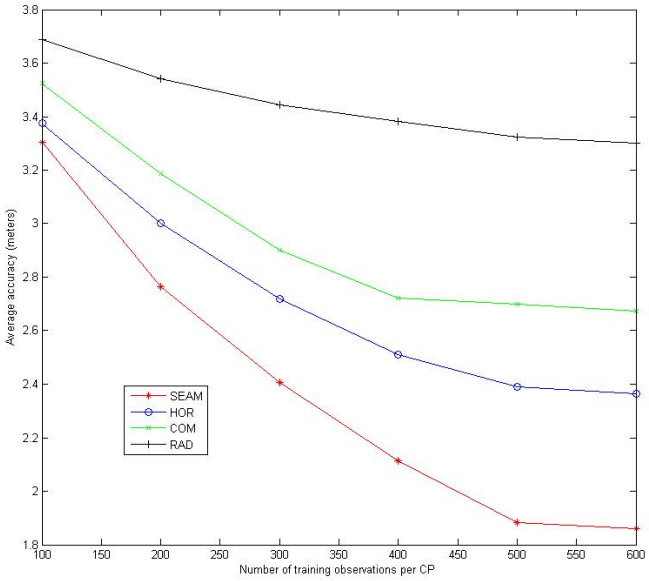


(a)

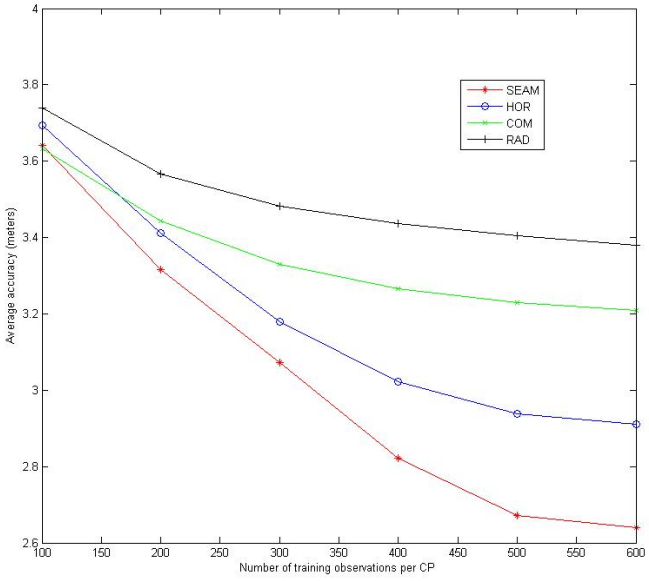


(b)

Figure 6.12: (a) Average accuracy (in meters) vs. number of APs in ES_1; (b) Average accuracy (in meters) vs. number of APs in ES_2

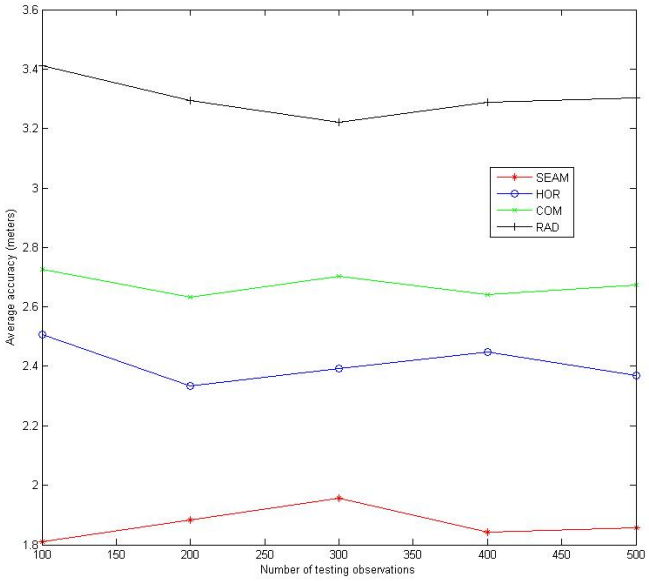


(a)

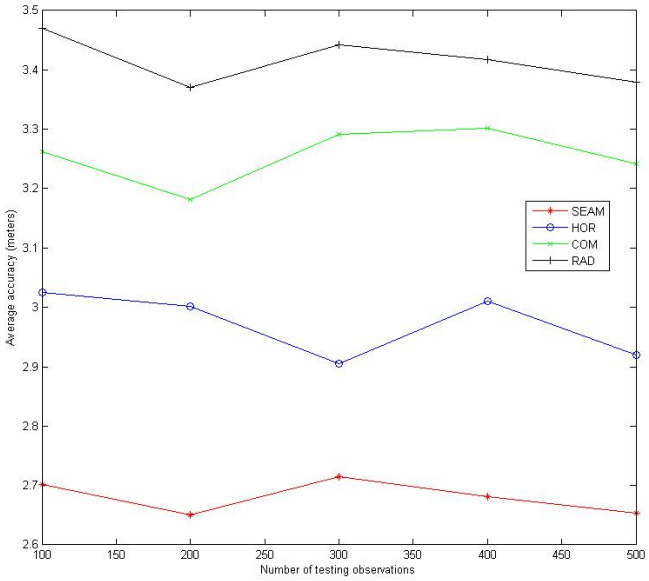


(b)

Figure 6.13: (a) Average accuracy (in meters) vs. number of RSSI training observations in ES_1; (b) Average accuracy (in meters) vs. number of RSSI training observations in ES_2



(a)



(b)

Figure 6.14: (a) Average accuracy (in meters) vs. number of RSSI testing observations in ES_1; (b) Average accuracy (in meters) vs. number of RSSI testing observations in ES_2

Chapter 7

Conclusions

Indoor localisation is still an unsolved problem. Many approaches have been developed to resolve this problem. Fusion approaches that were used to fuse two or more sensing modalities, e.g. image and RF data have been investigated. In this thesis the fusion of two complementary and different modalities is presented and as such represents the first localisation work that deals with these two technologies fused using a late fusion process. This can be regarded as a key novelty in the work presented.

There are four contributions in the technical work of this thesis. The first contribution of this thesis lies in building an actual fusion function for the two sensing modalities. The function shows great robustness in terms of variations of the both WLAN and image data. Its thresholds can be tuned and synchronized for a given application so different value of fusion performance can be obtained. Although it might appear that the improvement the fusion function gives is only incremental, it is important when any of the modalities fails (e.g. WLAN breaks down or if there exist significant occlusion in the images). The performance variation for the localisation to within an office obtained by using a variable number of CPs gives an interesting conclusion: whilst the best results are obtained by using 5 calibration points for each office, one can see that using only one calibration point produces reasonably good performance. The proposed fusion technique outperforms the other early and concept fusion techniques it was compared with. Furthermore, as the method does not depend

on the actual technology it would feasibly be employed when fusing some other sensing modalities thus broadening its scope of applications.

The second contribution is based on seamless indoor localisation of a user anywhere in space between preselected calibration points. A novel scaled probability approach based on a reduced number of selected CPs is presented. The approach is based on robust scaled likelihood functions that depend on angle and distance using linear interpolations. The location of a user is found by solving a system of two non-linear equations with two unknowns derived for a pair of CPs out of three that form a triangle. The centroid of the three solutions represents the deemed location of a user. The method was tested on very challenging data in a variety of conditions. It was compared with well-know localisation methods and demonstrated improved performance. Less number of CPs than in other methods were used thus making it very practical and easy to employ. This reduces the calibration time while the processing time is very similar to other approaches. Moreover, as the performance depends on frequency of CPs placed in the grid; accuracy and precision can be tuned and configured according to user needs. Lower resolution for the experimental setup does not need extra processing and moreover can be done in a shorter amount of time. The potential usefulness of this approach, beside user localisation, is envisaged in a range of ambient assisted living applications. Based on the directions given in section 6 this approach may show even greater robustness using adaptive weights and/or using more sophisticated classifiers. The results so far have already given some promising results.

