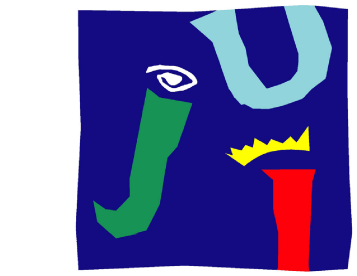


UNIVERSITAT JAUME·I

Received Signal Strength-based Indoor Positioning

Integrative Outlooks and Solutions

Germán Martín Mendoza Silva

Supervisors: Dr. Joaquín Huerta Guijarro
Dr. Joaquín Torres Sospedra

A dissertation presented for the degree of Doctor of Philosophy

Castelló de la Plana (Spain)
October 2019

Programa de Doctorado en Informática

Escuela de Doctorado de la Universitat Jaume I

Received Signal Strength-based Indoor Positioning

Integrative Outlooks and Solutions

Memoria presentada por Germán Martín Mendoza Silva para optar al grado de doctor por la Universitat Jaume I

Autor
Germán Martín Mendoza Silva

Directores
Joaquín Huerta Guijarro y Joaquín Torres Sospedra

Castelló de la Plana (Spain), octubre de 2019

Financiación recibida

Esta tesis ha sido financiada por la ayuda predoctoral PREDOC/2016/55 concedida por la Universitat Jaume I.

Resumen

Desde hace más de una década, se viene reconociendo cada vez más la necesidad de los Sistemas de Posicionamiento en Interiores (IPS, por sus siglas en inglés). No existe una tecnología actual capaz de proporcionar estimaciones precisas y confiables de posición para entornos de interior que sea aplicable a la mayoría de los entornos y aplicaciones. Como consecuencia, se han propuesto muchas soluciones que se basan en una tecnología específica y están destinadas a entornos y aplicaciones específicas. Entre las soluciones más populares están las basadas en fingerprinting, principalmente usando las señales de puntos de acceso WiFi y balizas Bluetooth Low Energy (BLE). En la implantación de un IPS basado en fingerprinting existen tres etapas: Descripción del entorno, ajuste del modelo, y evaluación del sistema.

Esta tesis se enfoca principalmente en las etapas de descripción del modelo y evaluación para sistemas basados en fingerprinting, aunque algunas de sus aportaciones son aplicables a otros sistemas. La primera aportación de la tesis es un meta-review que sirve como introducción a los sistemas de posicionamiento en interiores, el estado actual del arte, y una evaluación crítica sobre el impacto de las publicaciones sobre IPS. El meta-review es una revisión sistemática de surveys que tratan sistemas de posicionamiento aplicados a entornos interiores publicados en los últimos cinco años. La segunda aportación es dos datasets abiertos de mediciones de RSS de WiFi y BLE RSS recolectadas para fomentar la experimentación y la reproducibilidad en las soluciones para el posicionamiento en interiores. La tercera es un conjunto de guías y métodos para mejorar la recolección de muestras de señales WiFi y BLE. Las guías se enfocan en las herramientas y metodologías para llevar a cabo la recolección y la selección de posiciones. Los métodos permiten determinar la posición de puntos de acceso WiFi y usar una regresión que brinda mejores estimaciones de intensidad que otros modelos. La cuarta aportación de la tesis es un método de selección de posiciones para actualizar la caracterización del entorno. La quinta aportación de la tesis es un conjunto de guías y un procedimiento para realizar la evaluación de IPS. Las guías son resultado de experiencias en la evaluación de IPS en competiciones. El método proporciona una forma de medir el error de posicionamiento en la que la magnitud del error es medida como la distancia de la ruta que une la posición real y

la posición estimada. Además, se analiza la localidad del error asociado a un IPS.

Abstract

The need for Indoor Positioning Systems (IPS) has been increasingly recognized for more than a decade. There is no current technology capable of providing accurate and reliable position estimates for indoor environments that applies to most environments and applications. As a consequence, many solutions have been proposed that are based on a specific technology and are intended for specific environments and applications. Among the most popular solutions are those based on fingerprinting, mainly using signals from WiFi access points and Bluetooth Low Energy (BLE) beacons. In the implementation of an IPS based on fingerprinting, there are three stages: Environment description, model setup, and evaluation.

This dissertation focuses mainly on the stages of model description and evaluation for fingerprinting-based IPS, although some of its contributions apply to other systems. The first contribution of the dissertation is a meta-review that serves as an introduction to indoor positioning systems, the current state of the art, and a critical evaluation of the impact of IPS publications. The meta-review is the result of a systematic review of surveys that dealt with positioning systems applied to indoor settings in the last five years. The second contribution is two open datasets of WiFi and BLE RSS measurements intended to foster experimentation and reproducibility of indoor positioning solutions. The third contribution is a set of guides and methods to improve the collection of samples of WiFi and BLE signals. The guides focus on the tools and methodologies to carry out the collection and selection of positions. The methods allow determining the position of WiFi access points and provide a regression that produces better signal intensity estimates than other models. The fourth contribution of the dissertation is a method for selecting the positions the collection should be carried out to update the characterization of the environment. The fifth contribution of the dissertation is a set of guides and a procedure to perform the IPS evaluation. The guides are the result of experiences in the evaluation of IPS in competitions. The procedure provides a way to measure the positioning error in which the magnitude of the error is measured as the distance of the route between the actual position and the estimated position. Also, the locality of the error associated with an IPS is analyzed.

Dedication

To my favorite band: My mother, my wife and my daughter.

Acknowledgments

I thank God, for giving me strengths when I felt weak.

Joaquín (Huerta Guijarro), thanks to your support, I was able to start, continue, and finish this work. You have had a hopeful answer for every restless question. For that, I will be forever grateful.

Ximo (Joaquín Torres-Sospedra), you taught me how to be a scientist and how to be a better worker. Thank you for patiently lead my way through these years. Besides a supervisor, you have been a friend to me.

Everybody in Geotec, thank you for accepting me the way I am and despite saying 'no' countless times, still inviting me to join you to the morning's coffee breaks. You create an open and friendly environment that many would envy. You are a great team!

My family, mainly my wife and my daughter, thanks for your understanding and accepting to lose so many hours of my time, which I cannot make up for. We will grow, side by side, through life.

To anyone that helped me reach this point of my academic life in any way, you all have my most sincere gratitude.

List of Figures

List of Tables

Acronyms

AOA	Angle of Arrival
BLE	Bluetooth Low Energy
DOI	Digital Object Identifier
DR	Dead Reckoning
GPR	Gaussian Process regression
IDW	Inverse Distance Weighting
IMU	Inertial Measurement Unit
IoT	Internet of Things
IPS	Indoor Positioning System
LED	Light-Emitting Diode
LOS	Line of sight
NFC	Near Field Communication
NLOS	Non-Line of sight
PDR	Pedestrian Dead Reckoning
RF	Radio Frequency
RFID	Radio Frequency Identification
RMS	Root Mean Square
RSS	Received Signal Strength
SVR	Support Vector Regression
TDOA	Time Difference of Arrival
TOA	Time of Arrival
UJI	Universitat Jaume I
UWB	Ultra-Wideband
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
WSN	Wireless sensor networks

Contents

Chapter 1

Introduction

The first positioning system that we humans had was our mind, which works based on the information gathered first by our senses – mainly sight – and later by representations and devices created by ourselves. The information for that system comes from experience, either by traveling through the environment or by receiving instructions from, e.g., another person or a book [7]. Pieces of the information – landmarks, routes, and regions – are given some relevance and learned, and relations are established among them. The learned information, considering the invariant character of the sites and boundaries is used in conjunction to our heading – our facing direction – to get a sense of direction and position [4].

Our capacity to accurately determine our position at increasingly larger scales improved alongside technological advancements to quantify and describe our environment. We developed the ability to measure distance, angles and time and the mathematical tools to support them, like trigonometry. We improved landmarks recognition using lighthouses and telescopes. We built devices that helped us relating landmarks and measures, like the compass or the sextant. We created representations of the landmarks relation, like the maps. But despite those advancements, the only known positioning system was still our mind. The great leap came with the development of radio systems.

Experiments performed by Heinrich Hertz in 1888 discovered that it was possible to discover the source of a radio emission. Hertz also determined several other characteristics of radio propagation, including its propagation speed. The radio direction finder systems appeared along the 20^{th} century, as electronic circuits and time measurement systems improved. Eventually, position determination systems based on fixed terrestrial radio beacons appeared and provided different capabilities and accuracies. Examples of such systems include the British Decca Navigator System, the Russian CHAYKA, the USA LORAN, and the multinational Omega [15]. All of

them were gradually replaced by GPS, mainly after the "Selective Availabilit" that degraded accuracy for civilian applications, was discontinued from GPS in 2000. The ground-based fixed beacons were mostly replaced by a constellation of precisely-timed satellites that provided global positioning and navigational services.

It is difficult to imagine our lives without a positioning system giving us support. GPS or similar GNNS constellations have shaped our modern life. Many current activities, like transportation and large engineering projects, could not be possible without positioning systems. But beyond large-scale activities, our daily activities are also shaped by our automatically determined position. From direction indications in a tourism trip to personalized advertisements, many of us enjoy the benefits of the development of Location-Based Services (LBS).

LBS are defined as computer applications that provide their goal content adapted to the position and context of the device and the user [3, 12]. Apart from GNNS services, LBS has been boosted by smartphones popularity. An increasing number of the world's population owns a smartphone and has Internet connectivity. For some people, a smartphone is the first and the last device with which we interact every day – apart perhaps from the glasses to read their screens. People want their services provided in their smartphones, everything from guidance, banking and office tools to shopping, social networking, and entertainment. Furthermore, as we have our smartphones next to us at almost all times, they create excellent possibilities to deliver LBS. The global market knows that, and it is why LBS and their supporting technologies draw a lot of interest and investment [9, 1].

While the 20^{th} century saw the success of the global positioning systems, the 21^{st} century will see the success of Indoor Positioning Systems (IPSs) [6, 5, 16]. In developed countries, people spend most of their time indoors [8, 14]. As the whole world develops, this tendency will expand to other countries. However, position estimation indoors still lack a generally applicable, low-cost solution that allows LBS tailored for indoor scenarios.

1.1 Motivation

The position estimation of mobile application users in indoor environments is the subject of lots of works in academia, ranging from bachelor projects to doctoral dissertations to well-funded cross-national projects. The amount of publications on the subject, of varying quality and in a variety of media, is so large that it is difficult to fully survey them all. It is also difficult to find all the reasons behind this thirst for publishing about indoor positioning. Some likely answers may include a general urge for publication in academia; a variety of new publication media; the ease of implementing a simple and cheap positioning system, thus encouraging its further

research and publication; the elusiveness of a system that is accurate, cheap and applicable to many indoor environments at the same time; and the promise from industry of rapid adoption and billionaire investments for such golden system.

The golden IPS should ideally be used in the way that GNSS currently is for smartphones outdoors, but adapted to indoor requirement and challenges. It would be a technology, a technique, a method or their combination that feature a set of traits that can be inferred from the goal of obtaining the capabilities of GNNS for smartphones and from several common evaluation metrics for IPS [2, 13, 17, 18, 19], later presented in Chapter 2. The set of desirable traits is:

1. It can be used smoothly, without diverting the user main focus. The user would not need to switch to another activity for receiving position estimates.
2. It uses only sensors in current smartphones or at least in smartphones foreseen in a very near future (e.g., support for 5G).
3. It does not requires the deployment of devices whose acquisition or maintenance cost rivals with expected positioning benefits, even when the system is scaled up.
4. It achieves an accuracy in the order of tens of centimeters.
5. It has context-related accuracy predictability. For example, it is expected to obtain large positioning errors from a GNNS-based receiver inside a tunnel.
6. The privacy compromises may not become a high concern for users. Users feeling that their positions are shared with others may be reluctant to use a positioning service.

In the pursuit of the golden IPS, several technologies have been used to build system proposals that tried to achieve and sometimes achieved, some of the previous traits. Technologies like (visible or infrared) light, (ultra or audible) sound, (Earth or artificial) magnetic fields, computer vision, UWB, WiFi, BLE, RFID, NFC, Zigbee, GSM, have been used to build solutions that highly varied in complexity and cost as well as in accuracy and coverage. The iconic accuracy and coverage chart from the habilitation thesis by Mautz [10] from seven years ago still holds true.

The combination of all the previous traits sets the bar high for any proposal. The fist trait rules out proposals based on technologies like NFC or computer vision. The second trait discards those based on technologies that require signal direction measurement, like light-based or sound-based approaches. The third trait may exclude those based on technologies like pseudolites, UWB or artificial magnetic fields. The fourth trait rejects current proposals based on technologies like WiFi, BLE, RFID,

GSM or Zigbee. Although the last two are rather related to system characterization or design restrictions, the first four traits are enough to support the idea that current proposals are not definitive solutions for indoor positioning and research should continue until new disruptive technology may finally arrive, much like GPS did for outdoors.

Until the breakthrough may finally arrive, the technologies used in current proposals so far have margins for improvements. Of that many agree and though those achievable improvements may not be enough for the ultimate IPS, they may approach to the requirements of many LBS applications which, for example, do not require accuracies as low as stated by the accuracy trait. Among the proposed IPS, those based on the measured signal strength (RSS) from WiFi access points have been particularly popular. Apart from the accuracy trait, those proposals have complied to several degrees with the other traits. The smartphone-based WiFi RSS-based IPS have typical accuracies of 4 m to 7 m [11], which is enough for some applications that require positioning in large indoor areas.

Much like with the positioning process performed in our mind, an (indoor) positioning system required previous information of the environment to then compute position estimates. In the case of systems based on radio frequencies, that knowledge is the position and characteristics of the emitters or the characterization of how the emitters' signals are perceived across an environment. Once the features of the environment are known, a system needs to be trained for the environment. The model that translates feature indications to position estimation needs to be discovered or tuned. Finally, the system needs to be assessed, so that corrective actions may be taken if necessary. Therefore, an WiFi-based IPS construction may transit through three well-differentiated stages: Environment Characterization, Model Setup, and Evaluation. This dissertation focuses on the first and third stages, mainly for the case of smartphone-based WiFi RSS-based IPS, which typically apply the fingerprinting technique.

1.2 Contributions

Many studies in IPS focus on the application of data analysis and machine learning to use the hardware already present in modern indoor environments, instead of developing radical new technologies. Such focus is in part a result of the current development and high interest in data processing techniques. This dissertation is part of that flow, but it states boundaries on what may be expected from current indoor positioning systems while describing solutions that allow achieving practical implementations of indoor positioning systems. This dissertation provides an end-to-end overview of a smartphone-based WiFi RSS-based IPS, showing *practical improvements across the*

Environment Characterization and Evaluation stages of the construction of such IPS to enforce the thesis that the IPS improvement is an integrative process of multiple refinements. Therefore, refinements that improve upon the state of the art across the three stages are part of the contributions of this thesis. In particular, the contributions are:

1. [Chapter 2]A meta-review of IPS research, sharing insights on the abundance, significance and the provided actual improvement of academic publications. Also, the most prominent of the found works are highlighted and a general overview of IPS is provided.

2. [Chapter 3]Two RSS datasets, one from WiFi measurements and the other from BLE measurements. The datasets, which are freely and openly available, were collected, curated and described to foster IPS development and evaluation.

3. [Chapter 4]Refinements to the environment characterization for WiFi fingerprinting to reduce the effort required for radiomap creation. The effort reduction is addressed from tools that ease the collection, the selection of reference points and the enrichment of radio-maps using regression methods.

4. [Chapter 5]Refinements to robustness to changes in the environment. The robustness is dealt with the selection of survey points for radio-map update and the identification of the non-nominal situations that affects positioning.

5. [Chapter 6]Refinements to evaluation procedures that ease comparisons among IPS regarding their accuracy and that improves positioning errors measurement. The evaluation procedures are addressed regarding insights from experiences of IPS competition challenges. Also, a novel measurement procedure of positioning errors and the local analysis of positioning errors are provided.

Chapter 7 ends this dissertation by providing concluding remarks.

References

[1] ABIresearch©. *Retail Indoor Location Market Breaks US$10 Billion in 2020.* `https://www.abiresearch.com/press/retail-indoor-location-market-breaks-us10-billion-/`. Online; accessed 22/07/2019. 2015.

[2] Anahid Basiri et al. "Indoor location based services challenges, requirements and usability of current solutions". In: *Computer Science Review* 24 (May 2017), pp. 1–12. DOI: `10.1016/j.cosrev.2017.03.002`. URL: `https://doi.org/10.1016/j.cosrev.2017.03.002`.

[3] Allan Brimicombe and Chao Li. *Location-based services and geo-information engineering*. Vol. 21. John Wiley & Sons, 2009.

[4] Edward H Cornell, Autumn Sorenson, and Teresa Mio. "Human sense of direction and wayfinding". In: *Annals of the Association of American Geographers* 93.2 (2003), pp. 399–425.

[5] Street Fight. *What Comes Next for Indoor Navigation? Enterprise Success, SMB Struggles*. https://streetfightmag.com/2019/09/27/what-comes-next-for-indoor-navigation-enterprise-success-smb-struggles/. Online; accessed 01/10/2019. 2019.

[6] Forbes. *Everything You Need To Know About The New Apple U1 Chip*. https://www.forbes.com/sites/quora/2019/09/12/everything-you-need-to-know-about-the-new-apple-u1-chip/#1680cf825df6. Online; accessed 01/10/2019. 2019.

[7] Reginald G Golledge. "Place recognition and wayfinding: Making sense of space". In: *Geoforum* 23.2 (1992), pp. 199–214.

[8] NEIL E KLEPEIS et al. "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants". In: *Journal of Exposure Science & Environmental Epidemiology* 11.3 (July 2001), pp. 231–252. ISSN: 1559-064X. DOI: 10.1038/sj.jea.7500165. URL: https://doi.org/10.1038/sj.jea.7500165.

[9] Markets&Markets. *Indoor Location Market by Positioning Systems, Maps and Navigation, Location based analytics, Location based services, Monitoring and emergency services. Worldwide Market Forecasts and Analysis (2014-2019)*. http://www.researchandmarkets.com/reports/2570920. Online; accessed 22/07/2019. 2014.

[10] Rainer Mautz. "Indoor positioning technologies". en. Habilitation Thesis. ETH Zurich, Department of Civil, Environmental, Geomatic Engineering, Institute of Geodesy, and Photogrammetry, 2012. DOI: 10.3929/ethz-a-007313554. URL: http://hdl.handle.net/20.500.11850/54888.

[11] Antoni Pérez-Navarro et al. "Challenges of Fingerprinting in Indoor Positioning and Navigation". In: *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*. Elsevier, 2019, pp. 1–20. DOI: 10.1016/b978-0-12-813189-3.00001-0. URL: https://doi.org/10.1016/b978-0-12-813189-3.00001-0.

[12] Jonathan Raper et al. "A critical evaluation of location based services and their potential". In: *Journal of Location Based Services* 1.1 (2007), pp. 5–45.

[13] Wilson Sakpere, Michael Adeyeye Oshin, and Nhlanhla BW Mlitwa. "A State-of-the-Art Survey of Indoor Positioning and Navigation Systems and Technologies". In: *South African Computer Journal* 29.3 (Dec. 2017). DOI: 10.18489/sacj.v29i3.452. URL: https://doi.org/10.18489/sacj.v29i3.452.

[14] Monika Śmiełowska, Mariusz Marć, and Bożena Zabiegała. "Indoor air quality in public utility environments—a review". In: *Environmental Science and Pollution Research* 24.12 (2017), pp. 11166–11176.

[15] Cezary Specht, Adam Weintrit, and Mariusz Specht. "A history of maritime radio-navigation positioning systems used in Poland". In: *The Journal of Navigation* 69.3 (2016), pp. 468–480.

[16] J. Talvitie et al. "High-Accuracy Joint Position and Orientation Estimation in Sparse 5G mmWave Channel". In: *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*. May 2019, pp. 1–7. DOI: 10.1109/ICC.2019.8761910.

[17] Zain Bin Tariq et al. "Non-GPS Positioning Systems". In: *ACM Computing Surveys* 50.4 (Aug. 2017), pp. 1–34. DOI: 10.1145/3098207. URL: https://doi.org/10.1145/3098207.

[18] Ali Yassin et al. "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications". In: *IEEE Communications Surveys & Tutorials* 19.2 (2017), pp. 1327–1346. DOI: 10.1109/comst.2016.2632427. URL: https://doi.org/10.1109/comst.2016.2632427.

[19] Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. "A Survey of Indoor Localization Systems and Technologies". In: *IEEE Communications Surveys & Tutorials* (2019), pp. 1–1. DOI: 10.1109/comst.2019.2911558. URL: https://doi.org/10.1109/comst.2019.2911558.

Chapter 2

Positioning Systems for Indoor Scenarios

An accurate and reliable IPS applicable to most indoor scenarios has been sought for many years. The number of technologies, techniques, and approaches in general used in IPS proposals is remarkable. Such diversity, coupled with the lack of strict and verifiable evaluations, leads to difficulties for appreciating the true value of most proposals. This chapter first provides an introduction to the different technologies, techniques and some methods employed for indoor position estimation. Later, it focuses on WLAN-based fingerprinting, which is the main focus of this dissertation. The chapter describes the inner workings of WLAN-based fingerprinting, its variants, and known challenges. Finally, the chapter provides a critical discussion on the current state of IPS academic research from the results of a meta-review of 61 surveys on IPS from the last 5 years.

2.1 The Case of Indoor Positioning

Location and position are sometimes used in an interchangeable manner. Through this dissertation, however, the terms are given distinct meanings. *Positioning* is understood as the action of determining something's or someone's position. *Position* is the answer to the question "where something/someone is" [108]. For Pérez-Navarro et al. [101], "position corresponds to the coordinates of a specific point in a coordinate system", a definition that is also supported by Brena et al. [16]. The terms position and coordinate system are widely assumed to be defined with respect to the Earth surface. They can also be defined locally to cover only a given surface or volume, e.g., the floor of a building. The term *location* [101], or *place* [16], refers to a specific area or point that is used as a reference, i.e., a location can give context to a position.

Despite the term *Indoor Location Systems* appears largely in the indoor positioning literature, this dissertation always refers to *Indoor Positioning Systems* as its focus. Another important term is *navigation*. According to the USA Federal Radio Navigation Plan [129], navigation is "the process of planning, recording, and controlling the movement of a craft or vehicle from one place to another". This dissertation prefers to extend the subject of the action to any entity moving between two locations and explicitly state that the connecting route may have given requirements, thus favoring the definition from Pérez-Navarro et al. [101].

Positioning systems can be global or local. *Global positioning systems* can provide position estimations world-wide. The global positioning systems currently available are collectively called the Global Navigation Satellite System (GNSS). Each GNSS (sub)system provides positioning, navigation, and timing determination capabilities and employs individual constellations of satellites that operate in conjunction with ground stations [129]. The best-known GNSS system is GPS. GPS has had a great impact on the development of Location-Based Services (LBS) since it became fully operational in 1996 [108]. The Russian GLONASS, the European Galileo, and the Chinese Beidou are currently the other GNSS systems, although their coverage does not match that of GPS.

The GNSS-based positioning has had large success as a result of its availability, coverage, and the existence of receivers that are both cheap and compact-sized. However, it is not adequate for all scenarios and applications because of accuracy requirements and the degradation of satellite signals. *Local positioning systems* are the ones applied for those specific scenarios and applications where GNNS positioning is not appropriate. The locality or coverage of those positioning systems varies significantly. They range from systems based on networks of pseudolites that can cover very large areas to light-based systems that are typically applied to rooms. This dissertation focuses on local positioning systems intended to provide position estimation inside buildings, which are known as *Indoor Positioning Systems*.

The materials and structures of modern buildings may influence notably on the signal from GNNS. At indoors, those signals reach receivers with a level of degradation that makes the civilian-graded accuracy of GPS insufficient for many indoor applications. Furthermore, indoor environments are normally crowded with fixed and moving obstacles, including people. The obstacles interact in undesirable ways with the signals, causing reflections and absorption. Given that people are increasingly spending most of their time indoors [101], the large efforts devoted along the past 15 years to find new solutions to IPS seems reasonable.

Even if GNSS signals were not so significantly degraded by buildings, local positioning systems will still be needed for indoor LBS applications. Some applications require positioning accuracies far beyond those GNNS may provide. For example, a

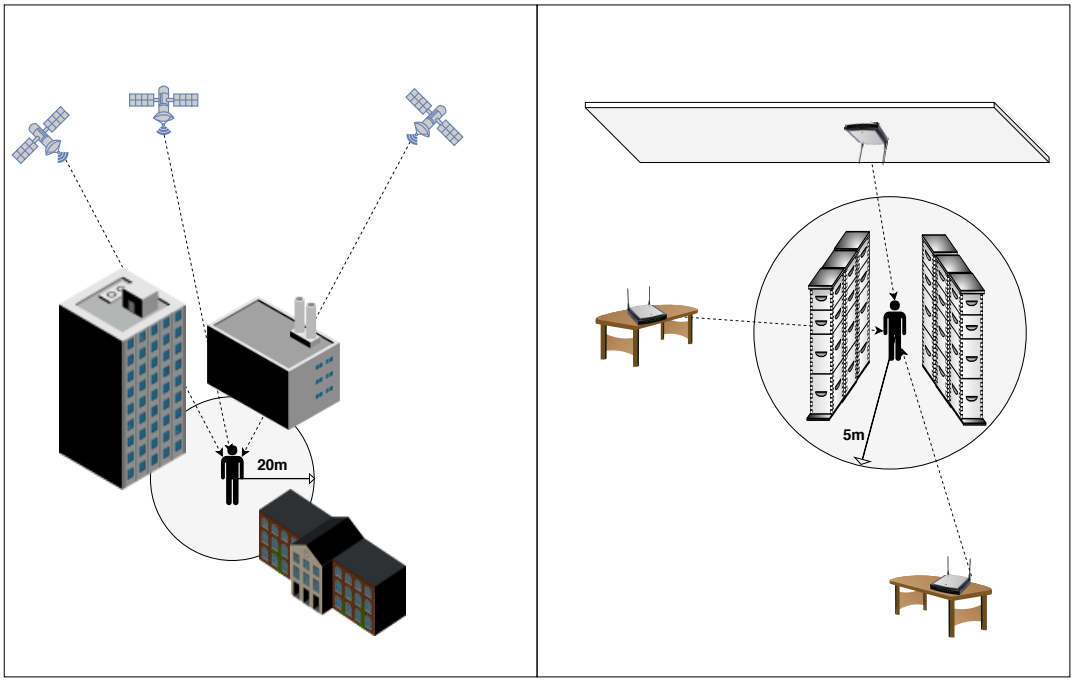


Figure 2.1: The magnitude of a positioning error matters differently to a person depending on the context.

robot performing precise operations and whose next move could not be determined in an off-line planning step will probably require a millimeter accuracy. Furthermore, the accuracy requirements for pedestrian applications may be related to the size of the target environment. The comfort thresholds of people regarding the positioning accuracy can be related to the free space they have around, as depicted in Figure 2.1. The typical accuracy for a smartphone's GNSS receptor under open sky conditions is 4.9 m [130], which is fine for someone at a park or wide street. However, that distance may be misleading in indoor environments. For example, a mean accuracy of 5 m would be insufficient to tell apart consecutive lanes among bookshelves in a library, which sometimes are less than 1 meter apart from each other. Therefore, an IPS applicable to most indoors environments and applications would preferably have mean accuracies around 1 m or 2 m.

Accuracy is the degree of conformance, or the closeness, between an estimated or measured position and the true position of a subject or object at a given time [129]. The above accuracy definition is wide. However, given the significant number of different existing (indoor) positioning systems, it is difficult to have a narrower definition or only one metric [57, 102]. Despite accuracy is of utmost importance, it

is not the only criterion taken into account for assessing an IPS. Coverage, complexity, robustness, scalability, cost, privacy and power consumption are metrics typically used for IPS evaluation and comparison [12, 123, 110, 152, 156]:

- Coverage refers to the range of the signals from the technology that supports an IPS [12]. A high coverage can translate into the IPS' applicability to large areas using a low number of emitters [156].
- Complexity refers to the efforts required for the construction, deployment or configuration of the hardware and software of the IPS [123, 152, 110].
- Robustness is the system's resilience to conditions beyond those that are considered nominal [123].
- Scalability relates to the system's ability to provide positioning for a large number of users in large spaces [123, 156].
- Cost refers to any kind of cost related to the positioning devices or the required infrastructure [12, 123].
- Privacy is related to system restrictions that avoid the collection of information that may be used to identify or track the users [12, 110, 152].
- Finally, the lower the power requirements, the better. In user devices, a low power requirement translates into a low battery drain. For the IPS infrastructure, a low power consumption may translate into that only small efforts are required for the maintenance of, e.g., battery-powered devices [12, 156].

2.2 An Overview of IPS Solutions

The high demand for IPS has driven an increasing number of research works along the past 15 years or more. The number of IPS proposals found in this period is very large, which is reflected in the number of references of recent IPS-related survey works. Some IPS-related reviews have over 200 references [17, 51, 122, 65, 71, 109] or even over 300 references [94]. The insights presented in the rest of this chapter are mainly based on 61 IPS-related survey works [1, 48, 54, 60, 84, 89, 116, 119, 150, 153, 3, 13, 17, 19, 73, 25, 26, 51, 70, 91, 83, 99, 118, 122, 132, 143, 12, 16, 21, 22, 32, 65, 80, 87, 95, 98, 110, 121, 123, 142, 152, 47, 71, 90, 94, 111, 112, 128, 140, 159, 161, 2, 58, 64, 72, 78, 23, 101, 109, 131, 134, 156] published between 2015 and 2019. Of them, 47 were published in periodic publication journals. Figure 2.2 presents the main details of the steps followed to identify the 61 survey works. Additionally, Table 2.1 shows the number of the selected surveys that were published each year.

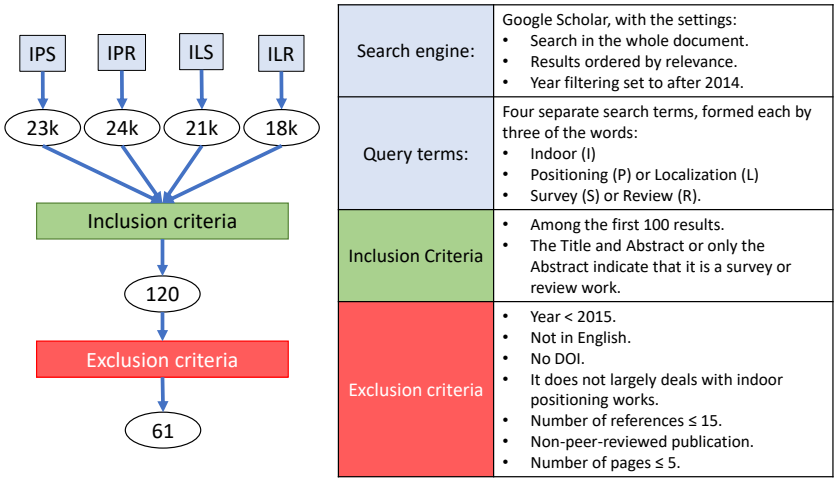


Figure 2.2: Surveys identification process followed for this work. The numbers inside the ellipses indicate the number of works obtained at each step.

Table 2.1: Publication year distribution of the selected surveys.

Year	2015	2016	2017	2018	2019
Number	12	16	13	13	7

The notable number of surveys has driven them into narrow focuses. However, some of them have addressed a broad spectrum of solutions applicable to indoor positioning [143, 12, 16, 21, 110, 152, 123, 47, 156]. There is not a clear consensus among this "general", or other, surveys on a strict taxonomy for IPS. The most commonly found classification attends to the underlying technology. The technology then determines which physical quantities the system can measure. The physical quantities are also used for IPS classification and are commonly addressed as the applied techniques. On top of technologies and techniques, there is a myriad of methods that, from the measured quantities, create position estimations. The methods also tend to be classified into range-based or range-free methods, depending on whether they estimate distances or angles to known landmarks like, e.g., the signal emitters.

The following list briefly introduces the most common techniques employed with technologies used in IPS. They are addressed in almost all of the selected surveys and they are often mentioned in the rest of this section. General surveys [83, 123, 94] include explanation and usages of other techniques like Received Signal Phase (RSF), Roundtrip Time of Flight (RTF) or Channel State Information (CSI).

- Time of Arrival (TOA). It measures the time of arrival of the signal from an

emitter, as recorded by the receiver. It is used for estimating the distance to each emitter, as the propagation speed of the signal (sound, radio frequencies) is known for the transmission medium (air).

- Time Difference of Arrival (TDOA). It is similar to TOA. It measures the differences in the time of arrival of signals from different emitters. It is used for estimating differences in distances to each emitter.

- Angle of Arrival (AOA). It refers to the angle at which the signal reaches the sensor. Angles are then used to obtain a position fix.

- Received Signal Strength (RSS). It is the intensity at which the signal from an emitter is measured. The signal strength decreases as the distance to the emitter increases, although their relation may be affected by attenuation and interference.

Some methods employed for position determination use the above techniques. For them, the technique employed for a solution determines how the position is estimated. They include the lateration, angulation and fingerprinting methods. Other methods do not depend on measuring signals from an emitter, like the hop count and vision analysis methods. The following list briefly describes the above mentioned position determination methods:

- Lateration. TOA, TDOA, and RSS are used for estimating distances to signal emitters. The estimated distances to a set of emitters are then used in what is called lateration to find the position estimate that best fit the set of distances (see Figure 2.3a). Lateration is called trilateration if three distances are used, while it is called multilateration if more than three are used.

- Angulation. The angles obtained in AoA are used to compute a likely fix on the target position, as shown in Figure 2.3b, in what is known as angulation.

- Fingerprinting. Sometimes called scene analysis, the fingerprinting encompasses two stages. In the first stage, also known as offline stage, the signal quantity of each detected emitter at a given time and position (a fingerprint) is measured at several places in the target scenario and stored to create a characterization of the signals in that scenario as comprehensive as possible. The collected database is called the training database. If the measured signals are radio frequencies (RF), it is also called radio map. The collection process of the database is called site survey, war-driving, radio map creation or training fingerprints collection. In the second stage, also known as online stage, the position corresponding to new measured signal quantities is estimated using the positions associated to

(a) Lateration (b) Angulation (c) Fingerprinting

Figure 2.3: Most common methods used in IPS.

the stored fingerprints that are the most similar when compared to the new measurements (see Figure 2.3c).

- Hop Count. A hop occurs when a packet passes from a network segment to the next one. The number of hops from known nodes is then used for (coarse) distance estimations [122] or to infer positions through a graph embedding problem [71].

- Vision Analysis. Vision analysis refers to the application of computer vision approaches to images gathered using some imaging technique like, e.g., cameras. The analyses detect relevant features in the scene that allow the estimation of the positions of entities in the scene or the position of the recording imaging device [6].

Both lateration and angulation are commonly classified as range-based – or ranging – methods, and they require the previous knowledge of the positions of the emitters. Fingerprinting, hop count and vision analysis are considered range-free methods.

Another relevant high-level classification is device-based or device-free positioning [143]. In the former, there is a positioning device that acts as an active agent of the positioning process by measuring or emitting a signal. The positioning device is normally a portable device like a smartphone or a tag. Device-based solutions are the most commonly addressed in surveys and IPS literature in general. Device-free positioning is found mainly in the form of presence detection or radar-like systems. The importance of the device-based vs device-free division is supported by the fact that one of the driven forces for the growth of LBS is the widespread usage of smartphones. Some surveys explicitly focus on IPS solutions applicable to smartphones [83, 22] and others on device-free positioning [95, 121]. The distinction on whether an IPS

is applicable to smartphones or not is necessary given that even modern smartphones have a limited number of sensory capabilities.

Another classification is whether the solution requires the deployment of dedicated infrastructure or not, being the former called an infrastructure-based solution and the later an infrastructure-free solution [101].

The underlying technology does matter and its particularities should be taken into account when creating an IPS. This section proceeds by providing a review of the technologies most commonly applied for IPS solutions. In each case, relevant references to the applied techniques and the applicability to range-based vs range-free or device-based vs device-free classification will be mentioned. The following subsections will comment on each of the most popular technologies used for IPS.

2.2.1 Light

Among the reviewed surveys, six of them specifically dealt with Visible Light Communication (VLC) IPS solutions [48, 25, 73, 80, 161, 2], which mainly considered LEDs as light sources. In addition, infrared – which could also be LED-based – is mentioned in general surveys [109, 16]. VLC IPS appear as device-based solutions, while infrared IPS may appear both as device-based or device-free (passive) solutions.

The VLC-based IPS rely on the idea that LED lighting is increasingly popular and LEDs can switch the intensity level in a way that is fast and imperceptible to the human eye. The intensity level switches are then used to encode information. Settings that have several LEDs emitters use a multiplexing protocol, either by frequency or time. The receiver is commonly equipped with a photodiode array that captures signal properties like RSS, TDOA, and AOA, or with a camera that takes images of the transmitters. For the former, positioning approaches like lateration, angulation or fingerprinting can be applied. For the later, the coordinates that the LEDs have in the image are translated into coordinates of the environment, usually using several images and additional support hardware like accelerometers. The mean accuracies reported by VLC-based IPS are measured in centimeters, with many of them under the 30 cm [25, 161, 2]. However, as stated by Do and Yoo [25], the reported results should be taken with caution as the test scenarios and conditions vary significantly among these solutions. Also, challenges like emitters time synchronization [48] and robustness to solar light [25] remain.

Infrared signals have been also used as light-based IPS, either as device-based or device-free solutions. A notable example of the former is the pioneering work of the Active Badge Location System [135], where a tag emitted a code carried by an infrared signal that was captured by a network of sensors. Examples of the later are Passive infrared (PIR) solutions, which could use infrared cameras for imaging or

thermopile arrays that provide AoA measurements [143, 16, 121, 109]. According to the performed searches for this work, there is no infrared-based IPS survey and the number of such system proposals is low when compared to other IPS technologies, which may be motivated by the LOS requirement of infrared-based solutions.

2.2.2 Computer Vision

Apart from their support in VLC-based or infrared-based positioning, vision techniques are also used in camera-based IPS without any supporting light framework apart from those common to any modern building. One [6] of the found surveys deals with visual odometry, although three of the general surveys [16, 123, 110] and a SLAM survey [17] also address vision-based IPS. Despite Simultaneous Localization and Mapping (SLAM) is mainly based on cameras for sensory input nowadays, one should not enclose SLAM into vision-based methods. SLAM is possible using, for example, laser scanners, sonars or odometric data provided by wheel encoders [17].

For device-based solutions, one of the most straightforward solutions is based on markers like, e.g., printed QR codes. Marker-based IPS may even provide continuous estimations if the markers are processed in a stream of images (video) of the scene and perspective is used for refinement [110]. Visual odometry is commonly applied to the input from one or several cameras. The cameras can be monocular, stereo or omnidirectional, to estimating the position positioning subject. The cameras motion and subject speed are usually determined by applying methods like feature tracking or optical flow [6, 17]. Visual odometry surpasses other odometry technologies regarding cost and accuracy [6] and is supported by modern computer vision techniques and the computation power found in robots and mobile devices, like in the applications of Google's ARcore [40].

For device-free solutions, it is common that several cameras are installed in the target environment. The collection of images is then used to identify the targets in conjunction with environment details recorded for the target scenario [123].

2.2.3 Sound

Only one [128] of the survey works was specific to acoustic signals applied to IPS. Three general reviews, however, provide some details on device-based sound-based IPS [143, 16, 110]. Those four surveys make a distinction between systems operating on ultrasound – frequencies beyond 20 KHz – and those using audible frequencies. They reference as pioneering works the "Active Bat" [136], "Cricket" [103] and "Dolphin" [36] systems for ultrasound, and the "BeepBeep" system [100] for audible frequencies. The ultrasound-based IPS generally use TOA or TDOA to perform lateration,

although RSS has also been used [16, 109], even for fingerprinting [143]. To improve the accuracy, the ultrasound pulses are complemented with RF pulses [103] or BLE advertisements [74].

Despite the acoustic solutions have reported accuracies of less than 30 cm [152], they present notable changes as sound speed is affected by temperature and humidity. Furthermore, they suffer from interference of bouncing pulses and, in the case of ultrasonic systems, they require specialized hardware that is expensive to deploy [128, 156]. Most of the acoustic IPS presented in the surveys refer mainly to device-based IPS. There are examples of device-free IPS in the form of audible sound source positioning, and as ultrasonic radars [123], although they face strong challenges due to noises, echoes, and multiple sound sources [95].

2.2.4 Magnetic Fields

The work of Pasku et al. [98] is the one survey found that focus on the magnetic fields IPS. It classifies the solutions into those that use the natural Earth magnetic field, those that use DC (static) artificial magnetic fields, and those that use AC (time-varying) artificial magnetic fields. Other general surveys, like Zafari, Gkelias, and Leung [156], do not make such distinction, while others like Brena et al. [16] do make the distinction but concentrates on those that only use the Earth magnetic field, as a result of considering that it is the approach followed by most modern solutions. Indeed, the selection made by [16] is supported by the facts that such systems do not incur in deployment costs, they are readily applicable to smartphones [22], and their reported accuracies are in the range of a few meters.

Magnetic IPS use strength variations in the measured magnetic field to infer a position estimate. In the case of the ambient (Earth) magnetic field, such variations are usually caused by steel structures in the target indoor scenario, requiring the creation of a database with the recorded variations of the magnetic field strength [101]. The created database is later used by a fingerprinting method to compute position estimations. One of the most known works using this approach is Magicol [120], which is referred by three surveys [22, 98, 123].

IPS based on the Earth magnetic field require a collection effort and typically have lower accuracies than those based on artificially generated magnetic fields. However, they have a low cost, a low system complexity, and a large operating range [22]. Those based on artificial fields require coil-based systems that are power hungry and operate at short ranges but they are able to produce accuracies in the range of several centimeters [98]. All magnetic-based IPS are device-based IPS solutions, to the best of this work's knowledge.

2.2.5 Dead Reckoning

Dead Reckoning (DR) is applied to device-based IPS. Among the found surveys, three of them [150, 140, 23] focus on DR, one [21] focuses on its combination with wireless positioning, and one [131] focuses on step length estimation, a major part of pedestrian Dead Reckoning (PDR). Other general surveys also address PDR [110, 123]. Dead Reckoning refers to the estimation of the current position of a target based on a previously known position (a fix) of it and measurements of quantities that are used to describe its movement, e.g., heading and speed [110]. Such measurements are commonly obtained using accelerometers that sense translations, i.e., provide the acceleration magnitude at each of the three axes; gyroscopes that sense rotations, i.e., provide roll, pitch and yaw measurements; and magnetometers (compasses, which are not inertial sensors) that give orientation regarding the Earth magnetic poles, i.e., provide field strength measurement along three axes.

Modern smartphones include most of these sensors and also have the computing capabilities to perform PDR. PDR works have also used units that assemble inertial sensors, which are called Inertial Measurement Units (IMUs). IMUs are mounted mainly in feet and legs, although it has been reported also for waist mounts [140, 23]. The shoe-mounted setting has been the most popular, given that mechanics involving walking process and the foot allow re-calibrations at every step applying the Zero-velocity UPdaTes (ZUPT) method [23]. Filtering algorithms like (Extended or Unscented) Kalman Filtering and Particle Filter (PF) that combine all information are the core of DR [110, 140].

The movement or trajectory is usually estimated using two approaches. The systems that use the first approach are called Inertial Navigation Systems (INSs), and they perform the integration of the sensor data. The systems that used the second approach are called Step and Heading Systems (SHSs), and they detect and quantify steps and their headings [110, 123, 24, 140]. If possible, other inputs like maps and constraints are applied using filters to improve the resulting accuracy [140]. PDR has a low cost, it does not require external references, and it has a high accuracy when new positions are not estimated far apart from the last fix. However, it suffers from accumulative errors or drifts [110]. That is why PDR is usually used coupled with other technologies that support IPS and provide periodical estimations that help in correcting drifts.

2.2.6 Ultra-Wideband (UWB)

UWB is highly acknowledged as an IPS technology. Among the selected surveys, three of them specifically focused on UWB [119, 3, 87]. Also, UWB is dealt at different lengths by almost all general or wireless networks-based IPS surveys. The

three UWB surveys agreed in using the USA Federal Communications Commission (FCC) definition of UWB, which states that it refers to RF signals whose bandwidth is greater than 20% of the center carrier frequency, or is greater than 500 MHz. That large bandwidth is related to a key characteristic that is acknowledged in many works: UWB works by emitting precisely-timed very short pulses, of ≈ 200 ps (pulse width), with very low transmission power [92]. The low transmission power avoids interference to WiFi, BLE or similar. The very short pulse modulation renders UWB almost immune to multipath issues. Given that the inter-pulse period is large enough to unambiguously perform multipath resolution, NLOS paths are detected after the main pulse detection [87]. Also, UWB has penetration capabilities considerably larger than WiFi and BLE, being LOS situations less relevant for it [3]. Furthermore, the energy consumption is lower than other WLAN technologies like Bluetooth or WiFi [87].

UWB device-based positioning requires the deployment of tags, both emitters and receivers. UWB emitters have been reported either in fixed or mobile configurations, with tags having different sizes and shapes and being mounted or worn at different places on the positioning subject, e.g., mounted on the feet or the head [16, 109]. The modern smartphones commonly have no support for UWB. With UWB, the positioning is performed using any of the RSS, ToA, AoA, or TDoA techniques, depending on the tags design and their resulting capabilities. Therefore, the reported methods used for positioning are fingerprinting, lateration or angulation [119, 3, 87]. Given the effort implied in the radio map creation for fingerprinting, it is the least used method for UWB. The reported accuracies are typically below the 20 cm [87], which makes UWB attractive for many applications, as long as they can afford the UWB tags, both in terms of cost and the application requirements. However, the short nominal range of UWB [87] and the cost of UWB equipment [3] makes the scalability a severe issue.

For device-free positioning, UWB has been used by applying the principle of radar [3], which is attractive given the UWB capability of penetrating through walls. In a room with UWB emitters and receivers, a positioning subject creates reflections of the signals that, using the TOA and TDOA techniques, can be used to estimate the subject's position [110, 123].

2.2.7 WiFi

WiFi, or Wi-Fi, is the IEEE standard 802.11 for WLAN [84]. WiFi-based positioning is sometimes addressed by the name of WLAN positioning [16, 110, 152], which is the result of WiFi being the default technology for setting up a WLAN. WiFi is mentioned as an IPS supporting technology in all selected surveys. What is more, to the best of this work's knowledge, it is acknowledged in all published IPS proposals.

From the selected surveys, one is devoted to time-based WiFi IPS techniques [84], four are completely devoted to WiFi fingerprinting [13, 51, 65, 142], one focus on WiFi CSI techniques [64], and one addresses device-free solutions based on WiFi [66]. The remaining surveys address WiFi IPS at varying lengths.

WiFi operates on the bands of 2.4 GHz and 5.0 GHz [84], with typical channel widths of 20 MHz, 40 MHz, and 80 MHz. The signals from bands of 2.4 GHz travel farther, while those from bands of 5.0 GHz have wider channels and are more robust to fast fading [28]. Works that report the usage of CSI, ToF and AoA techniques are not uncommon [156]. However, the main applied technique is RSS, given that is the technique applicable to many modern smartphones [22]. Despite its popularity, IPS based on WiFi have many challenges that arise from the alterations that the RF signals suffer in indoor environments. Also, the WiFi-based proposals use the existing WiFi networks to achieve a low cost solutions, but those networks are commonly deployed for communication purposes and not for positioning [156]. Solutions for WiFi RSS positioning may apply lateration based on a propagation model if the AP positions are known, which is also known as model-based approach [65]. However, the most popular approach is fingerprinting [16] because of its consistently better accuracy results in comparison to lateration. Given their direct applicability to many smartphones, WiFi-based IPS are mainly device-based solutions.

There are also device-free solutions [66] where WiFi signals are used to estimate the position of a subject that does not carry a WiFi receiver. Those solutions commonly perform anomaly or motion detection first and then determine the position of the entities using techniques like RSS or CSI [143]. Fingerprinting, link-based schemes, and Radio Tomographic Imaging are among the employed methods [143, 66].

2.2.8 Bluetooth Low Energy (BLE)

To the best of this work's knowledge, no survey focuses on BLE-based IPS currently, although surveys covering wider BLE topics do exist [59]. BLE-base positioning is briefly acknowledged in most of selected surveys. Some of them devoted entire sections to address BLE-based IPS [122, 12, 16, 22, 110, 156]. BLE follows Bluetooth classic in the usage of frequency hopping to communicate. The BLE emitters are commonly called BLE beacons. They are small, advertisement-emitting devices available in many configurations that are attractive for IPS given their cost, privacy, and a low footprint on the smartphones battery and network traffic [28, 22]. BLE advertises at three channels of 2 MHz of width in the 2.4 GHz band [28]. The small channel width translates into a larger fast fading effect than that seen for WiFi, even for the 2.4 GHz band [28]. Nevertheless, BLE is sometimes seen as the most suitable positioning technology for indoor navigation and tracking currently [12]. Its suitability is supported by the relatively low cost of BLE emitters, their very low power con-

sumption that let them run on batteries for months, and a generalized capability of modern smartphones to read their advertisements [16].

BLE shares many similarities with WiFi at the 2.4 GHz band [22], and thus it has been used for positioning by applying the RSS, AoA, and ToF techniques, being RSS the most often applied technique [156]. To the best of this work's knowledge, BLE has been used only for device-based IPS. BLE has a low detection range, typically under 20 m [12]. Such a short detection range reduces its applicability to device-free solutions. The accuracies achievable using BLE are typically higher than that of WiFi, which is related to the usually higher density of deployed emitters [22]. Despite BLE beacons are generally cheap, the scalability to large scenarios may be an issue if dense deployments are required [110].

2.2.9 Radio Frequency Identification (RFID) and Near Field Communication (NFC)

Only one conference proceedings work of the selected surveys focuses on RFID [118], although another two general surveys [16, 110] briefly address it. The main components of RFID are electronic tags that store some data – usually an ID – and readers capable of obtaining through RF the data from those tags. The tags can be passive, active or semi-passive. Passive tags use energy from the reader's signal to transmit their data. Active tags have batteries and broadcast their data periodically. Semi-passive tags broadcast their data only upon the detection of a reader's signal [16].

As readers are usually larger and more expensive than tags, one common setting in RFID-based IPS is to deploy a large number of tags across the environment and have the readers being part of or carried by the positioning subjects [110, 118]. The setting where readers are fixed and tags are carried by the positioning targets is more convenient for supporting large numbers of positioning subjects [16]. The detection range of the tags depends on whether they are active or passive. For passive tags, the detection range also depends on the reader's signal power. The reported accuracy in RFID varies significantly, and it depends on how dense the deployment of tags or readers had been, with some works reporting accuracies below 30 cm [16]. Most of the RFID-based IPS are based on simple proximity or RSS-based lateration [118, 16].

NFC-based IPS have been considered as a variant of RFID-based IPS [16]. NFC allows two devices, usually smartphones, to communicate while in touch or close proximity. A few recent IPS proposals [110] have harnessed the NFC capability of many modern smartphones in combination with the deployment of NFC tags across an environment. However, such systems have the inconvenience of requiring the active participation of the positioning subject, who is required to approach the smartphone

to the tag. Thus, such systems are unable to provide continuous position estimates [16, 110].

2.2.10 Other Technologies and Particular Cases

Cellular networks, in the form of GSM, LTE, 5G or other, are frequently mentioned but rarely addressed in detail in the selected surveys. A few surveys [143, 123, 71, 111] cite some works that employed cellular network signals to perform positioning by applying proximity, RSS fingerprinting, and observed TDOA lateration, but not necessarily for indoor positioning. Indeed, with reported accuracies that are commonly above 50 m, there are no IPS that are based solely on cellular networks signals to the best of this work's knowledge. Thus, this technology is used as a signal of opportunity in conjunction with others like WiFi, BLE and, FM.

Basiri et. al. [12] refers to other technologies like high sensitivity GNSS receivers, pseudolites, and tactile sensors. It explains that the first two are expensive and the third one is harder to manage in crowded scenarios. The tactile sensors placed on the floor may have real applications. Their sensing capability is provided by relative simply technologies like piezoelectric sensors, buttons or capacitive touch surfaces that can provide high accuracy in an unobtrusive way to the user. Xiao et. al. [143] comments on the usage of FM radio signals for positioning, being mainly applied using fingerprinting. Despite Xiao et. al. acknowledges advantages of FM-based positioning like low power consumption in smartphones and robustness to obstacles, it cast doubts on FM usage for IPS devoted to smartphones because of the lack of RSS readings availability.

Wireless Sensor Networks (WSNs) do not constitute a technology upon which IPS are supported. However, they are commonly mentioned in the selected surveys as they may be applied to indoor scenarios and network node position determination is a relevant problem for them. Although the positioning topic in WSN is most commonly addressed for large outdoor scenarios, there were two [89, 90] of the selected surveys that specifically addressed the topic for indoor environments, while another two [19, 112] focused significant parts of their contents for the indoor case. A WSN is a collection of nodes able to communicate among them – or at least to the nearest neighbors – and perform some sensing task [90]. If the position of some nodes is unknown, the communication technology or some extra ranging capability is usually used for determining the nodes' positions [112]. If a ranging capability is used, then the TOA, TDOA, AOA and RSS techniques are used with the angulation and lateration methods, among others. Ranging may be energy prohibitive for some WNS, which thus mostly rely on interpreting the connectivity information between the nodes, i.e., using hop count [71].

Much like WSN, a highly heterogeneous network of devices organized under the concept of the Internet of Things (IoT) uses the connectivity or ranging capabilities of their nodes for position determination [156]. However, in IoT, the node's connectivity is generally less restrictive than in WSN. Thus, IoT is expected to boost the device-free positioning to enable applications like smart environments [94]. Of the selected surveys, three [94, 111, 72] focused indoor positioning and IoT, while another three [19, 71, 156] addressed the combination of IoT and WSN at various lengths.

Despite it is briefly mentioned in many of the selected surveys, only two of them [143, 94] provide some details on ZigBee applications to IPS. ZigBee is a standard for WPAN focused on providing short-range, low power and low data rates communication. ZigBee is prone to interference from signals operating at the same frequency. It has been used for device-based IPS using the RSS, TDoA, ToA and AoA techniques [94]. It has also been used for device-free positioning by analyzing the signal fading induced by human movement [143].

2.3 WiFi and BLE RSS Fingerprinting for Smartphone-based IPS

The IPS solutions applied to smartphones normally use technologies that depend on the propagation of RF signals in indoor environments. The indoor environment, however, is not benevolent for RF signals. Multipath originates as a consequence of reflections of the signals on obstacles. The multipath appreciated outdoors at a large scale is more significant indoor, given that obstacles are abundant and close to each other. In multipath, a signal pulse reaches a receiver as several components that may have an additive or subtractive effect on the signal power [129]. The effect of multipath depends, at least in part, on the bandwidth of the signal pulse. As reviewed in Section 2.2, UWB is less affected than WiFi, which in turn in less affected than BLE. Additionally, all measurements are affected by the random noise that results from, e.g., thermal and circuit noise. Also, the collisions that result from using a shared medium are common [84]. Furthermore, the human body significantly attenuates the signals from the 2.4 GHz band [37], which leads to obtaining measurements that notably differ depending on the receiver device orientation when it is held by a person.

The above challenges, in conjunction with the common difficulty of knowing the actual position of RF emitters, hinders the obtainment of proper accuracies for lateration methods. Thus, RF RSS fingerprinting has become widely popular among solutions for smartphone-based IPS, mainly for those using the WiFi or BLE. Fingerprinting is applied using either deterministic, probabilistic or machine learning approaches [22, 65].

The deterministic approach commonly refers to the k-Nearest Neighbors (kNN) method or a variant of it. In kNN, the new fingerprint used for position estimation is compared for similarity against fingerprint values previously stored in a database. The positions associated with the k most similar fingerprints are used to infer a position for new fingerprint. Apart from the k value and the way that the k positions are combined, the distance metric used for determining similarity among fingerprints is an important aspect to consider in kNN [125].

In probabilistic approaches, the stored fingerprint values are used to compute the probability distribution of the signal of each emitter at each point. Those distributions are used later to select the most likely positions for the new fingerprint using Bayesian theory based on signal strength. Given the Gaussianity of RSS behavior is commonly acknowledged, the computation of the probability distribution normally determines the μ and σ parameters of a Gaussian distribution. However, some studies state that Gaussianity assumption does not necessarily hold [65]. Probabilistic approaches include Bayesian networks, expectation-maximization, Kullback-Leibler divergence, Gaussian processes, and conditional random fields [51].

The machine learning approaches harness the development of methods and tools created in recent years for the field of Artificial Intelligence. The radio map is used to train Support Vector Machines (SVM), Artificial Neural Network (ANN), among others [142], normally in a supervised fashion. Then, the trained models are commonly used for regression on 2D positions, and classification for floor or building estimates.

It is acknowledged that IPS based solely on WiFi fingerprinting can offer 2 m to 3 m of mean accuracy, although most common figures show 6 m to 7 m. BLE can provide better accuracies than WiFi, in the order of 1 m to 2 m, but larger values are also possible [88]. The variations in the previous numbers not only relate to specifics of the approach applied, but also to the characteristics of the environment and the collected training fingerprints. Indoor environments are not equally detrimental for RF fingerprinting, varying significantly in the amount and materials of the obstacles, influence of electronic equipment and the dynamics of the people moving around. Also, the number and disposition of the emitters are highly relevant [124, 142]. An increase in the number of emitters (WiFi AP or BLE beacons) found in a scenario can improve IPS accuracy [88]. Furthermore, relying only on distant emitters leads to little differentiability among fingerprints of distinct reference points, thus deteriorating the accuracy.

The number and distribution of reference points can also influence the obtained accuracy. On one hand, large mean errors are expected for sparse collection over large environments [127], assuming that no densification strategy is used. On the other hand, the mean error tends to be small in small environments that are densely surveyed [126]. Additionally, the devices that measure the signals – the chipsets in

smartphones – have different levels of sensibility [51] and identify the signal strength in distinct ways [142]. Thus, the strength reported for the same signal may have distinct values for distinct devices. Furthermore, one device may measure and report the RSS of a signal while another device may not report any value for the same signal because the signal strength is beyond its sensibility threshold. The RSS values measured at a given point are also affected by the device orientation [142], which is a result of the type and disposition of the antennas in the device and the partial absorption of the WiFi and BLE signals by the human body.

The previous difficulties have long been known for studies on WiFi, and later BLE, fingerprinting IPS solutions. To cope with them, the most straightforward strategy is to collect for the target environment a large database of fingerprints that covers as much as possible the given scenario, that has as many as possible reference points and that collects several fingerprints at several directions using distinct devices. However, such collection is very time consuming. Apart from being cumbersome, it is not affordable sometimes. Therefore, one of the main challenges for (particularly WiFi) fingerprinting positioning has been the effort reduction for collecting the training fingerprints. The proposed solutions for alleviating the collection efforts mainly include:

- do the collection using crowdsourcing,
- apply a propagation model to estimate the expected RSS values, or
- perform a small site survey and apply regression techniques to enrich the radio map,

The crowdsourcing approach harness the potential of the explicit or implicit participatory actions of users [65]. In the explicit modality, the users manually tag the positions of the recorded measurements. The users' participation can be during the offline stage or the online stage. During the offline, the participants supply the tags required to have an operational positioning service [65]. During the online stage, participants only supply position tags when the system is unable to provide an estimate [54]. In the implicit modality, the user is not requested to provide position labels. Instead, the labeling is performed by the positioning system whenever is possible using occasional fixes provided, e.g., by the GNNS receiver. Also, the IPS can infer fixes using heuristics or learning methods based on the existing radio map and inertial measurements [54, 65].

The approach of creating the radio map by applying a propagation model is attractive and has been proposed [142]. However, this approach requires either previous information on the positions and broadcasting parameters of the emitters, or to estimate that information using a few samples. The radio map densification using regression has been effectively used in several studies using regressors like linear

regression, nonlinear Gaussian Process, Gaussian Kernel Learning, and augmented path-loss model, Support Vector Regression, and Random Forest [41, 52, 65].

Apart from the radio map creation, another acknowledged challenge in WiFi or BLE fingerprinting based IPS is the energy consumption reduction [101]. WiFi scans are very energy demanding. BLE scans demand less energy than WiFi scans. Furthermore, the deterministic approaches have to deal with large databases for large scenarios, which implies performing a large number of comparisons of vectors of potentially large dimensionality. Pérez-Navarro et al. [101] and He et. all. [51] mention several approaches to deal with the former issue. Khalajmehrabadi et. al. [65] mention approaches to deal with large fingerprint databases. Those approaches are mainly based on techniques like clustering by Binary AP Coverage, K-means clustering, Affinity propagation, and spectral clustering.

Other acknowledged challenges are to detect and adapt to changes in the infrastructure that affect the positioning [51], which is also related to radio map construction; the detection of outliers [65]; the device diversity issue, i.e., the adaptation to heterogeneous devices [65]; the proper evaluation of IPS proposals [22, 101]; and the accuracy improvement. Solutions based solely on WiFi or BLE have known accuracy limitations [101]. Accuracy improvement is the challenge most often addressed. It is not commonly addressed for approaches based solely on WiFi or BLE, but for solutions that combine several technologies and techniques. For example, WiFi or BLE is normally combined with PDR, and the estimations are corrected using filters like Kalman or particles filter [142] and map-matching [71, 101].

2.4 Discussion on IPS Current State

Researchers working in the field of indoor positioning have the notion that one of the traits of the current status of IPS is that there is no clear prevalent technology or method for IPS. The variety of environments and applications make it difficult to find a general solution applicable to most situations. Some surveys include evaluation metrics to compare the reviewed work and inform the reader about the characteristic of the most notable works selected by the survey's authors. Accuracy, cost, and scalability are among the most common metrics found in the surveys, as mentioned in Section 2.2.

Another trait of the current status of IPS is the large number of solutions that have been proposed along the years, while only a few have had large significance. To deeper explore this trait, this work compiled and linked the publications referenced by the surveys introduced in Section 2.2. The linkage is based on DOIs [97]. The idea is to construct an undirected graph that links the surveys and the publications referenced by them, using the DOI associated with each of them as link information.

This work acknowledges that this approach does not include every single academic paper. There are old papers, papers published in low profile conferences, technical reports, patents, books and online resources that do not have DOIs assigned to them. During the curation, there were publications [77, 11, 53, 67, 56, 117, 145, 33, 14, 30] with no DOI that, some of them, were cited 11 times in the selected surveys. Most of the publications with no DOI were cited two or three times in the surveys. The DOI approach was followed given that it is relatively simple and allows for automatic analysis. Furthermore, most recent academic papers published in known resources are given a DOI.

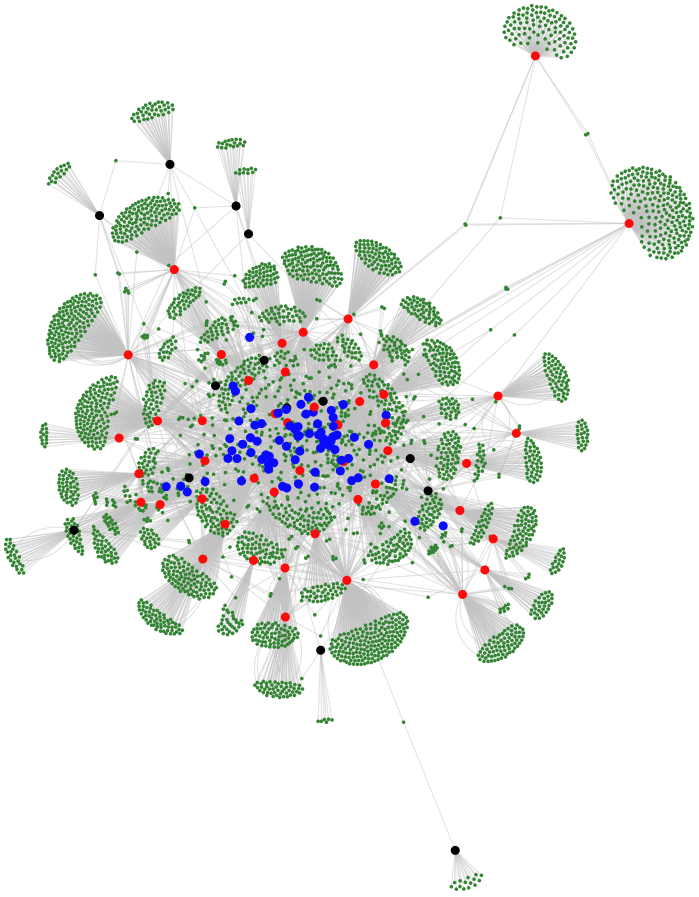


Figure 2.4: Graph of surveys and their referenced works. The selected surveys are identified by red (journal-published) and black (conference proceedings-published) dots. The publications that they referenced are represented by blue (more than 5 citations by the surveys) and dark green (5 or less citations by the surveys) dots.

The curation was performed using automatic and manual means. The automatic means included web scrapping for the references extraction and DOI discovery using

the Crossref's Link References matching tool [104]. After the application of automatic means, a manual revision was performed to correct inconsistencies produced both by the web scrapping and by the DOI discovery processes. The processing of the 61 selected surveys resulted in 3,901 unique works, including surveys distinct from the 61 selected. Figure 2.4 shows the resulting graph applied to all surveys, visualized using a force-directed approach [35]. Some surveys stand apart from others and lie in the outer parts of the image. Those divergent surveys are those with narrow focuses like SLAM [17], visual odometry [6], RFID [118], multidimensional scaling techniques [112], and those published in conference proceedings [90, 153, 60, 1]. It is of relevance to remark that some conferences apply a strict page limit, which in most of cases make the author to shorten the list of references to the most relevant ones. The surveys lying close to the center of the graph are those that reference works that are commonly referenced by other works. Those surveys mainly address WiFi-based solutions [13, 51], solutions that avoid or reduce the site survey efforts [54, 58], inertial sensors [150], WLAN-based solutions [65], solutions specific to pedestrians or smartphones [21, 22] and general surveys [143, 16, 152].

Figure 2.4 indicates with blue dots the works that were cited more than 5 times by the selected surveys. The number of citations of a work is determined by the degree of the node that represents it in a graph, i.e., the number of its incident edges. The number of highlighted works is 75. It is a rather low number compared to the total number of referenced works, even without taking into account the selected surveys. Indeed, the mean number of citations of a work without considering the selected surveys is 1.45. Such low number is significant not only because it hints at that most publications are cited only once, but also because the considered works also included surveys – which were published before 2015 and therefore were not among the selected surveys for this study.

Table 2.2 presents the number of citations of non-survey works, i.e., those that were not among the selected surveys, they do not include in their title the words "survey" or "review", and were not known to be a survey. Most of the referenced publications are only cited once in the selected surveys, while some of them are cited twice. The number of works that are cited more than five times is very small. The distribution in the number of citations might suggest that (1) the surveys tend to be very specific and barely intersect in content, (2) the proposed solutions quickly get obsolete or (3) it is hard for surveys in IPS to assess the relevance of the proposed solutions. The weight of the first conjecture seems to be low, given that despite some of the surveys are very focused, like Saeed et. al. [109], most of them deal with several techniques, which should result in a large number of intersecting works. The weight of the second conjecture seems more significant than the one from the first conjecture. Despite the obtained accuracies and the acknowledged challenges [101] have remained similar for several years [54], authors may try to incorporate the latest proposals to add value

to their surveys, regardless of the actual importance of the proposals. The third conjecture is supported by works on IPS that plead for better evaluation procedures and reproducible research [102]. With so many proposed solutions, the value of a work is not always correctly assessed without a proper evaluation that includes tests on publicly available data featuring well-known characteristics. The lack of usage of a common objective evaluation framework may lead to unworthy citations or simply the dispersion appreciated in Table 2.2 and Figure 2.4.

Table 2.2: Number of citations of non-survey works.

Citations	Percentage
1	80.0 %
2	11.3 %
3	3.9 %
4	2.0 %
5	1.1 %
$\geq$6	1.7 %

The works that had the largest number of citations were classics of the IPS literature: RADAR [10], Horus [155], LANDMARK [93] and Zee [107]. RADAR (27 citations, year 2000) is the best known of the first IPS solutions that applied WiFi-based fingerprinting, so it is commonly cited when a proposal deals with RSS fingerprinting, mainly for WiFi or BLE. Horus (16 citations, year 2005) is also one of the best known first applications of WiFi fingerprinting. The difference between RADAR and Horus is that the former applied a deterministic approach, while the latter used a probabilistic one. LANDMARK (15 citations, year 2003) is an early IPS based on RFID. Zee (15 citations, year 2012) is an early solution to the radio construction effort, which is one of the most acknowledge challenges of fingerprinting. It successfully combines PDR with map matching to make the WiFi radiomap to grow with self-tagged samples. Other works that have more than 10 citations are related to IPS based on UWB [38]; PDR and map-matching [137]; a combination of light, sound, WiFi and inertial sensor inputs [9]; WiFi without site survey [18] and [149]; a combination of WiFi, magnetic, PDR and inertial sensor inputs [133]; visible light [69]; and RF and ultrasound [103]. Apart from the value of their respective contributions, the previous works may be highly cited because they were early proposals of IPS using solutions that were novel at their times.

To analyze the focus of the selected surveys regarding technology, they were assigned a category. The selected categories and the number of surveys that belong to each category are presented in Table 2.3. Despite the data from the table may suggest that light-based IPS are as popular as WiFi-based IPS, many of the surveys

Table 2.3: Number of selected survey per technology.

Technology	Surveys
Light	6
WiFi	6
PDR	4
UWB	3
Magnetic	1
Sound	1
RFID	1
Vision	1
Several	38

that were classified in the category "Several" do not deal with light-based IPS, either because they are restricted to network-based technologies or technologies that are fully supported by smartphones. Furthermore, the category "Several" include surveys that are not devoted WiFi but focus on topics that are very relevant to WiFi-based IPS, like crowdsourcing, fingerprinting, radiomap construction, and IPS for smartphones. To further explore the focus of the current IPS research, this study extracted the 53 non-survey works most cited (more than 5 citations) in the selected surveys. Figure 2.5 shows the result of taking the titles of the extracted works [135, 136, 10, 103, 7, 154, 93, 62, 46, 82, 34, 38, 155, 50, 27, 100, 137, 49, 141, 9, 96, 18, 106, 55, 61, 43, 20, 42, 139, 31, 160, 115, 133, 149, 107, 76, 75, 4, 29, 147, 105, 113, 138, 146, 114, 63, 157, 86, 69, 68, 81, 151, 144] and creating a word cloud that highlights the occurrence frequency of each word in the titles of the extracted works. Punctuation signs as well as words non-relevant for comparison like 'indoor', 'system' or 'and' were removed.

The image presented in Figure 2.5 gives further support to the notion that WiFi is the technology most prevalent among IPS proposals. WiFi seems to be followed by light and wireless technologies – in which WiFi is also included. The 53 non-survey most-cited works were published before the year 2017, and almost half of them belong to the period 2011-2013. The trend in their publication years was expected given they are mostly pioneers in the indoor positioning field which have maintained their research value. It is difficult to forecast how long will they remain valuable. WiFi, BLE, and light seem to be the front-runners until a new technology – at least cheaper than UWB – becomes widely available for its use with smartphones. The chosen reviews that focused on light are from the years 2016-2019, which highlights the increasing importance that is given to this technology. Regarding WiFi, as long as smartphones keep being able to perform WiFi scanning at reasonable rates to

Figure 2.5: Word cloud of the works most cited (more than five citations) by the selected surveys.

provide near real-time positioning [5], research on WiFi-based IPS will persist given it is a technology that requires no infrastructure apart from the pervasive AP already in place. BLE, which provides better estimates than WiFi, will also remain on the choices for IPS. Indeed, it is popular among IPS provider companies.

The final batch of analyses from this section included the number of citations of the 53 non-survey most-cited works according to Google Scholar[39]. Figure 2.6 contrast the number of citations according to the selected survey and the number of citations according to Google Scholar. Also, a Pearson correlation test was applied to the two citation measures. The correlation results indicate a moderate, positive and significant correlation between the two citation measures. However, some works with a small number of citations in the surveys may also have a large number of citations in Google Scholar. An example is the work of Arulampalam et. al. [7], which provided a tutorial on particle filters. That work has only 9 citations in the surveys but over 10,000 citations in Google Scholar. Despite its applicability to IPS, it does not focus on IPS and it is relevant for a broad spectrum of topics related to positioning and navigation. The work that introduced the RADAR IPS [10] has fewer citations in Google Scholar that the work of Arulampalam et. al. [7], despite having 27 citations in the selected surveys.

The trend visible in the scatter plot from Figure 2.6 shows that works with a large number of citations in the surveys tend to have a large number of citations in Google Scholar. An additional search for works in indoor positioning was performed in Google Scholar to find works with a large number of citations that were not referenced in

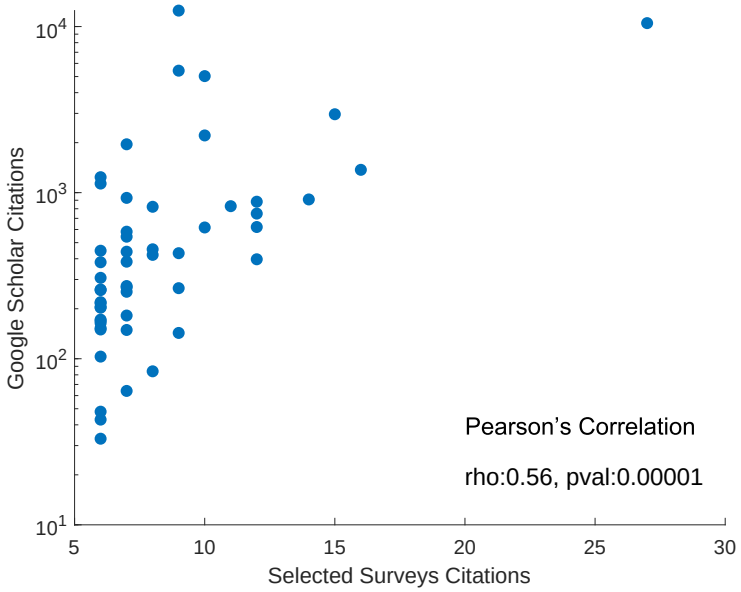


Figure 2.6: Number of citations in selected surveys vs citation from Google Scholar for the 53 non-survey most-cited works. A logarithmic transformation was applied to the y-axis to reduce the represented distance among data points.

the selected surveys. The search resulted in a few works published before 2011 with around 200 to 400 citations according to Google Scholar [44, 79, 8, 45, 85, 158, 148, 15]. Numbers of citations of 200 to 400 are not highly significant given that the median value of Google Scholar citations for the 53 works is almost 400. Thus, a work that has a large number of citation from surveys is likely to have a large number of citations in Google Scholar.

Figure 2.7 explores the relation of the number of citations according to the selected surveys with the years of publication of the extracted works. Figure 2.7 shows that, regarding the selected surveys, the number of citations of a work does not depend on the number year since its publication. The works corresponding to data points 35, 39, 40, 41 and 47 [18, 133, 149, 107, 69], published in or after 2010, have more citations than all other works apart from those corresponding to the data points 8, 19 and 29 [10, 93, 155]. Further analyses regarding the number of citations from Google Scholar showed a steady decrease in the number of citation along the years. Such decrease creates difficulties to assess the importance of a published work. Thus, Figure 2.8 presents the mean number of citations per year in Google Scholar. Figure 2.7 and Figure 2.8 used the same numbers for identifying works. Some works that Figure 2.7 showed to be relevant, like those from data points 8, 19, 39, and 40, are still noticeable in Figure 2.8. However, the mean number of citations per year seems to decrease along the years, thus suggesting that the mean number of citation per year for a published

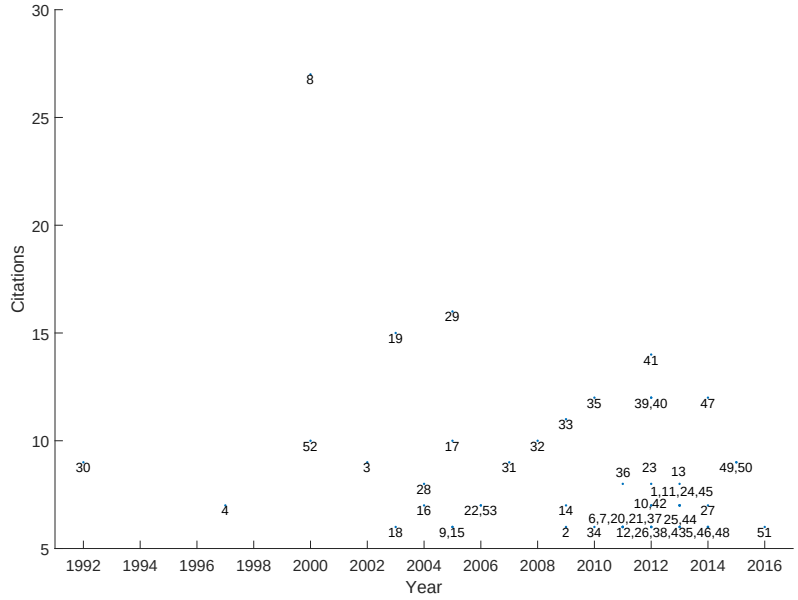


Figure 2.7: Citations in the selected surveys of the 53 extracted works. The x-axis represent the year of publication of the work.