Two additional contributions are based on novel image-based localisation and a novel tracking algorithm. The third contribution extends hierarchical vocabulary tree approach and is based on robust tuning the cluster centers of the hierarchical vocabulary tree of the SURF feature descriptors. Cluster centers were calculated recursively using the previously calculated cluster centers. The results outperformed the standard vocabulary tree approach.

The last, fourth, contribution represents a simple Viterbi-based multiplestate model based on a simple Hidden Markov Model states. A novel transitional probability function converts times of locations visited into probabilities. Also it takes into account confidences

at specific locations when using any of the single modalities or fusion. The combination of both sources increase the tracking performance, thus making the difference between them notably reduced.

The methods proposed in this thesis move on from the current state of the art in several directions. First, the thesis proposes an image-based localisation method that outperforms standard hierarchical vocabulary tree approach by using a fine tuning of the cluster centers. Also, the thesis proposes more robust and more accurate fusion function that better fuses WLAN and image data and is used in localisation and tracking. It outperforms the previous works in both localisation and tracking. A novel tracking approach introduces function that converts times of visited locations into probabilities. So far only time analysis was used with a simple transformation function. The last but not the least is the algorithm for localising the user anywhere in space between CPs. The approach outperforms three other competitive methods, is fast to compute, and robust to changes in environment.

In the future work heterogeneous localisation can be performed using different devices, e.g. laptop and smartphone or tablet. This means using different devices in training and testing phase and building robust mapping functions that would reliably map image and WLAN data from one to another feature space or from the first to the second device. Thus, heterogeneous localisation would make this approach more robust and sought after. Another direction would be to use this approach together with other sensing modalities such as audio and GPS. Audio would introduce more descriptive component into the system and aid to hearing impaired people. GPS can be used to make this localisation system seamless outdoor as well. Thus, it can be used when GPS is weak or unreliable outdoors (tunnels and urban canyons) and switching to other two sensing modalities. All these sensing modalities are nowadays commonly available on any smartphone device. Also in the future work two classification methods will be analyzed: Support Vector Machines (SVMs) and neural networks. They tend to perform better with continuous and multidimensional features but performance also depends on the dataset type as well. For neural network models and SVMs, a larger dataset size is required for achieving its maximal accuracy. Eventually,

adaptive and confidence-based weighting in the multisensor fusion can be used to achieve better accuracy and/or precision.

Support Vector Machines (SVMs) are relatively novel supervised machine learning technique. SVMs are based on a margin - either side of selected hyperplane which separates two data classes [33, 76]. Choosing the maximal margin and making the largest possible distance between the separating hyperplane and the instances on either side reduces an upper bound of the expected generalisation error. Assume some training data D, a set of n points of the form $(\mathbf{x}_i, c_i)$ where c_i is either 1 or -1, indicating the class to which the point $\mathbf{x}_i$ belongs. Each $\mathbf{x}_i$ is a p-dimensional vector. The main objective is to find the maximum-margin hyperplane that separates the points having $c_i = 1$ from those having $c_i = -1$. Training the SVM is done by solving an N dimensional QP problem, where N is the number of samples in the training dataset. As supervised machine learning techniques are applicable in numerous domains. A choice of algorithm always depends on the actual task [72, 48, 76].

An Artificial Neural Network (ANN) is a system that consists of many parts which are highly interconnected and work synchronously in order to solve specific problems. These neural networks can be used for detecting and localising patterns, usually when that cannot be achieved using other classification techniques [175, 127, 182]. Using first experience and/or training data ANNs are able to learn how to perform classification. As ANNs processings can be performed in parallel it significantly saves amount of time needed for such operations. A simple artificial neural network can be modelled as a device (neuron) that has many inputs and one single output (shown in fig 7.1) [115]. There are two working modes: the training and the using mode. For specific input values, in the training mode, the neuron can be trained to fire (or not).

When it is ready to be used and when a learned data pattern at the input is acknowledged, its associated output is now the present output. If the input data pattern is not recognized in the set of learned detected patterns then the system decides whether to fire. For every node there is a set of training instances; some of them can fire (the 1-taught set

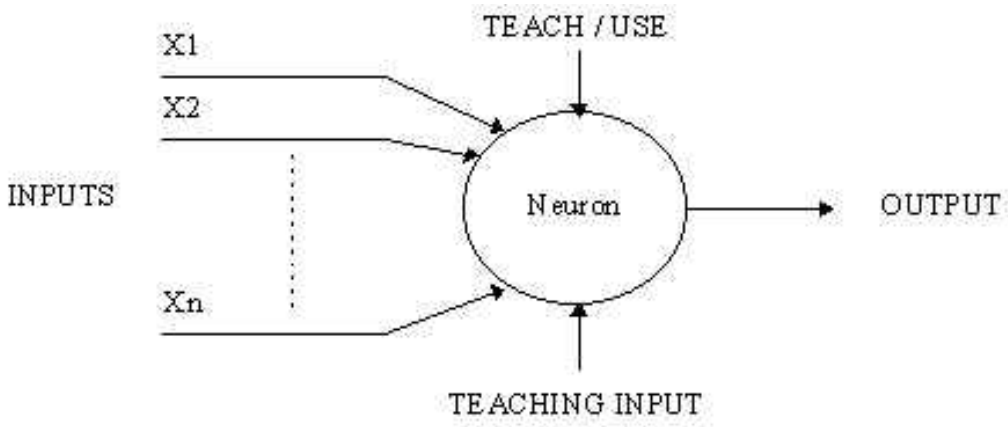


Figure 7.1: A simple neuron [182]

of patterns) while some others prevent from doing the same thing (the 0-taught set). After we receive new data patterns, they can fire if there are more common elements with data patterns in the 1-taught set than in the 0-taught set. If one can not decide the system is in so called *undefined* state. Therefore, the rule can respond meaningfully to data patterns which are not processed in the training part.

Multisensor fusion based on adaptive and confidence-based weighting is in general terms complex and difficult task. The current methods have shown problems in weight specification especially in query-based and in fusion approach where weights that connect different modalities are functions of different classes (e.g., users, tools, players, etc.). Many approaches use weights that were previously trained and calculated using large training datasets. The main problems are their poor performance and lack of scalability. Novel adaptive learning weighting framework for multisensor fusion is given to overcome these difficulties [99]. There is an adaptive scheme that learns on the fly all weights that are needed in fusion process. K-nearest neighbour approach selects dynamically all these weights. Its main characteristics are:

1) one doesn't need pre-defined classes

2) without fixing parameter values one can accurately learn query weights for multisensor fusion

3) the training examples are determined in classification process without noise and following the rules of semantics

Appendix A

WLAN and image data capture

A.1 Introduction

In this Appendix we provide additional technical detail on how WLAN and image data were captured and pre-processed for our experimental set-ups. The WLAN infrastructure is widely used to provide indoor user localisation without installing additional equipment. A wireless network interface card collects WLAN signals and acts as a sensor device. Initial data analysis is important for location fingerprinting. The properties of the received signal strength values (RSSI) are crucial to understand the characteristics of WLAN features in general. As they effectively depend on location, understanding their physical characteristics can have an important effect on the development of indoor localisation system.

A digital image is a numeric representation of a real world scene. Often a finite set of values is observed giving so called raster images. Its pixels are given as a two-dimensional array of small integers (so called raster map) thus making calculations very expensive. Users process raster images through many kinds of image formats. Raw image format enables all data to be used by the user (rather than losing some data due to compression) and these can be taken using some digital cameras. In this chapter devices and software used in WLAN and image data collection are presented. InSSIDer software installed on a laptop is used for WLAN RSSI data collection while the images are extracted from video using FFmpeg

software.

A.2 WLAN data collection

Initialization of a WLAN station starts when a device finds an AP that is available for connection. When the connection to that AP becomes unstable, the station must find another AP that can establish a more reliable connection. Scanning is defined as looking for an AP. One can differentiate active from passive scanning. In active scanning for each channel on which an AP operates, a probe request is sent by the station and the station waits until a response is received [114, 75]. Based on these probe responses, the station then decides the best AP to which it can connect. Every channel is listened on for beacons by the station. The beacons are sent from the transmitter regularly by access points [20]. An AP may take 100 ms to send a beacon while an AP responds to a probe request within 20 ms which makes active scanning more used and efficient than passive scanning. The number of APs identified before and reported after the scanning process varies due to the type of microprocessor on the host device, environmental interference, etc [117]. During the scanning process, a station can not receive or send any payload data. In this thesis active scanning was used to obtain RSSI values from the corresponding APs using InSSIDer software.