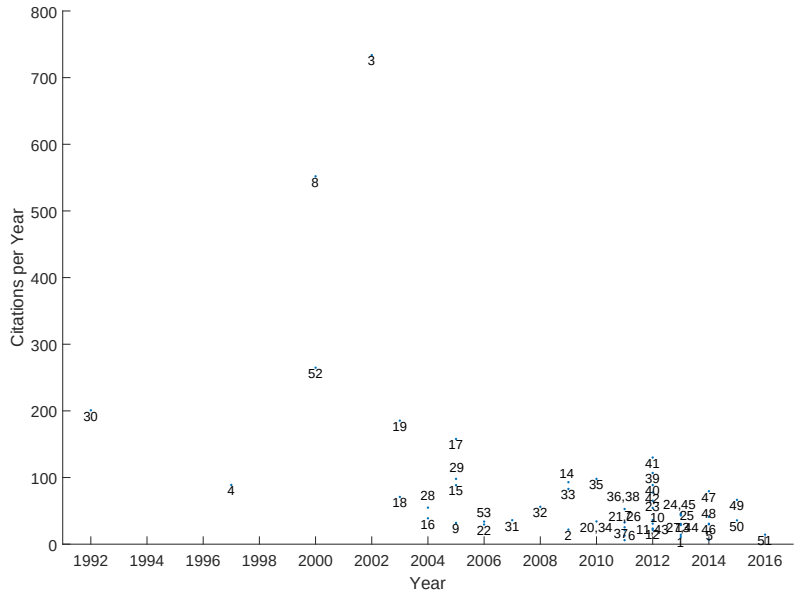


Figure 2.8: Mean number of citations per year in Google Scholar of the 53 extracted works. The x-axis represent the year of publication of the work.

work may indeed increase every year.

According to analyses presented in this section, the set of IPS-related surveys

seems to be a good mean to compile the works that had fundamental contributions to the IPS topic. However, each of them incorporates many works whose appreciated relevance is not shared among other surveys or even among the IPS literature in general.

2.5 Conclusions

This chapter provided an introduction to Indoor Positioning Systems and provided a meta-review of surveys published in recent years. The introduction to IPS and the meta-review addressed key concepts and the main technologies that currently support indoor positioning systems, briefly stating the techniques and methods applied in each of them. The chapter also compiled the most-recognized challenges in indoor positioning and deepened into positioning based on WiFi and BLE fingerprinting. The meta-review allowed a discussion on the influence of publications as seen using citation from the analyzed surveys and Google Scholar. The materials provided in this chapter allow the reader to use it as a handy collection that provides links to more specialized surveys. Also, the analyses from the meta-review show that most of the cited papers within a review are not disruptive. However, 75 works were identified to have more than 5 citations in the selected surveys, which are considered well-known or disruptive works.

References

[1] Stephan Adler et al. "A survey of experimental evaluation in indoor localization research". In: *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Oct. 2015. DOI: 10.1109/ipin.2015.7346749. URL: https://doi.org/10.1109/ipin.2015.7346749.

[2] Milad Afzalan and Farrokh Jazizadeh. "Indoor Positioning Based on Visible Light Communication". In: *ACM Computing Surveys* 52.2 (May 2019), pp. 1–36. DOI: 10.1145/3299769. URL: https://doi.org/10.1145/3299769.

[3] Abdulrahman Alarifi et al. "Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances". In: *Sensors* 16.5 (May 2016), p. 707. DOI: 10.3390/s16050707. URL: https://doi.org/10.3390/s16050707.

[4] Moustafa Alzantot and Moustafa Youssef. "CrowdInside: automatic construction of indoor floorplans". In: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems - SIGSPATIAL '12*. ACM Press, 2012. DOI: 10.1145/2424321.2424335. URL: https://doi.org/10.1145/2424321.2424335.

[5] Android Police. *Android started heavily throttling Wi-Fi scanning in Pie, Google confirms it's here to stay.* https://www.androidpolice.com/2019/07/10/android-started-heavily-throttling-wi-fi-scanning-in-pie-google-confirms-its-here-to-stay/. Online; accessed 22/07/2019.

[6] Mohammad O. A. Aqel et al. "Review of visual odometry: types, approaches, challenges, and applications". In: *SpringerPlus* 5.1 (Oct. 2016), p. 1897. ISSN: 2193-1801. DOI: 10.1186/s40064-016-3573-7. URL: https://doi.org/10.1186/s40064-016-3573-7.

[7] M.S. Arulampalam et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on Signal Processing* 50.2 (2002), pp. 174–188. DOI: 10.1109/78.978374. URL: https://doi.org/10.1109/78.978374.

[8] Abdalkarim Awad, Thorsten Frunzke, and Falko Dressler. "Adaptive distance estimation and localization in WSN using RSSI measures". In: *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007)*. IEEE. 2007, pp. 471–478.

[9] Martin Azizyan, Ionut Constandache, and Romit Roy Choudhury. "SurroundSense: Mobile Phone Localization via Ambience Fingerprinting". In: *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*. MobiCom '09. Beijing, China: ACM, 2009, pp. 261–272. ISBN: 978-1-60558-702-8. DOI: 10.1145/1614320.1614350. URL: http://doi.acm.org/10.1145/1614320.1614350.

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

[11] Victor Bahl and Venkat Padmanabhan. *Enhancements to the RADAR User Location and Tracking System*. Tech. rep. MSR-TR-2000-12. Microsoft Research, Feb. 2000, p. 13. URL: https://www.microsoft.com/en-us/research/publication/enhancements-to-the-radar-user-location-and-tracking-system/.

[12] Anahid Basiri et al. "Indoor location based services challenges, requirements and usability of current solutions". In: *Computer Science Review* 24 (May 2017), pp. 1–12. DOI: 10.1016/j.cosrev.2017.03.002. URL: https://doi.org/10.1016/j.cosrev.2017.03.002.

[13] Chaimaa BASRI and Ahmed El Khadimi. "Survey on indoor localization system and recent advances of WIFI fingerprinting technique". In: *2016 5th International Conference on Multimedia Computing and Systems (ICMCS)*. IEEE, Sept. 2016. DOI: 10.1109/icmcs.2016.7905633. URL: https://doi.org/10.1109/icmcs.2016.7905633.

[14] Alan Bensky. *Wireless positioning technologies and applications*. Artech House, 2016.

[15] Joydeep Biswas and Manuela Veloso. "Wifi localization and navigation for autonomous indoor mobile robots". In: *2010 IEEE international conference on robotics and automation*. IEEE. 2010, pp. 4379–4384.

[16] Ramon F. Brena et al. "Evolution of Indoor Positioning Technologies: A Survey". In: *Journal of Sensors* 2017 (2017), pp. 1–21. DOI: 10.1155/2017/2630413. URL: https://doi.org/10.1155/2017/2630413.

[17] Cesar Cadena et al. "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age". In: *IEEE Transactions on Robotics* 32.6 (Dec. 2016), pp. 1309–1332. DOI: 10.1109/tro.2016.2624754. URL: https://doi.org/10.1109/tro.2016.2624754.

[18] Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N. Padmanabhan. "Indoor Localization Without the Pain". In: *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*. MobiCom '10. Chicago, Illinois, USA: ACM, 2010, pp. 173–184. ISBN: 978-1-4503-0181-7. DOI: 10.1145/1859995.1860016. URL: http://doi.acm.org/10.1145/1859995.1860016.

[19] Tashnim J.S. Chowdhury et al. "Advances on localization techniques for wireless sensor networks: A survey". In: *Computer Networks* 110 (Dec. 2016), pp. 284–305. DOI: 10.1016/j.comnet.2016.10.006. URL: https://doi.org/10.1016/j.comnet.2016.10.006.

[20] Jaewoo Chung et al. "Indoor location sensing using geo-magnetism". In: *Proceedings of the 9th international conference on Mobile systems, applications, and services - MobiSys '11*. ACM Press, 2011. DOI: 10.1145/1999995.2000010. URL: https://doi.org/10.1145/1999995.2000010.

[21] Alejandro Correa et al. "A Review of Pedestrian Indoor Positioning Systems for Mass Market Applications". In: *Sensors* 17.8 (Aug. 2017), p. 1927. DOI: 10.3390/s17081927. URL: https://doi.org/10.3390/s17081927.

[22] Pavel Davidson and Robert Piche. "A Survey of Selected Indoor Positioning Methods for Smartphones". In: *IEEE Communications Surveys & Tutorials* 19.2 (2017), pp. 1347–1370. DOI: 10.1109/comst.2016.2637663. URL: https://doi.org/10.1109/comst.2016.2637663.

[23] Estefania Munoz Diaz, Dina Bousdar Ahmed, and Susanna Kaiser. "A Review of Indoor Localization Methods Based on Inertial Sensors". In: *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*. Elsevier, 2019, pp. 311–333. DOI: 10.1016/b978-0-12-813189-3.00016-2. URL: https://doi.org/10.1016/b978-0-12-813189-3.00016-2.

[24] L. E. Díez et al. "Step Length Estimation Methods Based on Inertial Sensors: A Review". In: *IEEE Sensors Journal* 18.17 (Sept. 2018), pp. 6908–6926. ISSN: 1530-437X. DOI: 10.1109/JSEN.2018.2857502.

[25] Trong-Hop Do and Myungsik Yoo. "An in-Depth Survey of Visible Light Communication Based Positioning Systems". In: *Sensors* 16.5 (May 2016), p. 678. DOI: 10.3390/s16050678. URL: https://doi.org/10.3390/s16050678.

[26] Felis Dwiyasa and Meng-Hiot Lim. "A survey of problems and approaches in wireless-based indoor positioning". In: *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Oct. 2016. DOI: 10.1109/ipin.2016.7743591. URL: https://doi.org/10.1109/ipin.2016.7743591.

[27] Frédéric Evennou and François Marx. "Advanced Integration of WiFi and Inertial Navigation Systems for Indoor Mobile Positioning". In: *EURASIP Journal on Advances in Signal Processing* 2006.1 (Apr. 2006). DOI: 10.1155/asp/2006/86706. URL: https://doi.org/10.1155/asp/2006/86706.

[28] Ramsey M. Faragher and Robert K. Harle. "Towards an Efficient, Intelligent, Opportunistic Smartphone Indoor Positioning System". In: *Navigation* 62.1 (Mar. 2015), pp. 55–72. DOI: 10.1002/navi.76. URL: https://doi.org/10.1002/navi.76.

[29] Ramsey Faragher and Rob Harle. "SmartSLAM-an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing". In: *ION GNSS*. Vol. 13. 2013, pp. 1–14.

[30] Ramsey Faragher and Robert Harle. "An analysis of the accuracy of bluetooth low energy for indoor positioning applications". In: *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*. Vol. 812. 2014, pp. 201–210.

[31] Chen Feng et al. "Received-Signal-Strength-Based Indoor Positioning Using Compressive Sensing". In: *IEEE Transactions on Mobile Computing* 11.12 (Dec. 2012), pp. 1983–1993. DOI: 10.1109/tmc.2011.216. URL: https://doi.org/10.1109/tmc.2011.216.

[32] Andre Filipe Goncalves Ferreira et al. "Localization and Positioning Systems for Emergency Responders: A Survey". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2836–2870. DOI: 10.1109/comst.2017.2703620. URL: https://doi.org/10.1109/comst.2017.2703620.

[33] Brian Ferris, Dieter Fox, and Neil Lawrence. "WiFi-SLAM Using Gaussian Process Latent Variable Models". In: *Proceedings of the 20th International Joint Conference on Artifical Intelligence*. IJCAI'07. Hyderabad, India: Morgan Kaufmann Publishers Inc., 2007, pp. 2480–2485. URL: http://dl.acm.org/citation.cfm?id=1625275.1625675.

[34] E. Foxlin. "Pedestrian Tracking with Shoe-Mounted Inertial Sensors". In: *IEEE Computer Graphics and Applications* 25.6 (Nov. 2005), pp. 38–46. DOI: 10.1109/mcg.2005.140. URL: https://doi.org/10.1109/mcg.2005.140.

[35] Thomas MJ Fruchterman and Edward M Reingold. "Graph drawing by force-directed placement". In: *Software: Practice and experience* 21.11 (1991), pp. 1129–1164.

[36] Yasuhiro Fukuju et al. "DOLPHIN: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment." In: *WSTFES* 3 (2003), p. 53.

[37] S. Garcia-Villalonga and A. Pérez-Navarro. "Influence of human absorption of Wi-Fi signal in indoor positioning with Wi-Fi fingerprinting". In: *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Oct. 2015, pp. 1–10. DOI: 10.1109/IPIN.2015.7346778.

[38] S. Gezici et al. "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks". In: *IEEE Signal Processing Magazine* 22.4 (July 2005), pp. 70–84. DOI: 10.1109/msp.2005.1458289. URL: https://doi.org/10.1109/msp.2005.1458289.

[39] Google. *Google Scholar*. http://scholar.google.com. Online; accessed 22/07/2019.

[40] Google Inc. *Google's ARcore*. https://developers.google.com/ar/. Online; accessed 22/07/2019.

[41] Rafał Górak and Marcin Luckner. "Modified Random Forest Algorithm for Wi–Fi Indoor Localization System". In: *Computational Collective Intelligence*. Ed. by Ngoc Thanh Nguyen et al. Cham: Springer International Publishing, 2016, pp. 147–157. ISBN: 978-3-319-45246-3.

[42] Abhishek Goswami, Luis E. Ortiz, and Samir R. Das. "WiGEM: a learning-based approach for indoor localization". In: *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies on - CoNEXT '11*. ACM Press, 2011. DOI: 10.1145/2079296.2079299. URL: https://doi.org/10.1145/2079296.2079299.

[43] Brandon Gozick et al. "Magnetic Maps for Indoor Navigation". In: *IEEE Transactions on Instrumentation and Measurement* 60.12 (Dec. 2011), pp. 3883–3891. DOI: 10.1109/tim.2011.2147690. URL: https://doi.org/10.1109/tim.2011.2147690.

[44] J-S Gutmann and Christian Schlegel. "Amos: Comparison of scan matching approaches for self-localization in indoor environments". In: *Proceedings of the First Euromicro Workshop on Advanced Mobile Robots (EUROBOT'96)*. IEEE. 1996, pp. 61–67.

[45] Ismail Guvenc, Chia-Chin Chong, and Fujio Watanabe. "NLOS identification and mitigation for UWB localization systems". In: *2007 IEEE Wireless Communications and Networking Conference*. IEEE. 2007, pp. 1571–1576.

[46] Andreas Haeberlen et al. "Practical robust localization over large-scale 802.11 wireless networks". In: *Proceedings of the 10th annual international conference on Mobile computing and networking - MobiCom '04*. ACM Press, 2004. DOI: 10.1145/1023720.1023728. URL: https://doi.org/10.1145/1023720.1023728.

[47] Anum Hameed and Hafiza Anisa Ahmed. "Survey on indoor positioning applications based on different technologies". In: *2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. IEEE, Nov. 2018. DOI: 10.1109/macs.2018.8628462. URL: https://doi.org/10.1109/macs.2018.8628462.

[48] Naveed Ul Hassan et al. "Indoor Positioning Using Visible LED Lights". In: *ACM Computing Surveys* 48.2 (Nov. 2015), pp. 1–32. DOI: 10.1145/2835376. URL: https://doi.org/10.1145/2835376.

[49] Janne Haverinen and Anssi Kemppainen. "Global indoor self-localization based on the ambient magnetic field". In: *Robotics and Autonomous Systems* 57.10 (Oct. 2009), pp. 1028–1035. DOI: 10.1016/j.robot.2009.07.018. URL: https://doi.org/10.1016/j.robot.2009.07.018.

[50] M. Hazas and A. Hopper. "Broadband ultrasonic location systems for improved indoor positioning". In: *IEEE Transactions on Mobile Computing* 5.5 (May 2006), pp. 536–547. DOI: 10.1109/tmc.2006.57. URL: https://doi.org/10.1109/tmc.2006.57.

[51] Suining He and S.-H. Gary Chan. "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons". In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 466–490. DOI: 10.1109/comst.2015.2464084. URL: https://doi.org/10.1109/comst.2015.2464084.

[52] Noelia Hernández et al. "Continuous Space Estimation: Increasing WiFi-Based Indoor Localization Resolution without Increasing the Site-Survey Effort". In: *Sensors* 17.1 (2017). ISSN: 1424-8220. DOI: 10.3390/s17010147. URL: https://www.mdpi.com/1424-8220/17/1/147.

[53] Jeffrey Hightower and Gaetano Borriello. "Location sensing techniques". In: *IEEE Computer* 34.8 (2001), pp. 57–66.

[54] A.K.M. Mahtab Hossain and Wee-Seng Soh. "A survey of calibration-free indoor positioning systems". In: *Computer Communications* 66 (July 2015), pp. 1–13. DOI: 10.1016/j.comcom.2015.03.001. URL: https://doi.org/10.1016/j.comcom.2015.03.001.

[55] Joseph Huang et al. "Efficient, generalized indoor WiFi GraphSLAM". In: *2011 IEEE International Conference on Robotics and Automation*. IEEE, May 2011. DOI: 10.1109/icra.2011.5979643. URL: https://doi.org/10.1109/icra.2011.5979643.

[56] F. Ijaz et al. "Indoor positioning: A review of indoor ultrasonic positioning systems". In: *2013 15th International Conference on Advanced Communications Technology (ICACT)*. Jan. 2013, pp. 1146–1150.

[57] ISO Central Secretary. *Information technology – Real time locating systems – Test and evaluation of localization and tracking systems*. en. Standard ISO/IEC 18305:2016. Geneva, CH: International Organization for Standardization, 2016. URL: https://www.iso.org/standard/62090.html.

[58] Beakcheol Jang and Hyunjung Kim. "Indoor Positioning Technologies Without Offline Fingerprinting Map: A Survey". In: *IEEE Communications Surveys & Tutorials* 21.1 (2019), pp. 508–525. DOI: 10.1109/comst.2018.2867935. URL: https://doi.org/10.1109/comst.2018.2867935.

[59] K. E. Jeon et al. "BLE Beacons for Internet of Things Applications: Survey, Challenges, and Opportunities". In: *IEEE Internet of Things Journal* 5.2 (Apr. 2018), pp. 811–828. DOI: 10.1109/JIOT.2017.2788449.

[60] Sun Jian et al. "A Survey and Application of Indoor Positioning Based on Scene Classification Optimization". In: *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*.

IEEE, Aug. 2015. DOI: 10.1109/uic-atc-scalcom-cbdcom-iop.2015.282. URL: https://doi.org/10.1109/uic-atc-scalcom-cbdcom-iop.2015.282.

[61] Soo-Yong Jung, Swook Hann, and Chang-Soo Park. "TDOA-based optical wireless indoor localization using LED ceiling lamps". In: *IEEE Transactions on Consumer Electronics* 57.4 (Nov. 2011), pp. 1592–1597. DOI: 10.1109/tce.2011.6131130. URL: https://doi.org/10.1109/tce.2011.6131130.

[62] K. Kaemarungsi and P. Krishnamurthy. "Properties of indoor received signal strength for WLAN location fingerprinting". In: *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.* IEEE, 2004. DOI: 10.1109/mobiq.2004.1331706. URL: https://doi.org/10.1109/mobiq.2004.1331706.

[63] Georg Kail et al. "Robust asynchronous indoor localization using LED lighting". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, May 2014. DOI: 10.1109/icassp.2014.6853922. URL: https://doi.org/10.1109/icassp.2014.6853922.

[64] Laxima Niure Kandel and Shucheng Yu. "Indoor Localization Using Commodity Wi-Fi APs: Techniques and Challenges". In: *2019 International Conference on Computing, Networking and Communications (ICNC).* IEEE, Feb. 2019. DOI: 10.1109/iccnc.2019.8685501. URL: https://doi.org/10.1109/iccnc.2019.8685501.

[65] Ali Khalajmehrabadi, Nikolaos Gatsis, and David Akopian. "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges". In: *IEEE Communications Surveys & Tutorials* 19.3 (2017), pp. 1974–2002. DOI: 10.1109/comst.2017.2671454. URL: https://doi.org/10.1109/comst.2017.2671454.

[66] D. Konings et al. "Device-Free Localization Systems Utilizing Wireless RSSI: A Comparative Practical Investigation". In: *IEEE Sensors Journal* 19.7 (Apr. 2019), pp. 2747–2757. ISSN: 1530-437X. DOI: 10.1109/JSEN.2018.2888862.

[67] Hakan Koyuncu and Shuang Hua Yang. "A survey of indoor positioning and object locating systems". In: *IJCSNS International Journal of Computer Science and Network Security* 10.5 (2010), pp. 121–128.

[68] Swarun Kumar et al. "Accurate indoor localization with zero start-up cost". In: *Proceedings of the 20th annual international conference on Mobile computing and networking - MobiCom '14.* ACM Press, 2014. DOI: 10.1145/2639108.2639142. URL: https://doi.org/10.1145/2639108.2639142.

[69] Ye-Sheng Kuo et al. "Luxapose: Indoor Positioning with Mobile Phones and Visible Light". In: *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*. MobiCom '14. Maui, Hawaii, USA: ACM, 2014, pp. 447–458. ISBN: 978-1-4503-2783-1. DOI: 10.1145/2639108.2639109. URL: http://doi.acm.org/10.1145/2639108.2639109.

[70] Luan D. M. Lam, Antony Tang, and John Grundy. "Heuristics-based indoor positioning systems: a systematic literature review". In: *Journal of Location Based Services* 10.3 (July 2016), pp. 178–211. DOI: 10.1080/17489725.2016.1232842. URL: https://doi.org/10.1080/17489725.2016.1232842.

[71] Christos Laoudias et al. "A Survey of Enabling Technologies for Network Localization, Tracking, and Navigation". In: *IEEE Communications Surveys & Tutorials* 20.4 (2018), pp. 3607–3644. DOI: 10.1109/comst.2018.2855063. URL: https://doi.org/10.1109/comst.2018.2855063.

[72] Bahareh Lashkari et al. "Crowdsourcing and Sensing for Indoor Localization in IoT: A Review". In: *IEEE Sensors Journal* 19.7 (Apr. 2019), pp. 2408–2434. DOI: 10.1109/jsen.2018.2880180. URL: https://doi.org/10.1109/jsen.2018.2880180.

[73] Steven De Lausnay et al. "A survey on multiple access Visible Light Positioning". In: *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*. IEEE, Aug. 2016. DOI: 10.1109/emergitech.2016.7737307. URL: https://doi.org/10.1109/emergitech.2016.7737307.

[74] Patrick Lazik et al. "ALPS: A Bluetooth and Ultrasound Platform for Mapping and Localization". In: *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. SenSys '15. Seoul, South Korea: ACM, 2015, pp. 73–84. ISBN: 978-1-4503-3631-4. DOI: 10.1145/2809695.2809727. URL: http://doi.acm.org/10.1145/2809695.2809727.

[75] Binghao Li et al. "How feasible is the use of magnetic field alone for indoor positioning?" In: *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Nov. 2012. DOI: 10.1109/ipin.2012.6418880. URL: https://doi.org/10.1109/ipin.2012.6418880.

[76] Fan Li et al. "A reliable and accurate indoor localization method using phone inertial sensors". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM Press, 2012. DOI: 10.1145/2370216.2370280. URL: https://doi.org/10.1145/2370216.2370280.

[77] Liqun Li et al. "Epsilon: A Visible Light Based Positioning System". In: *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. Seattle, WA: USENIX Association, 2014, pp. 331–343. ISBN: 978-1-931971-09-6. URL: https://www.usenix.org/conference/nsdi14/technical-sessions/presentation/li.

[78] Ziwei Li et al. "Machine-Learning-Based Positioning: A Survey and Future Directions". In: *IEEE Network* 33.3 (May 2019), pp. 96–101. DOI: 10.1109/mnet.2019.1800366. URL: https://doi.org/10.1109/mnet.2019.1800366.

[79] Konrad Lorincz and Matt Welsh. "Motetrack: A robust, decentralized approach to rf-based location tracking". In: *International Symposium on Location-and Context-Awareness*. Springer. 2005, pp. 63–82.

[80] Junhai Luo, Liying Fan, and Husheng Li. "Indoor Positioning Systems Based on Visible Light Communication: State of the Art". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2871–2893. DOI: 10.1109/comst.2017.2743228. URL: https://doi.org/10.1109/comst.2017.2743228.

[81] Dimitrios Lymberopoulos et al. "A realistic evaluation and comparison of indoor location technologies". In: *Proceedings of the 14th International Conference on Information Processing in Sensor Networks - IPSN '15*. ACM Press, 2015. DOI: 10.1145/2737095.2737726. URL: https://doi.org/10.1145/2737095.2737726.

[82] D. Madigan et al. "Bayesian indoor positioning systems". In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*. IEEE, 2005. DOI: 10.1109/infcom.2005.1498348. URL: https://doi.org/10.1109/infcom.2005.1498348.

[83] Halgurd S. Maghdid et al. "Seamless Outdoors-Indoors Localization Solutions on Smartphones". In: *ACM Computing Surveys* 48.4 (Feb. 2016), pp. 1–34. DOI: 10.1145/2871166. URL: https://doi.org/10.1145/2871166.

[84] Ahmed Makki et al. "Survey of WiFi positioning using time-based techniques". In: *Computer Networks* 88 (Sept. 2015), pp. 218–233. DOI: 10.1016/j.comnet.2015.06.015. URL: https://doi.org/10.1016/j.comnet.2015.06.015.

[85] Guoqiang Mao, Brian DO Anderson, and Barış Fidan. "Path loss exponent estimation for wireless sensor network localization". In: *Computer Networks* 51.10 (2007), pp. 2467–2483.

[86] Alex T. Mariakakis et al. "SAIL: single access point-based indoor localization". In: *Proceedings of the 12th annual international conference on Mobile systems, applications, and services - MobiSys '14*. ACM Press, 2014. DOI: 10.1145/2594368.2594393. URL: https://doi.org/10.1145/2594368.2594393.

[87] Fazeelat Mazhar, Muhammad Gufran Khan, and Benny Sällberg. "Precise Indoor Positioning Using UWB: A Review of Methods, Algorithms and Implementations". In: *Wireless Personal Communications* 97.3 (Aug. 2017), pp. 4467–4491. DOI: 10.1007/s11277-017-4734-x. URL: https://doi.org/10.1007/s11277-017-4734-x.

[88] Germán Martín Mendoza-Silva et al. "BLE RSS Measurements Dataset for Research on Accurate Indoor Positioning". In: *Data* 4.1 (2019). ISSN: 2306-5729. DOI: 10.3390/data4010012. URL: https://www.mdpi.com/2306-5729/4/1/12.

[89] Hetal P. Mistry and Nital H. Mistry. "RSSI Based Localization Scheme in Wireless Sensor Networks: A Survey". In: *2015 Fifth International Conference on Advanced Computing & Communication Technologies*. IEEE, Feb. 2015. DOI: 10.1109/acct.2015.105. URL: https://doi.org/10.1109/acct.2015.105.

[90] M. Shahkhir Mozamir, Rohani Binti Abu Bakar, and Wan Isni Soffiah Wan Din. "Indoor Localization Estimation Techniques in Wireless Sensor Network: A Review". In: *2018 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*. IEEE, Oct. 2018. DOI: 10.1109/i2cacis.2018.8603685. URL: https://doi.org/10.1109/i2cacis.2018.8603685.

[91] Ph. Müller et al. "A survey of parametric fingerprint-positioning methods". In: *Gyroscopy and Navigation* 7.2 (Apr. 2016), pp. 107–127. DOI: 10.1134/s2075108716020061. URL: https://doi.org/10.1134/s2075108716020061.

[92] Cam Nguyen and Meng Miao. "Fundamentals of UWB Impulse Systems". In: *Design of CMOS RFIC Ultra-Wideband Impulse Transmitters and Receivers*. Cham: Springer International Publishing, 2017, pp. 7–24. ISBN: 978-3-319-53107-6. DOI: 10.1007/978-3-319-53107-6_2.

[93] L. M. Ni et al. "LANDMARC: indoor location sensing using active RFID". In: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).* Mar. 2003, pp. 407–415. DOI: 10.1109/PERCOM.2003.1192765.

[94] George Oguntala et al. "Indoor location identification technologies for real-time IoT-based applications: An inclusive survey". In: *Computer Science Review* 30 (Nov. 2018), pp. 55–79. DOI: 10.1016/j.cosrev.2018.09.001. URL: https://doi.org/10.1016/j.cosrev.2018.09.001.

[95] Sameera Palipana, Bastien Pietropaoli, and Dirk Pesch. "Recent advances in RF-based passive device-free localisation for indoor applications". In: *Ad Hoc Networks* 64 (2017), pp. 80–98. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2017.06.007. URL: http://www.sciencedirect.com/science/article/pii/S1570870517301257.

[96] Jun-geun Park et al. "Growing an organic indoor location system". In: *Proceedings of the 8th international conference on Mobile systems, applications, and services - MobiSys '10*. ACM Press, 2010. DOI: 10.1145/1814433.1814461. URL: https://doi.org/10.1145/1814433.1814461.

[97] Norman Paskin. "Digital object identifier (DOI®) system". In: *Encyclopedia of library and information sciences* 3 (2010), pp. 1586–1592.

[98] Valter Pasku et al. "Magnetic Field-Based Positioning Systems". In: *IEEE Communications Surveys & Tutorials* 19.3 (2017), pp. 2003–2017. DOI: 10.1109/comst.2017.2684087. URL: https://doi.org/10.1109/comst.2017.2684087.

[99] Ling Pei et al. "A Survey of Crowd Sensing Opportunistic Signals for Indoor Localization". In: *Mobile Information Systems* 2016 (2016), pp. 1–16. DOI: 10.1155/2016/4041291. URL: https://doi.org/10.1155/2016/4041291.

[100] Chunyi Peng et al. "BeepBeep: A High Accuracy Acoustic Ranging System Using COTS Mobile Devices". In: *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*. SenSys '07. Sydney, Australia: ACM, 2007, pp. 1–14. ISBN: 978-1-59593-763-6. DOI: 10.1145/1322263.1322265. URL: http://doi.acm.org/10.1145/1322263.1322265.

[101] Antoni Pérez-Navarro et al. "Challenges of Fingerprinting in Indoor Positioning and Navigation". In: *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*. Elsevier, 2019, pp. 1–20. DOI: 10.1016/b978-0-12-813189-3.00001-0. URL: https://doi.org/10.1016/b978-0-12-813189-3.00001-0.

[102] F. Potortì et al. "Evaluation of Indoor Localisation Systems: Comments on the ISO/IEC 18305 Standard". In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sept. 2018, pp. 1–7. DOI: 10.1109/IPIN.2018.8533710.

[103] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. "The Cricket Location-support System". In: *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*. MobiCom '00. Boston, Massachusetts, USA: ACM, 2000, pp. 32–43. ISBN: 1-58113-197-6. DOI: 10.1145/345910.345917. URL: http://doi.acm.org/10.1145/345910.345917.

[104] Publishers International Linking Association Inc. *Crossref's Link References tool*. https://search.crossref.org/references. Online; accessed 22/07/2019.

[105] Valentin Radu and Mahesh K. Marina. "HiMLoc: Indoor smartphone localization via activity aware Pedestrian Dead Reckoning with selective crowdsourced WiFi fingerprinting". In: *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, Oct. 2013. DOI: 10.1109/ipin.2013.6817916. URL: https://doi.org/10.1109/ipin.2013.6817916.

[106] Mohammad Shaifur Rahman, Md. Mejbaul Haque, and Ki-Doo Kim. "High precision indoor positioning using lighting LED and image sensor". In: *14th International Conference on Computer and Information Technology (ICCIT 2011)*. IEEE, Dec. 2011. DOI: 10.1109/iccitechn.2011.6164805. URL: https://doi.org/10.1109/iccitechn.2011.6164805.

[107] Anshul Rai et al. "Zee: Zero-effort Crowdsourcing for Indoor Localization". In: *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. Mobicom '12. Istanbul, Turkey: ACM, 2012, pp. 293–304. ISBN: 978-1-4503-1159-5. DOI: 10.1145/2348543.2348580. URL: http://doi.acm.org/10.1145/2348543.2348580.

[108] Jonathan Raper et al. "A critical evaluation of location based services and their potential". In: *Journal of Location Based Services* 1.1 (2007), pp. 5–45.

[109] Nasir Saeed et al. "A State-of-the-Art Survey on Multidimensional Scaling Based Localization Techniques". In: *IEEE Communications Surveys & Tutorials* (2019), pp. 1–1. DOI: 10.1109/comst.2019.2921972. URL: https://doi.org/10.1109/comst.2019.2921972.

[110] Wilson Sakpere, Michael Adeyeye Oshin, and Nhlanhla BW Mlitwa. "A State-of-the-Art Survey of Indoor Positioning and Navigation Systems and Technologies". In: *South African Computer Journal* 29.3 (Dec. 2017). DOI: 10.18489/sacj.v29i3.452. URL: https://doi.org/10.18489/sacj.v29i3.452.

[111] Mahbubeh Sattarian et al. "Indoor navigation systems based on data mining techniques in internet of things: a survey". In: *Wireless Networks* 25.3 (May 2018), pp. 1385–1402. DOI: 10.1007/s11276-018-1766-4. URL: https://doi.org/10.1007/s11276-018-1766-4.

[112] Fernando Seco, Antonio R. Jimenez, and Pekka Peltola. "A Review of Multidimensional Scaling Techniques for RSS-Based WSN Localization". In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, Sept. 2018. DOI: 10.1109/ipin.2018.8533748. URL: https://doi.org/10.1109/ipin.2018.8533748.

[113] Moustafa Seifeldin et al. “Nuzzer: A Large-Scale Device-Free Passive Localization System for Wireless Environments”. In: *IEEE Transactions on Mobile Computing* 12.7 (July 2013), pp. 1321–1334. DOI: 10.1109/tmc.2012.106. URL: https://doi.org/10.1109/tmc.2012.106.

[114] Souvik Sen et al. “Avoiding multipath to revive inbuilding WiFi localization”. In: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services - MobiSys '13*. ACM Press, 2013. DOI: 10.1145/2462456.2464463. URL: https://doi.org/10.1145/2462456.2464463.

[115] Souvik Sen et al. “You are facing the Mona Lisa: spot localization using PHY layer information”. In: *Proceedings of the 10th international conference on Mobile systems, applications, and services - MobiSys '12*. ACM Press, 2012. DOI: 10.1145/2307636.2307654. URL: https://doi.org/10.1145/2307636.2307654.

[116] Jianga Shang et al. “Improvement Schemes for Indoor Mobile Location Estimation: A Survey”. In: *Mathematical Problems in Engineering* 2015 (2015), pp. 1–32. DOI: 10.1155/2015/397298. URL: https://doi.org/10.1155/2015/397298.

[117] Guobin Shen et al. “Walkie-Markie: Indoor Pathway Mapping Made Easy”. In: *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*. nsdi'13. Lombard, IL: USENIX Association, 2013, pp. 85–98. URL: http://dl.acm.org/citation.cfm?id=2482626.2482636.

[118] Jian Shen, Chen Jin, and Dengzhi Liu. “A Survey on the Research of Indoor RFID Positioning System”. In: *Cloud Computing and Security*. Springer International Publishing, 2016, pp. 264–274. DOI: 10.1007/978-3-319-48674-1_24. URL: https://doi.org/10.1007/978-3-319-48674-1_24.

[119] Guowei Shi and Ying Ming. “Survey of Indoor Positioning Systems Based on Ultra-wideband (UWB) Technology”. In: *Wireless Communications, Networking and Applications*. Springer India, Oct. 2015, pp. 1269–1278. DOI: 10.1007/978-81-322-2580-5_115. URL: https://doi.org/10.1007/978-81-322-2580-5_115.

[120] Y. Shu et al. “Magicol: Indoor Localization Using Pervasive Magnetic Field and Opportunistic WiFi Sensing”. In: *IEEE Journal on Selected Areas in Communications* 33.7 (July 2015), pp. 1443–1457. ISSN: 0733-8716. DOI: 10.1109/JSAC.2015.2430274.

[121] Shaufikah Shukri and Latifah Munirah Kamarudin. "Device free localization technology for human detection and counting with RF sensor networks: A review". In: *Journal of Network and Computer Applications* 97 (2017), pp. 157–174. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2017.08.014. URL: http://www.sciencedirect.com/science/article/pii/S1084804517302928.

[122] Ashraf Tahat et al. "A Look at the Recent Wireless Positioning Techniques With a Focus on Algorithms for Moving Receivers". In: *IEEE Access* 4 (2016), pp. 6652–6680. DOI: 10.1109/access.2016.2606486. URL: https://doi.org/10.1109/access.2016.2606486.

[123] Zain Bin Tariq et al. "Non-GPS Positioning Systems". In: *ACM Computing Surveys* 50.4 (Aug. 2017), pp. 1–34. DOI: 10.1145/3098207. URL: https://doi.org/10.1145/3098207.

[124] Joaquín Torres-Sospedra and Adriano Moreira. "Analysis of sources of large positioning errors in deterministic fingerprinting". In: *Sensors* 17.12 (2017), p. 2736.

[125] Joaquín Torres-Sospedra et al. "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems". In: *Expert Systems with Applications* 42.23 (2015), pp. 9263–9278. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2015.08.013. URL: http://www.sciencedirect.com/science/article/pii/S0957417415005527.

[126] Joaquín Torres-Sospedra et al. "Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning". In: *Mobile Information Systems* 2016 (2016).

[127] Joaquín Torres-Sospedra et al. "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems". In: *2014 international conference on indoor positioning and indoor navigation (IPIN)*. IEEE. 2014, pp. 261–270.

[128] J. Ureña et al. "Acoustic Local Positioning With Encoded Emission Beacons". In: *Proceedings of the IEEE* 106.6 (June 2018), pp. 1042–1062. ISSN: 0018-9219. DOI: 10.1109/JPROC.2018.2819938.

[129] US Departments of Defense, Homeland Security and Transportation. *US Federal Radionavigation Plan*. Tech. rep. United States. Dept. of Defense, 2017. URL: https://www.navcen.uscg.gov/pdf/FederalRadioNavigationPlan2017.pdf.

[130] Frank Van Diggelen and Per Enge. "The worlds first gps mooc and worldwide laboratory using smartphones". In: *Proceedings of the 28th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2015)*. 2015, pp. 361–369.

[131] Melanija Vezocnik and Matjaz B. Juric. "Average Step Length Estimation Models' Evaluation Using Inertial Sensors: A Review". In: *IEEE Sensors Journal* 19.2 (Jan. 2019), pp. 396–403. DOI: 10.1109/jsen.2018.2878646. URL: https://doi.org/10.1109/jsen.2018.2878646.

[132] Quoc Duy Vo and Pradipta De. "A Survey of Fingerprint-Based Outdoor Localization". In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 491–506. DOI: 10.1109/comst.2015.2448632. URL: https://doi.org/10.1109/comst.2015.2448632.

[133] He Wang et al. "No Need to War-drive: Unsupervised Indoor Localization". In: *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*. MobiSys '12. Low Wood Bay, Lake District, UK: ACM, 2012, pp. 197–210. ISBN: 978-1-4503-1301-8. DOI: 10.1145/2307636.2307655. URL: http://doi.acm.org/10.1145/2307636.2307655.

[134] Jindong Wang et al. "Deep learning for sensor-based activity recognition: A survey". In: *Pattern Recognition Letters* 119 (Mar. 2019), pp. 3–11. DOI: 10.1016/j.patrec.2018.02.010. URL: https://doi.org/10.1016/j.patrec.2018.02.010.

[135] Roy Want et al. "The Active Badge Location System". In: *ACM Trans. Inf. Syst.* 10.1 (Jan. 1992), pp. 91–102. ISSN: 1046-8188. DOI: 10.1145/128756.128759. URL: http://doi.acm.org/10.1145/128756.128759.

[136] Andy Ward, Alan Jones, and Andy Hopper. "A new location technique for the active office". In: *IEEE Personal communications* 4.5 (1997), pp. 42–47.

[137] Oliver Woodman and Robert Harle. "Pedestrian Localisation for Indoor Environments". In: *Proceedings of the 10th International Conference on Ubiquitous Computing*. UbiComp '08. Seoul, Korea: ACM, 2008, pp. 114–123. ISBN: 978-1-60558-136-1. DOI: 10.1145/1409635.1409651. URL: http://doi.acm.org/10.1145/1409635.1409651.

[138] Kaishun Wu et al. "CSI-Based Indoor Localization". In: *IEEE Transactions on Parallel and Distributed Systems* 24.7 (July 2013), pp. 1300–1309. DOI: 10.1109/tpds.2012.214. URL: https://doi.org/10.1109/tpds.2012.214.

[139] Kaishun Wu et al. "FILA: Fine-grained indoor localization". In: *2012 Proceedings IEEE INFOCOM*. IEEE, Mar. 2012. DOI: 10.1109/infcom.2012.6195606. URL: https://doi.org/10.1109/infcom.2012.6195606.

[140] Yuan Wu et al. "A Survey of the Research Status of Pedestrian Dead Reckoning Systems Based on Inertial Sensors". In: *International Journal of Automation and Computing* 16.1 (Sept. 2018), pp. 65–83. DOI: 10.1007/s11633-018-1150-y. URL: https://doi.org/10.1007/s11633-018-1150-y.

[141] Henk Wymeersch, Jaime Lien, and Moe Z. Win. "Cooperative Localization in Wireless Networks". In: *Proceedings of the IEEE* 97.2 (Feb. 2009), pp. 427–450. DOI: 10.1109/jproc.2008.2008853. URL: https://doi.org/10.1109/jproc.2008.2008853.

[142] Shixiong Xia et al. "Indoor Fingerprint Positioning Based on Wi-Fi: An Overview". In: *ISPRS International Journal of Geo-Information* 6.5 (Apr. 2017), p. 135. DOI: 10.3390/ijgi6050135. URL: https://doi.org/10.3390/ijgi6050135.

[143] Jiang Xiao et al. "A Survey on Wireless Indoor Localization from the Device Perspective". In: *ACM Computing Surveys* 49.2 (June 2016), pp. 1–31. DOI: 10.1145/2933232. URL: https://doi.org/10.1145/2933232.

[144] Bo Xie et al. "LIPS: A Light Intensity–Based Positioning System for Indoor Environments". In: *ACM Transactions on Sensor Networks* 12.4 (Sept. 2016), pp. 1–27. DOI: 10.1145/2953880. URL: https://doi.org/10.1145/2953880.

[145] Jie Xiong and Kyle Jamieson. "ArrayTrack: A Fine-grained Indoor Location System". In: *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*. nsdi'13. Lombard, IL: USENIX Association, 2013, pp. 71–84. URL: http://dl.acm.org/citation.cfm?id=2482626.2482635.

[146] Chenren Xu et al. "SCPL: indoor device-free multi-subject counting and localization using radio signal strength". In: *Proceedings of the 12th international conference on Information processing in sensor networks - IPSN '13*. ACM Press, 2013. DOI: 10.1145/2461381.2461394. URL: https://doi.org/10.1145/2461381.2461394.

[147] Sungwon Yang et al. "FreeLoc: Calibration-free crowdsourced indoor localization". In: *2013 Proceedings IEEE INFOCOM*. IEEE, Apr. 2013. DOI: 10.1109/infcom.2013.6567054. URL: https://doi.org/10.1109/infcom.2013.6567054.

[148] Zheng Yang and Yunhao Liu. "Quality of trilateration: Confidence-based iterative localization". In: *IEEE Transactions on Parallel and Distributed Systems* 21.5 (2009), pp. 631–640.

[149] Zheng Yang, Chenshu Wu, and Yunhao Liu. "Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention". In: *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. Mobicom '12. Istanbul, Turkey: ACM, 2012, pp. 269–280. ISBN: 978-1-4503-1159-5. DOI: 10.1145/2348543.2348578. URL: http://doi.acm.org/10.1145/2348543.2348578.

[150] Zheng Yang et al. "Mobility Increases Localizability". In: *ACM Computing Surveys* 47.3 (Apr. 2015), pp. 1–34. DOI: 10.1145/2676430. URL: https://doi.org/10.1145/2676430.

[151] Zhice Yang et al. "Wearables Can Afford: Light-weight Indoor Positioning with Visible Light". In: *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '15*. ACM Press, 2015. DOI: 10.1145/2742647.2742648. URL: https://doi.org/10.1145/2742647.2742648.

[152] Ali Yassin et al. "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications". In: *IEEE Communications Surveys & Tutorials* 19.2 (2017), pp. 1327–1346. DOI: 10.1109/comst.2016.2632427. URL: https://doi.org/10.1109/comst.2016.2632427.

[153] Mohamad Yassin and Elias Rachid. "A survey of positioning techniques and location based services in wireless networks". In: *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*. IEEE, Feb. 2015. DOI: 10.1109/spices.2015.7091420. URL: https://doi.org/10.1109/spices.2015.7091420.

[154] M.A. Youssef, A. Agrawala, and A. Udaya Shankar. "WLAN location determination via clustering and probability distributions". In: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).* IEEE Comput. Soc, 2003. DOI: 10.1109/percom.2003.1192736. URL: https://doi.org/10.1109/percom.2003.1192736.

[155] 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*. ACM Press, 2005. DOI: 10.1145/1067170.1067193. URL: https://doi.org/10.1145/1067170.1067193.

[156] Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. "A Survey of Indoor Localization Systems and Technologies". In: *IEEE Communications Surveys & Tutorials* (2019), pp. 1–1. DOI: 10.1109/comst.2019.2911558. URL: https://doi.org/10.1109/comst.2019.2911558.

[157] Weizhi Zhang, M. I. Sakib Chowdhury, and Mohsen Kavehrad. "Asynchronous indoor positioning system based on visible light communications". In: *Optical Engineering* 53.4 (Apr. 2014), p. 045105. DOI: 10.1117/1.oe.53.4.045105. URL: https://doi.org/10.1117/1.oe.53.4.045105.

[158] Yiyang Zhao, Yunhao Liu, and Lionel M Ni. "VIRE: Active RFID-based localization using virtual reference elimination". In: *2007 International Conference on Parallel Processing (ICPP 2007)*. IEEE. 2007, pp. 56–56.

[159] Xiaolei Zhou et al. "From one to crowd: a survey on crowdsourcing-based wireless indoor localization". In: *Frontiers of Computer Science* 12.3 (May 2018), pp. 423–450. DOI: 10.1007/s11704-017-6520-z. URL: https://doi.org/10.1007/s11704-017-6520-z.

[160] Zhou Zhou. "Indoor positioning algorithm using light-emitting diode visible light communications". In: *Optical Engineering* 51.8 (Aug. 2012), p. 085009. DOI: 10.1117/1.oe.51.8.085009. URL: https://doi.org/10.1117/1.oe.51.8.085009.

[161] Yuan Zhuang et al. "A Survey of Positioning Systems Using Visible LED Lights". In: *IEEE Communications Surveys & Tutorials* 20.3 (2018), pp. 1963–1988. DOI: 10.1109/comst.2018.2806558. URL: https://doi.org/10.1109/comst.2018.2806558.

Chapter 3

Collected RSS Datasets

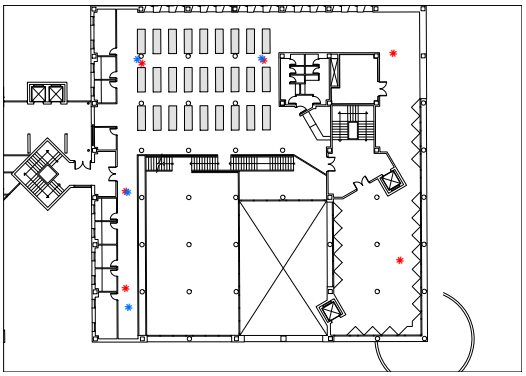
Data collected for testing purposes is of utmost importance for the development and tuning of many systems, included IPS based on WiFi or BLE fingerprinting. Datasets are also used for credible comparisons among IPS. Thus, it is common that competition challenges or evaluation platforms include one or several datasets upon which the IPS are gauged. In recent years, several WiFi and BLE datasets have been collected and made publicly available [27, 13, 12, 2]. The relevance of a dataset lies in its usefulness to test specific aspects of an IPS. For example, the UJIIndoorLoc dataset [27] – used in the Track 3 of the IPIN 2015 competition [24] – and the datasets from the Track 3 of the 2016 and 2017 IPIN competitions considered large environments, several collection devices and subjects, with the main goal of making the positioning a challenge, and challenging their positioning was [24, 26, 25].

There is more to evaluate than the ability of an IPS to deal with large, sparsely described environments. This chapter addresses two collected datasets to show goals different from comparing the accuracy of several IPS: (1) the robustness of WiFi-based IPS to changes in the environment and (2) high accuracy for BLE-based IPS. The collected datasets not only provided measurements and their collection tags – e.g., position and time – but also facilitated the data loading and selection, incorporated environment descriptors and were shipped with analyses that constituted starting points into further research on the goals for which the datasets were compiled. This chapter provides examples of the way that open datasets should be collected and provided in order to test WiFi-based or BLE-based IPS proposals and to foster their development.

3.1 Long-Term WiFi Fingerprinting Dataset

The WiFi long-term dataset [18] was created and made publicly available to ease the study of long-term and short-term WiFi RSS variations, to spur the creation of methods able to cope with those variations and to test robust IPS proposals. The dataset provides sets of WiFi RSS collected by a trained person that assured trustworthy position labels. Several fingerprints were collected per each collection point. The collection process spanned 15 months originally – as published in [19] – but it was later extended to 25 months. The collection points covered small areas among the bookshelves from two floors (the 3^{rd} and 5^{th}) of the north wing of the UJI's Library building (see Figure 3.1a). The combined collection area of both floors covered 308.4 m^2. The dataset was provided with supporting materials that included descriptions of the environment and the nearby WiFi APs (Figure 3.1b). The collection was performed following the campaign-based approach and collection software that Chapter 4 will later describe. The person that performed the collection went through an ordered list of positions, for which the person had to face a specific direction and collect a specific number of fingerprints (samples).

(a) Library 3rd floor

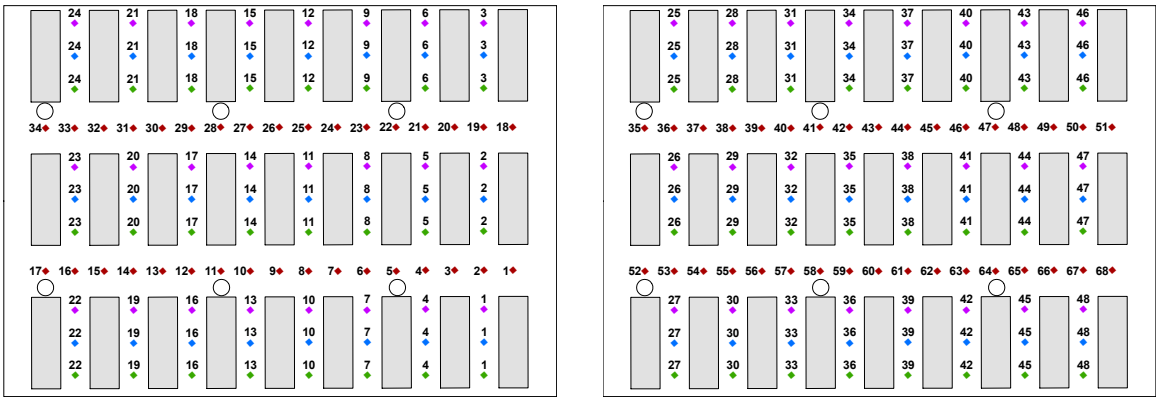


(b) Wireless network devices

Figure 3.1: Library Environment: (a) a picture of the 3rd floor collection area that shows the bookshelves and the stairs that connect the two floors; and (b) the network devices close to the collection area. The red asterisks represent the 3^{rd} floor's devices, and blue asterisks represent 5^{th} floor's devices.

The collected data was organized into training and test (Test-01, Test-02, Test-03, Test-04, and Test-05) sets, being each set the result of performing a collection campaign. Figure 3.2 shows the collection positions for a floor. The collection was performed by following two collection series: direct and reverse. The complete order of samples gathered for a month and a set was (1) direct series 3^{rd} floor; (2) reverse series 3^{rd} floor; (3) direct series 5^{th} floor; and (4) reverse series 5^{th} floor. Regarding

directions, Test-01, Test-02, and Test-03 sets were collected facing the "Up" (direct series) and "Down" (reverse series) directions. Test-04 and Test-05 sets were collected facing the "Left" (direct series) and "Right" (reverse series) directions.



(a) Direct direction (b) Reverse direction

Figure 3.2: Positions in the direct (a) and reverse (b) collection directions in a floor. Bookshelves are represented by rectangles. Collection positions of training, Test-01, and Test-05 sets are represented by blue diamonds. Collection positions of Test-02, Test-03, and Test-04 sets are represented by while green, pink and red diamonds, respectively. The numbers indicate the collection order within each set.

The supporting materials also include the coordinates that define the geometries of the bookshelves. Such supporting information is important for several aspects of an IPS. In the example presented in Figure 3.3, the bookshelves geometries helped in ruling out some areas of the scenario for densification of the training dataset through regression. Those geometries can also be used to correct position estimates toward valid areas. Furthermore, the high RSS values seen in the left bottom part of the image can be attributed to LOS conditions, which are inferred because the geometries indicate that there is a lane free of obstacles between that part of the area and what appears to be the position of the AP.

The collection of the WiFi dataset is dense and contains several samples in two opposite directions at each reference point, which allows the study of short-term variations. Additionally, the analysis of long-term variations is also allowed by a collection performed across 25 months. The WiFi signals detected in a scenario constitute a dynamic environment. The signals from the (fixed) APs deployed in a scenario for connectivity are not the only ones detected. It is possible – and likely – to detect signals of temporal APs created at smartphones – which have low detection ranges – and signals of fixed APs from neighboring buildings – which detection ranges that may be large. In the case of residential areas, APs from neighboring areas may

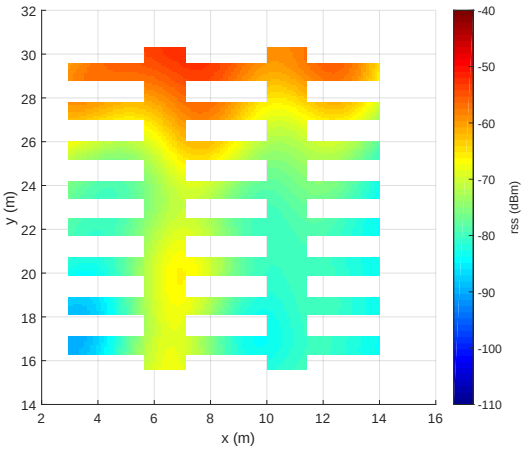


Figure 3.3: SVR regression of mean of the six samples corresponding to each position, with with map mask, for training dataset enrichment.

alternate between off and on states along a day. Therefore, a large variability is to be expected across the time for detected APs. The detection of new APs is must be considered nominal, i.e., an IPS should be prepared to handle such cases.

Table 3.1 presents the intermittence found in the collection area in the Library. The values shown in the table do not account for all APs detected in the dataset, given that only training sets for the collection Month 1 were used for their computation. In only 20 days, 182 different APs were detected, a number that is notable given the size of the collection area. The set number ("Number" column in Table 3.1) also indicates the order in which the sets were collected. It is significant that even in the 5^{th} collection campaign – 14 days after the first campaign – the number of new detected APs was almost 20. For each campaign, the sum of the number of the APs no longer seen with respect to the previous campaign ("Gone" column), the number of APs that started being seen again ("Returned" column) and the number of APs being gone for a second time ("Re-Gone" column) is always over 15 after collection campaign number 04, depicting the large intermittence that is to be expected for similar scenarios.

The APs deployed in a scenario for connectivity (fixed APs) may also have a dynamic behavior, as presented in Figure 3.4. Figure 3.4 shows some APs that were only detected in one month, others whose detection was intermittent, and a notable change in network configuration in Month 12. The MAC or SSID configurations of some devices changed between Months 11 and 12, which is the reason why some of the APs consistently seen in most of the months previous to Month 12 were no longer detected, and new APs were continuously observed onward. The non-detection during a campaign realization of some fixed APs may be nominal behavior if it is the

Table 3.1: Statistics of presence of APs, using measurements from the 15 training datasets for Month 1.

Number	Total	New	Gone	Returned	Re-Gone	Seen	Since 1st Day
01	77	77	0	0	0	77	0
02	97	22	2	0	0	99	1
03	118	23	4	2	0	122	5
04	106	9	23	2	0	131	6
05	127	19	12	15	1	150	14
06	119	4	14	8	6	154	14
07	126	5	4	9	3	159	15
08	120	3	6	6	9	162	15
09	112	4	6	8	14	166	18
10	125	6	4	15	4	172	18
11	119	4	7	6	9	176	18
12	115	1	8	16	13	177	18
13	127	4	1	15	6	181	19
14	124	0	4	15	14	181	20
15	119	1	2	9	13	182	20

result of temporal conditions – like the RSS of a far away AP dropping below the sensor's sensibility. However, the disappearance of (fixed) APs deployed in a scenario for connectivity, as seen along the time in a whole area and not at a single position, is not nominal and an IPS should devise countermeasures to avoid affectation to position estimates, as Chapter 5 will later explain.

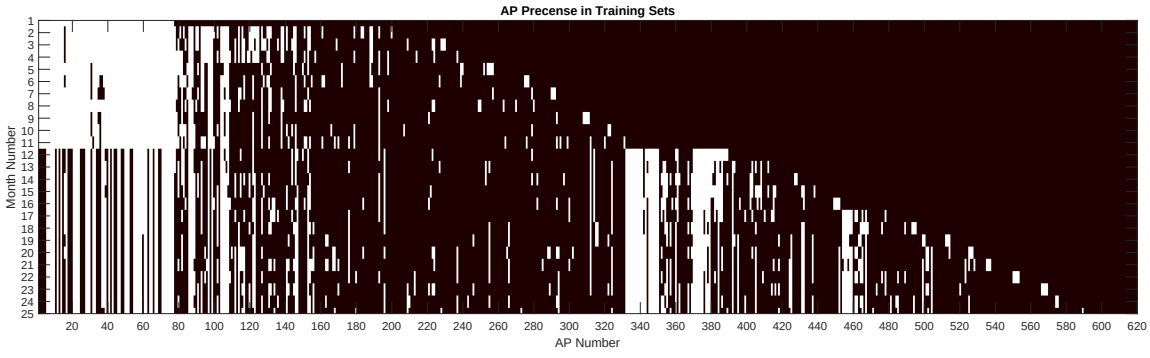


Figure 3.4: AP ephemerality, as seen using the Training-01 datasets from each month. The white narrow rectangles indicate the detection of an AP in a month. Some APs do not appear in this figure because they were are only detected in test datasets.

3.1.1 IPS Robustness

The long period that spanned the collection makes the Library dataset adequate for testing the behavior of IPS against network changes and long-term variations of the WiFi signal environment. To exemplify such analyses, six fingerprinting-based IPS were selected and tested against the test datasets of each collection month. The six IPS are positioning methods found in literature and they were selected because of their popularity and simplicity. The radio-map used for fingerprinting was the Training-01 dataset. The method's hyper-parameters – if any – were determined in simple preliminary experiments, considering the accuracy of the method for the first month of data. The six tested methods are:

1. Rand: It returns a randomly-chosen position (x, y, floor). It was only provided as a reference measure of the lower expected accuracy measure and floor detection rate that an IPS may have.

2. Prob: It is a known probabilistic method [28] that uses as position estimate the centroid of k positions l (x, y, floor) from the radiomap. The k positions are those from the samples with the highest values for the probability $P(l|s)$, with s being the operational fingerprint and with $P(s|l) = \prod P(s_i|l)$, where s_i is the RSS value of the i^{th} detected AP. The computation of $P(s_i|l)$ followed that of Berkvens et. al. [3]:

$$P(s_i|l) = \int_{s_i+0.5}^{s_i-0.5} \mathcal{N}(\mu_l, \sigma^2{}_l)\mathrm{d}r, \tag{3.1}$$

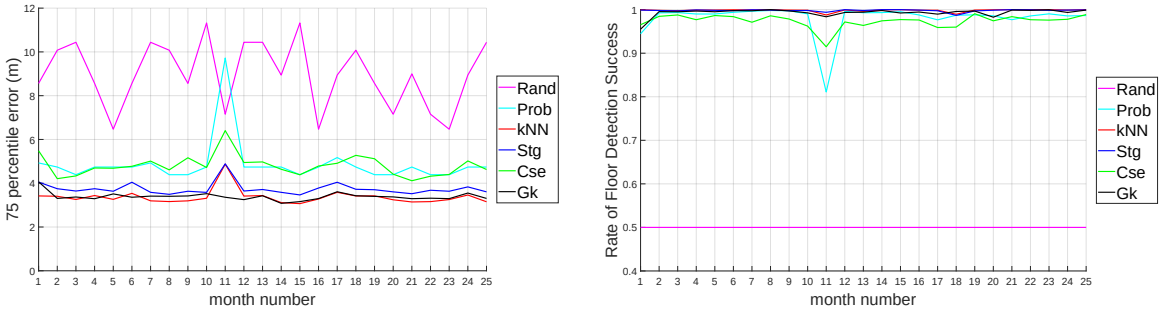
   where r refers to a variable in the RSS domain, $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2; μ_l and $\sigma^2{}_l$ are functions of l, the position, and are calculated for each point - position and direction of collection.

3. kNN: It is a known deterministic method [1] that finds the k closest samples in the fingerprint space to the operational fingerprint. It provides as the 2D position estimate the centroid of the positions (x, y) of those k closest samples. The value used for k was 9 and the Euclidean distance was selected as fingerprint distance.

4. Stg: This method is based on the proposal by Marques et. al. [14]. The method selects samples in the radio-map whose APs with the s strongest RSS are the same APs in the operational fingerprint with the s strongest RSS. The kNN method explained above is applied to the selected samples. The parameter values are $s = 3$ and $k = 5$.

5. CSE: It is based on the method proposed in Hernández et. al. [9], which applies SVR over the radiomap collected for a floor to enrich it. SVR was used here as provided by [15], using a Radial Basis Function (RBF) kernel and performing predictor data standardization. Per each AP and position, the RSS difference between the radiomap and the operational fingerprint is computed. That RSS difference is used to compute a score. Positions whose RSS difference is zero get the highest scores. The score gets smaller as the difference increases, and it is zero beyond a margin m (a value of 4 was used here). The scores of each AP are summed up to obtain a general score for each position. A map mask is applied to discard unfeasible positions. Scores are computed independently for each floor. The position (x, y, floor) with the highest general score is used as the position estimate. The map mask was computed here using the geometries of the library bookshelves.

6. Gk: This method is based on the method by Roos et. al. [23], proposed in the form of a kernel density estimator (KDE). It computes the likelihood of the RSS at each fingerprint position. It assumes a model that follows a Normal distribution with the sample mean at each reference point and a constant standard deviation for all observations. The position estimate is computed as the centroid of the positions corresponding to the k highest likelihood values.



(a) 2D positioning error metric.

(b) Floor detection success ratio.

Figure 3.5: Positioning accuracy of each tested IPS.

Figure 3.5 presents the positioning performance results of the tested methods. Figure 3.5a presents the 2D accuracy, evaluated in terms of the 75^{th} percentile of Euclidean distance. The Euclidean distance was used for simplicity. The 75^{th} percentile was favored over the mean given the optimistic view of the later and the fact that the 75^{th} has been adopted as the positioning accuracy metric by the IPIN Track 3 Competitions [26, 25]. Figure 3.5b presents the floor detection success rate. For the Prob, kNN, Stg and Gk methods, the floor estimate is provided as the mode of the floors from the closest samples.

The kNN, Gk and Stg methods provide the best results regarding 2D accuracy and floor detection success. Despite the kNN method surpasses the Gk method in 2D accuracy in most months, the Gk method is the most stable along the whole period. Furthermore, it is the most robust against the changes of the network configuration that occurred between the collection months 11 and 12. In general, the methods provided accuracy results that are in line with those reported in the literature for WiFi-based fingerprinting approaches [8]. Regarding the floor detection success rate, the kNN, Gk and Stg methods are the best-performing ones. These methods had little affectation by the changes in the network configuration, being the Stg method is the least affected. Compared to the other methods, the Prob and CSE performed poorly regarding both the 2D accuracy and the floor detection rate.

3.2 BLE RSS Dataset

The publicly available BLE RSS dataset [17] was collected for research on RSS-based, smartphone-based, fine-grained, indoor positioning techniques. The dataset provides several sets of BLE RSS measurements, which were carefully labeled with their collection details. The dense BLE beacon deployment in the collection areas allow positioning accuracies below 3 m. Besides the BLE RSS measurements, beacon deployment descriptions and environment information were provided. Also, dataset information includes a set of MATLAB ©scripts that help in data handling and perform demonstrative analyses. The analyses are intended to show how the BLE RSS measurements can be used in indoor positioning research. Also, they highlight important concerns that may affect BLE RSS-based positioning methods.

The collection of RSS measurements was performed in two zones from the UJI, in Castellón, Spain. The zone is an area among bookshelves in the Library building – hereinafter, Library. The other zone is an area of the office space of the Geotec Research Group – hereinafter, Geotec. Data that describe the environment, the collection, and the beacon deployment are presented in the Table 3.2. In Table 3.2, the values presented by the "MDMS" column are the maximum distance from the position of a collection point to another collection point that is the closest to it in the space. the values presented by the "MDMB" column are described as the "MDMS" column, but regarding the position of beacons instead of the position of collection points.

Twenty two BLE beacons were deployed in the bookshelves of an area in the Library to support a positioning service for users looking for a book. Figure 3.6a presents their deployment positions. The beacons were placed inside the enclosed top of the wooden shelves for security concerns, as presented in Figure 3.6b. Therefore, there were no line-of-sight situations between beacons and collection points in this

Table 3.2: General description of each collection zone.

	Geotec	Library
number of beacons	22	24
number of samples	2652	2100
number of campaigns	1	2
smartphones	A5	A5,BQ,S6
involved subjects	1	3
consecutive collection samples	13	6
transmission powers (dBm)	$-20,-12,-4$	-20
area (m^2)	151.07	176.02
area covered by tall obstacles (m^2)	8.70	68.73
area covered by samples (m^2)	137.51	121.65
area covered by beacons (m^2)	138.08	172.9
MDMS (m)	2.34	1.79
MDMB (m)	3.07	3.71

environment. The dense beacon deployment assured that the area covered by them included all bookshelves and the deployed beacons were as far apart from each other as possible. The shelves height is 2.35 m. Given that the ceiling is about 2.60 m above the floor, the beacons were close to it. The collection was divided into two campaigns. In one of the campaigns, the subjects faced the up or down directions at each position (blue squares in Figure 3.6a). In the other campaign, the subjects faced the left or right directions at each position (orange circles in Figure 3.6a). Each collection campaign was performed three times, each time using a different smartphone and by a different subject. The smartphones were a BQ Aquaris X5 Plus, a Samsung Galaxy S6 (SM-G920F) and a Samsung Galaxy A5 2017 (SM-A520F), hereinafter BQ, S6, and A5, respectively.

Twenty four BLE beacons were deployed in the Geotec zone, in the positions presented in Figure 3.7a. The beacons were attached to the ceiling tiles, as shown by Figure 3.7b. Given that the environment contained only four tall furniture pieces, every collection point had line-of-sight situations with more than three beacons at the same time, if human-body blockage is not considered. The collection was performed following one campaign, in which with the subject faced the up or down directions, using one of the smartphones used for the Library zone. The campaign was performed three times, each time with the beacons configured for a particular transmission power.

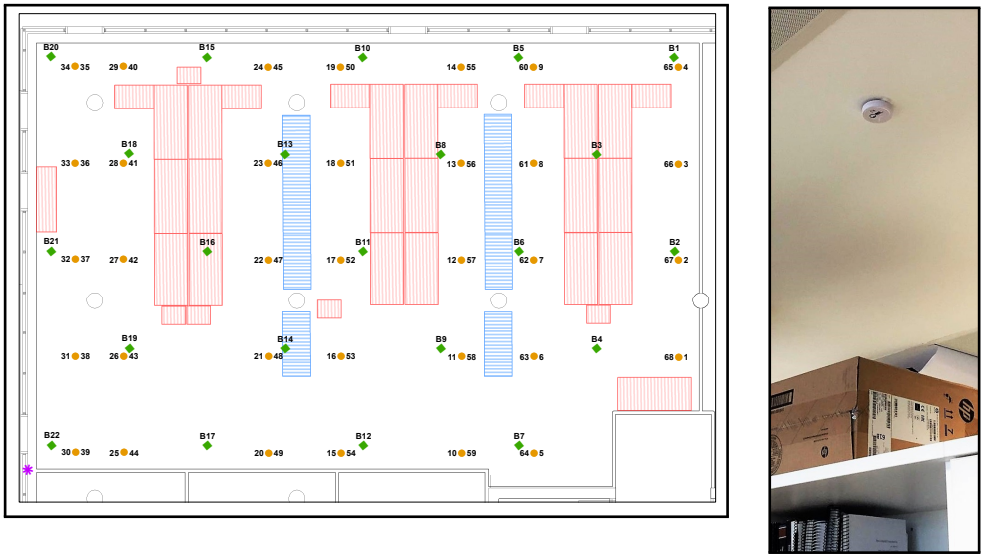
(a) Library Deployment (b) Deployed beacon

Figure 3.6: The 5^{th} floor Library zone. In (a), green diamonds and the magenta asterisk represent the position of beacons and the local coordinates origin, respectively. The orange circles and blue squares represent the collection positions where the subject faced the left-right directions and those where they faced the up-down directions, respectively. The pale blue rectangles represent bookshelves. In (b), a shot that shows how beacons were deployed inside the top of bookshelves.

3.2.1 Environment and Device: Signal Intensities and Advertisements Loss

The dense bookshelf placement in the Library environment has a notable influence is signal strength. There were no line-of-sight situations between beacons and a smartphone in this zone. As presented in Figure 3.8, the drop in signal strength as the distance to a beacon increases is more significant in the Library zone than in the Geotec zone. In the latter, there are only four tall shelves and the height of other furniture pieces is below the altitude at which a person usually holds a smartphone. Thus, the furniture in this zone does not significantly influence the reported beacon RSS.

Table 3.3 shows that the range at which a beacon is detected not only depends on the transmission power but also depends on the collection smartphone and the deployment environment. The first row in the table shows the maximum RSS value detected for a beacon at any collection point. The signals strengths reported by the S6 smartphone were weaker than those reported by the A5 smartphone in the



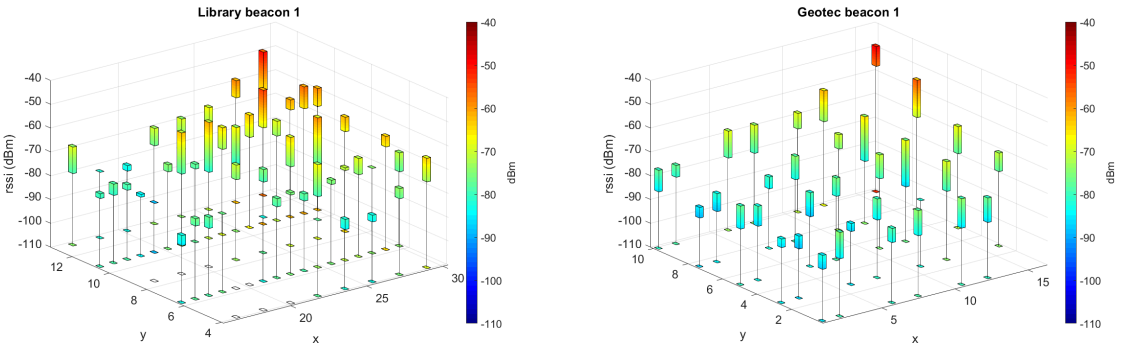
(a) Geotec Deployment (b) Deployed beacon

Figure 3.7: Geotec office zone. In (a), green diamonds, orange circles and the magenta asterisk represent the position of beacons, collection positions and the local coordinates origin, respectively. The blue horizontal-lines-filled rectangles represent tall furniture that create NLOS situations. In (b), a shot to show how beacons were deployed in the ceiling tiles.

Library zone. The A5 smartphone could detect stronger signals in the Library than in Geotec. Despite the non-existence of line-of-sight situations in the Library, the minimum (3D) distance between collection points and beacons was smaller than in the Geotec zone. The second row (MDP70) in the table displays the maximum distance at which a beacon was detected with a signal strength over -70 dBm. The distances are consistent with the maximum detected RSS values. The third row (MDHT) in the table presents the maximum distance at which a beacon was consistently detected in more than half of fingerprints at a collection point. The S6 smartphone was notably better at detecting beacon advertisements than the A5 smartphone, despite it registered weaker values of RSS.

Table 3.3: Beacons detection range.

	Library (S6)	Library (A5)	Geotec (A5)
Max detected RSS (dBm)	-52	-33	-44
MDP70 (m)	0.47	3.92	2.75
MDHT (m)	4.50	1.47	1.16

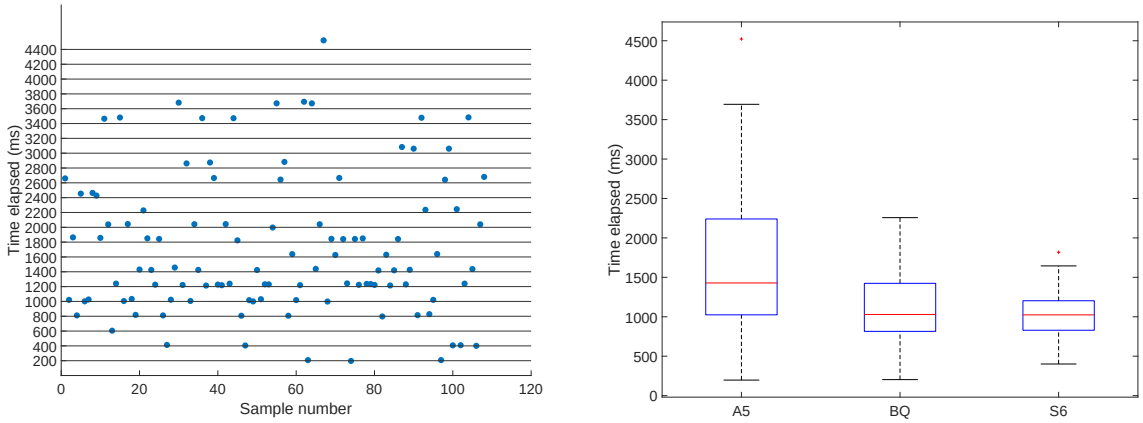


(a) Library Zone (b) Geotec Zone

Figure 3.8: Difference between zones in the RSS values as detected by the A5 smartphone, with beacon transmission power set to −12 dBm.

BLE beacons can usually be configured to send advertisements at given time intervals. However, the time spans between two consecutive detections reported by a smartphone are usually higher than the interval setting configured in the beacons [10, 7]. It is not only that the receptor do not detect some advertisements when it is beyond the beacon range, or that the time increases because of some tiny extra processing delays, but that some smartphones cannot properly detect all advertisements. Figure 3.9a shows the time spans between advertisement detections for the smartphone A5 and beacon 1 in the Library zone. All beacons were configured to advertise every 200 ms. However, the number of two consecutive advertisements detected with a time difference close to 200 ms is very low. Figure 3.9b shows that the detection capability of the smartphone A5 is low when compared to the BQ and S6 smartphones.