A.2.1 InSSIDer

InSSIDer is WLAN network scanner software for Microsoft Windows produced by MetaGeek, LLC [6]. Like NetStumbler, InSSIDer is free but unlike NetStumbler, it is open source software (Apache 2.0 license) and currently an alpha version of InSSIDer is available for installation on Linux. InSSIDer works very similarly to NetStumbler: the program immediately begins scanning for and displaying information on the WLANs it finds after it has been launched. InSSIDer uses the wireless network card to measure signal strength and channel selection of available wireless APs (its interface is shown in figure A.1). InSSIDer has a Channel Graph so one can see which APs are overlapping and causing interference

[6]. The Channel Graph also shows how neighboring channels overlap. InSSIDer can be updated to refresh its display at a certain speed i.e. sampling speed. Each display area shows available APs, their public name and related information [6]. The Time Chart plots time along the bottom and signal strength along the side. The Channel Graph shows available APs by channel along the bottom and the signal strength along the side. It uses the same negative vertical scale as the Time Chart. The Network Table shows available APs' details. Each AP has a different color that is used throughout the three displays. One can also exclude uninteresting APs from the charts as well [6]. Signal strength is a good wireless AP proximity indicator - signal strength increases as you get closer. Operating systems can function as APs and wireless clients simultaneously. Details about the APs detection using InSSIDer are displayed in several columns as explained below in table A.1.

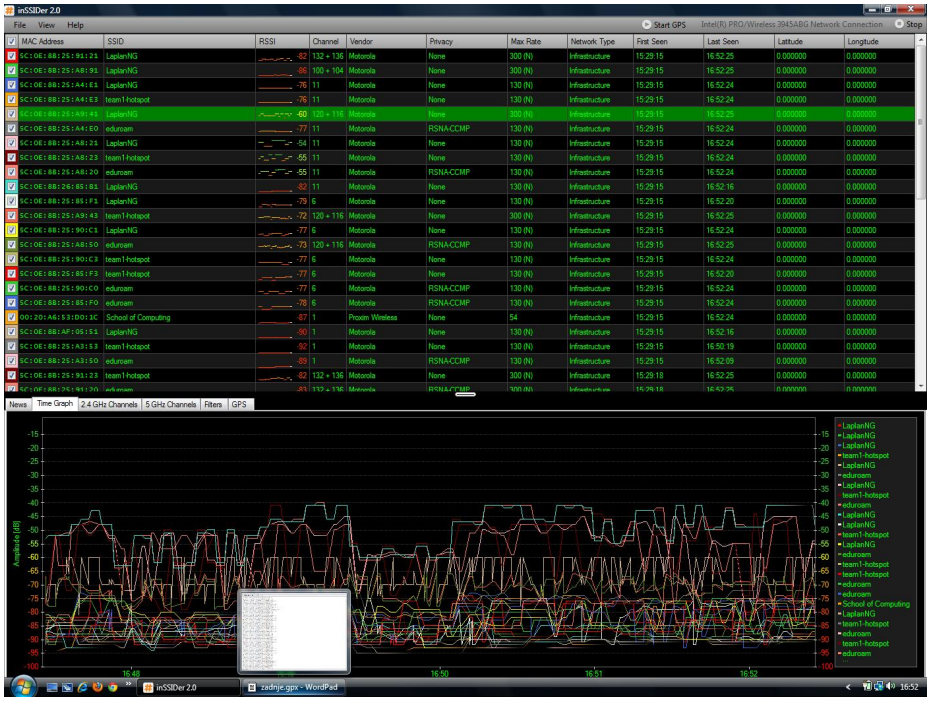


Figure A.1: InSSIDer interface

InSSIDer is straightforward to use. In terms of finding and reporting on existing WLANs in user vicinity, it vastly outperforms WLAN scanners [6].

Abbreviation	Description
MAC Address	the MAC address of the AP
SSID	the service set identifier or name of the WLAN
RSSI	the received signal strength indicator
Channel	the channel that the WLAN is operating on
Distributor	the manufacturer of the WLAN card
Privacy	the type of encryption the WLAN uses to secure its transmissions
Max Rate	the theoretical maximum data transmission rate for the WLAN
Network Type	only two choices here: infrastructure or ad hoc
First Seen	displays the time when InSSIDer first detected the WLAN
Last Seen	displays the time when InSSIDer last saw the WLAN
Latitude and Longitude	used in conjunction with InSSIDers GPS functionality

Table A.1: The functionality of InSSIDer that appears in the user interface

A.2.2 Scanning and RSSI properties

Localisation of a WLAN station is determined based on the RSSI of probe requests detected by an AP and forwarded via their registered WLAN controllers to the location appliance [136]. Regular and consistent probing of the network is important to obtain good WLAN location estimation. This process starts when clients (WLAN stations) issue probe requests to discover the existence of WLAN networks in their surrounding. An unassociated client may generate probe requests quite regularly, while clients that are associated to a network will issue probe requests less often [2].

Signal strengths are expressed in a negative value of signal loss and denoted with dBm. RSSIs are expressed as the absolute values of signal strengths and are expressed in dBm as well. The RSSIs were captured using InSSIDer software, together with their unique MAC address, signal to noise ratio (SNR), channel information and other information.

This step also includes filtering weak signals. In the case given here it consists of fingerprints whose RSSI values are greater than 80 dBm meaning that we have very weak signals that need to be removed from further analysis. It might be the case they will not appear in a scan at another given time. Also stronger signals whose RSSIs are lower than 70 dBm are most likely to be always present. Even then, at some measured location RSSI value can be bigger than 70 dBm thus meaning the signal is relatively weak. Because the

WLAN signal carrier is not consistent itself, it is hard to identify a location without a fixed signal strength. Signal strengths vary over time, caused by multipath effects, and these signal strengths themselves are not consistent either. However, by averaging the signal strength values, and by creating a search space (a range) around this average, it is possible to predict a location. Each location is represented with a series of fingerprints. The table of fingerprinting vectors at a location was built and an example is shown in figure A.2.

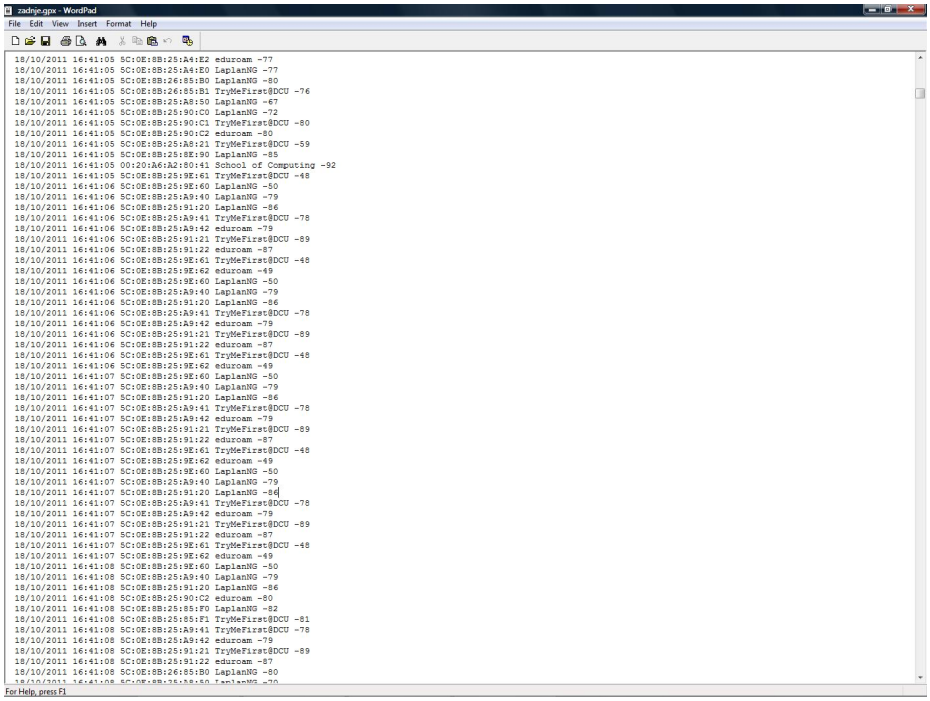


Figure A.2: Output file example

It consists of the MAC address of the network as well as the corresponding time stamp and RSSI value. The most important fields include the location ID from the table location, which is used in any other table. This coincides with the fingerprint, as the fingerprint is the node itself. There should be a reasonable amount of caution exercised when fingerprinting as the process assumes that RSSIs are dependent on physical objects in the space due to multipath and that they change over time. In addition, their transmitters, the routers or Media Access Control units (MAC), can be replaced as well. The uniqueness is necessary to distinguish APs with the same SSID, as otherwise measuring using SSIDs will result in

collated signal strengths. The InSSIDer's open source code was changed to collect RSSI signals at a fixed sampling rate. At each location, active scanning is carried out with sampling interval of a 1 second. Those locations were then tagged with location information, and merged into one file. The data were collected and the experiments were performed on Dell Inspiron laptop with Intel Core 2 Duo Processor T5250 (2.0 GHz, 2 MB L2 cache, 667 MHz FSB), memory of 2×2048 MB, 667 MHz Dual Channel DDR2 SDRAM, SATA Hard Drive with 450GB (5.400rpm) and Intel PRO/Wireless 3945ABG card. A 1 second interval showed good, stable, RSSI values as it can fluctuate in a short time. Signal strengths change between 39 dBm and 45 dBm, caused by both direct and indirect signals at the place of reception due to multipath effects. It also detects the stability of APs: at a given location, if at least 40% of total number of APs are present, this can be considered as relatively stable (see fig. A.3). This will eliminate unstable APs, which are usually outlying routers.

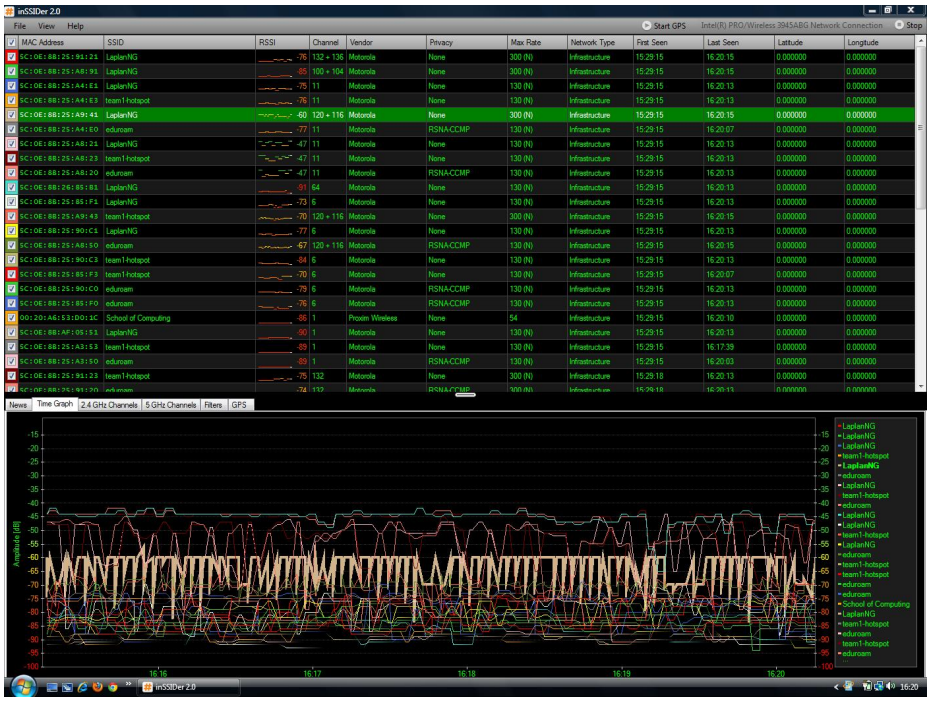


Figure A.3: Stable and unstable APs shown on InSSIDer interface

It has been seen that upon using recorded values from the 14 - 17 stable APs within reach, often false locations have been provided. Further conclusions can be drawn regarding

the localisation algorithm and the database.

A.3 Image acquisition

FFmpeg is a useful and reliable audio/video extractor and converter used to gather and process information and data from a live audio or video data source [4, 3, 5]. It successfully extracts from arbitrary sampling rates and is able to rescale video in real time. The program takes an arbitrary amount of files as input (streams of a network, grabbing devices, regular files, pipes, etc.), specified by the i value, and outputs an arbitrary amount of files specified by an output filename [4, 5]. Each input or output file can consist of an arbitrary number of different type of streams (video/audio/data/attachment/subtitle/). Using the -map function one can choose which streams from input should be forwarded to output. This can be done automatically as well. In the options, input files can be referred to using their indices. This means that by 0 we denote the the first input file, by 1 the second one etc. Thus, for example, $3:4$ denotes the fifth stream in the fourth input file. A general rule says that options are applied to the next file of interest. It highlightens the importance of the order and one can apply many times the same option. For each the process is repeated for the next input or output file. This is not the case with the global parameters given at the very beginning. The input and output files should not be mixed; firstly all input files should be specified and all output files afterwards. Between two consecutive input or output files the whole process is performed and but first a reset needs to be done.

Extracting images from a video is an automatic process in which frames are extracted first and afterwards transformed into images in a way that each image corresponds to particular frame. The number of frames per second can be controlled, as well as image format, the frame rate and in case user wants a particular image resolution, it can be done by changing the size of the frame which will be given below. The most general way of extraction images from a video [4, 3, 5] is given with:

[shredder12]$ ffmpeg -i inputfile.avi -r 1 -f image2 image-%kd.jpeg

-r defines now many frames will be extracted per second. Thus it sets video frame rate

(a) (b)

(c) (d)

Figure A.4: Examples of visual sensing given in (a),(b),(c) and (d); sample images collected of office spaces

where each frame represents an image. The default value is 25 but any number can be chosen. Here in this work $r = 2$ is used.

In the previous command one can define different flags. k defines image format used in the extraction process. If $k = 2$ then the images will be in the format e.g. image-01.jpeg, image-02.jpeg or if $k = 3$ it will be e.g. image-001.jpeg, image-002.jpeg. Here $k = 3$ is used.

With -s flag one sets the size of extracted image. By default the image corresponds to the video's resolution. The image resolution used and processed in this thesis is 640×480. One stream of every type only is added from the input file to the every output file. It can be either video, audio or subtitle [4]. For video the highest resolution stream is chosen. If there are more streams with the same type the one with the lowest numbered stream is chosen. In the input string the numerical options take a string that represents a number, which may contain one of the International System number postfixes. There are stream specifiers, options that are applied per-stream, that with great precision specify which stream(s) a

given option belongs to. A colon usually separates stream specifier from the option name. -codec:c:3 ac3 option contains c:3 stream specifier. If a stream spec. matches more streams, the process is done to all of them. An empty stream specifier can match every stream [4, 3, 5].

A.4 Conclusion

In this Appendix we provide useful supplementary details on data collection which in spite of its characteristics managed to be performed successfully using the state-of-the-art software solutions. Although the data collection (RSSI data and images) was performed on two different devices (laptop and camera respectively) in near future this will be feasible using e.g. wearable cameras which would reduce both financial cost and hardware requirements. The data reliably represent challenging everyday environments. The challenging datasets are formed using the data and used to test extensively localisation methods proposed in chapter 5 and chapter 6.

References

[1] Ekahau - WiFi tracking systems, RTLS and WLAN site survey.

[2] Cisco Location-Based Services Architecture. *http://www.cisco.com/Mobility/wifich3.html*, 2012.