Closer distances to an emitter correlate to higher variance in the reported RSS values for the 2.4 GHz networks. Also, some works stated that close distances to a beacon correlate to low detection delays [4]. The results of a Pearson correlation test between detection delays and distance to beacons are presented in Table 3.4 for the A5 smartphone. The correlation was individually tested for each beacon, and only the ρ and p-value for the beacon with the most interesting values are presented in Table 3.4. The results are not statistically significant. However, they suggest that the distance to a beacon is not a key factor in the observed detection delays of advertisements. Also, given that the consecutive measurements of a point were collected facing the same direction, orientation is likely to play only a small influence on detection delays, at least for our settings and analyses.



(a) Detection delays A5 (b) Detection delays, each phone

Figure 3.9: Times elapsed between two consecutive beacon detections while active registration of BLE advertisements. Measurements correspond beacon 1 broadcasting at −12 dBm in the Library zone. In (**a**), the $t\times200$ ms ($t \geq 1$) detection delay pattern is shown. In (**b**), a comparison of detection capability among the three smartphones is presented.

Table 3.4: Pearson correlation test between detection delays and distance to a beacon. Values correspond to the beacon reporting the ρ value farthest from zero, considering data captured by the A5 smartphone and a −12 dBm beacon power setting).

Zone	$\boldsymbol{\rho}$	p-Value	Beacon
Geotec	−0.14	0.04	2
Library	−0.14	0.12	18

3.2.2 Positioning Accuracy

BLE RSS-based positioning methods are commonly affected by factors like the environment layout, beacon advertisement configuration, beacon deployment positions, and the user smartphone. Some of the methods found in IPS literature address one or several of those factors [22]. The Weighted Centroid (WC) and the k-Nearest Neighbors (kNN) methods were tested to explore the previous factors in this work. The WC is used in BLE-based positioning because the positions of the deployed beacons are known. The kNN method is commonly used in fingerprinting for indoor positioning, mainly for WiFi signals.

The WC method was previously addressed for BLE-based positioning, sometimes with other names [7, 21]. The WC method estimates a 2D position for a BLE operational fingerprint by computing the mean of the positions of the $1 \leq k \leq n$ detected

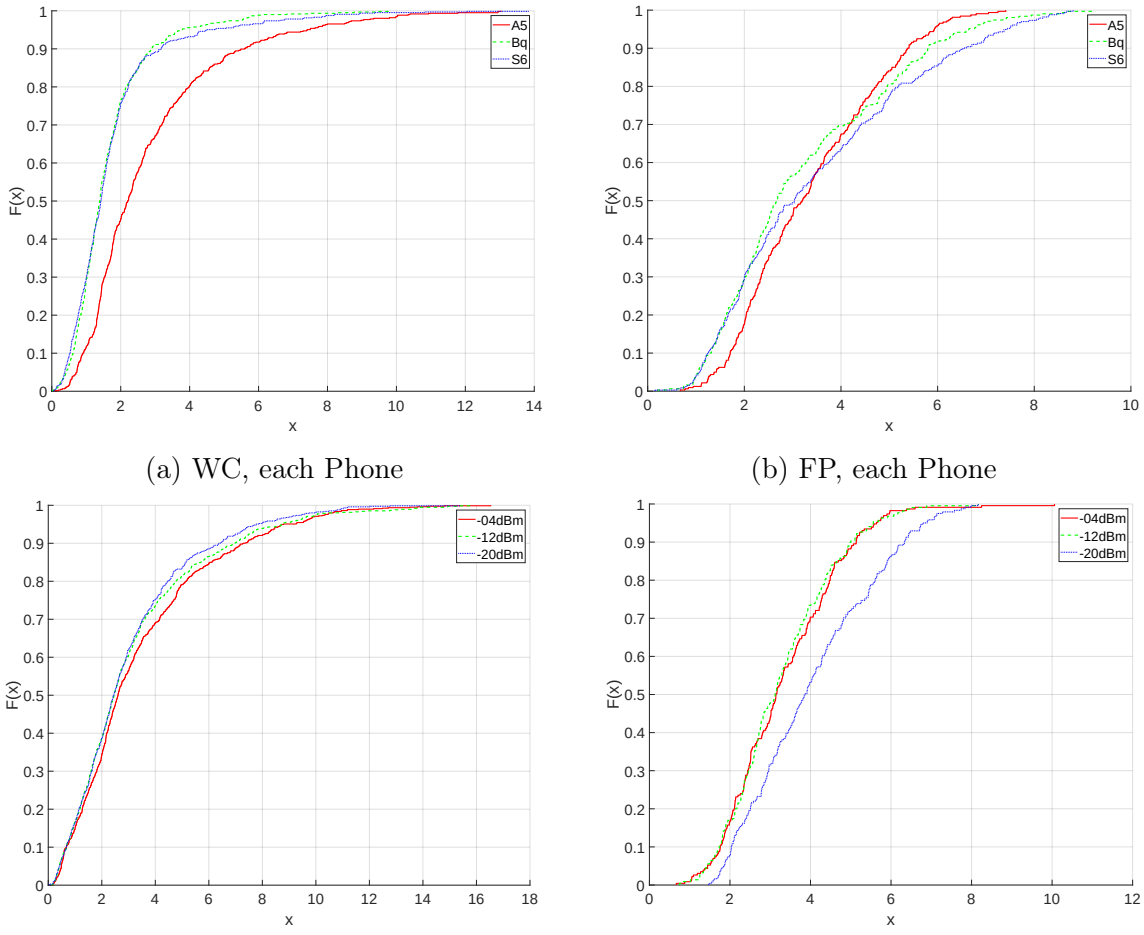
beacons, being n the number of beacons deployed in the target area. The k selected beacons are those with the highest RSS values in the fingerprint. The kNN method [1], hereinafter FP, is very popular in RSS-based indoor positioning. It requires a previously collected set of training fingerprints with associated position tags. The description of this method was previously provided in Section 3.1. The sets used for training and test purposes in the experiments shown later in this section are those suggested by the scrips shipped with the BLE measurements.

The accuracy of an RSS-based positioning method is computed in terms of the positioning errors. In the work presented in this section, the error was computed as the 2D Euclidean distance between the estimated position and the actual position where a fingerprint was collected. The accuracy metric mainly used here is the 75 percentile of the positioning error, following the same reasons presented in Section 3.1. The comparison of plots showing the empirical cumulative distribution functions is also used.

The positioning error analyzes used the RSS data from the Library zone to address the influence of distinct phone models (Figure 3.10a,b). The analyzes that use the data from the Geotec zone addressed distinct beacon transmission powers (Figure 3.10c,d).

Considering the WC method, the positioning errors are more noticeable when using data from A5 than when using data from BQ and S6. However, the errors are similar among the three phones when the FP method is used. In general, the smartphone that provided the BLE measurements most suitable for positioning is the BQ. The effect of the selected transmission power on the accuracy depended on the position method. Using data captured with the -20 dBm power setting led to better accuracy for the WC method, and a notably worse accuracy for the FP method, than when using data of the other two powers. The accuracy obtained when using the data collected at the -12 dBm transmission power was similar to that obtained for the data collected at the -4 dBm transmission power. Across the explored power configurations and collection smartphones, the WC method had maximum positioning errors larger than the FP method. However, for percentiles below the 75^{th}, the WC method provides positioning errors lower than the FP method. The smartphone employed for samples collection is likely to have an important impact on the positioning accuracy. Also, very weak transmission powers should be avoided. The positioning accuracy for fingerprinting is poor, even below that reported for WiFi in Section 3.1 for the same environment. BLE signals are more affected by fast fading than WiFi signals – its known solution is addressed later. Furthermore, the test positions for the dataset addressed in Section 3.1 included positions closer to the training positions than those from the BLE data described in this section.

A proper value of the parameter k should be found for the WC and FP methods



(a) WC, each Phone

(b) FP, each Phone

(c) WC, each Power

(d) FP, each Power

Figure 3.10: Positioning accuracy for the WC and the FP methods.. For (a) and (b), data was captured in the Library zone. For (c) and (d), data was captured in the Geotec zone.

before assessing their positioning accuracy. In WC, the k parameter represents the number of the beacons most strongly detected in the operational fingerprint whose positions are use used to compute the position estimate. In FP, k is the number of the training fingerprints most similar to the operational fingerprint whose positions are used in the estimation. For the WC method in Library zone, the difference on accuracy among several values of k was negligible, as long as the values were greater than 3. Therefore, and considering the detection delays issue explored in the previous section, three configurations were tested:

- All: all beacons detected in a fingerprint are used. All estimations are used to

compute the accuracy metric.

- K10: $k = 10$. Ten was the k value providing the best accuracy. Notice that the number of beacons detected in a fingerprint can be lower than 10. All estimations are used to compute the accuracy metric.
- KF10: $k = 10$. Estimations for samples with less than 10 detected beacons are considered unfeasible and thus, not used to compute the accuracy metric.

The results of applying the previous three configurations on the positioning accuracy are shown in Table 3.5. The value of k has no significant influence on the WC's accuracy for this environment. The layout of the Library zone determined that there was no beacon in the environment whose detection range was larger than the detection range of other beacons. The results for the configuration KF10 indicate that the accuracy for samples with 10 or more detected beacons id better than for the rest of the samples. The positioning method could be adapted so that it does not provide estimates for the latter samples. Also, the positioning method may associate to a position estimate a certainty indicator based on the number of beacons detected for a sample.

Table 3.5: Influence of the number of detectable beacons in WC accuracy.

	A5	BQ	S6
Error(m)-All	3.51	1.95	2.01
Error(m)-K10	3.51	1.94	2.00
Error(m)-KF10	2.63	1.88	1.83
Ratio-KF10	0.44	0.82	0.83

The FP method used $k = 72$. For this method, the number of detected beacons is relevant for both training and operational (test) fingerprints. The values presented in Table 3.6 help in exploring the relationship between the accuracy in FP and the number of coincident beacons detected in both the operational and training fingerprints. The number of coincident beacons was low, or very low in the A5 smartphone, with an accuracy lower than the one obtained with WC. The last two rows of Table 3.6 present the results of the Pearson correlation test between the positioning error and the number of coincident beacons for each test sample. The correlation is large and significant for the BQ and S6 smartphones. It is large but not significant for the A5 smartphone. This correlation may explain why the accuracy for FP is lower than for WC. Thus, a solution that reduces the negative effect of advertisement detection delays should be applied to both training and operational fingerprints.

Measurement windows have been proposed as solutions for advertisement detection delays [7]. As Chapter 4 will later explain, the collection of BLE measurements

Table 3.6: Influence of the number of detectable beacons in FP accuracy.

	A5	BQ	S6
Error (m)	4.43	4.61	4.88
Max number of coincident beacons	4	12	11
ρ	-0.80	-0.95	-0.93
p-value	0.2	≈ 0	≈ 0

required to create a fingerprint abstraction. A fingerprint in the BLE dataset is the result of grouping the measurements collected during a time window. However, that time window is small (1 s) and all previously seen records are discarded. Table 3.7 presents the effect that applying fixed-size moving windows over the fingerprints that belong to the same collection point had on the position accuracy. A collection point is uniquely identified by its 2D position and the direction that the subject faced when performing the collection. The RSS value that corresponds to a beacon in a window of fingerprints was computed either as the mean or the last value of the beacon's RSS values in those fingerprints. As presented in Table 3.7, the larger the window size, the better the accuracy. The improvements may be a result of the point-based collection procedure followed for the BLE RSS dataset. Also, in some cases, the accuracy difference between the FP and WC decreases.

Table 3.7: Buffering results using data for the A5 phone and power -12 dBm. For realistic comparisons, position estimation were only computed for test sets in WC.

	Mean				Last			
	Library		Geotec		Library		Geotec	
Windows size	FP	WC	FP	WC	FP	WC	FP	WC
none	4.46	3.46	4.16	3.44	4.46	3.46	4.16	3.44
2	3.77	2.58	3.91	2.69	3.85	2.56	3.93	2.69
3	3.43	2.32	3.49	2.59	3.41	2.35	3.52	2.49
4	2.95	2.29	3.38	2.61	3.03	2.31	3.47	2.45
5	2.83	2.28	3.34	2.48	2.87	2.30	3.55	2.37

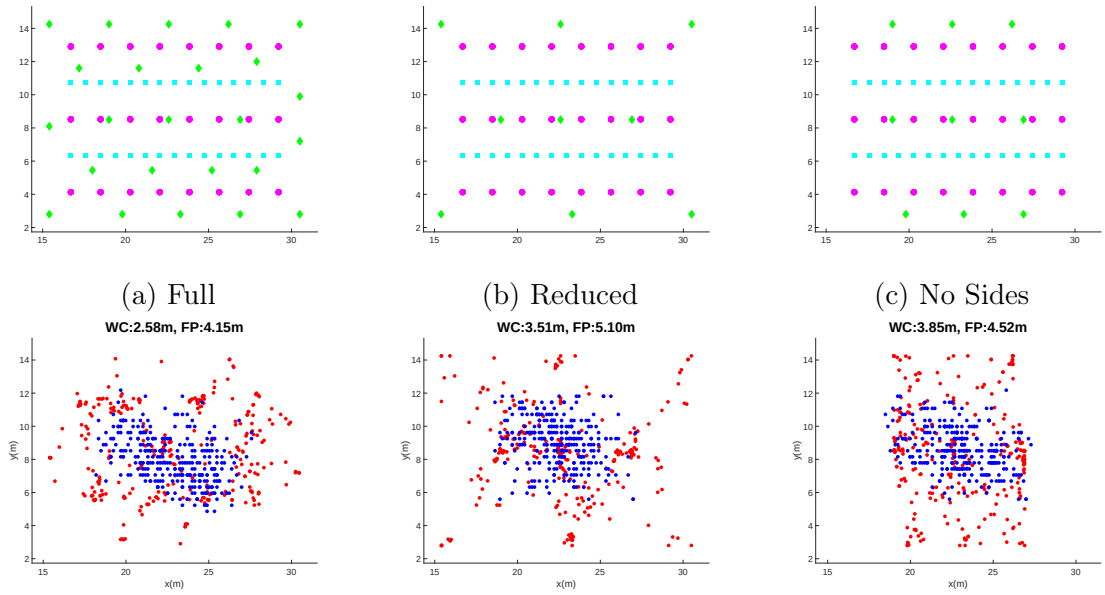
3.2.3 Beacon Deployment, Estimations Positions and Robustness

A high density of WiFi routers in the target area improves the accuracy of WiFi-based IPS. A large density of beacons also benefits BLE-based IPS [11]. BLE beacons have lower costs and are easier to deploy than WiFi routers. However, largely increasing

the number of beacons for a deployment is usually not practical. The chosen number of beacons and their placement depends on factors like deployment restrictions in the target area, the beacon range, the positioning method, and the obstacles found in the environment. Beacon deployment choice is a known problem that several works found in the BLE IPS literature have addressed [7, 5, 6], showing the complexity of the problem and only a few placement guidelines.

The beacon deployments used during the collection of the BLE RSS dataset were dense in the two zones. The scripts provided with the BLE RSS data allow selecting several distinct subsets of the deployed beacons. Those sets can be considered alternative deployments. Figure 3.11 shows the position estimates and their accuracy using the WC and FP methods for three distinct beacon deployments. The first deployment (see Figure 3.11a) included all beacons deployed in the Library zone. The positioning methods used a buffering window of size 2 to improve the results. The estimates were computed only for test points. Notice that the k parameter in the FP method had a value lower than the one used in previous experiments. The new k value provided an accuracy value worse than the previous k value. However, the new k value was chosen to avoid a heavy concentration of the estimates in the center of the target zone. The second deployment (shown in Figure 3.11b) included only nine beacons. The accuracy decreased by about 1 m in the two methods. Notice that the FP estimated positions remained clustered in the center of the area. However, the positions estimated by the WC changed and wide areas appeared near the borders and away from beacons where the WC method could not place position estimates. The third deployment (presented in Figure 3.11c) also had only nine beacons. However, no beacon was placed in the left and right sides of the area. The WC method could not achieve proper position estimates for samples in areas near the sides. The lack of beacons deployed near the boundary of the target area did not affect the FP method in these experiments. The FP accuracy was improved in the third deployment. The new layout decreased the mean distance between beacons and collection points. Thus, the number of beacons detected in training and test fingerprints was larger.

The BLE beacons are devices that commonly run on batteries. Thus, unless active monitoring and fast battery replacement are done, is likely that one or more beacons are not broadcasting at a given time when the power in their batteries is close to the exhaustion. The robustness of WC and FP methods to situations where one to six beacons stop broadcasting is presented in Figure 3.12. The WC method has better accuracies than the FP method, as previously seen, but it is also more affected than the FP when the number of active beacons decreases. The largest reported errors should correspond to cases where the removed beacons are close to the boundaries of the collection area. The FP method is not heavily affected by the positions of the removed beacons. Thus, the accuracy changes are not as drastic as for the WC method.



(a) Full (b) Reduced (c) No Sides

(d) Full (e) Reduced (f) No Sides

Figure 3.11: Estimated positions and accuracy (d–f) for three deployments (a–c) using the WC ($k = 10$) and FP ($k = 12$) methods. In the deployment figures, green diamonds, cyan squares and magenta circles represent the position of beacons, test and training samples, respectively. In the position estimation figures, red circles and blue circles represent the estimations provided by the WC and FP methods, respectively. Data was collected in the Library zone using the S6 smartphone.

3.3 Conclusions

This chapter described two RSS datasets collected to foster BLE and WiFi IPS development, credible evaluations and reproducibility. The WiFi dataset allows testing the robustness of a WiFi IPS to short and long term variations of the WiFi signals. The BLE dataset allows testing BLE IPS, with the goal of obtaining high positioning accuracies, using several beacon configurations, environments and collection smartphones. The description provided in this chapter exemplifies how open RSS datasets are collected and how they should be described to highlight the goals of their collection. Chapter 4 and Chapter 5 use the WiFi dataset described in this chapter.

The materials presented in this chapter are supported by two publications [20, 16].

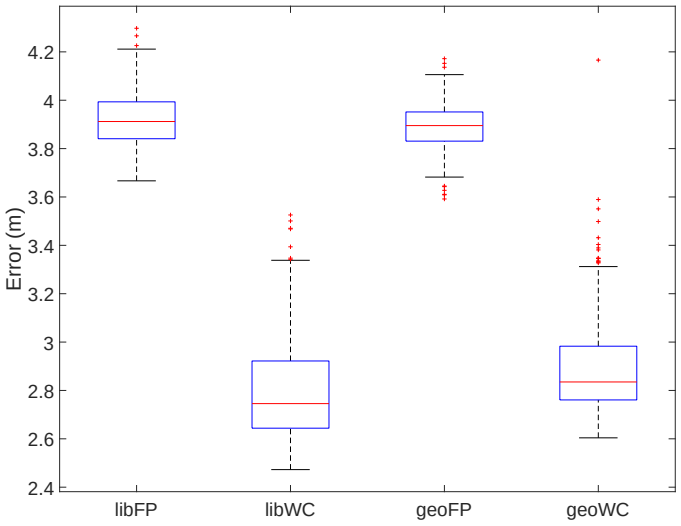


Figure 3.12: Positioning accuracies when 1–6 beacons stop sending advertisements (A5 smartphone).

References

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

[2] Paolo Baronti et al. "Indoor Bluetooth Low Energy Dataset for Localization, Tracking, Occupancy, and Social Interaction". In: *Sensors* 18.12 (2018). ISSN: 1424-8220. DOI: 10.3390/s18124462. URL: https://www.mdpi.com/1424-8220/18/12/4462.

[3] Rafael Berkvens, Herbert Peremans, and Maarten Weyn. "Conditional Entropy and Location Error in Indoor Localization Using Probabilistic Wi-Fi Fingerprinting". In: *Sensors* 16.10 (2016), p. 1636. ISSN: 1424-8220. DOI: 10.3390/s16101636. URL: http://www.mdpi.com/1424-8220/16/10/1636.

[4] K. Bouchard, R. Ramezani, and A. Naeim. "Features based proximity localization with Bluetooth emitters". In: *2016 IEEE 7th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*. Oct. 2016, pp. 1–5. DOI: 10.1109/UEMCON.2016.7777845.

[5] Jan Budina et al. "Method of iBeacon Optimal Distribution for Indoor Localization". In: *Modeling and Using Context*. Ed. by Henning Christiansen, Isidora Stojanovic, and George A Papadopoulos. Cham: Springer International Publishing, 2015, pp. 105–117. ISBN: 978-3-319-25591-0.

[6] Manuel Castillo-Cara et al. "An Analysis of Multiple Criteria and Setups for Bluetooth Smartphone-Based Indoor Localization Mechanism". In: *Journal of Sensors* 2017 (Oct. 2017), pp. 1–22. ISSN: 16877268. DOI: 10.1155/2017/1928578. URL: https://www.hindawi.com/journals/js/2017/1928578/.

[7] Ramsey Faragher and Robert Harle. "Location fingerprinting with bluetooth low energy beacons". In: *IEEE Journal on Selected Areas in Communications* 33.11 (Nov. 2015), pp. 2418–2428. ISSN: 07338716. DOI: 10.1109/JSAC.2015.2430281. URL: http://ieeexplore.ieee.org/document/7103024/.

[8] Suining He and S.-H. Gary Chan. "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons". In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 466–490. DOI: 10.1109/comst.2015.2464084. URL: https://doi.org/10.1109/comst.2015.2464084.

[9] Noelia Hernández et al. "Continuous Space Estimation: Increasing WiFi-Based Indoor Localization Resolution without Increasing the Site-Survey Effort". In: *Sensors* 17.1 (2017). ISSN: 1424-8220. DOI: 10.3390/s17010147. URL: https://www.mdpi.com/1424-8220/17/1/147.

[10] K. E. Jeon et al. "BLE Beacons for Internet of Things Applications: Survey, Challenges, and Opportunities". In: *IEEE Internet of Things Journal* 5.2 (Apr. 2018), pp. 811–828. DOI: 10.1109/JIOT.2017.2788449.

[11] Pavel Kriz, Filip Maly, and Tomas Kozel. "Improving Indoor Localization Using Bluetooth Low Energy Beacons". In: *Mobile Information Systems* 2016 (Apr. 2016), pp. 1–11. ISSN: 1875905X. DOI: 10.1155/2016/2083094. URL: http://www.hindawi.com/journals/misy/2016/2083094/.

[12] Patrick Lazik et al. "ALPS: A bluetooth and ultrasound platform for mapping and localization". In: *Proceedings of the 13th ACM conference on embedded networked sensor systems*. ACM. 2015, pp. 73–84.

[13] Elena Simona Lohan et al. "Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning". In: *Data* 2.4 (2017). ISSN: 2306-5729. DOI: 10.3390/data2040032. URL: http://www.mdpi.com/2306-5729/2/4/32.

[14] N Marques, F Meneses, and A Moreira. "Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning". In: *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. Nov. 2012, pp. 1–9. DOI: 10.1109/IPIN.2012.6418937.

[15] MathWorks®. *Support Vector Machine Regression, in MATLAB® R2017b and Statistics and Machine Learning Toolbox™*. 2017. URL: https://es.mathworks.com/help/stats/support-vector-machine-regression.html (visited on).

[16] Germán Martín Mendoza-Silva et al. "BLE RSS Measurements Dataset for Research on Accurate Indoor Positioning". In: *Data* 4.1 (2019). ISSN: 2306-5729. DOI: 10.3390/data4010012. URL: https://www.mdpi.com/2306-5729/4/1/12.

[17] Germán Martín Mendoza-Silva et al. *BLE RSS meaurements database and supporting materials*. Zenodo repository https://zenodo.org/record/1066041 (accessed Aug 2019). Dec. 2018. DOI: 10.5281/zenodo.1618692.

[18] Germán Martín Mendoza-Silva et al. *Long-Term Wi-Fi fingerprinting dataset and supporting material*. Zenodo repository https://zenodo.org/record/1066043 (accessed Aug 2019). Nov. 2017. DOI: 10.5281/zenodo.1478083.

[19] Germán Martín Mendoza-Silva et al. "Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning". In: *Data* 3.1 (2018). ISSN: 2306-5729. DOI: 10.3390/data3010003. URL: http://www.mdpi.com/2306-5729/3/1/3.

[20] Germán Martín Mendoza-Silva et al. "Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning". In: *Data* 3.1 (2018). ISSN: 2306-5729. DOI: 10.3390/data3010003. URL: https://www.mdpi.com/2306-5729/3/1/3.

[21] Mario Muñoz-Organero, Pedro J Muñoz-Merino, and Carlos Delgado Kloos. "Using bluetooth to implement a pervasive indoor positioning system with minimal requirements at the application level". In: *Mobile Information Systems* 8.1 (2012), pp. 73–82.

[22] Vicente Canton Paterna et al. "A bluetooth low energy indoor positioning system with channel diversity, weighted trilateration and kalman filtering". In: *Sensors (Switzerland)* 17.12 (Dec. 2017), p. 2927. ISSN: 14248220. DOI: 10.3390/s17122927. URL: http://www.mdpi.com/1424-8220/17/12/2927.

[23] Teemu Roos et al. "A Probabilistic Approach to WLAN User Location Estimation". In: *International Journal of Wireless Information Networks* 9.3 (2002), pp. 155–164. ISSN: 1068-9605. DOI: 10.1023/A:1016003126882.

[24] Joaquín Torres-Sospedra et al. "A realistic evaluation of indoor positioning systems based on Wi-Fi fingerprinting: The 2015 EvAAL–ETRI competition". In: *Journal of ambient intelligence and smart environments* 9.2 (2017), pp. 263–279.

[25] Joaquín Torres-Sospedra et al. "Off-Line Evaluation of Mobile-Centric Indoor Positioning Systems: The Experiences from the 2017 IPIN Competition". In: *Sensors* 18.2 (2018). ISSN: 1424-8220. DOI: 10.3390/s18020487. URL: https://www.mdpi.com/1424-8220/18/2/487.

[26] Joaquín Torres-Sospedra et al. "The Smartphone-Based Offline Indoor Location Competition at IPIN 2016: Analysis and Future Work". In: *Sensors* 17.3 (2017). ISSN: 1424-8220. DOI: 10.3390/s17030557. URL: https://www.mdpi.com/1424-8220/17/3/557.

[27] J Torres-Sospedra et al. "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems". In: *International Conference on Indoor Positioning and Indoor Navigation*. 2014, pp. 261–270. DOI: 10.1109/IPIN.2014.7275492.

[28] 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*. ACM Press, 2005. DOI: 10.1145/1067170.1067193. URL: https://doi.org/10.1145/1067170.1067193.

Chapter 4

WiFi and BLE Data Collection for IPS Fingerprinting

The first step to setup an IPS based on fingerprinting is usually the creation of a training database. This step, in turn, requires additional tasks that may determine the quality of the resulting database. This chapter addresses the tasks involved in creating the fingerprint database. First, it addresses the tools and procedures to perform the collection and the curation of the resulting data. Later, it focuses on the selection of the reference point where the collection is carried out and the position determination of APs relevant for positioning. Finally, it addresses the database enrichment using regression methods.

4.1 Collection Planning, Execution and Curation

The collection of a significant number of WiFi or BLE samples requires a notable amount of time. The time depends not only on the number of samples but also on the hardware and the software used for collection. The familiarity that the subject has with the collection process and the target environment is also a key factor. The hardware, commonly the smartphone model, determine the speed at which WiFi or BLE scans are performed. The software normally takes into account the operating system and the collection application. The operating system, most usually Android and iOS, may impose timing restrictions on the scan activities of the applications. The collection application should not impose additional timing restrictions on the collection speed. However, it may ease the collection process by having good usability and providing relevant indications. Regarding the collection subject, a person unfamiliar with a task would perform it slower, and presumably poorer, than a person that is familiar with it. Also, the process of determining one's position in a place is

not trivial for many people. Thus, the subject familiarity can decrease the collection time.

Once the signal measurements have been collected, the data should be extracted from the collection devices and errors or inconsistencies should be eliminated. Finally the collected data is transformed to the format desired for the training database of an IPS or a public database. In the latter case, meta-data should be also created.

This section describes a set of campaigns-based planning, collection, and processing tools created to ease the gathering of WiFi and BLE samples. The tools help to reduce the effort required to have the training database for WiFi or BLE fingerprinting-based IPS. Also, this section shares brief insights into the data curation for fingerprinting.

4.1.1 Campaign-based WiFi and BLE Collection

Samples collection for fingerprinting is carried out by IPS users, volunteers or IPS creators/deployers. In the latter case, the subject's goal is to gather data with accurate position tags and at positions previously determined to be the most relevant for a proper signal characterization of the environment. Having too many positions would require too much effort; while too few positions may lead to a poor positioning accuracy.

The collection may be point-based or route-based:

- In a point-based collection, the subject walks to a position, starts the measurement process for that position, and when the measurement is done, the subject repeats the process for the reaming positions.
- In a route-based collection, the measurement process starts at the first position and only stops at the last position.

Even for point-based collections, the positions should normally be visited in a specific order. Also, the direction at which the subject has to face when recording a fingerprint is relevant.

The predefined plan of the collection is called here a *campaign*. The plan contains instructions on how to conduct the collection. Performing the collection following a campaign is also useful for volunteers that help crowdsource the measurements. As it happens with the citizen science approach [45], some factors affect the quality and quantity of the contributions. Because the data collection tasks for WiFi-based or BLE-based positioning are relatively simple, the main challenge for them is the tag quality assurance. The process of taking WiFi sample is not error-free even when performed by trained subjects. Thus, the web and smartphone applications that this

section presents ease the signal measurement collection not only to reduce the required time but also the tagging errors. The tools allow the steps of planning, gathering, and processing of campaign-based, point-based collections of WiFi and BLE samples.

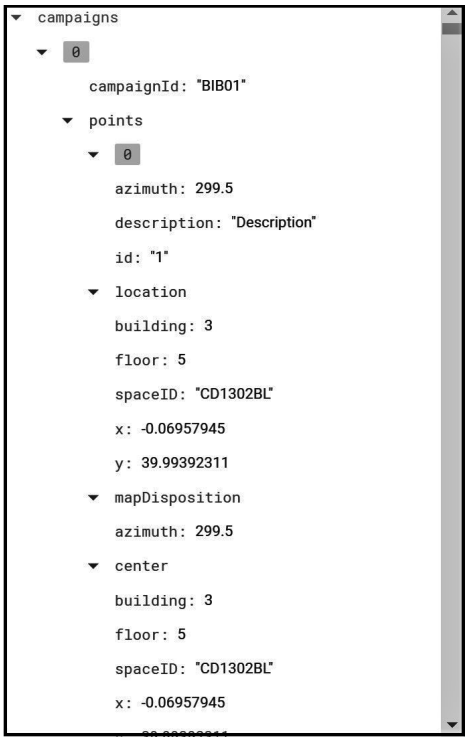
The workflow to perform a collection with the tools is as follows:

1. A user creates a campaign using the planning tool.
2. That campaign is then used to guide the collection process using the collection tools.
3. Once the collection for a campaign is completed, the data can be uploaded using the collection app, thus becoming available for its revision, formatting, and download in the visualization tool.

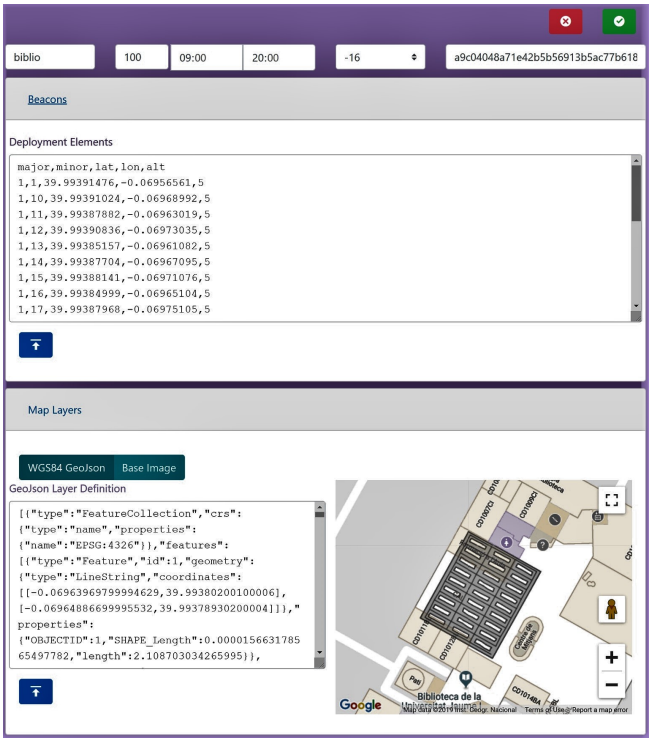
The planning tool is a web application that let registered users define all the properties that a collection campaign has. The campaign is defined using a JSON text, which is supplied to the web application. Figure 4.1a shows part of the data included in a campaign definition. The most important part of a campaign definition is the list of points where the collection is to be carried out. Apart from where it should be collected, the definition of each reference point contains information that would help the subject in performing the collection. The point information includes general data like the point id and description. Also, the information includes the direction at which the subject will have to face to perform the collection, and the center, zoom, and rotation that the app will set when the subject selects the point. The most laborious part of the campaign JSON construction is the specification of the point positions. This step can be done with a part of the web application (presented in Figure 4.1b) that allows users to put points on a map, keeping the record of the point order. Once the points are located on the map, their positions and orders can be exported to later incorporate them into the campaign's description.

For WiFi samples collection, a campaign definition is enough for setting up the collection. However, BLE collections are assumed to be linked to a previous BLE beacon deployment. Thus, the collector app can be instructed to only gather advertisements from the managed BLE devices. Figure 4.1c shows the interface for defining a new deployment. The required information includes beacon ids, positions, advertising parameters, and map-based information.

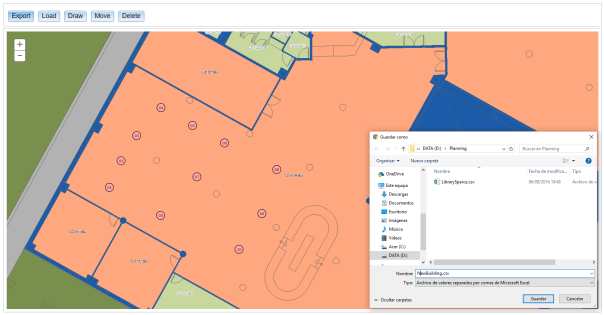
Once a campaign has been defined, a subject may select the campaign in the collection tools and carry out the campaign. The collection tools are two Android smartphone apps with interfaces very similar to each other. One of them is devoted to the collection of BLE measurements, while the other is intended for the collection of WiFi samples. The apps were used to collect the WiFi and BLE databases described in Chapter 3.



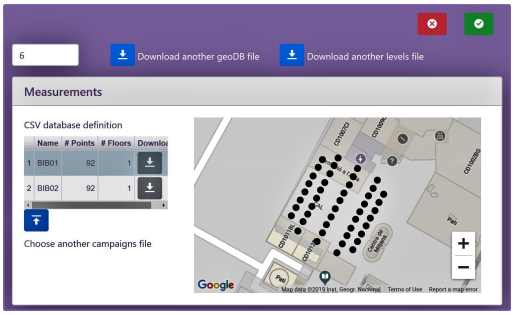
(a) Campaign definition data



(c) BLE beacon deployment definition



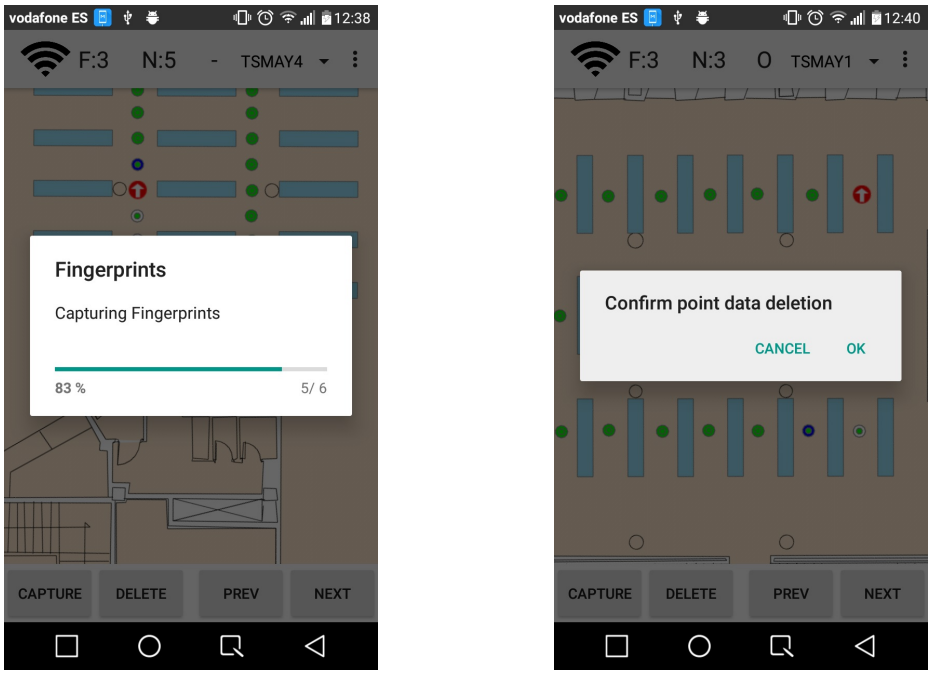
(b) Point positions definition



(d) Campaign definition interface

Figure 4.1: Screen shots of the web tools for deployment and campaign definition.

In the case of BLE, Android provides two ways of notifying BLE advertisements: anytime an advertisement is received, or when a batch of advertisements is collected. Tests performed during the application development suggested that the batch approach leads to the detection of fewer advertisements. Thus, the former approach was chosen. Giving that the advertisement collection was not batch-driven, a fingerprint abstraction was created, so that all measurements received in a window of 1 s were placed into a fingerprint. If a beacon was detected more than once in a



(a) Capture in progress (b) Measurement deletion

Figure 4.2: Scree shots of the Wi-Fi campaign-based application.

fingerprint's window time, only its mean RSS value was recorded.

In the case of WiFi, Android only allows the batch approach. Apart from varying scanning times in different smartphone models and Android versions, the test performed during the application development showed two important concerns. The first concern is that an initiated WiFi scan may be delayed or interrupted when a smartphone has an active WiFi connection and the connection switches from one AP to another. The second concern is that for some smartphones and Android versions the buffering is so aggressive that the system returns almost instantly a previously buffered fingerprint when a scan is requested. The two concerns were solved using time protection mechanisms and by giving the apps the ability to work offline. The time protection mechanisms are based on timers that control the maximum and minimum times allowed for a collecting a batch of fingerprints. Figure 4.2 shows examples of the usage of the WiFi samples collection app.