[3] Extract Images From a Video, or Create a Video from Images using FFmpeg. In *FFmpeg Tutorial*, 2012.

[4] FFmpeg Documentation - http://ffmpeg.org/ffmpeg.html#synopsis. 2012.

[5] How to extract images from a Video using FFmpeg - http://linuxers.org/tutorial/how-extract-images-video-using-ffmpeg. 2012.

[6] InSSIDer - http://www.metageek.net/products/inssider. 2012.

[7] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. Building Rome in a day. *ACM Commun.*, 54(10):105–112, October 2011.

[8] U. Ahmed, A. Gavrilov, Sungyoung Lee, and Young-Koo Lee. Context-aware fuzzy artmap for Received Signal Strength based location systems. In *International Joint Conference on Neural Networks, IJCNN 2007.*, pages 2740–2745, aug. 2007.

[9] Lan F. Akyildiz, Welljan Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. A survey on sensor networks, 2002.

[10] A.S. Al-Ahmadi, A.I. Omer, M.R. Kamarudin, and T.A. Rahman. Multi-floor indoor positioning system using bayesian graphical models. *Progress In Electromagnetics Research B*, 25:241–259, 2010.

[11] M. Al-Hames, C. Lenz, S. Reiter, J. Schenk, F. Wallhoff, and G. Rigoll. Robust multi-modal group action recognition in meetings from disturbed videos with the asynchronous hidden markov model. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 213–216, sep. 2007.

[12] A.M. Alattar. Wipe scene change detector for use with video compression algorithms and mpeg-7. *IEEE Transactions on Consumer Electronics*, 44(1):43–51, feb. 1998.

[13] B. Alavi and K. Pahlavan. Analysis of undetected direct path in time of arrival based UWB indoor geolocation. In *62nd IEEE Conference on Vehicular Technology*, volume 4, pages 2627–2631, sep. 2005.

[14] P.F. Alcantarilla, Kai Ni, L.M. Bergasa, and F. Dellaert. Visibility learning in large-scale urban environment. In *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6205–6212, may 2011.

[15] Lora Aroyo, Rogier Brussee, Lloyd Rutledge, Peter Gorgels, Natalia Stash, and Yiwen Wang. Personalized Museum Experience: The Rijksmuseum Use Case. In *Museums and the Web*, San Francisco, USA, April 11-14, 2007.

[16] Clemens Arth, Daniel Wagner, Manfred Klopschitz, Arnold Irschara, and Dieter Schmalstieg. Wide area localization on mobile phones. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR, pages 73–82, 2009.

[17] Ronald Azuma, Jong Weon Lee, Bolan Jiang, Joo Park, Jun Park, Soya You, and Ulrich Neumann. Tracking in unprepared environments for augmented reality systems. *Computers Graphics*, 23:787–793, 1999.

[18] Paul Bach-y Rita, Mitchell E. Tyler, and Kurt A. Kaczmarek. Seeing with the brain. *International Journal of Human-Computer Interaction*, 15(2):285–295, 2003.

[19] P. Bahl and V.N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *INFOCOM 2000: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 775–784, 2000.

[20] Menelaos Bakopoulos, Sofia Tsekeridou, Eri Giannaka, Zheng-Hua Tan, and Ramjee Prasad. Mobile video annotation for enhanced rich media communication during emergency handling. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, ISABEL '11, pages 1–32, 2011.

[21] G. Balakrishnan and G. Sainarayanan. Stereo image processing procedure for vision rehabilitation. *Appl. Artif. Intell.*, 22:501–522, July 2008.

[22] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and Sazali Yaacob. Wearable real-time stereo vision for the visually impaired. *Engineering Letters*, 14(2):6–14, 2007.

[23] J. Bardwell. Converting signal strength percentage to dbm values. *White paper*, 2(3):203–217, November 2002.

[24] Herbert Bay, Tinne Tuytelaars, Van Gool, and Luc. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, August 2008.

[25] Jeffrey S. Beis and David G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1000–1007, 1997.

[26] G. Beliakov and J. Warren. Appropriate choice of aggregation operators in fuzzy decision support systems. *IEEE Transactions on Fuzzy Systems*, 9(6):773–784, dec. 2001.

[27] F. Bellavia, M. Cipolla, D. Tegolo, and C. Valenti. An evolution of the non-parameter Harris Affine Corner Detector: A distributed approach. In *International Conference on Parallel and Distributed Computing, Applications and Technologies*, pages 18–25, dec. 2009.

[28] Samy Bengio. An Asynchronous Hidden Markov Model for Audio-Visual Speech Recognition. In *Advances in Neural Information Processing Systems - NIPS 15*, 2003.

[29] E. Berry, A. Hampshire, J. Rowe, S. Hodges, N. Kapur, P. Watson, G. Browne, G. Smyth, K. Wood, and A. M. Owen. The neural basis of effective memory therapy in a patient with limbic encephalitis. *Journal of Neurology, Neurosurgery & Psychiatry*, 80(11):1202–1205, November 2009.

[30] Saad Biaz and Yiming Ji. A survey and comparison on localisation algorithms for wireless ad hoc networks. *Int. J. Mob. Commun.*, 3(4):374–410, May 2005.

[31] Sogne Brilingaite, Peter Jensen, and Nora Kulyte. Indoor route analyses as context in different services. In *IIS '09: Proceedings on Indoor information systems*, pages 127–136, 2004.

[32] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web*, pages 107–117, 1998.

[33] Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[34] Daragh Byrne, Aiden R. Doherty, Gareth J. F. Jones, Alan F. Smeaton, Sanna Kumpulainen, and Kalervo Järvelin. The SenseCam as a tool for task observation. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*, pages 19–22, 2008.

[35] Daragh Byrne, Aiden R. Doherty, Gareth J. F. Jones, Alan F. Smeaton, Sanna Kumpulainen, and Kalervo Järvelin. The SenseCam as a tool for task observation.

In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*, pages 19–22, 2008.

[36] Daragh Byrne, Aiden R. Doherty, Cees G. Snoek, Gareth G. Jones, and Alan F. Smeaton. Validating the detection of everyday concepts in visual lifelogs. In *Proceedings of the 3rd International Conference on Semantic and Digital Media Technologies: Semantic Multimedia*, SAMT '08, pages 15–30, 2008.

[37] Roberto Cabeza, Yonatan S. Mazuz, Jared Stokes, James E. Kragel, Marty G. Woldorff, Elisa Ciaramelli, Ingrid R. Olson, and Morris Moscovitch. Overlapping parietal activity in memory and perception: Evidence for the attention to memory model. *J. Cognitive Neuroscience*, 23(11):3209–3217, 2011.

[38] Q. Cai, A. Mitiche, and J.K. Aggarwal. Tracking human motion in an indoor environment. In *International Conference on Image Processing*, volume 1, pages 215–218, oct. 1995.

[39] Paul Castro, Patrick Chiu, Ted Kremenek, and Richard R. Muntz. A probabilistic room location service for wireless networked environments. In *Proceedings of the 3rd international conference on Ubiquitous Computing*, UbiComp '01, pages 18–34, 2001.

[40] Xiaoyong Chai and Qiang Yang. Reducing the calibration effort for probabilistic indoor location estimation. *IEEE Transactions on Mobile Computing*, 6(6):649–662, June 2007.

[41] S.D. Chitte, S. Dasgupta, and Zhi Ding. Distance estimation from received signal strength under log-normal shadowing: Bias and variance. *IEEE Signal Processing Letters*, 16(3):216–218, mar. 2009.

[42] O Chum, J Matas, and J Kittler. Locally optimized RANSAC. *Pattern Recognition*, 2781(7):236–243, 2003.

[43] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *IEEE International Conference on Computer Vision*, 2007.

[44] Ondrej Chum, Andrej Mikulík, Michal Perdoch, and Jiri Matas. Total recall II: Query expansion revisited. In *CVPR*, pages 889–896, 2011.

[45] Ondrej Chum, Michal Perdoch, and Jiri Matas. Geometric min-hashing: Finding a (thick) needle in a haystack. In *CVPR*, pages 17–24, 2009.