The collection apps made the collection process less cumbersome, more accurate and faster by showing collection instructions to the subject. The application let the subject choose which of the available campaign to work with. The subjects did not have to carry hard-copy instructions to follow a well-planned campaign. They just needed to focus on the map and instructions from the applications. They could finish

one campaign and then switch to another one, and upload the campaigns' data at the time of their choosing.

The subjects that perform a collection may mainly include two types of position tagging errors:

- They may place themselves at wrong positions.
- They may wrongly indicate their current position on the map for collection applications that require indications from the users.

With the two collection apps shown in this work, the second kind of error is no longer present because subjects are not required to indicate their current floor or tap their position on the application's map. The application's design reduces the likeliness of the first kind of error by:

- Map's zoom and orientation adjustments at each position.
- Indications of the facing direction at each position.
- Actions for traversing the list of positions and deleting the captures that the user may consider erroneous.
- Indication of the point that follows the current one in the capture process.

As confirmed with the collection of the Library WiFi database described in Chapter 3, the collection apps let employ most of the collection time to fingerprint capture. For example, for a campaign having 96 fingerprint batches, each batch composed of six fingerprints, the mean time to complete the campaign was 50 minutes. As the mean time for capturing one batch was 30 seconds, a mean of only 2 minutes was used for displacements, orientation, and other tasks. Among the factors influencing the efficient time utilization are:

- Capture positions corresponding to the same floor were close to each other,
- The user did not have to check hard copy instructions nor indicate the current capture position or the floor to the application,
- An indication of the following position to capture was provided, which let subjects use the batch capture time in spatially locating themselves on the following destination and help them to stay focused on the task, and
- The termination of the batch capture process was announced by sound and vibration actions in order to regain the subject's attention.

The value of the campaign-driven applications lies not only in improving the collection experience and reducing the errors but also in its potentials for crowdsourcing. Assisting the subject when collecting samples is also useful since the subjects have only to go to the places shown on the map. The responsibility of selecting the reference points may require additional time and produce stress in the people that collect the data. For the Library WiFi dataset, no samples had to be discarded after a campaign was completed. In comparison, the curators of the UJIIndoorLoc dataset [52] had to discard 6% fingerprints of its training set and edit 15% of them. The changes to the collected fingerprints were results from errors acknowledged by subjects and inconsistencies in the order of temporal marks.

Figure 4.3: Web tool for samples visualization. The point clustering changes with the zoom dynamically.

Once the collection of a campaign is complete, the subjects may upload the collected data. The uploaded measurements are then available at the data downloading and visualization tool. The data is stored in JSON and can be downloaded in JSON or CSV formats. Before downloading the data, the web application presents the user with a visualization of where the samples were collected (see Figure 4.3). The tool also allows reviewing the attributes of each sample, like the signal intensities and the subject who performed the collection.

4.1.2 Data Curation

Once the data of WiFi or BLE samples are collected, the process of creating the radio for an IPS or a public database for sharing goes beyond a quick review of where the collection was performed. In the IPS context, fingerprint data curation

involves transforming the data a specific format, apply cleaning operations, and correct inconsistencies. The format transformation can be done using automatic means. However, the cleaning and inconsistencies correction operations should be performed under careful supervision.

Android allows filtering the advertisements of BLE beacons that will be reported. For example, in the case of the iBeacon protocol, the UUID code can be used as a filter. The BLE collection app previously described linked a campaign to a specific BLE deployment. Thus, it performed additional filtering using the minor and major characteristics from the iBeacon protocol.

However, Android does not incorporate a filtering capability for WiFi scans. Furthermore, it is commonly difficult to know in advance all the fixed AP that will be seen in a scenario. Therefore, the collected data could be cleaned from WiFi APs that are not known to be deployed in the target scenario or whose behavior may affect the IPS performance. For example, in the UJI WiFi dataset presented in Chapter 3, a significant number of the APs detected along the collection time showed large variability in their detection. Those APs with large variability could be filtered out. The filtering can also be done using the MAC or SSID of the detected APs if they follow some pattern for the target scenario. For example, in the case of UJI, the AP deployed by the university's IT service broadcast with SSIDs that include either the string "Eduroam" or "UJI".

The buffering that Android smartphones commonly perform of WiFi signals can be a difficulty for positioning. The RSS values that Android report corresponds to measurements performed in previous instants. The times that those instants correspond to depend on the smartphone model and Android version. The UJI long-term WiFi dataset, which was addressed in Chapter 3, provided the opportunity to analyze such issue. The dataset was created through a point-based collection. Table 4.1 presents the mean point-wise RSS difference value among pairs of samples taken for each point. The sample pair 1–2 has the largest difference (1 dBm more than the others), which indicates that the first sample may be buffered measurements taken by smartphone's software before the subject arrived at the collection position. In the UJI long-term WiFi, six samples were collected for each position and direction. If the dataset is to be used as a radio map for an IPS, the first sample could be dropped, thus reducing the buffering artifacts.

The inconsistencies in the data are commonly the result of position tagging errors. For example, two samples collected in positions only a few meters away from each other should be similar. They should share, at least, a few of the APs that each sample detected. Also, the collection is commonly performed following an order, mainly if the subject follows a campaign-based approach. Thus, changes in the order, mainly those showing large divergences, should be verified with the subject. Order

Table 4.1: Point-wise mean RSS difference between pairs of samples in the WiFi dataset presented in Chapter 3. The measurement correspond to AP 7, 3^{rd} floor, Month 15, "direct" direction from sets Test01, Test02, Test03 and Test04.

Sample Pair	1–2	2–3	3–4	4–5	5–6
Mean Difference (dBm)	3.19	2.06	1.82	1.74	1.95

changes may even invalidate a complete collection for a campaign if the timing order was relevant.

4.2 Reference Point Selection

Samples collection is known as one of the main challenges of WiFi fingerprinting. It requires a significant effort to collect a representative radio map for large areas like, for example, several buildings in a university campus. Chapter 2 mentioned the solutions for reducing the required collection effort. Recalling, these solutions are (1) crowdsource the collection to volunteers, (2) apply a propagation model to estimate the expected RSS values, or collect samples at a reduced number of reference points and then apply a regression technique that densifies the radio map.

Crowdsourced signal data is very valuable and is the only affordable option for some cases. However, the reliability of position tags and the poor distribution of sample positions are usual concerns. The application of a propagation model to estimate the expected RSS values may provide high-quality data for the radiomap. However, the model setup requires very precise information on obstacles in the environment and the broadcasting parameters of the emitters. Furthermore, the RSS estimations may fail to account for variations caused by presence of people or the effect that the body of the collection subject has on the RSS.

An alternative to avoid this drawback is performing a quality collection at only a few reference points. Samples with reliable position tags are valuable for IPS tuning and evaluation. Some previous works addressed the reference point position determination. In Kanaris et al. [18], the authors proposed an algorithm that suggested a collection's sample size given a small preliminary set of measurements. They suggested the definition of a grid of locations in a target area and randomly choose locations in the amount determined by the sample size calculation. Jung, Moon, and Han [17] analyzed placement in under assumptions of specific positions of test samples and using simulations.

This section focuses on determining the positions of reference points most convenient for WiFi fingerprint collection. The distribution analyses take into account the

environment geometries and the experiments are performed using the data from two publicly available datasets. The experiments show how the distribution of training fingerprints, the environment structure and the detected intensities of the APs affect the positioning and the regressions used the reduce the collection efforts.

4.2.1 Materials

The two datasets used in the experiments are denoted in this section the Library and the Mannheim datasets. The Library dataset is a subset of the dataset [37] that was collected in two floors of the UJI's Library building during 25 cumulative months. Six fingerprints were collected per each reference point and each of the two directions at which the collection subject was facing. The subset contains only one training set and test sets that correspond to the first collection month. Also, as the data contained information about a 620 AP, a selection of the 48 most relevant APs was performed (as done in [53]) to ease the analyses and reduce the noise created by many intermittent APs. The collection area is a relatively small environment that covered about $15 \times 10\,\mathrm{m}^2$. The average distance between reference points is about 2 m. Chapter 3 provided more details about this dataset.

The Mannheim dataset [21] was collected in the Mannheim University. The collection area comprises a medium-scale environment, covering about $50 \times 36\,\mathrm{m}^2$ of corridors of a university department. The fingerprints are on a 1.5 m grid [20] and the positions of 11 APs are known. The data contained 110 fingerprints per reference point. Out of 110 samples, 10 were randomly selected to ease the analyzes and have a number of samples that is closer to that of the Library dataset. Both the Mannheim and the Library datasets provided their position tags using a local coordinates system that allow distance computation using the Euclidean distance.

Figure 4.4 shows the environments where the two datasets were collected. The representations of structural barrier were created from floor plans. In those representations, thicker barriers that are usually more opaque to WiFi signals were drawn in black color, while lighter shades of gray were used to draw thinner and less opaque to WiFi signals barriers. Figure 4.4 also presents the distribution of training and test reference points, as well as the position of some known APs close to the collection area.

4.2.2 Experiments

The experiments studied the influence that the distribution of collection points and the AP strength in an environment have on the accuracy of an IPS and the quality of regressions used to reduce collection efforts.

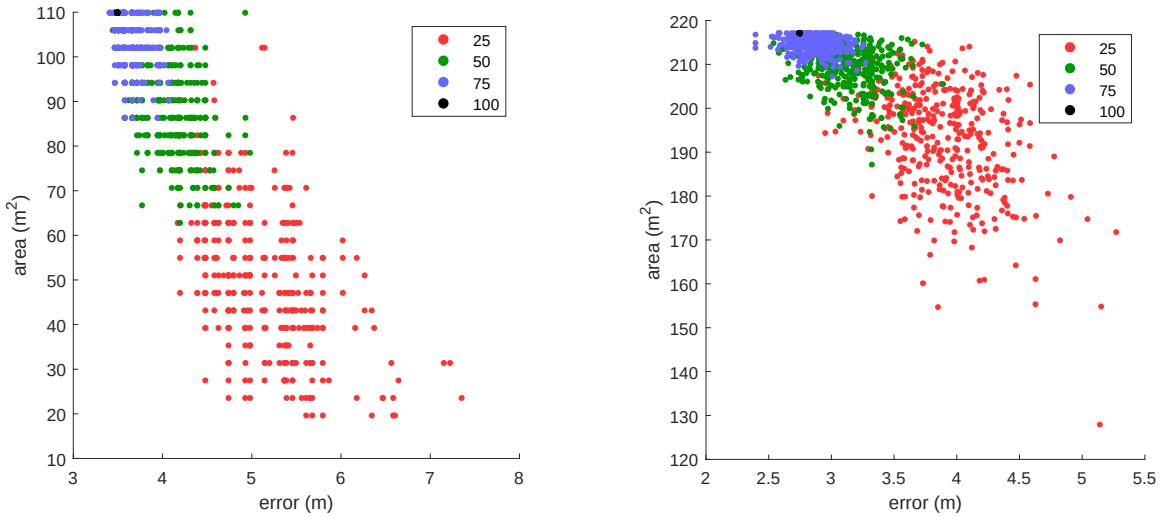
(a) Library 5^{th} floor (b) Mannheim

Figure 4.4: Environments for analyses of reference point selection. Blue and magenta squares represent training and test reference points, respectively. Known AP positions are identified by numbered orange circles.

Influence of Position Distribution on Positioning Accuracy

The goal of sample collection in WiFi fingerprinting is to achieve a proper signal characterization of the target environment. The number and distribution of the collected samples are strongly related to the quality of that characterization and thus to the accuracy of the target IPS, as shown in Figure 4.5 for the two addressed environments. The experiments for Figure 4.5 addressed one of the environments and used a percentage of the original training points for the selected environment. They were repeated 400 times. A kNN method computed position estimations for all test positions, with k set to the best performing value for the experiment data. Each dot in the figure represents the 75^{th} percentile (Q_3) value of the positioning error and the convex hull's area of the used training points. Recall that the training points for an experiment are a subset (percentage) of the original training points. The error for an estimate was computed as the Euclidean distance between the estimate and the ground truth positions. The area was computed for the convex hull, intersected with the concave hull for the original set of training point, to avoid considering non-valid zones like, e.g., positions out of the buildings.

The kNN method can provide position estimates only within the convex hull of the training reference points. Thus, the larger the area covered by the subset of training points, the better the chances that the test points lie inside the convex hull of the training points, and thus the better the chances that the kNN can provide feasible estimates for all test points. Fingerprinting rests on the notion that samples taken at positions close to each other should have RSS values more similar to each other than to the RSS values of samples taken farther away. Thus, and as widely shown



(a) Library 5th floor (b) Mannheim

Figure 4.5: Relation between positioning accuracy and the area covered by training points. The colors of the dots represent the percentage of the original training points used for an experiment.

by previous studies[14], increasing the density of training reference points may lead to a better IPS accuracy. However, Figure 4.5 shows that, besides the point density and the covered area, the layout of points is highly relevant. Apart from the 100% case of training points, the positioning accuracy can vary from almost 1 m to more than 2 m for the same density and covered area in the two environments. The largest differences are observed for the lowest point densities.

Placing training reference points close to the inner boundary of the collection area would maximize the covered area and assure that test positions are located inside the convex of the training positions. Finding the positions that define the boundary shape is a trivial task and automatic procedures like computing an alpha-shape exist [8]. The density and distribution of reference points depend on the environment and the collection effort that can be afforded. The collection is only possible at feasible locations. It is impossible to collect samples inside a wall. The feasible positions determine the pool where reference points are selected from. Depending on the effort afforded for collection, the radio map may need to be enriched. The estimation of fingerprint values for the remaining positions in the pool can be done using regression.

One strategy for creating the set of reference points is to first add reference points lying close to environment boundary and later add a number of points mp that maximize the mean minimum distance among the points in the set. In kNN, the estimated position is commonly computed as the centroid of the positions of the most similar

samples in the training dataset. Thus, maximizing the minimum distance among the training reference points reduces the areas without position estimates produced by kNN. Such an even distribution of point also benefits regressions as it provides intermediate positions that help explain non-linear behaviors. The value of mp may be dictated by the affordable collection effort. For low values of mp – like those below 20 – a brute force approach may be applied to determine the mp positions of the reference points. For large mp values, a Monte Carlo approach [41] can be used. This work used an optimization approach based on agents moving under repulsion forces [7].

To explore the convenience of using the previous training points distribution, a Pearson correlation test was applied between the mean minimum distance and the positioning accuracy for several distributions of training points. The tests were performed separately for each of the two environments. The tests used 400 random sets of training reference points that had the same number of points and included the shape boundary points. The position estimations were obtained with the kNN, using the best k for the training set, and all test samples. Table 4.2 presents the correlation results. The negative correlation between the mean minimum distance and the positioning accuracy is not statistically significant. For the Library environment, the negative weak to moderate correlation appears only for large sets, and it is statistically significant for them. The correlation is consistently negative for all set sizes in the Mannheim environment. However, its statistical significance does not show a clear pattern. The results from Table 4.2 suggest that the distribution of the inner reference points proposed above is beneficial for environments that are large or have relatively dense collections. Despite it is desirable to avoid the existence of non-positionable zones, alternative distributions may be preferable for other environments.

Influence of AP Strength on Positioning Accuracy

It is known that the signal strength decreases as the distance to the emitter increases [11]. The decrease follows a logarithmic behavior. Thus, the signals from an emitter vary more in the area close to the emitter. A good radio map should grasp as much as the signal variations in the environment as possible, and a set of reference points that includes points close to the emitter might incorporate much of those variations. To test the importance of incorporating such points in the training points, the correlation between the positioning accuracy and the closeness to the seen APs of a set of points was performed. The ideal way to measure closeness to a seen AP would be by determining the distance to it. However, such distance determination is a process that requires to know the actual position of the AP. The position estimation of an AP is normally error-prone. Thus, the closeness was considered using the RSS values. It is expected that the closer an AP, the higher the RSS values. Therefore, the maximum

Table 4.2: Correlation between the mean minimum distance between training points and the Q_3 value of positioning accuracy.

	Library				Mannheim			
%	mean k	rho	pval	Q3 acc	mean k	rho	pval	Q3 acc
0.25	2	0.004	0.932	4.740	3	-0.144	0.004	3.898
0.30	2	0.075	0.135	4.480	3	-0.069	0.167	3.562
0.35	2	0.122	0.015	4.390	3	-0.120	0.016	3.403
0.40	2	0.091	0.068	4.171	3	-0.086	0.086	3.319
0.45	2	0.024	0.628	3.971	3	-0.146	0.003	3.193
0.50	2	0.060	0.233	3.661	3	-0.131	0.009	3.146
0.55	2	0.036	0.470	3.576	3	-0.130	0.010	3.071
0.60	3	-0.036	0.471	3.576	3	-0.077	0.122	2.990
0.65	3	-0.124	0.013	3.505	3	-0.085	0.088	2.966
0.70	3	-0.200	0.000	3.466	3	-0.075	0.134	2.926
0.75	4	-0.313	0.000	3.428	4	-0.152	0.002	2.900
0.80	4	-0.395	0.000	3.390	4	-0.138	0.006	2.874
0.85	4	-0.302	0.000	3.322	4	-0.056	0.267	2.864
0.90	4	-0.279	0.000	3.318	4	-0.131	0.009	2.833

RSS value of each AP in a set of training points was computed. Then, the closeness value for a set of training points was the Q_2 value of the previous set of maximum AP RSS values. The accuracy value of a set of training points was the Q_3 value of positioning errors.

The closeness and accuracy values were computed for 400 random sets of training reference points obtained in the same way that those used in previous experiments of this section. The number of reference points of each set was 25% of the number of points in the original training set. The correlation between closeness and accuracy values was separately computed for each environment. The correlation results, presented in Table 4.3, were statistically significant. The low to moderate negative correlation indicates that the higher accuracies tend to be associated with distributions of training points with large closeness values. Thus, the results suggest the convenience of distributing some training reference points in zones of the collection area that are close to the APs.

As presented in Figure 4.5, the training sets considering only 25% of available training reference points did not achieve the positioning accuracies achieved by sets that considered larger percentages of reference points. While a reduction in the number of reference points to only 25% is desirable, the reduction in the positioning accuracy is not. To deal with such (expected) reduction, it is common to apply a

Table 4.3: Correlation between between the median maximum RSS across training points and the Q_3 value of positioning accuracy. Training random sets contained 25% of the available training points.

Library		Mannheim	
rho	pval	rho	pval
-0.37	≈ 0	-0.28	≈ 0

regression or interpolation technique that densifies the radio map. The notions of previous experiments regarding the placement of sampling points in the environment boundaries and the maximization of the mean minimum distance among sampling points also benefit the regression or interpolation results. Interpolation is the most benefit from having sets of sampling points whose convex hull covers most of the environment. Also, regression and interpolations may benefit from the even distributions created by the maximization of the mean minimum distance among sampling points, because it provides intermediate positions that help in creating non-linear models.

Influence of AP Strength on RSS Regressions

The following experiments addressed the notion of the convenience of having more reference points close to nearby APs in order to improve the RSS regression or interpolation results. The goodness of a regression or an interpolation applied to radio map densification is normally assessed by the difference between the estimated RSS and their actual values. The interpolation methods used in the experiments were Natural Neighbours [48], (Bi)Cubic Interpolation [58, 56] and Inverse Distance Weighting [47]. Natural Neighbours and Cubic Interpolation are interpolations based on triangulation and the implementation used here was that provided in MATLAB [34]. An implementation of Inverse Distance Weighting was developed for this work. The regression methods used in the experiments were SVM [54], Gaussian Process [57], Generalized Linear Models [44], Decision Trees [2] and Ensembles of Decision Trees [10]. The implementations for all regression methods were those provided by MATLAB [32, 31, 29, 33, 30]. The interpolation and regression methods, hereinafter only called regression methods, were applied using training point to fit the model and test point to computed RSS estimates. The mean RSS value for an AP and a reference point was used to train the regression model for an AP and to later compute the regression residuals. The residuals are the AP-wise absolute difference between RSS estimates provided by the regression and the actual RSS used for training.

Let $S_j = \{s_1, \ldots, s_n\}$ and $R_j = \{r_{j,1}, \ldots, r_{j,n}\}$ be two sets, where n is the number of APs detected in an environment. The value s_i was computed as the mean RSS

value of the i^{th} AP in the environment, considering all reference points. The value $r_{j,i}$ was computed as the mean of the residual values obtained for the i^{th} AP applying the j^{th} regression method in the environment. The values for the signal strength and regression residuals used for the correlation test in an environment are the sets $S = \{S_1 + \ldots + S_m\}$ and $R = \{R_1 + \ldots + R_m\}$, where m is the number of regression methods and the + operator performs the orderly concatenation of the sets. $S_i = S_j$ for all i and j.

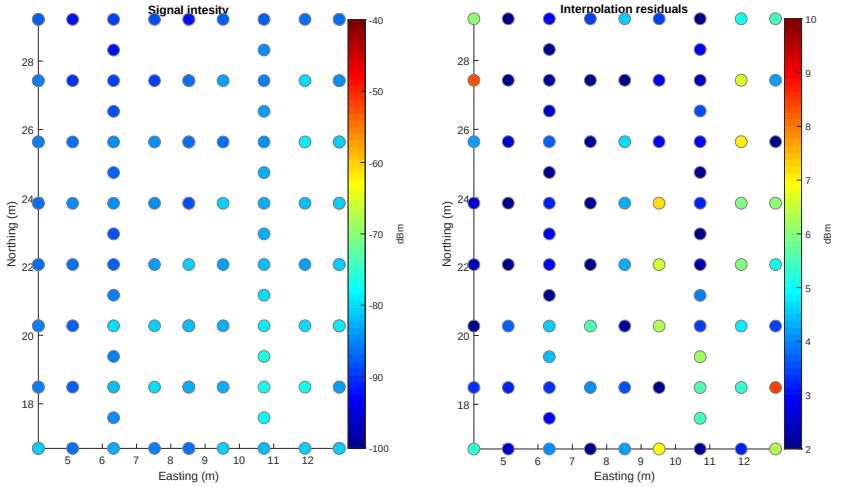
Table 4.4 shows the correlation values between signal strength and regression residuals for each environment. The correlation is statistically significant for the two environments. The correlation magnitude is weak for the Mannheim environment but notable for the Library environment. The higher the median value of the signal strength in the environment, the larger the residuals of the regressions. The correlation difference between the two environments is a likely result of the dimensions of the environments. The Mannheim environment is large, and thus the detected signal intensities for an AP can be very strong in some areas and very weak at some other areas. Very strong and very weak signal intensities are not detected for the same AP in the Library environment.

Table 4.4: Correlation between mean values of signal strength in the environment and mean values of regression residuals.

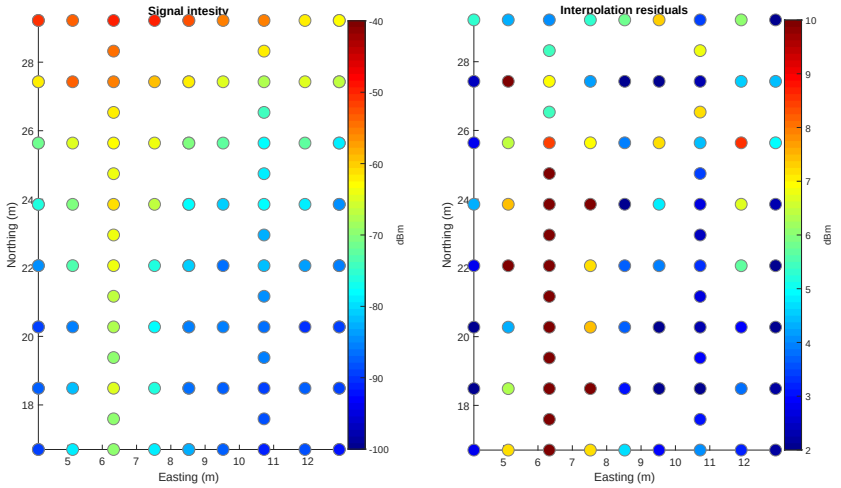
Environment	ρ	p-value
Library	0.88	≈ 0
Mannheim	0.24	≈ 0

Figure 4.6 presents a further test of the relation between the strength with which an AP is seen in an environment and the regression goodness. The exploration was performed for two APs in the Library environment. One of the APs was detected only with weak RSS values, while in the other strong RSS values were detected. The charts presented for each AP include the median value for the RSS values of the AP at each reference point and the median value of the regression residuals at each reference point. Figure 4.6a shows regression residuals of moderate values for the weak AP, while Figure 4.6b shows regression residuals for the strong AP that are not only notably larger than those for the weak AP but also mainly situated in a specific zone of the environment. The charts suggest that for weak (presumably far away APs) the regression precise only a few samples to train a model, as the APs signals are only weakly affected by the environment. However, the strength values of signals from APs near the target environment heavily depend on the LOS and NLOS situations.

Table 4.5 presents the spatial auto-correlation test as obtained by Morans'i [42] for



(a) Weak AP



(b) Strong AP

Figure 4.6: Mean value of residuals distribution compared to mean value of RSS for the Library environment.

the two antennas addressed in Figure 4.6. The table suggests that for APs strongly seen across the environment the distribution of regression residuals is not random and tends to organize in clusters; while for APs weakly seen in the environment the distribution of residuals is likely random. As stated in the literature [11], the environment influence is less significant for weak than for strong signals. Furthermore, the signal in free space follows a logarithmic decay, i.e., the farther from the AP the slower the decay rate. The tested regression models fail to account for a spatial process induced by the environment for strong signals. Thus, samples are required in

zones of LOS and NLOS with respect to strong (and thus nearby) APs, given that the RSS values in those two situations can be significantly different.

Table 4.5: Spatial auto-correlation (Morans'i) of regression residuals.

Behavior	Q_2 of RSS	Q_2 of residuals	z-score	p-value
Weak AP	-83	4	0.920	0.357
Strong AP	-74	6	7.702	≈ 0

Given the moderate correlation obtained in some of the analyses, and that the experiments were only performed in two environments, a reference point position determination method is not proposed. However, such determination method may have the following steps:

1. Place some reference points in the boundaries.
2. Distribute the rest of point maximizing the mean minimum distance among reference points.
3. Adjust the distribution to have some points closer to nearby APs.
4. Tend to LOS situations, assuring to place points in LOS and NLOS situations.

This work recommends the previous method steps as a set of guidelines that follow after the results of the analyses provided in this section. The most common approach of placing the reference points on a grid does not take into account the environment characteristics. The guidelines suggest adapting the sampling positions to the environment and highlight the importance of knowing the position of nearby antennas. Thus, the following section addresses the AP position determination.

4.3 AP Position Estimation

As seen in Section 4.2, the knowledge of the WiFi AP positions may be relevant for deciding where to collect samples for creating a fingerprinting database. Also, for the proximity and lateration techniques, the position of emitters is a fundamental piece of information. Discovering the WiFi AP position is also important for management tasks, such as detecting rogue APs. The knowledge of AP positions, however, may be unknown for IPS managers. APs may be hidden inside inaccessible places, and their positions or broadcasting parameters may be difficult to obtain the IT service of the target institution.

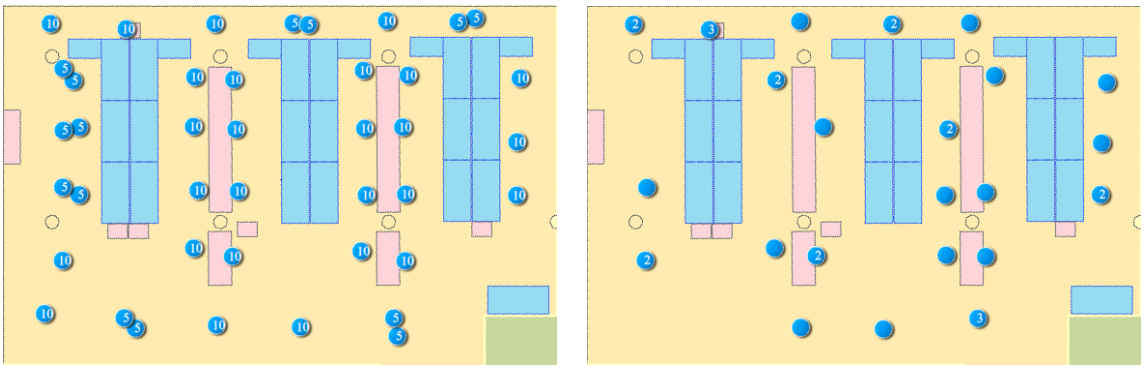
Several research works have proposed the AP position determination based on its RSS measurements [4, 1, 13, 22, 23, 61, 6, 5, 24, 16, 55, 43, 62]. Some of those methods have reported AP positioning accuracies from 2 m to 15 m [13, 24, 16, 62]. This section describes the proposal of a WiFi AP positioning method as well as experiments that assessed the accuracy of the proposed method and other methods known in the literature. Given the low accuracy that in general is possible to obtain for AP position determination, this section also presents a method proposal that qualifies whether the proposed AP positioning method is providing a good estimation. The later method is called the Situation Goodness method.

4.3.1 Materials

The WiFi dataset collected at the GEOTEC's laboratory [51] was used to perform the experiments. The dataset was collected in two separate moments by two subjects using a Samsung S3 (Android 4.3) smartphone and a G Spirit (Android 5.0.1) smartphone, respectively. The Samsung S3 smartphone detected 2.4 GHz and 5.0 GHz networks, while the G Spirit was only able to detect those at 2.4 GHz. The laboratory is a typical office working environment that includes wooden and metallic furniture, working computers, typical office equipment, people working and moving. The laboratory was not intentionally prepared in any way for WiFi-related experiments. Also, the lab occupies 260 m^2. The dataset is composed of two sets. One of them contains samples meant to be used for training a model, while the other was collected for test purposes. The dataset used in the experiments was created using the two set by applying the following steps:

1. Combine the training and validation sets. The sets were combined in order to have a large number of samples to work with.
2. Select samples corresponding to APs whose positions were known. The selected APs are the ones located inside the GEOTEC's lab and knowing their positions was mandatory for the evaluation of the AP positioning methods.
3. Use only the samples taken using the Samsung S3 smartphone. The G Spirit smartphone did not detect 5.0 GHz networks. Thus, its samples were ruled out to have numbers of samples similar for each AP.

The new dataset, hereinafter called the root dataset, was in turn used to create new datasets by applying uniform random elimination in order to create different sampling alternatives. This strategy led to obtain 10 datasets D10, D20, ..., D100 that contain 10, 20, ..., 100 percent of the root dataset samples, respectively. Figure 4.7a shows the root dataset and Figure 4.7b shows an example of a dataset containing only 10% of the root dataset samples.



(a) 100% (b) 10%

Figure 4.7: Examples of dataset used in the experiments. The encircled numbers represent the number of samples at a position or very close to it.

4.3.2 WiFi AP positioning methods

The experiments on WiFi AP positioning tested five methods: the Weighted Centroid (WC), the Interpolation Contours Centroid (Interpolation), and the method proposed by Koo & Cha [23] (Linear), by Ji et al. [16] (Monte Carlo), and by Zhao et al. [62] (Gradient). The first two methods were the implementation of a well-known method and method proposed by this work. The last three methods were selected from literature given their claimed accuracy.

Weighted Centroid

The WC method is intuitive and it has been used as a baseline method for comparison with new AP position determination proposals [4, 1, 13, 23, 61, 6, 16, 25, 62]. The reasons behind its popularity include its simplicity, its low computational complexity, and its relatively low AP positioning error in some known situations.

For a set of n AP's signal strengths measurements with known positions, the weighted centroid method estimates the AP position using Equation 4.1.

$$P_a = \frac{\sum_{i=1}^{n} P_i W_i}{\sum_{i=1}^{n} W_i} \tag{4.1}$$

P_i and W_i represent the position and the calculated weight of the i^{th} measurement. P_a denotes the resulting AP position. The weights represent the importance of each measurement position. It is calculated based on known or estimated distances, or directly from the RSS value so that the highest weights correspond to the strongest RSS values. The weight calculation used in this work was presented in Lohan et

al. [26] and it is shown in Equation 4.2, where R_i denotes the RSS value of the i_{th} measurement.

$$W_i = 10^{\frac{R_i}{10}} \tag{4.2}$$

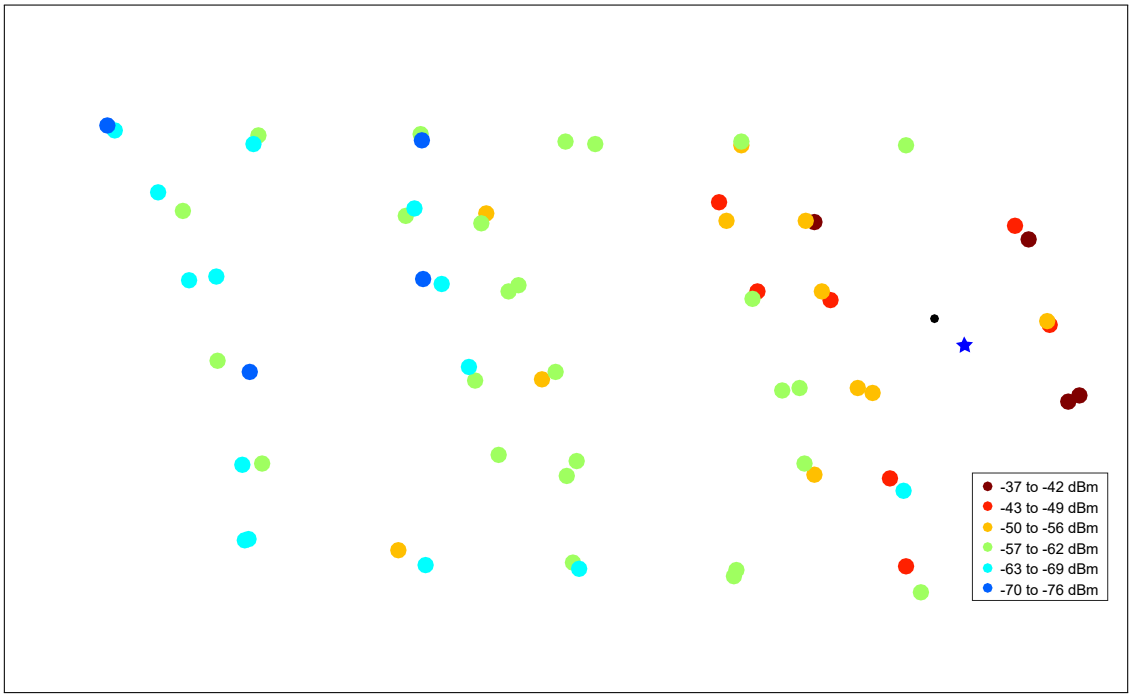


Figure 4.8: AP position estimated using the WC method. The AP position estimation, the black point, is closest to the strongest RSS values. The blue star represents the real AP position.

The distribution of the samples in relation to the AP position may heavily affect the accuracy of the method. If there are measurement taken around the real AP position or they are close to the AP, the method can provide high accuracies, as in the case shown in Figure 4.8. However, if the measurements were taken only towards one side of the AP, and none of them is close to the AP, the position estimate is poor. Despite those facts, many studies have used the weighted centroid method as a baseline method for comparisons in situations where the collection positions are not favorable for the method to provide good results.

Interpolation Contours Centroid

Interpolation and extrapolation methods have already been used in the enrichment of WiFi radio map for RSS-based indoor positioning [9, 49, 15]. Interpolation methods

normally produce better estimates than the extrapolation ones [49]. The method proposed here uses interpolation results to spot regions of the target area where the signal intensities are the highest. Later, the method uses these regions to estimate the position of the AP. This method, however, can only estimate the position of an AP inside the collection area. Interpolations only provide function value estimates for within the convex hull of the original set of positions. The regions selected in this method are those with the highest intensity, i.e., the ones closest to AP position. Thus, they can be used to get AP position estimates similar to the WC method for cases where this method perform the best, and better results than it for other cases.

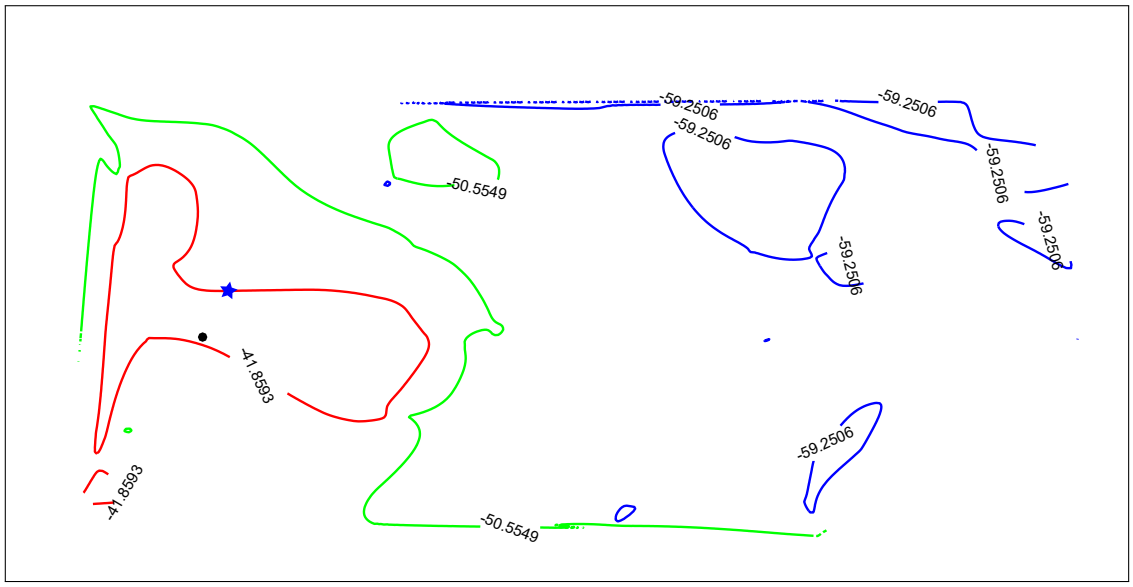


Figure 4.9: AP position estimation using the interpolation contours centroid method. The AP position estimation (red contours' centroid) is shown using the black point and the real AP position using a blue star.