[46] Li Cong and Weihua Zhuang. Non-line-of-sight error mitigation in mobile location. In *Infocom 2004*, pages 7–11, 2004.

[47] Ciaran O Connaire, Keith Fogarty, Connor Brennan, and Noel O'Connor. User Localisation using Visual Sensing and RF Signal Strength. In *ImageSense 2008 - Workshop on Applications, Systems, and Algorithms for Image Sensing, at the 6th ACM Conference on Embedded Networked Sensor Systems*, 2008.

[48] Nello Cristianini and John Shawe-Taylor. *An introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, March 2000.

[49] Zhu Daixian. SIFT algorithm analysis and optimization. In *International Conference on Image Analysis and Signal Processing (IASP)*, pages 415–419, apr. 2010.

[50] Ju-Yong Do, M. Rabinowitz, and P. Enge. Performance of Hybrid Positioning System Combining GPS and Television Signals. pages 556–564, apr. 2006.

[51] Aiden Doherty. Providing effective memory retrieval cues through automatic structuring and augmentation of a lifelog of images. In *Seminar at the Multimedia Information Retrieval Group, Yahoo! Research*, 2008.

[52] Aiden Doherty, Pauly-Takacs K, Cathal Gurrin, Moulin C, and Alan F. Smeaton. Three Years of SenseCam Images - Observations on Cued Recall. In *SenseCam Symposium at the 39th Annual Meeting of the Society for Neuroscience*, 2009.

[53] Aiden R. Doherty, Ciarán Ó Conaire, Michael Blighe, Alan F. Smeaton, and Noel E. O'Connor. Combining image descriptors to effectively retrieve events from visual lifelogs. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, MIR, pages 10–17, 2008.

[54] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay. Clustering large graphs via the singular value decomposition. *Mach. Learn.*, 56(3):9–33, June 2004.

[55] C. Durieu, H. Clergeot, and F. Monteil. Localization of a mobile robot with beacons taking erroneous data into account. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1062–1068, may 1989.

[56] E.D. Eade and T.W. Drummond. Unified Loop Closing and Recovery for Real Time Monocular SLAM. In *Proc. BMVC*, pages 6.1–6.10, 2008.

[57] P. Enge, T. Walter, S. Pullen, Changdon Kee, Yi-Chung Chao, and Yeou-Jyh Tsai. Wide area augmentation of the Global Positioning System. *Proceedings of the IEEE*, 84(8):1063–1088, aug. 1996.

[58] Shih-Hau Fang and Tsungnan Lin. Principal component localization in indoor wlan environments. *IEEE Transactions on Mobile Computing*, 11(1):100–110, jan. 2012.

[59] Shih-Hau Fang, Tsungnan Lin, and Kun-Chou Lee. A novel algorithm for multipath fingerprinting in indoor wlan environments. *IEEE Transactions on Wireless Communications*, 7(9):3579–3588, september 2008.

[60] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, jun. 2003.

[61] Rowanne Fleck and Geraldine Fitzpatrick. Supporting collaborative reflection with passive image capture. In *Supplementary Proceedings of COOP '06*, pages 41–48, 2006.

[62] J. Fodor, J.-L. Marichal, and M. Roubens. Characterization of the ordered weighted averaging operators. *IEEE Transactions on Fuzzy Systems*, 3(2):236–240, may. 1995.

[63] F. Forno, G. Malnati, and G. Portelli. Design and implementation of a bluetooth ad hoc network for indoor positioning. *IEE Proceedings - Software*, 152(5):223–228, oct. 2005.

[64] Zhouyu Fu and A. Robles-Kelly. An instance selection approach to multiple instance learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 911–918, jun. 2009.

[65] G. Gates. The reduced nearest neighbor rule. *IEEE Transactions on Information Theory*, 18(3):431–433, may. 1972.

[66] Jim Gemmell, Aleks Aris, and Roger Lueder. Telling stories with mylifebits. In *ICME*, pages 1536–1539. IEEE, 2005.

[67] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat. Vision and RFID data fusion for tracking people in crowds by a mobile robot. *CVIU*, 114(6):641–651, June 2010.

[68] Stuart Golden and Steve Bateman. Sensor Measurements for WiFi Location with Emphasis on Time-of-Arrival Ranging. *IEEE Transactions on Mobile Computing*, 6(10):1185–1198, October 2007.

[69] Iryna Gordon and David G Lowe. What and where: 3D Object Recognition with Accurate Pose. *Lecture Notes in Computer Science*, 4170/2006(1):6782, 2004.

[70] K. Gowda and G. Krishna. The condensed nearest neighbor rule using the concept of mutual nearest neighborhood (corresp.). *IEEE Transactions on Information Theory*, 25(4):488–490, jul. 1979.

[71] Thomas Guckenbiehl, Heiko Milde, Bernd Neumann, and Peter Struss. Meeting re-use requirements of real-life diagnosis applications. In *PROC. XPS-99: KNOWLEDGE-BASED SYSTEMS, SPRINGER, LNAI 1570*, 1999.

[72] Steve R. Gunn. Support vector machines for classification and regression. Technical report, May 1998.

[73] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the Tenth ACM International Conference on Mobile Computing and Networking (MOBICOM 2004)*, pages 70–84, Philadelphia, PA, Sept.-Oct. 2004.

[74] A. Harter and A. Hopper. A distributed location system for the active office. *IEEE Network*, 8(1):62–70, jan/feb 1994.

[75] N.S.A. Hassan, S. Hossain, N.H.A. Wahab, S.H.S. Ariffin, N. Fisal, L.A. Latiff, M. Abbas, and Choong Khong Neng. An Indoor 3D Location Tracking System Using RSSI. In *Sixth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, pages 323–328, dec. 2010.

[76] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf. Support vector machines. *Intelligent Systems and Their Applications*, 13:18–28, 1998.

[77] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, aug. 2001.

[78] Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, August 2001.

[79] Todd D. Hodes, Randy H. Katz, Edouard Servan-Schreiber, and Lawrence Rowe. Composable ad-hoc mobile services for universal interaction. In *Proceedings of the 3rd annual ACM/IEEE international conference on Mobile computing and networking*, MobiCom '97, pages 1–12, 1997.

[80] Steve Hodges, Emma Berry, and Ken Wood. SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, 19(7):685–696, October 2011.

[81] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Kenneth R. Wood. SenseCam: A Retrospective Memory Aid. In *Ubicomp*, pages 177–193, 2006.

[82] S. Ichitsubo, T. Furuno, T. Taga, and R. Kawasaki. Multipath propagation model for line-of-sight street microcells in urban area. *IEEE Transactions on Vehicular Technology*, 49(2):422–427, mar. 2000.

[83] Arnold Irschara, Christopher Zach, and Horst Bischof. Towards Wiki-based Dense City Modeling. In *ICCV*, pages 1–8, 2007.

[84] K. Iwatsuka, K. Yamamoto, and K. Kato. Development of a guide dog system for the blind people with character recognition ability. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, volume 1, pages 453–456, aug. 2004.

[85] Songmin Jia, Jinbuo Sheng, and K. Takase. Obstacle recognition for a service mobile robot based on RFID with multi-antenna and stereo vision. In *International Conference on Information and Automation*, pages 125–130, jun. 2008.

[86] Liangxiao Jiang, H. Zhang, and Zhihua Cai. A Novel Bayes Model: Hidden Naive Bayes. *IEEE Transactions on Knowledge and Data Engineering*, 21(10):1361–1371, oct. 2009.

[87] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 90–96, jun. 2004.

[88] S. Karaman, J. Benois-Pineau, R. Meandgret, V. Dovgalecs, J.-F. Dartigues, and Y. Gaeandstel. Human daily activities indexing in videos from wearable cameras for monitoring of patients with dementia diseases. In *20th International Conference on Pattern Recognition (ICPR)*, pages 4113–4116, 2010.

[89] Eirini Karapistoli, Ioannis Gragopoulos, Ioannis Tsetsinas, and Fotini-Niovi Pavlidou. UWB Technology to Enhance the Performance of Wireless Multimedia Sensor Net-

works. In *12th IEEE Symposium on Computers and Communications (ISCC)*, pages 57–62, July 2007.

[90] Yan Ke and R. Sukthankar. PCA-SIFT: a more distinctive representation for local image descriptors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 506–513, jun. 2004.

[91] Thomas King, Stephan Kopf, Thomas Haenselmann, Christian Lubberger, and Wolfgang Effelsberg. Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses. In *Proceedings of the 1st international workshop on Wireless network testbeds, experimental evaluation & characterization*, WiNTECH '06, pages 34–40, 2006.

[92] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, Nara, Japan, November 2007.

[93] Martin Klepal, Maarten Weyn, Warsun Najib, Inge Bylemans, Sigit Wibowo, Widyawan Widyawan, and Bimo Hantono. OLS: opportunistic localization system for smart phones devices. In *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds (MobiHeld)*, August 2009.

[94] P Krishnamurthy. Position location in mobile environments. In *Proceedings of NSF Workshop on Context-Aware Mobile Database Management (CAMM)*. Providence, RI, 2002.

[95] F. Kristensen and W. J. Maclean. Real-time extraction of maximally stable extremal regions on an fpga. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 165–168, 2007.

[96] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Lydia E. Kavraki, and Dan S. Wallach. Robotics-based location sensing using wireless Ethernet. *Wireless Networks*, 11:189–204, January 2005.

[97] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Guillaume Marceau, Lydia E. Kavraki, and Dan S. Wallach. Robotics-based location sensing using wireless ethernet. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, MobiCom '02, pages 227–238, 2002.

[98] Luke Ledwich and Stefan Williams. Reduced SIFT features for image retrieval and indoor localisation. In *Australian Conference on Robotics and Automation*, 2004.

[99] Wen-Yu Lee, Po-Tun Wu, and Winston Hsu. Adaptive learning for multimodal fusion in video search. In *PCM '09: Proceedings of the 10th Pacific Rim Conference on Multimedia*, pages 659–670, 2009.

[100] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 775–781, june 2005.

[101] D. Letourneau, F. Michaud, J.-M. Valin, and C. Proulx. Textual message read by a mobile robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 2724–2729, oct. 2003.

[102] Li Li and J.L. Krolik. Target tracking in uncertain multipath environments using viterbi data association. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–7, july 2011.

[103] Xiang-Yang Li and Peng-Jun Wan. Constructing minimum energy mobile wireless networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(4):55–67, October 2001.

[104] Xiaowei Li, Changchang Wu, Christopher Zach, Svetlana Lazebnik, and Jan-Michael Frahm. Modeling and recognition of landmark image collections using iconic scene

graphs. In *Proceedings of the 10th European Conference on Computer Vision: Part I*, ECCV '08, pages 427–440, 2008.

[105] Yunpeng Li, Noah Snavely, and Daniel P. Huttenlocher. Location recognition using prioritized feature matching. In *Proceedings of the 11th European conference on Computer vision: Part II*, ECCV'10, pages 791–804, 2010.

[106] T. Lindeberg. Junction detection with automatic selection of detection scales and localization scales. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 924–928, nov. 1994.

[107] H. Liu, H. Darabi, P. Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 37(6):1067–1080, nov. 2007.

[108] X. Liu and J. Samarabandu. An edge-based text region extraction algorithm for indoor mobile robot navigation. *IEEE International Conference on Mechatronics and Automation*, 2:701–706, 2005.

[109] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, 1982.

[110] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.

[111] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999.

[112] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999.

[113] Juergen Luettin and Stephane Dupont. Continuous audio-visual speech recognition. In *Proceedings of the 5th European Conference on Computer Vision (ECCV)*, volume 2, pages 657–673, 1998.

[114] G. Lui, T. Gallagher, Binghao Li, A.G. Dempster, and C. Rizos. Differences in RSSI readings made by different WiFi chipsets: A limitation of WLAN localization. In *International Conference on Localization and GNSS (ICL-GNSS)*, pages 53–57, june 2011.

[115] Xiaolong Ma and Konstantin K. Likharev. Global reinforcement learning in neural networks. *IEEE Transactions on Neural Networks*, 18(2):573–577, 2007.

[116] Steve Mann, Jason Huang, Ryan Janzen, Raymond Lo, Valmiki Rampersad, Alexander Chen, and Taqveer Doha. Blind navigation with a wearable range camera and vibrotactile helmet. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 1325–1328, New York, NY, USA, 2011.

[117] Hamid Mehmood and Nitin K. Tripathi. Optimizing artificial neural network-based indoor positioning system using genetic algorithm. *International Journal of Digital Earth*, 2(1):1–27, 2007.

[118] Carlos Merino and Majid Mirmehdi. A framework towards real-time detection and tracking of text. In *Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2007)*, pages 10–17, September 2007.

[119] Michael J. Meyer, Terry Jacobson, Maria E. Palamara, Elizabeth A. Kidwell, Robert E. Richton, and Giovanni Vannucci. Wireless enhanced 911 service thus making it a reality. *Bell Labs Technical Journal*, 1(2):188–202, 1996.

[120] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1792–1799, oct. 2005.

[121] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, oct. 2005.

[122] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1):43–72, November 2005.

[123] H. Misra, H. Bourlard, and V. Tyagi. New entropy based combination rules in HMM/ANN multi-stream asr. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 741–744, apr. 2003.

[124] T. Miyaki, T. Yamasaki, and K. Aizawa. Tracking persons using particle filter fusing visual and WiFi localizations for widely distributed camera. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 225–228, sep. 2007.

[125] Thomas P. Moran and Paul Dourish. Introduction to this special issue on context-aware computing. *Hum. Comput. Interact.*, 16(2):87–95, December 2001.

[126] Patrik Moravek, Dan Komosny, David Girbau Sala, and Antoni Lazaro Guillen. Received signal strength uncertainty in energy-aware localization in wireless sensor networks. In *International Conference on Environment and Electrical Engineering (EEEIC)*, pages 538–541, may. 2010.

[127] N. J. S. Morch, U. Kjems, L. K. Hansen, C. Svarer, I. Law, B. Lautrup, S. Strother, and K. Rehm. Visualization of neural networks using saliency maps. In *IEEE International Conference on Neural Networks*, volume 4, pages 2085–2090, 1995.

[128] D. Musicki and W. Koch. Geolocation using TDOA and FDOA measurements. In *11th International Conference on Information Fusion*, pages 1–8, jun. 2008.

[129] Hesam Najafi, Yakup Genc, and Nassir Navab. Fusion of 3D and appearance models for fast object detection and pose estimation. In *ACCV*, pages 415–426, 2006.

[130] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 652–659, 7 2004.

[131] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[132] David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2161–2168, Washington, DC, USA, 2006.

[133] T. Oskiper, Han-Pang Chiu, Zhiwei Zhu, S. Samarasekera, and R. Kumar. Multi-modal sensor fusion algorithm for ubiquitous infrastructure-free localization in vision-impaired environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1513–1519, oct. 2010.

[134] K. Pahlavan, Xinrong Li, and J.P. Makela. Indoor geolocation science and technology. *IEEE Communications Magazine*, 40(2):112–118, feb 2002.

[135] A.S. Paul and E.A. Wan. RSSI-based indoor localization and tracking using sigma-point kalman smoothers. *IEEE Journal of Selected Topics in Signal Processing*, 3(5):860–873, oct. 2009.

[136] Thomas Prunnel. Aggressive scanning in WLAN networks. *McGraw-Hill*, 2011.

[137] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. volume 77, pages 257–286, feb 1989.

[138] T.S. Rappaport, J.H. Reed, and B.D. Woerner. Position location using wireless communications on highways of the future. *IEEE Communications Magazine*, 34(10):33–41, oct 1996.

[139] M. Redzic, C. O'Conaire, C. Brennan, and N. O'Connor. A hybrid method for indoor user localisation. In *4th European Conference on Smart Sensing and Context (EuroSSC)*, 2009.

[140] Milan Redzic, Conor Brennan, and Noel E. O'Connor. Dual-sensor fusion for indoor user localisation. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 1101–1104, New York, NY, USA, 2011. ACM.