This work has used for the proposed method the Natural Neighbor interpolation [48]. The steps of the proposed method are the following:

1. Calculate an interpolated grid of positions using Natural Neighbor interpolation.

2. Choose three intensity levels – high, medium and low – that split RSS values into 4 equally spaced intervals.

3. Compute the contours corresponding to the highest intensity level. Contours are calculated using the Marching Squares algorithm [27], as implemented in MATLAB [28].

4. The estimated position of the AP is the centroid of the points from the contours determined in the third step.

Figure 4.9 shows the result of estimating an AP position using the proposed method on a set of RSS values. This method is more computationally complex than the WC. Is complexity depends on the number of interpolated points.

Methods from Literature

Linear: The method presented in Koo & Cha [23] uses an equation system that relates distances to an AP and its measured intensities at several positions. Each equation is originally based on the path loss model. The equations are transformed to linear equations based on the transformations proposed by Savvides et al. [46]. The resulting system of linear equations only depends on the measurements' positions and intensities, and on one linearization coefficient that does not affect the final AP position estimate. The authors tested their method through simulations and the error estimates they presented were higher than 5 m.

Monte Carlo: In Ji et al. [16], a simulation-like (Monte Carlo test) approach is used to test a set of possible positions of an AP to find the one that better fits the measured RSS for the AP. For each possible AP position, and based on measured RSS values, the path-loss model parameters are estimated by solving a system of linear equations. Those parameters are in turn used to obtain an estimated RSS value at each measurement point. The difference between the expected RSS and the measured RSS values is then used as a metric to evaluate the explored position for the AP. The authors tested their method using simulations, achieving an accuracy below 10 m in 95% of cases. With the real samples, the method achieved an accuracy below 12 m in 90% of cases.

Gradient: The authors in Zhao et al. [62] proposed a method for AP positioning based on determining RSS gradients on a rectangular grid of positions. The gradient direction for each position was determined using discrete derivative masks using the four neighbors of each point. The authors then applied k-means clustering on the computed gradients. The head of the cluster with the highest number of collection points was used as the estimate position of AP. Also, the clustering results were used to identify and remove gradient outliers. Their method had a 1.5 m mean AP positioning error on experimentally collected data, outperforming the WC algorithm and the gradient approach without clustering.

AP positioning method evaluation

Each positioning experiment consisted in applying one of the five AP positioning methods to one dataset. As that strategy involves random elimination, the evaluation for a method proceeded as follows:

1. Repeat 50 times:
   (a) Create the datasets D10, D20, ..., D100 according to the elimination strategy.
   (b) Run the chosen method against each dataset and store the positioning error results (distance from the estimated position to the real one for each antenna) in E10, E20, ..., E100, respectively.
   (c) Store the results from the previous step into Ac10, Ac20, ..., Ac100, respectively.
2. Calculate the mean values M10, M20, ..., M100 from values accumulated in the previous step.

Table 4.6 describes the results obtained by applying the above steps. Despite the evaluation process considered to repeat 50 times the same experiment to take into account some random value selections – like in the elimination strategy or the Monte Carlo method – the mean error presented in Table 4.6 does not steadily increase as the number of measurement points decreases. Although experiments with a higher number of repetitions can be performed, it can be inferred that the above fact means that none of this method is heavily affected by the sampling density, as long as the point's distribution remains practically the same. However, considering the overall mean error, the best two are the WC and Interpolation methods. If we consider the overall variance for these two methods, the WC is better.

Table 4.7 deepens into the fact that the WC method provide better positioning results when fingerprints have been collected around the target AP. The column "Centroid (surrounded)" presents the mean positioning error for such cases, i.e., considering only APs for which there are fingerprints taken around them. The overall positioning error, which consider all datasets, for the "surrounded" cases is 3 m better than in the "non-surrounded" cases.

4.3.3 Situation Goodness method

The Situation Goodness method combines the Interpolation method and the WC method. The proposed method explores the distribution of the highest intensity level

Table 4.6: Mean estimation error of the selected AP positioning methods. OME state for Overall Mean Error and OV for Overall Variance.

Dataset	Centroid	Monte Carlo	Linear	Interpolation	Gradient
D10	3.833	8.185	7.529	3.457	11.937
D20	4.697	8.059	7.260	3.962	13.030
D30	4.447	8.119	7.266	4.310	13.636
D40	4.232	8.011	7.159	2.943	13.998
D50	3.749	7.962	7.198	3.595	14.256
D60	3.823	8.002	7.109	3.910	14.415
D70	3.978	7.886	7.078	3.381	14.605
D80	4.254	7.812	7.098	3.296	14.420
D90	4.115	7.776	7.102	3.446	14.391
D100	3.988	7.742	7.057	2.826	14.555
OME	4.112	7.955	7.186	3.513	13.924
OV	0.091	0.022	0.020	0.208	0.724

Table 4.7: Differences in mean AP positioning error for the WC method. In the column headers,"surrounded" indicates cases where there are samples taken around the AP, and "non-surrounded" cases where no samples are taken around the AP. OME states for Overall Mean Error.

Dataset	Centroid (surrounded)	Centroid (non-surrounded)
D10	2.170	5.164
D20	2.751	6.254
D30	1.972	6.427
D40	2.567	5.564
D50	1.278	5.725
D60	2.059	5.235
D70	2.277	5.339
D80	2.036	6.027
D90	1.922	5.869
D100	1.924	5.639
OME	2.095	5.724

contours computed in the Interpolation method. Specifically, the method tests for an even distribution of contours around the AP position estimated by the WC method.

Figure 4.10 presents two examples of the contours distribution with respect to an estimated AP position. Figure 4.10a presents an example where the measurements do not surround the AP – which is located at the bottom, outside the convex hull of the sample positions. For that example, the highest-level regions lie on one side of the AP position estimated by the WC method. In Figure 4.10b, the highest-level contours the AP estimated position. In this example, the WC estimate is very close to the real WiFi device position that broadcasts the AP.

The distribution of the highest-level contours in relation to an AP position is explored using geometric analyses. The analyses output a likelihood value that indicates whether the WC method provides a solution that is close to the actual AP position. The method performs the following steps:

1. Calculate the weighted centroid (WC) AP position estimation $p_{wc} = (x_{wc}, y_{wc})$.

2. Calculate the highest level regions contours, as set of points $C = C_1, \ldots, C_n$, where $C_i = c_{i1} = (x_{i1}, y_{i1}), \ldots, c_{im} = (x_{im}, y_{im})$.

3. Calculate the directions (angles) of all contour points in C, taking p_{wc} as origin, as well as the distances to p_{wc}. The direction of the contour point $c_{ij} = (x_{ij}, y_{ij})$, which is the j^{th} point of i^{th} contour, is calculated as $v_{ij} = \tan^{-1}((y_{wc} - y_{ij})/(x_{wc} - x_{ij}))$. The distance associated to c_{ij} is calculated as the Euclidean distance between c_{ij} and p_{wc}.

4. Group the contours' points according to their directions in windows with a small size ($\pi/60$).

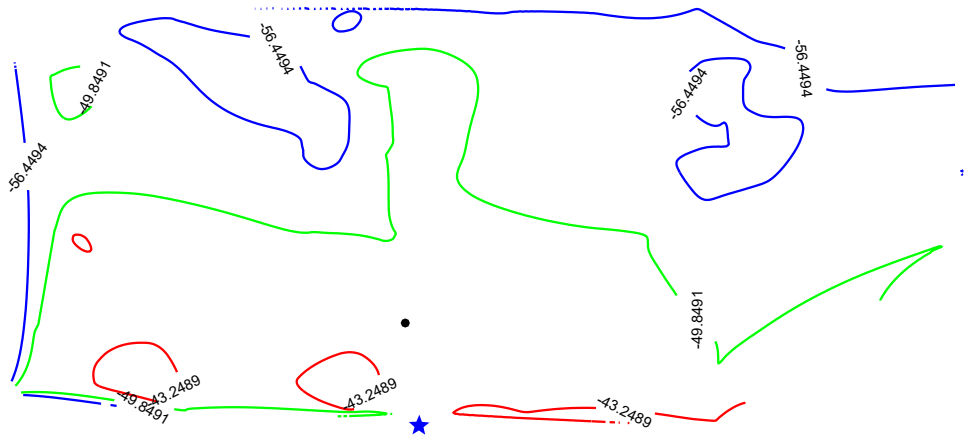
5. For each pair of opposing windows (W_f, W_b), e.g., $Wf = [0, \pi/60]$ and $Wb = [\pi, \pi + \pi/60]$, do:

   (a) Calculate s_{wf} and s_{wb} as the sum of all distances (calculated in step 3) of points in W_f and W_b, respectively.

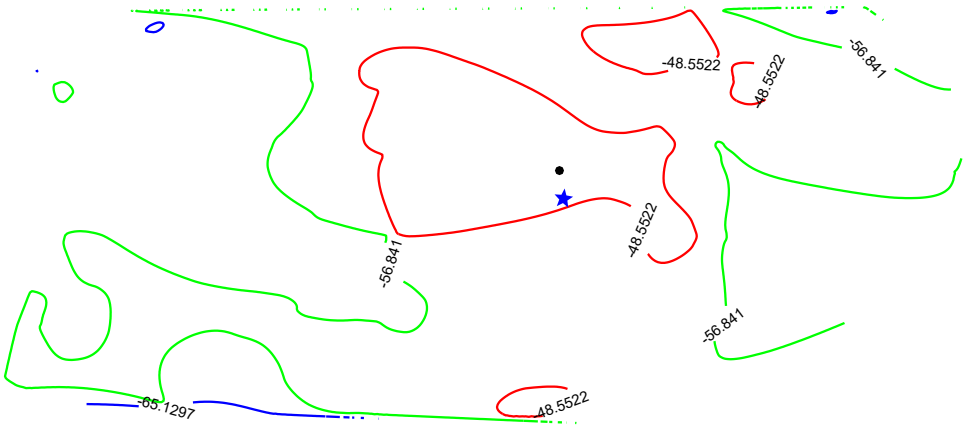
   (b) Determine the dominance d_{wf} and d_{wb} of windows W_f and W_b, respectively, as follows:

      i. When $(s_{wf}/s_{wb}) > 2$, $d_{wf} = 1$ and $d_{wb} = 0$.
      ii. When $(s_{wf}/s_{wb}) < 0.5$, $d_{wf} = 0$ and $d_{wb} = 1$.
      iii. When $0.5 \leq (s_{wf}/s_{wb}) \geq 2$, $d_{wf} = 1$ and $d_{wb} = 1$.

6. Fill the dominance gaps. When two windows whose dominance is 1 are separated by no more than 3 (grace gap) windows whose dominance is 0 (gap windows), set the dominance of the separating windows to 1.



(a) Measures to one side of the AP



(b) Measures around an AP

Figure 4.10: Examples of intensity level contours. The highest-level contours are drawn in red. The black dots represent the AP position estimation provided by the WC. The real AP position is shown by blue stars.

7. Set u as the number of opposing windows whose dominance is different and t as the total number of windows whose dominance is 1.

8. Provide the output likelihood as: $prob = 1 - u/t$.

The descriptions previously provided for the WC and Interpolation methods indicate how the results from steps (1) and (2) are to be computed, respectively. The angles and distances obtained in the step (3) are used to determine how balanced is the distribution of the points in the highest-level contours wit respect to the AP position estimated by the WC method. For the analyses, the contour points are grouped into fixed-size windows in step (4). The window's size presented in that step, $(\pi/60)$, was chosen because it is relatively small and round. Values smaller than $(\pi/60)$ were also tested and provided equally good results.

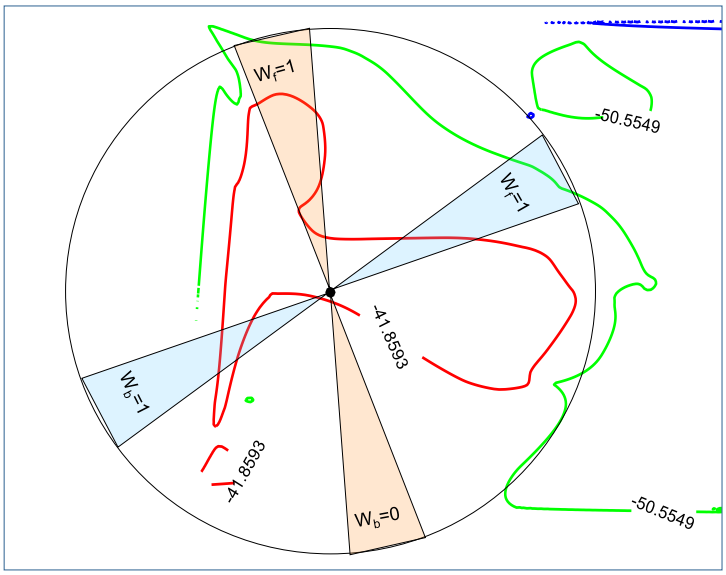


Figure 4.11: Examples of opposing windows and the meaning of dominance. Triangles of equal color represent opposing windows. None of the two blue windows is dominant, but one of the orange ones is.

The step (5) associates values called dominances to each window. Each of these values indicates, for a particular pair of opposing windows (opposing sets of directions), towards where the highest level areas are located (see Figure 4.11). To determine the winning window, the sum of distances from the window's points to the centroid position is computed. If the sum of a window is twice or more the opposing window's sum, the first window is the dominant one.

The contours can be very irregular. Thus, step (6) fills the gaps between dominant windows that are close to each other, i.e., the windows corresponding to those gaps are made dominant. The value determining the maximum size of gaps is called grace gap. This value is directly related to the window size. The grace gap value of 3 was found to work well for a window size of $(\pi/60)$.

Step (7) uses dominance values to find unbalanced situations. The unbalanced situations are those where one window is dominant and the opposing one is not. The count of unbalanced situations is used in step 8 to determine the intended likelihood.

Situation Goodness method evaluation

The evaluation for the Situation Goodness method followed an approach similar to that used for the evaluation of AP positioning methods. In particular, the evaluation used the following steps:

1. Repeat 50 times:
   (a) Create the datasets D10, D20, ..., D100 according to the elimination strategy.
   (b) Run the method against each dataset and store the probability results in P10, P20, ..., P100, respectively.
   (c) Accumulate the results from the previous step into Ac10, Ac 20, ..., Ac100, respectively.
2. Calculate the mean values M10, M20, ..., M100, and variance values V10, V20, ..., V100 from values accumulated in the previous step.

The evaluation results are presented in Table 4.8 and Table 4.9. Hereinafter, we call "surrounded cases" the cases in which the WC method provides a good AP position estimation – as a result of the existence of measurements taken around the AP. The other cases are called here "non-surrounded cases". In Table 4.8 and Table 4.9, the column "WCG" indicates the surrounded cases with a "yes" value and non-surrounded cases with a "no" value.

Table 4.8 shows that the Situation Goodness method provides likelihood values that are always below or equal to 0.3 for non-surrounded cases. The lowest likelihood values correspond to the datasets with the highest number of measurements. The likelihood values are almost always above 0.3 for the surrounded cases. For datasets from D20 to D100, it is always above 0.3. However, if the dataset has a very low number of samples, the likelihood value is not conclusive for identifying the surrounded cases. The largest likelihood values for the surrounded cases correspond to the datasets with the highest number of measurements.

The mean likelihood values for APs 70, 74, 80 and 84 are similar. Those APs are likely broadcast from a single device. Two of them are broadcast at 2.4 GHz and the other two at 5.0 GHz. Considering the surrounded cases, the antenna 68 is the one with lower likelihood value for the largest dataset. Antennas 95, 96 and 97 are

Table 4.8: Mean values of the likelihood provided by the Situation Goodness method. The column ID represents the ID of the AP. Column WCG indicates whether the WC estimate was a good solution.

ID	D10	D20	D30	D40	D50	D60	D70	D80	D90	D100	WCG
67	0.19	0.29	0.27	0.28	0.30	0.24	0.22	0.28	0.26	0.13	no
68	0.18	0.36	0.42	0.43	0.42	0.44	0.50	0.53	0.52	0.52	yes
70	0.14	0.22	0.15	0.19	0.09	0.13	0.13	0.09	0.08	0.00	no
74	0.14	0.17	0.21	0.21	0.15	0.20	0.19	0.11	0.19	0.00	no
80	0.15	0.20	0.17	0.17	0.12	0.13	0.14	0.10	0.10	0.00	no
84	0.16	0.17	0.18	0.15	0.11	0.11	0.14	0.10	0.12	0.00	no
95	0.42	0.54	0.59	0.53	0.62	0.63	0.70	0.71	0.66	0.85	yes
96	0.17	0.31	0.36	0.40	0.43	0.48	0.42	0.46	0.51	0.66	yes
97	0.46	0.65	0.71	0.71	0.79	0.82	0.82	0.86	0.83	0.91	yes

placed on top of desks, while antenna 68 is attached to the ceiling. The antenna 67 is located inside of a room adjacent to the collection area.

Table 4.9: Variance values of the likelihood provided by the Situation Goodness method. The column ID represents the ID of the AP. Column WCG indicates whether the WC estimate was a good solution.

ID	D10	D20	D30	D40	D50	D60	D70	D80	D90	D100	WCG
67	0.08	0.08	0.08	0.06	0.07	0.06	0.04	0.05	0.04	0.00	no
68	0.05	0.08	0.07	0.06	0.07	0.05	0.02	0.02	0.02	0.00	yes
70	0.04	0.07	0.05	0.06	0.01	0.02	0.02	0.01	0.01	0.00	no
74	0.06	0.07	0.08	0.10	0.04	0.07	0.07	0.04	0.06	0.00	no
80	0.05	0.06	0.06	0.06	0.02	0.02	0.03	0.01	0.01	0.00	no
84	0.06	0.06	0.07	0.06	0.03	0.03	0.06	0.03	0.04	0.00	no
95	0.11	0.11	0.08	0.08	0.07	0.05	0.04	0.04	0.06	0.00	yes
96	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.02	0.03	0.00	yes
97	0.12	0.08	0.06	0.07	0.02	0.02	0.01	0.01	0.02	0.00	yes

Table 4.9 presents the variability of the likelihood values obtained through 50 experiment repetitions. No variation was found for the dataset D100 because no sample elimination was performed on it. The variance values are no higher than 0.07 for datasets D50 to D90. Also, the mean likelihood values for those datasets (see Table 4.8) were consistently above 0.3 for surrounded cases and consistently below or equal to 0.3 for non-surrounded cases. Thus, the threshold value of 0.3 can be used

to tell apart surrounded cases from non-surrounded cases in this environment.

Table 4.10: Positioning mean errors on dataset D100 when the Situation Goodness method is applied to combine the Linear and WC methods.

ID	Linear	W.C.-Linear	WCG
67	15.021	15.021	no
68	7.778	5.202	yes
70	5.677	5.677	no
74	5.766	5.766	no
80	5.592	5.592	no
84	5.847	5.847	no
95	3.483	0.640	yes
96	10.212	0.808	yes
97	4.135	1.045	yes

This work used the Situation Goodness method to combine the WC method with another AP positioning method. Given a set of AP RSS measurements, if the likelihood given by the Situation Goodness method is lower than 0.31, use the main method. The main was either Monte Carlo, Linear, Interpolation or Gradient. Otherwise, use the WC method. Table 4.10 presents the positioning errors when the above combination of methods is applied to dataset D100. The WC method was combined with the Linear method. The likelihood provided by the Situation Goodness method was correctly used to identify the surrounded cases for an AP. The AP position estimate provided by the WC was used instead of the one provided by the Linear method.

Table 4.11 presents the positioning errors when the AP positioning evaluation is performed using the method combination explained above. When compared with the data presented in Table 4.6, Table 4.11 shows that for the Interpolation method, the mean positioning error did not change significantly. However, the mean error is greatly reduced for the Monte Carlo, Linear and Gradient methods, which have poor performance on the tested datasets.

4.4 Regressions

Section 4.2 showed that regression goodness is affected by whether the estimates are computed for a strongly or weakly seen AP, i.e. if the AP is nearby the target scenario or not. Section 4.2 also showed that the selection of positions for samples collection should tend to the position of the AP. Thus, the knowledge of the AP position for

Table 4.11: Positioning mean errors on dataset D100 when the Situation Goodness method is applied to combine the Linear and WC methods. OME state for Overall Mean Error and OV for Overall Variance.

Dataset	WC	WC – Monte Carlo	WC – Linear	WC – Interpolation	WC – Gradient
D10	4.681	6.510	6.220	4.455	8.280
D20	4.417	5.865	5.800	3.873	9.228
D30	4.196	5.245	5.622	3.755	9.134
D40	4.183	5.300	5.385	3.590	9.936
D50	4.028	5.031	5.460	3.507	9.895
D60	4.033	4.839	5.220	3.461	10.202
D70	4.116	4.808	5.119	3.415	9.756
D80	4.131	4.748	5.354	3.304	10.321
D90	4.097	4.755	5.223	3.158	10.372
D100	3.988	4.278	5.066	2.799	11.092
OME	4.187	5.138	5.447	3.532	9.822
OV	0.045	0.411	0.124	0.197	0.615

creating a radio map is useful. Section 4.3 addressed the position estimation for an AP based on signal measurements. The section explored several methods in an environment in which all APs were relatively close or inside to the environment. Despite the benevolent environment characteristics and the dense considered collection, the AP positioning errors were above 3 m. Thus, Section 4.3 explored the detection of whether an AP was inside or very close to the collection area, which is the case where the Weighted Centroid method provides its best results.

This section builds on the results from Section 4.2 and Section 4.3, to provide better regression results from samples collected at only a few reference points. The analyses explained in this section harness the guides provided by Section 4.2. Also, the analyses harness the possibility of knowing whether the AP is close to or inside to the collection area provided by Section 4.3 and a map representation of the target environment.

Regressions or interpolations are important solutions for reducing the RSS collection effort. The radio map densification using regression has been effectively used in several studies using regressors like linear regression, nonlinear Gaussian Process, Gaussian Kernel Learning, Support Vector Regression, and Random Forest [12, 15, 19]. Regressions have also been combined with the path loss model [59]. Furthermore, studies of several interpolation and extrapolation methods applied to radio maps exist [49].

4.4.1 Materials

The experiments of this section were performed using the materials from the Library environment already described in Section 4.2. Recall that the Library environment was the environment whose RSS values were the most correlated with positioning errors and regression residuals. Specifically, the residuals were particularly large for APs that, according to their RSS values, were inside or close to the collection area. Furthermore, the spatial distribution of those residuals was not random, suggesting that the positions of the reference points were insufficient to fully model the collected RSS values.

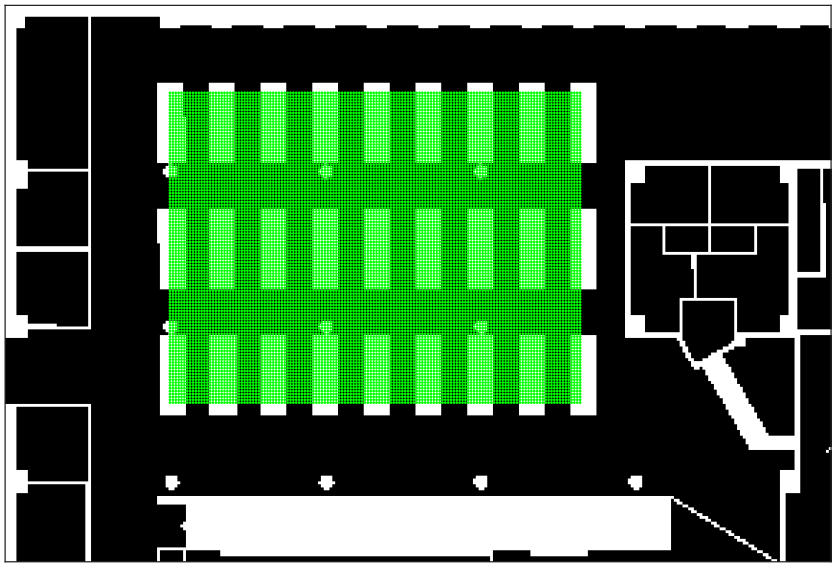


Figure 4.12: Environment and regression estimates positions. Black and white pixels indicate free space and obstacles, respectively. The green small dots represent the positions where the RSS values of an AP were estimated using regression.

This section uses map information to achieve a better model for the Library environment, as presented in Figure 4.12. The image represents all boundaries and obstacles found inside the scenario. The image was created with a low level of details. However, its map scale and pixel density allow the direct transformation from pixel distance to real-world distance. The image only contains two values (zero or 255) that represent either free space or obstacles. The white rectangles in the center of the image are bookshelves. The WiFi RSS for the Library environment were collected among those bookshelves. The green dots indicate the positions where regressions are applied in the new analyses in order to create regression visualizations.

Map gathering and map transformation imply extra efforts. However, they are important for the collection, as explained in Section 4.1); for radio-map enrichment

[15]; for estimation corrections using map matching [60]; and for IPS evaluation, as Chapter 3 addressed. Furthermore, the user applications that employ a positioning service normally display a map that provides context to the position.

The experiments of Section 4.2 followed the division in training and test sets suggested by the Library dataset. This section combines all sets used in Section 4.2 into a single one.

4.4.2 Experiments

Table 4.4 from Section 4.2 depicted a strong correlation between the signal strength of an AP in the Library environment and the regression residuals. Table 4.12 presents here the regression residuals of the experiment performed for Table 4.4 for each AP and regression method. The values presented in the table are for the Natural Neighbors (NAT) interpolation, Cubic (CUB) interpolation, Inverse Distance Weighting (IDW) interpolation, Support Vector Machine (SVM) regression, Gaussian Process regression (GPR), Generalized Linear Models (GLM) regression, Decision Trees regression (DTR), and Ensembles (ENS) of Decision Trees regression.

The APs whose mean RSS values are stronger than -80 dBm have large mean residuals. In particular, the AP 49 is the one with the strongest mean RSS value and the largest mean residual values. The APs with the IDs 33, 34, and 53 are the ones with the lowest mean residual values. Despite they are not the APs with the lowest mean RSS values, their mean RSS values are -88.6 dBm, -88.4 dBm, and -87.7 dBm, respectively, which are close to the minimum value of -89.6 dBm. Thus, most of the tested methods seem to be appropriate for regression on RSS values for weakly detected APs. However, careful attention is required for APs strongly detected in the target environment. Considering all APs, Natural Neighbors interpolation, Support Vector regression, and Gaussian Process regression provide, as a mean, better results than the other methods. However, if only the APs strongly seen in the environment are considered, it is difficult to choose a proper regression method. The Natural and Cubic interpolation methods and the Gaussian Process regression share the best results among the strong APs. The Support Vector regression provides low mean residual results for some strong APs.

Given the results from Table 4.12, the new regression model was intended to either made use of the Gaussian Process regression and Support Vector regression. Regression goes beyond interpolation in their prediction capability. They can cope with models more complex than an interpolation and provide estimations for positions beyond the convex hull of training positions. Natural Neighbors interpolation was chosen over the Cubic interpolation as the baseline method for comparisons given that the former provided lower mean residuals than the latter considering all APs.

Table 4.12: Mean value of regression residuals by AP and regression method for the Library environment. The RSS values are the mean value detected for the AP. The highlighted rows corresponds to APs strongly seen in the environment.

AP ID	RSS	NAT	CUB	SVR	IDW	GPR	GLM	DTR	ENS
01	-80.5	3.4	3.5	3.3	3.5	3.3	3.2	3.8	3.4
02	-83.8	3.8	3.8	3.6	3.7	4.2	4.2	3.9	5.2
03	-82.4	2.9	3.0	2.8	3.2	2.8	2.8	2.8	2.8
04	-81.6	3.9	4.0	3.8	4.0	3.7	3.7	4.2	4.2
06	-80.3	3.5	3.7	3.6	3.6	3.2	3.2	4.2	3.5
07	-82.2	2.8	3.0	2.9	2.8	2.6	2.6	2.9	2.9
08	-81.5	4.0	4.3	3.7	4.1	3.6	3.6	3.9	4.1
09	-84.1	4.7	4.2	4.1	4.0	4.3	4.3	4.1	5.1
11	-84.4	3.1	3.0	3.0	2.8	2.9	2.9	3.2	3.1
12	-84.5	2.9	2.9	2.7	2.6	2.9	2.9	3.2	2.7
15	**-71.6**	4.3	4.5	4.9	4.6	4.4	4.9	4.9	5.3
17	**-71.7**	4.3	4.6	4.6	4.8	4.2	5.0	4.8	5.3
19	-86.5	2.4	2.2	2.1	2.2	2.7	2.8	2.9	2.9
22	-86.8	2.3	2.2	2.3	2.2	2.7	2.7	2.3	2.3
25	-82.7	3.3	3.4	3.2	3.3	3.1	3.1	3.4	3.2
27	-85.1	2.8	2.9	2.6	2.8	2.6	2.6	2.8	3.2
28	-82.9	3.4	3.4	3.2	3.2	3.3	3.2	3.2	3.4
30	-85.8	3.1	3.4	2.8	3.3	2.8	2.8	3.4	3.5
33	-88.6	1.8	1.7	1.6	1.8	1.7	1.7	2.0	1.8
34	-88.4	1.8	1.8	1.7	1.8	1.7	1.7	2.1	1.9
41	-89.7	2.5	2.6	1.9	2.6	2.0	2.0	2.8	2.4
42	-89.6	2.4	2.5	2.1	2.5	2.1	2.1	2.7	2.4
47	-86.9	1.7	1.7	2.0	1.6	2.1	2.1	2.0	2.0
48	-86.6	1.9	1.8	1.8	1.7	2.1	2.1	2.2	2.2
49	**-70.2**	5.3	5.5	6.0	5.6	5.3	5.6	5.5	6.1
50	-87.9	2.0	1.9	2.0	2.0	1.9	2.0	2.3	2.3
51	**-71.1**	5.1	5.2	5.1	5.4	5.1	5.3	5.5	5.7
52	**-71.2**	5.5	5.2	5.6	5.2	5.2	5.5	5.9	5.6
53	-87.7	1.8	1.9	1.6	1.8	1.9	1.9	1.8	1.8
54	**-71.5**	6.7	6.6	7.7	6.2	7.1	7.1	6.5	6.7
63	-84.0	2.4	2.3	2.6	2.5	2.4	2.4	3.0	2.4
64	-84.1	2.3	2.2	2.3	2.8	2.4	2.4	3.0	2.9
69	**-73.0**	4.4	4.5	4.4	5.5	4.1	6.1	5.8	5.2
71	**-74.7**	5.0	4.7	4.7	5.1	4.5	5.9	5.4	4.9
mean	–	3.33	3.35	3.30	3.38	3.26	3.42	3.60	3.61

Also, the mean residual value of the Natural Neighbors interpolation was lower than the one for the Cubic interpolation for more than half of the strong APs.

The regression model presented in this section uses information on the AP position and a map from the target environment. This section relaxes the goal of finding a position estimate for an AP presented in Section 4.3. Instead, the new goal is to

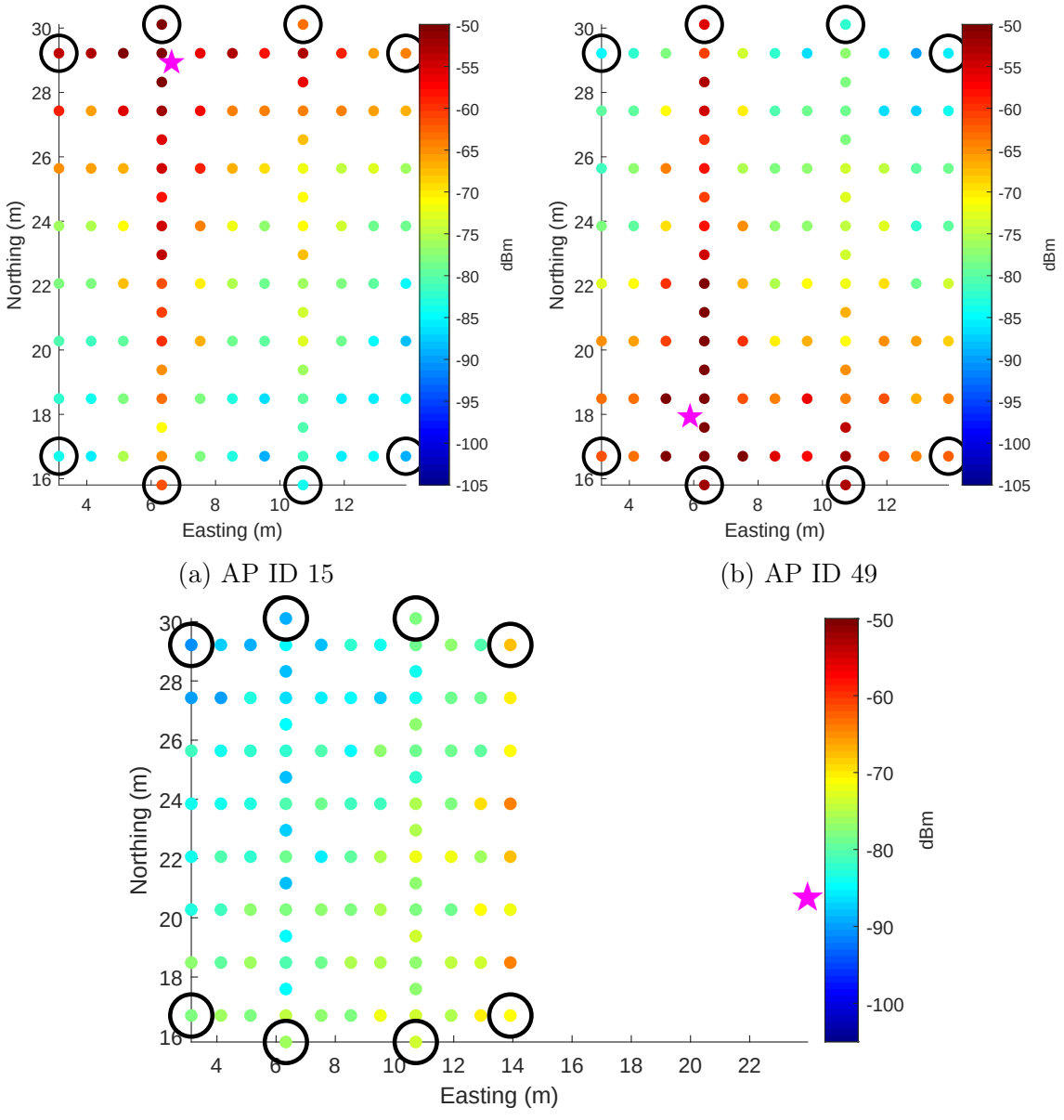
have a reference position. The reference position is used as a raw indication of where the AP is. The position of APs inside or very close to the collection area can be determined with relatively high accuracy. Either the Weighted Centroid method or the Interpolation method presented in Section 4.3 can be used in such cases. Furthermore, a quick walk in the target environment carrying a smartphone with an AP strength measurement application could suggest an approximate position of an AP.

The accuracy of AP positioning method is low for APs that are away from the collection area. Thus, an indication of the direction of the AP in relation to the collection area is preferred. Determining whether an AP is within the collection area could be done using the Situation Goodness test presented in Section 4.3 if a relatively dense sample collection is available. If a dense collection effort is not affordable, a sparse collection effort can be followed by a quick walk in the target environment carrying a phone can again be a useful strategy. APs away from the collection area are typically detected with maximum intensities weaker than -60 dBm. The reference AP position can be placed a few meters away from the center of environment limits where the AP is most strongly detected.

Figure 4.13 presents three examples of reference AP positions, along with mean RSS values per reference point of the AP. The APs shown in Figure 4.13 will later be used to show the application of the proposed regression model. The APs with IDs 15 and 49 are inside the collection area. Their positions shown in the figure are about half a meter and more than a meter away from the actual device positions, respectively. The position of the device that emitted the AP with ID 8 was unknown. The position shown in the figure is a raw approximation of the direction where the AP is. Nevertheless, such reference position is useful, as later analyses will show.

In the regressions used for Table 4.4, the predictor variables were the components of a position and the response variable was the RSS value associated with that position. In the model that this section proposes, the predictor variables include the position components but also variables whose values depend on the reference position of the AP and the environment map information. Moreover, the values of the response variable are determined as $\log_{10}(-RSS)$. If est is an estimate provided by the regression model, the RSS estimate is computed as $-(10^{est})$. The positions of points used for training and testing the model are expressed in the coordinate system of the Library dataset. Thus, their coordinates had to be transformed into image coordinates before applying the proposed model. The following definitions assume that the positions are expressed in image coordinates.

Let $rp = (rp_x; rp_y)$ be the position of a reference point used for training the model. Let $ap = (ap_x; ap_y)$ be the position of the AP targeted for regression. Let $B_{rp} = \{(x_1, y_1), \ldots, (x_k, y_k)\}$ be the line that connects rp and ap. The cell positions that constitute the line are determined using the Bresenham's line algorithm [3].The



(a) AP ID 15

(b) AP ID 49

(c) AP ID 8

Figure 4.13: Mean RSS values per reference point and device reference positions of three APs. The position of the AP's device is indicated with a star. Circles highlight the reference points whose values where used to train regression models.

values of predictor variables for rp are:

$$P_{rp} = \{rp_x, rp_y, \frac{d_{rp}+1}{2}, F_{rp}\}, \tag{4.3}$$

where d_{rp} is the Euclidean distance between rp and ap and $F_{rp} = \{f_1, \ldots, f_k, \ldots, f_n\}$. The value f_i is computed as:

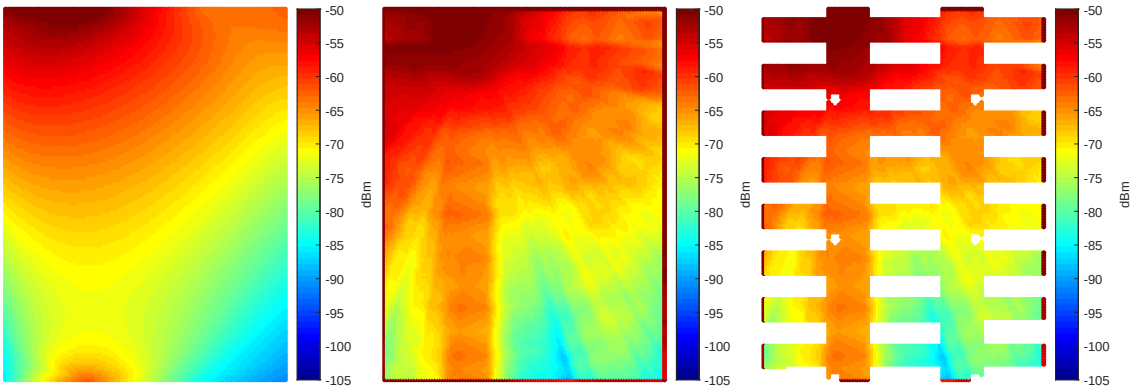
$$f_i = \begin{cases} \log_2(2 + 255 - Im(x_i, y_i)) & \text{for } 1 \leq i \leq k \\ 0 & \text{for k} < i \leq n \end{cases} \tag{4.4}$$

where x_i and y_i are the position components of the i^{th} point in B_{rp}, Im is the image representation of the environment, and $Im(x_i, y_i)$ is the cell value in the image Im whose position is (x_i, y_i). The value of n is the maximum number of points that may have a line connecting the positions of the AP and a point in the environment representation. If ap lies beyond the environment represented by Im, the image is enlarge applying a padding of zeros. In other words, $Im(x, y) = 0$ for all (x, y) that lies beyond the environment representation.