[141] Gerhard Reitmayr and Tom W. Drummond. Going out: Robust tracking for outdoor augmented reality. In *Proc. ISMAR 2006*, pages 109–118, Santa Barbara, CA, USA, October 22–25 2006.

[142] G. Ritter, H. Woodruff, S. Lowry, and T. Isenhour. An algorithm for a selective nearest neighbor decision rule. *IEEE Transactions on Information Theory*, 21(6):665–669, nov. 1975.

[143] D. Robertsone and R. Cipolla. An image-based system for urban navigation. In *Proc. BMVC*, pages 84.1–84.10, 2004.

[144] P Rong and M.L. Sichitiu. Angle of arrival localization for wireless sensor networks. In *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON)*, volume 1, pages 374–382, sep. 2006.

[145] Teemu Roos, Petri Myllyma, Henry Tirri, Pauli Misikangas, and Juha Sieva. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*, 9(3):155–164, 2002.

[146] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Fast image-based localization using direct 2D-to-3D matching. *IEEE International Conference on Computer Vision (ICCV)*, 0:667–674, 2011.

[147] G. Schindler, M. Brown, and R. Szeliski. City-scale location recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, june 2007.

[148] Stephen Se, David Lowe, and Jim Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21:735–758, 2002.

[149] Mark Shefel, Peter Vainer, and Igor Timko. Nearest neighbor approach in route grids. In *ALS '04: Proceedings of the 15th ACM international symposium on Advances in location systems*, pages 1–8, 2004.

[150] Sung-Tsun Shih, Kunta Hsieh, and Pei-Yuan Chen. An improvement approach of indoor location sensing using active RFID. In *First International Conference on Innovative Computing, Information and Control (ICICIC)*, volume 2, pages 453–456, aug. 2006.

[151] T. Shimbashi, Y. Kokubo, and N. Shirota. Region segmentation using edge based circle growing. In *International Conference on Image Processing (ICIP)*, volume 3, pages 65–68, oct. 1995.

[152] Josef Sivic and Andrew Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 1470–1477, 2003.

[153] Iryna Skrypnyk and David G. Lowe. Scene modelling, recognition and tracking with invariant image features. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 110–119, 2004.

[154] F. Sottile, R. Giannantonio, M.A. Spirito, and F.L. Bellifemine. Design, deployment and performance of a complete real-time zigbee localization system. In *1st IFIP Wireless Days*, pages 1–5, nov. 2008.

[155] Justin Stook. M.Sc. Thesis - http://igitur-archive.library.uu.nl/student-theses/2012-0316-200434/thesis_justin_stook_v1.6.pdf.

[156] Chunqiao Tan, Benjiang Ma, and Xiaohong Chen. Intuitionistic fuzzy geometric aggregation operator based on fuzzy measure for multi-criteria group decision making. In

Sixth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), volume 4, pages 545–549, aug. 2009.

[157] M. Tanaka and H. Goto. Autonomous text capturing robot using improved dct feature and text tracking. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, pages 1178–1182, 2007.

[158] T. Tatschke. Early sensor data fusion techniques for collision mitigation purposes. In *Intelligent Vehicles Symposium*, pages 445–452, 2006.

[159] David L. Tennenhouse, Jonathan M. Smith, W. David Sincoskie, David J. Wetherall, and Gary J. Minden. A survey of active network research. *IEEE Communications Magazine*, 35:80–86, 1997.

[160] R. Tesoriero, J. Gallud, M. Lozano, and V. Penichet. Using active and passive RFID technology to support indoor location-aware systems. *IEEE Transactions on Consumer Electronics*, 54(2):578–583, may. 2008.

[161] I. Tomek. An experiment with the edited nearest-neighbor rule. *IEEE Transactions on Systems, Man and Cybernetics*, 6(6):448–452, jun. 1976.

[162] V. Torra. Learning weights for the quasi-weighted means. *IEEE Transactions on Fuzzy Systems*, 10(5):653–666, oct. 2002.

[163] Antonio Torralba, Kevin P. Murphy, William T. Freeman, and Mark Rubin. Context-based vision system for place and object recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, pages 273–280, 2003.

[164] S. Van den Berghe, M. Weyn, V. Spruyt, and A. Ledda. Fusing camera and WiFi sensors for opportunistic localization. In *The Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM)*, pages 169–174, 2011.

[165] S. Venkatesh and R.M. Buehrer. Non-line-of-sight identification in ultra-wideband systems based on received signal statistics. *Microwaves, Antennas Propagation, IET*, 1(6):1120–1130, dec. 2007.

[166] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, 2001.

[167] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10:91–102, January 1992.

[168] Mark Weiser. Some computer science issues in ubiquitous computing. *ACM Communications*, 36:75–84, July 1993.

[169] Kamin Whitehouse, Chris Karlof, and David Culler. A practical evaluation of radio signal strength for ranging-based localization. *ACM Mobile Computing and Communications Review (MC2R): Special Issue on Localization Technologies and Algorithms*, pages 41–52, January 2007.

[170] K. Witrisal and P. Meissner. Performance Bounds for Multipath-assisted Indoor Navigation and Tracking (MINT). In *International Conference on Communications (ICC)*, Ottawa, Canada, 2012.

[171] J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization by combining an image-retrieval system with monte carlo localization. *IEEE Transactions on Robotics*, 21(2):208–216, 2005.

[172] Martin Wollmer, Marc Al-Hames, Florian Eyben, Bjorn Schuller, and Gerhard Rigoll. A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing (NEUCOM)*, 73:366–380, 2009.

[173] Lizhong Wu, S.L. Oviatt, and P.R. Cohen. Multimodal integration-a statistical view. *IEEE Transactions on Multimedia*, 1(4):334–341, dec. 1999.

[174] Jianxiong Xiao, Jingni Chen, Dit-Yan Yeung, and Long Quan. Structuring visual words in 3D for arbitrary-view object localization. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, pages 725–737, 2008.

[175] A. Yamazaki, T. B. Ludermir, and M. C. P. de Souto. Global optimization methods for designing and training neural networks. In *Proceedings of the 7th Brazilian Symposium on Neural Networks*, pages 136–141, 2002.

[176] Qiang Yang, Sinno Jialin Pan, and Vincent Wenchen Zheng. Estimating location using WiFi. *IEEE Intelligent Systems*, 23(1):8–13, jan. 2008.

[177] Jaegeol Yim, Seunghwan Jeong, Jaehun Joo, and Chansik Park. Development of Kalman filters for WLAN based tracking. In *Second International Conference on Future Generation Communication and Networking Symposia (FGCNS)*, volume 5, pages 1–6, dec. 2008.

[178] Jie Yin, Qiang Yang, and Lionel M. Ni. Learning adaptive temporal radio maps for signal-strength-based location estimation. *IEEE Transactions on Mobile Computing*, 7(7):869–883, July 2008.

[179] Taebok Yoon and Jee-Hyong Lee. Goal and path prediction based on user's moving path data. In *ICUIMC '08: Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 475–480, 2008.

[180] Moustafa Youssef and Ashok Agrawala. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, MobiSys '05, pages 205–218, New York, NY, USA, 2005. ACM.

[181] J.M.D. Zampieron. *Self-localization in ubiquitous computing using sensor fusion.* Rochester Institute of Technology, 2006.

[182] Pablo Zegers and Malur K. Sundareshan. Trajectory generation and modulation using dynamic neural networks. *IEEE Transactions on Neural Networks*, 14(3):520–533, May 2003.

[183] Vasileios Zeimpekis, George M. Giaglis, and George Lekakos. A taxonomy of indoor and outdoor positioning techniques for mobile location services. *SIGecom Exch.*, 3(4):19–27, December 2002.

[184] Wei Zhang and Jana Kosecka. Image based localization in urban environments. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, pages 33–40, 2006.

[185] Zhiwei Zhu, Taragay Oskiper, Supun Samarasekera, Rakesh Kumar, and Harpreet S. Sawhney. Real-time global localization with a pre-built visual landmark database. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:1–8, 2008.