The set F_{rp} in Equation 4.3 is a representation of the obstacles between rp and ap. Thus, the model is trained to learn their influence in the signal propagation. This work did not differentiate among distinct types of obstacle materials for simplicity. However, Equation 4.4 allows for obstacle values representation from 1 to 255. Setting appropriate values that describe the opaqueness of each material to WiFi signals requires careful consideration and possible experimental measurements. Equation 4.3 includes half of the distance between rp and ap. Using the actual value of the distance significantly decreased the obstacles influences in the model. The number of variables presented in Equation 4.3 depends on the environment and the AP position.

The regression model is applied using a regression method. Support Vector and Gaussian Process regressions were tested. Support Vector regression, using a linear kernel function, provided the lowest regression residuals. The regression model was tested using for training only the reference points that defined the boundary of the collection area. Figure 4.13 showed the positions of those reference points using black circles. They represent less than 8% of all available reference points and allow the application of an interpolation method as a baseline method for comparisons. The remaining reference points were used as test points, i.e., to compute regression residuals. The experiments only included APs that had measurements for all reference points. Most of them were APs strongly detected in the environment.

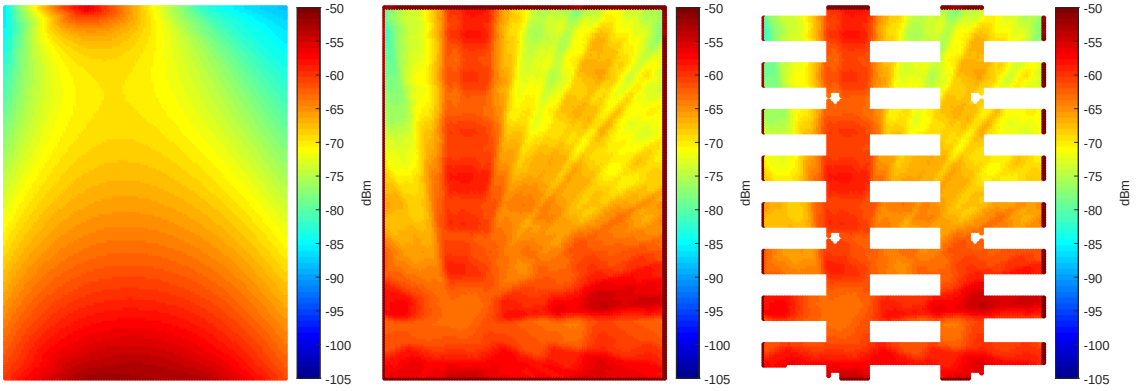
Figure 4.14 presents the regression estimates for the AP 15 provided by a method that combines Natural Neighbour interpolation and Gradient Extrapolation (Baseline), as provided by MATLAB [35], and by Support Vector Machine regression using the proposed regression model. The points used for training were those presented in Figure 4.13a. Given the small number of training points, the two regressions performed remarkably well for the considered AP and the environment. However, the proposed regression (Model) represents LOS situations better. Also, as depicted in Figure 4.14c, the proposed regression can capture the influence of bookshelves in the



(a) Baseline (b) Model (c) Model and obstacles

Figure 4.14: Regression results for AP ID 15.

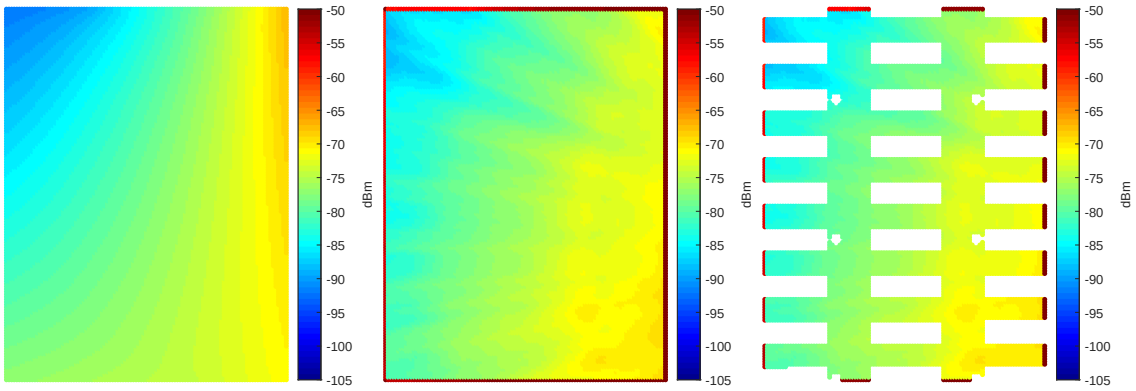
environment. The images presented in Figure 4.14b and Figure 4.14c were smoothed using 9 pixels square windows convolution. The device that broadcast the AP was inside the collection area.



(a) Baseline (b) Model (c) Model and obstacles

Figure 4.15: Regression results for AP ID 49.

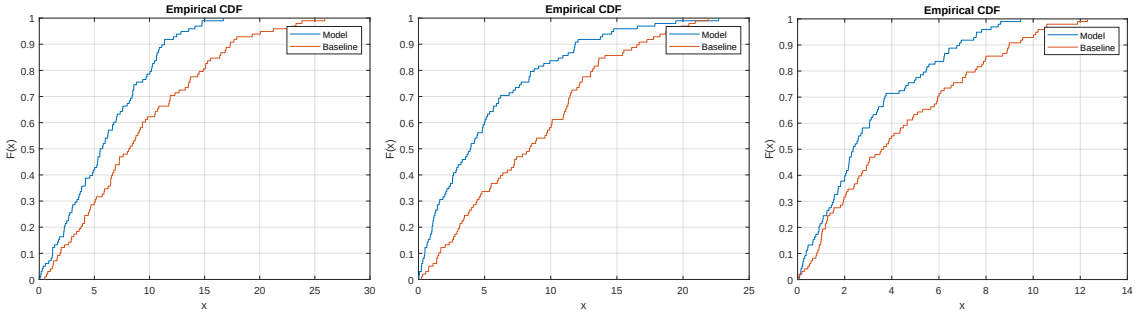
Figure 4.15 presents the regression estimates for the AP 49. The device of this AP was also inside the collection area. The points used for training were those presented in Figure 4.13b. The reference position of the AP was farther away from the training points than for AP 15. Thus, the model predicted for the reference AP position an RSS value weaker than the actual value. Also, the model predicted values too strong for the right top corner of the image. However, the improvement in estimates that correspond to LOS situations surpasses the accuracy losses near the AP and in the right top corner.



(a) Baseline (b) Model (c) Model and obstacles

Figure 4.16: Regression results for AP ID 8.

Figure 4.16 presents the regression estimates for the AP 8. The device of this AP was outside the collection zone and had a mean RSS of −81.5 dBm in Table 4.12. Thus, the reference AP position was placed a few meters away from the collection area, as presented in Figure 4.13c. The difference between Baseline and Model regression is not as significant as it was for the previous two analyzed APs. Position-based regression models like those used for Table 4.12 already provided low regression residuals for weakly detected APs. As explained before, the environment has less effect on the propagation of weak AP signals than on strong AP signals. The proposed model captures such behavior, and thus its estimates depend more on training point positions and the distance to the AP than on the obstacles' influence.



(a) AP ID 15 (b) AP ID 49 (c) AP ID 8

Figure 4.17: CDF plot of regression residuals from baseline and model. The red series represent the baseline method and blue series the proposed model.

Figure 4.17 presents the regression residuals obtained using the proposed model and the baseline method on the test reference points for the APs 15, 49 and 8. The

residuals obtained for the proposed model are consistently better than those from the baseline. For AP 15, the maximum residual value was about 10 dBm smaller in the proposed model than in the baseline. For AP 49, the maximum residual values were similar for the two approaches. However, the proposed model performed notably better than the baseline regarding percentiles between the 25^{th} and 75^{th}. For AP 8, the difference in residual values is less notable than for the previous two AP, which is in part a result of notably lower residual values.

Table 4.13: 75^{th} percentile of regression residuals.

AP ID	1	6	8	15	17	49	51	52	54	69
Model	6.7	6.8	4.7	8.8	8.0	7.8	9.5	8.4	8.5	11.5
Baseline	10.7	10.3	6.6	13.6	13.2	12.2	13.2	10.1	11.2	16.2

Table 4.13 present the 75^{th} percentile of regression residuals for the proposed model and the baseline method. The results are provided for some APs. The APs were those that had measurements available for all (106) reference points. The APs included all the APs strongly detected in the environment but the AP 71 (which had measurements for 105 points) and one weakly seen AP (AP 8). The three APs used for the previous analyses are included in this table. The proposed method performs better than the baseline for all selected APs.

The proposed regression model is recommended for regression on AP strongly seen in the target environment, though it could also be applied to weaker APs as long as their collected signal samples may indicate a likely direction of the position of the AP. Further research is still needed to test the proposed method in larger and more complex environments. Furthermore, its performance should be assessed in a large environment where the AP lies in the center of the collection area.

4.5 Conclusions

This chapter presented tools, guides, and methods to perform the collection of BLE or WiFi samples for fingerprinting. The chapter first addressed a methodology of performing collection organized in campaigns, presenting the tools that improve the collection and insights gained during their development and usage. The chapter then discussed the selection of positions for WiFi samples collection, presenting guides that were supported by experimental results. In addition, the chapter provided methods to estimate the position of a WiFi AP and to evaluate whether the AP is likely inside the area where a set of samples was collected. Finally, the chapter described a regression model that reduce residuals in the case of strongly detected AP. The

materials presented in this chapter represent valuable contributions that can be used in a collective or individual manner to reduce the collection efforts, which is one of the main challenges of fingerprinting.

Some materials presented in this chapter are supported by four publications [40, 39, 38, 36]. Other materials from this chapter were used as contributions in one publication [50].

References

[1] J Blumenthal et al. "Weighted Centroid Localization in Zigbee-based Sensor Networks". In: *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*. Oct. 2007, pp. 1–6. DOI: 10.1109/WISP.2007.4447528.

[2] L. Breiman et al. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984. ISBN: 9780412048418. URL: https://books.google.es/books?id=JwQx-WOmSyQC.

[3] Jack E Bresenham. "Algorithm for computer control of a digital plotter". In: *IBM Systems journal* 4.1 (1965), pp. 25–30.

[4] Yu-Chung Cheng et al. "Accuracy Characterization for Metropolitan-scale Wi-Fi Localization". In: *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*. MobiSys '05. New York, NY, USA: ACM, 2005, pp. 233–245. ISBN: 1-931971-31-5. DOI: 10.1145/1067170.1067195.

[5] Youngsu Cho et al. "Improved Wi-Fi AP position estimation using regression based approach". In: *Proc. of the International Conference on Indoor Positioning and Indoor Navigation*. 2012.

[6] Y Cho et al. "WiFi AP position estimation using contribution from heterogeneous mobile devices". In: *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*. 2012, pp. 562–567. DOI: 10.1109/PLANS.2012.6236928.

[7] Andrew Crooks. "Agent-based modeling and geographical information systems". In: *Geocomputation: a practical primer. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage* (2015), pp. 63–77.

[8] Herbert Edelsbrunner, David Kirkpatrick, and Raimund Seidel. "On the shape of a set of points in the plane". In: *IEEE Transactions on information theory* 29.4 (1983), pp. 551–559.

[9] Santiago Ezpeleta et al. "RF-Based Location Using Interpolation Functions to Reduce Fingerprint Mapping". en. In: *Sensors* 15.10 (Oct. 2015), pp. 27322–27340. ISSN: 1424-8220. DOI: 10.3390/s151027322. URL: http://www.mdpi.com/1424-8220/15/10/27322/htm.

[10] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232.

[11] A Goldsmith. "Path Loss and Shadowing". In: *Wireless Communications*. Cambridge University Press, 2005. Chap. 2, pp. 25–62. ISBN: 9780521837163. URL: https://books.google.es/books?id=n-3ZZ9i0s-cC.

[12] Rafał Górak and Marcin Luckner. "Modified Random Forest Algorithm for Wi–Fi Indoor Localization System". In: *Computational Collective Intelligence*. Ed. by Ngoc Thanh Nguyen et al. Cham: Springer International Publishing, 2016, pp. 147–157. ISBN: 978-3-319-45246-3.

[13] Dongsu Han et al. "Access Point Localization Using Local Signal Strength Gradient". In: *Passive and Active Network Measurement: 10th International Conference, PAM 2009, Seoul, Korea, April 1-3, 2009. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 99–108. ISBN: 978-3-642-00975-4. DOI: 10.1007/978-3-642-00975-4_10.

[14] Suining He and S.-H. Gary Chan. "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons". In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 466–490. DOI: 10.1109/comst.2015.2464084. URL: https://doi.org/10.1109/comst.2015.2464084.

[15] Noelia Hernández et al. "Continuous Space Estimation: Increasing WiFi-Based Indoor Localization Resolution without Increasing the Site-Survey Effort". In: *Sensors* 17.1 (2017). ISSN: 1424-8220. DOI: 10.3390/s17010147. URL: https://www.mdpi.com/1424-8220/17/1/147.

[16] M Ji et al. "A novel Wi-Fi AP localization method using Monte Carlo path-loss model fitting simulation". In: *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. Sept. 2013, pp. 3487–3491. DOI: 10.1109/PIMRC.2013.6666752.

[17] S. H. Jung, B. Moon, and D. Han. "Performance Evaluation of Radio Map Construction Methods for Wi-Fi Positioning Systems". In: *IEEE Transactions on Intelligent Transportation Systems* 18.4 (Apr. 2017), pp. 880–889. ISSN: 1524-9050. DOI: 10.1109/TITS.2016.2594479.

[18] Loizos Kanaris et al. "Sample Size Determination Algorithm for fingerprint-based indoor localization systems". In: *Computer Networks* 101 (2016), pp. 169–177. ISSN: 1389-1286. DOI: http://dx.doi.org/10.1016/j.comnet.2015.12.015.

[19] Ali Khalajmehrabadi, Nikolaos Gatsis, and David Akopian. "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges". In: *IEEE Communications Surveys & Tutorials* 19.3 (2017), pp. 1974–2002. DOI: 10.1109/comst.2017.2671454. URL: https://doi.org/10.1109/comst.2017.2671454.

[20] T. King, T. Haenselmann, and W. Effelsberg. "On-demand fingerprint selection for 802.11-based positioning systems". In: *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*. June 2008, pp. 1–8. DOI: 10.1109/WOWMOM.2008.4594839.

[21] Thomas King et al. *CRAWDAD dataset mannheim/compass (v. 2008-04-11)*. Downloaded from https://crawdad.org/mannheim/compass/20080411. Apr. 2008.

[22] Jahyoung Koo and Hojung Cha. "Autonomous construction of a WiFi access point map using multidimensional scaling". In: *Pervasive Computing*. Springer, 2011, pp. 115–132.

[23] Jahyoung Koo and Hojung Cha. "Localizing WiFi access points using signal strength". In: *IEEE Communications Letters* 15.2 (Feb. 2011), pp. 187–189. ISSN: 10897798. DOI: 10.1109/LCOMM.2011.121410.101379.

[24] Jahyoung Koo and Hojung Cha. "Unsupervised locating of WiFi access points using smartphones". In: *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 42.6 (2012), pp. 1341–1353.

[25] I Nizetic Kosovic and Tomislav Jagust. "Enhanced Weighted Centroid Localization Algorithm for Indoor Environments". In: *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering* 8.7 (2014), pp. 1184–1188.

[26] Elena Simona Lohan et al. "Received signal strength models for WLAN and BLE-based indoor positioning in multi-floor buildings". In: *Localization and GNSS (ICL-GNSS), 2015 International Conference on*. IEEE. 2015, pp. 1–6.

[27] C Maple. "Geometric design and space planning using the marching squares and marching cube algorithms". In: *Geometric Modeling and Graphics, 2003. Proceedings. 2003 International Conference on*. 2003, pp. 90–95. DOI: 10.1109/GMAG.2003.1219671.

[28] MathWorks. *MATLAB contourc function*. https://www.mathworks.com/help/matlab/ref/contourc.html. 2019.

[29] MathWorks. *MATLAB fitglm function*. https://ch.mathworks.com/help/stats/fitglm.html. 2019.

[30] MathWorks. *MATLAB fitrensemble function*. https://ch.mathworks.com/help/stats/fitrensemble.html. 2019.

[31] MathWorks. *MATLAB fitrgp function*. https://ch.mathworks.com/help/stats/fitrgp.html. 2019.

[32] MathWorks. *MATLAB fitrsvm function*. https://ch.mathworks.com/help/stats/fitrsvm.html. 2019.

[33] MathWorks. *MATLAB fitrtree function*. https://ch.mathworks.com/help/stats/fitrtree.html. 2019.

[34] MathWorks. *MATLAB griddata function*. https://ch.mathworks.com/help/matlab/ref/griddata.html. 2019.

[35] MathWorks. *MATLAB scatteredinterpolant function*. https://www.mathworks.com/help/matlab/ref/scatteredinterpolant.html. 2019.

[36] Germán Martín Mendoza-Silva et al. "BLE RSS Measurements Dataset for Research on Accurate Indoor Positioning". In: *Data* 4.1 (2019), p. 12.

[37] Germán Martín Mendoza-Silva et al. *Long-Term Wi-Fi fingerprinting dataset and supporting material*. Nov. 2017. DOI: 10.5281/zenodo.1066041. URL: https://doi.org/10.5281/zenodo.1066041.

[38] Germán Martín Mendoza-Silva et al. "Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning". In: *Data* 3.1 (2018). ISSN: 2306-5729. DOI: 10.3390/data3010003.

[39] Germán M Mendoza-Silva et al. "Situation Goodness Method for Weighted Centroid-Based Wi-Fi APs Localization". In: *Progress in Location-Based Services 2016*. Springer, 2017, pp. 27–47.

[40] Germán M Mendoza-Silva et al. "Solutions for signal mapping campaigns of Wi-Fi networks". In: *Proceedings of the VII Jornadas Ibéricas de Infraestructuras de Datos Espaciales*. 2016.

[41] Nicholas Metropolis et al. "The beginning of the Monte Carlo method". In: *Los Alamos Science* 15.584 (1987), pp. 125–130.

[42] Patrick AP Moran. "Notes on continuous stochastic phenomena". In: *Biometrika* 37.1/2 (1950), pp. 17–23.

[43] Seung Yeob Nam. "Localization of Access Points Based on Signal Strength Measured by a Mobile User Node". In: *Communications Letters, IEEE* 18.8 (2014), pp. 1407–1410.

[44] John Ashworth Nelder and Robert WM Wedderburn. "Generalized linear models". In: *Journal of the Royal Statistical Society: Series A (General)* 135.3 (1972), pp. 370–384.

[45] Oded Nov, Ofer Arazy, and David Anderson. "Scientists@ Home: what drives the quantity and quality of online citizen science participation?" In: *PloS one* 9.4 (2014), e90375.

[46] Andreas Savvides, Chih-Chieh Han, and Mani B Strivastava. "Dynamic fine-grained localization in ad-hoc networks of sensors". In: *Proceedings of the 7th annual international conference on Mobile computing and networking*. ACM. 2001, pp. 166–179.

[47] Donald Shepard. "A two-dimensional interpolation function for irregularly-spaced data". In: *Proceedings of the 1968 23rd ACM national conference*. ACM. 1968, pp. 517–524.

[48] Robin Sibson. "A brief description of natural neighbour interpolation". In: *Interpreting multivariate data* (1981).

[49] J Talvitie, M Renfors, and E S Lohan. "Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization With Indoor Wireless Signals". In: *IEEE Transactions on Vehicular Technology* 64.4 (2015), pp. 1340–1353. ISSN: 0018-9545. DOI: 10.1109/TVT.2015.2397598.

[50] Joaquín Torres-Sospedra et al. "Lessons Learned in Generating Ground Truth for Indoor Positioning Systems Based on Wi-Fi Fingerprinting". In: *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*. Elsevier, 2019, pp. 45–67.

[51] Joaquín Torres-Sospedra et al. "Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning". In: *Mobile Information Systems* 2016 (2016).

[52] Joaquín Torres-Sospedra et al. "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems". In: *2014 international conference on indoor positioning and indoor navigation (IPIN)*. IEEE. 2014, pp. 261–270.

[53] J. Torres-Sospedra et al. "Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates". In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sept. 2018, pp. 1–8. DOI: 10.1109/IPIN.2018.8533791.

[54] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995. ISBN: 0-387-94559-8.

[55] Saeed Varzandian, Hasan Zakeri, and Sadjaad Ozgoli. "Locating WiFi access points in indoor environments using non-monotonic signal propagation model". In: *Control Conference (ASCC), 2013 9th Asian*. IEEE. 2013, pp. 1–5.

[56] Debbie Watson. *Contouring: a guide to the analysis and display of spatial data*. Vol. 10. Elsevier, 1992.

[57] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. MIT Press Cambridge, MA, 2006.

[58] TY Yang. *Finite element structural analysis*. Vol. 2. Prentice-Hall Englewood Cliffs, NJ, 1986.

[59] Simon Yiu et al. "Wireless RSSI fingerprinting localization". In: *Signal Processing* 131 (2017), pp. 235–244.

[60] Francisco Zampella, Antonio Ramon Jimenez Ruiz, and Fernando Seco Granja. "Indoor positioning using efficient map matching, RSS measurements, and an improved motion model". In: *IEEE Transactions on Vehicular Technology* 64.4 (2015), pp. 1304–1317.

[61] Zengbin Zhang et al. "I am the antenna: accurate outdoor ap location using smartphones". In: *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM. 2011, pp. 109–120.

[62] Fang Zhao et al. "An RSSI gradient-based AP localization algorithm". In: *China Communications* 11.2 (Feb. 2014), pp. 100–108. ISSN: 16735447. DOI: 10.1109/CC.2014.6821742.

Chapter 5

WiFi Radio Map Maintenance

The solutions for updating the radio map for WiFi fingerprinting are sometimes treated as part of the solutions for radio map construction. The reason for placing both steps into one set of solutions is that for some IPS the radio map continuously evolves, getting either enriched or updated as new measurements are available. Usually, such continuous updates are performed using crowdsourcing. However, for other IPS, the update or maintenance of the radio map is a well-differentiated step. For them, periodic radio map updates are usually required, ideally only where they are needed. This chapter is devoted to finding the set of positions most convenient to perform periodic radio map updates under nominal situations. The chapter also briefly addresses when changes in the environment are to be considered nominal or not.

5.1 Periodic Radio Map Update

The crowdsourcing approach has been praised for radio map collection and update. However, some WiFi fingerprinting IPS prefer collections or updates that assure the quality of position tags and an even distribution of the samples collected across the target environment. There is still a solution for reducing the collection or update efforts for these systems: to collect measurements only at some positions and then use regression to estimate measurements for the remaining positions. The regression methods have been used for reducing the calibration effort in fingerprinting for more than 15 years [12, 14]. The proposals found in IPS literature include Linear interpolation [26]; Radial Basis interpolation [12, 5]; Gaussian Process regression [29]; and Support Vector Machine regression (SVR) [8]. Some studies particularly focused on the spatial relations of measurements and the spatial characteristics of the environment for regression. They included methods widely used in spatial analysis like Inverse Distance Weighting (IDW) and Kriging [14, 16, 9]; Voronoi Tessellation [13];

Sparsity Rank Singular Value Decomposition (SRSVD) [7]; and particular heuristics [3].

The studies that have proposed regression approaches generally show that the estimates made by their proposals can be used instead of a significant percentage of the actual measurements. To test the benefit of their approaches, they specify elimination procedures to drop some of the original positions. However, those elimination procedures are not intended as suggested strategies for determining collection positions. Using random positions for collection is a common approach [1], despite the distribution of the positions is very important [15] for radio map construction. The importance of the collection positions is intuitive and acknowledged in other topics for other phenomena. Specifically, several works address the optimal (or quasi-optimal) placement of sensors that best measure a given phenomenon [24, 10]. A set of WiFi measurements collected at known positions by a subject can also be understood as a set of sensor measurements. Therefore, choosing the best positions where a subject should collect the WiFi measurements can be considered as a placement optimization for a set of sensors.

This section merges the notions of sensor placement and regression applied to WiFi samples in a method that helps in deciding the positioning where new samples for a radio map should be collected. The section addresses sensor placement as an optimization problem.

5.1.1 Positions Set Determination for Radio Map Update

The experiments that demonstrate regression proposals commonly transform a relatively dense dataset of RSS measurements into sparser new datasets through elimination strategies. The regression proposals are then applied to the new datasets to obtain RSS estimates for the positions of the samples that were removed. The goodness of a regression is usually evaluated considering two measures. The first measure is the regression residuals. Residuals are the absolute AP-wise difference in RSS values between the removed measurements and their corresponding estimates. The second measure is the difference in the accuracy metric between when an IPS is trained using the original dataset and when it is trained using a dataset with a high percentage of removed samples. The elimination strategy is an important factor in the results obtained in regression evaluations [26]. The regression performance reported in the literature varies significantly, from discrete but reasonably results of 50% samples reduction [5] to impressive results of 5% samples reduction [7] with very small regression residuals or very little affectation to positioning accuracy.

Most of the regression proposals found in IPS literature indicate how much they can reduce the number of required collection positions. However, they do not com-

monly provide a methodology for determining the number of distinct collection positions for a given environment. Furthermore, they do not state how to determine where those positions should be. An intuitive approach is to choose the number of collection points as a function of the target area size and randomly determine their positions in that area. Some studies have used similar approaches. In Kanaris et al. [11], the authors determined the sample size given a small preliminary set of measurements. They suggested to randomly choose positions from a grid in the number determined by the sample size calculation. Specifically, for the case of a radio map update, collecting measurements at random positions in a target area is a common approach [1]. Indeed, depending on the update frequency, the crowdsource approaches can be also considered a strategy of collecting update measurements at random positions.

Regarding the elimination strategies, the work of Krumm and Platt [12] proposed an elimination strategy consisting of running a k-means clustering algorithm, and selecting only the k positions nearest the k cluster centroids. Other studies have resembled in their proposed elimination strategies the types of positions absences that may happen in regular collection processes, like random isolated absent points, zones with a higher or lower percentage of elimination than others [5] or random blocks of absent points [26]. These approaches, however, have some drawbacks that are experimentally shown later in this section. It is almost intuitive that neither the number of reference points selected for radio map collection nor their actual distribution should be chosen randomly without any restriction. Besides, a distribution of uniformly spaced positions does not account for the environment and its influence on the WiFi signals. Deciding the number and distribution of the collection positions for WiFi fingerprinting is not a trivial task.

The problem of sensor placement, related to sensor selection or activation, is a well-known problem for wireless sensor networks. The sensor selection problem can be stated as choosing a set of k measurements from a set of m possible sensor measurements, which minimizes the error in the estimation of some parameters [10]. The approaches to deal with the sensor placement/selection problem vary depending on the usage of the sensor measurements [24]. Specifically, some studies have proposed approaches for the case of using the sensor measurements for estimating a field of values [10, 23, 25]. It is unfeasible to explore the whole solution space of possible positions given the combinatorial nature of the problem [10]. If there is a set of 24 possible positions in the target environment, and measurements are to be taken at 12 of those positions, the number of different possible sets of positions is $\binom{24}{12} = 2,704,156$. If the number of measurement positions is not already decided, the number of possible combinations rises to $2^{24} = 16,777,216$.

This work suggests that the solutions for the sensor placement problem can also be applied to finding the set of k positions from m possible ones, where the WiFi measurements will be collected. The k positions are the best ones to use for training

a regression model that will be later used for estimating the signal intensities for the remaining positions. What is more, this work does not consider a fixed number of positions and instead obtains a compromise between the number of positions and the goodness of the regression. Instead of harnessing some property of the target problem or forcing some form for the solution for the sensor placement problem [10, 23, 25], this work uses an optimization strategy based on genetic algorithms. Sensor placement optimization has already been addressed using genetic algorithms [28, 17], even for indoor acoustic positioning [17].

Genetic Programming for Position Set Determination

The approach proposed in this study uses a genetic optimization algorithm to find a set of positions that includes only a small number of positions without compromising the regression quality. The explanation presented here for genetic algorithms, as well as the implementation used in the experiments, are based on Mitchell [22].

The genetic algorithms try to efficiently find solutions to problems that have huge spaces of candidate solutions. Each candidate solution for a problem is called an individual. Commonly, an individual is encoded as a bit string, where each bit represents the presence ("1") or absence ("0") of a trait. These algorithms start with a population of random individuals, and iteratively evolve it. The population of each iteration is called a generation. Each generation is the result of applying genetic operators on the previous generation. The selection operator selects pairs of individuals whose traits are combined using a crossover operator to produce offspring. A fitness value is computed for every individual in a generation. Individuals with higher fitness values are more likely to be chosen by the selection operator. A mutation operator is applied to the offspring to produce subtle changes in the resulting traits. Some of the new individuals can be randomly discarded, but the population size is maintained.

In this work's proposal, the set of all positions $L = \{l_1, ..., l_n\}$ from an initial, dense collection represents the possible traits that each individual may have. The set of positions L has associated WiFi RSS measurements $D = \{r_{ijk}\}$. Assume a function $fmap(A, B) \rightarrow C$ so that A is a set of RSS values, B is a set of positions and C is the set of RSS values in A associated to positions in B. Then, $D_{l_p} = fmap(D, \{l_p\}) = \{r_{pjk}\}$ are the RSS measurements associated to the position l_p. An individual represents a subset L_I of L. The size of the population, as well as the number of generations considered for population evolution, are parameters of the algorithm that are presented in the description of the experiments. The fitness value calculation of an individual was designed so to penalize large subsets and large differences between actual and estimated RSS values. Specifically, the fitness computation steps are:

1. Fit regressions f_a, for every detected AP a, using L_I and their associated mea-

surements $fmap(D, L_I)$.

2. Use regressions f_a to estimate RSS values $E = \{\bar{r}_{ia}\}$ for positions of $\hat{L_I} = L - L_I$.

3. Compute the AP-wise and position-wise RSS absolute differences between E and $fmap(D, \hat{L_I})$. Let MRD be the maximum value of those differences.

4. The individual's fitness is $(MRD + 2MRD\frac{ab}{tb})^{-1}$, where ab and tb are the number of "1" bits and the total number of bits, respectively. If for some reason the target number of positions is already predefined, say k, the individual's fitness can become $(MRD + 2MRD|ab - k|)^{-1}$.

After a given number of generations, the individual with the highest fitness value, called the elite individual, could be chosen as the set of positions where WiFi samples should be collected. The elite individual represents the set that has so far achieved the best compromise between a small number of positions and little degradation of the regression goodness. The genetic algorithm does not guarantee that the elite individual would be the optimal solution for a given problem but is a fair alternative to an exhaustive search given the combinatorial nature of the problem.

This work's main goal is determining the suitable positions for conducting periodic WiFi radio map updates. The elite individual may represent a solution that is overfitted for the initial measurements. Therefore, in this work, the selected positions are those whose occurrence frequency is above a threshold in the final population. The resulting set of position is called here the *solution set*. Figure 5.1 shows an example of the position's frequency for a population of (200) evaluated sets of positions after 200 iterations. The number of traits, i.e., the number of positions in the initial, dense collection is 24. The bit frequency represents how often a position is found in sets of positions. If a high frequency threshold of 0.9 is chosen, the solution set would be $\{1, 2, 3, 18, 19, 21, 22, 23\}$.

In summary, the steps needed for selecting the positions where the periodic update measurements are to be collected are:

1. Collect a relatively dense WiFi RSS training database.

2. Use a genetic algorithm, such as the one described in pages eight and nine of Mitchell [22], using the fitness function described above in this section to determine the positions' frequency in the population of sets of positions.

3. Choose as the solution set the positions with frequencies above a certain threshold.

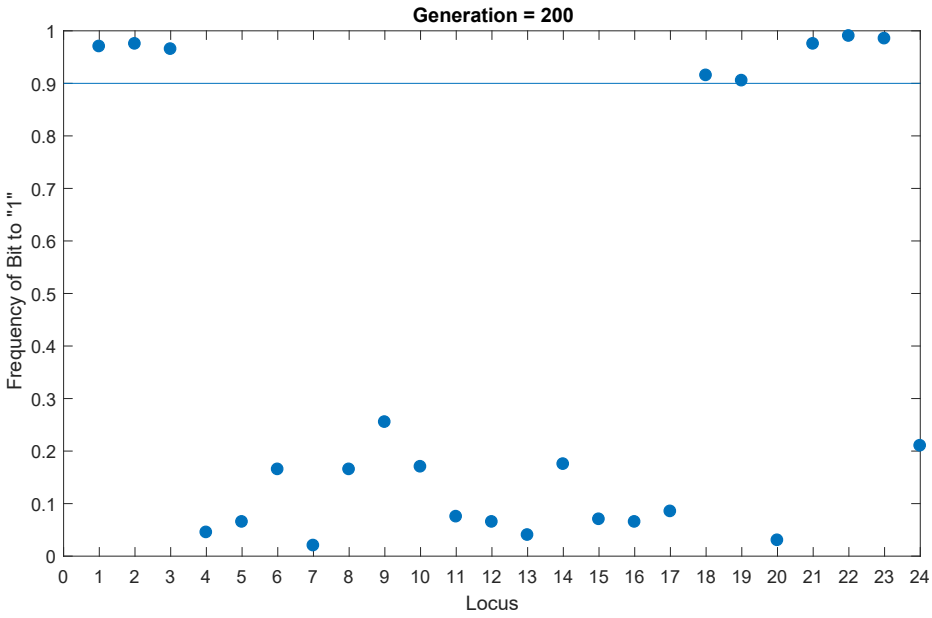


Figure 5.1: Positions (bits) frequency. The blue dots represent how often a position has been included in individuals of generation 200. The blue line represents the frequency threshold.

Besides suggesting a good placement for the measurement positions, the proposed approach can be also used for testing the performance of regression methods. By computing an individual's fitness as $MRD + 2MRD\frac{ab}{tb}$, the genetic algorithm would determine a compromise between a large number of positions and a high RSS absolute difference. The set resulting from choosing the n highest fitness sets of positions can be used as a challenging test for evaluating the performance of regression models.

5.1.2 Experiments

The proposed approach was tested on a subset of the Library dataset [21]. The Library dataset that collected in two floors of the UJI's Library building during 25 cumulative months. Six fingerprints were collected per each position and each of the two directions that the collection subject faced. The subset contains only one training set and the test sets that correspond to the 3^{rd} floor of the first ten collection months, which were the data available by the time of realization of the work presented in this section. The collection area is a relatively small environment that covered about 15×10 m. Chapter 3 provided more details about the Library dataset.

Most of the experiments presented in this section used only the training sets (24

positions). The test sets were only used for the evaluation using the kNN fingerprinting presented at the end of this section. For data dimensionality reduction, the APs detected in less than 5% of the fingerprints were removed.

The set determination method was defined without establishing any explicit restriction for the regression to use. However, the regression method should enable both interpolation and extrapolation, because there may be target measurement positions lying outside the convex hull of the positions in the solution set. This implicit restriction is also important because the extrapolation usage is mandatory for environments that, at collection time, contain areas where measurements cannot be taken – e.g., because of a meeting in an office.

Providing recommendations on regression methods for WiFi fingerprinting was not among the goals of this section (for recommendations see Chapter 4). Thus, only a few regression methods were tested: Inverse Distance Weighting (interpolation and extrapolation); Radial Basis function interpolations like those in Ezpeleta et. al. [5], combinations of Linear, Nearest Neighbors, and Natural Neighbors interpolation with Linear and Gradient extrapolation as provided by MATLAB [18]; and Support Vector Machine regression (SVR) as provided MATLAB [19]. The chosen regression method was SVR, using a Radial Basis Function (RBF) kernel and performing predictor data standardization. SVR provided the best results regarding regression residuals and because it has been successfully used in previous studies [8]. A regression method evaluation for a given environment is suggested before making a choice.

For evaluating the goodness of each set of positions regarding regression estimates, a metric that quantifies the RSS estimation error was defined. The metric is the maximum value of the regression residuals. The maximum value was preferred over other measures – e.g., the mean – that may mask high RSS differences that are significant for distance-based techniques like kNN-based fingerprinting.

Evaluation for the Initial Month

The experiments considered three strategies for the selection of collection positions. The strategies were applied to the training set corresponding to the first month of the WiFi RSS dataset. The strategies are:

1. Random Sets of Points,
2. Manually-defined Sets of Points,
3. Optimized Set of Points.

The first approach considers differently sized sets of random positions. The second approach uses sets manually defined by an expert. The third approach finds a set

of positions that establishes a compromise between the set's size and the regression goodness. The following subsections provide more details about each approach and its evaluation.

Selecting a random set of positions is an intuitive approach. The algorithm proposed in Kanaris et. al. [11] may allow determining the number of measurement positions. This work instead explored several numbers of positions, ranging from 6 to 18 points, which accounts for 25% to 75% of the 24 total positions in the target area, and they represent reasonable effort reductions. Table 5.1 presents the maximum and minimum of the RSS error metric previously defined. The experiment for each number of points was repeated 200 times.

Table 5.1: Minimum and maximum values of RSS error metric for sets of randomly chosen positions.

Set Size	Metric Min (dBm)	Metric Max (dBm)
6	19.13	44.74
8	15.71	43.01
10	12.12	41.35
12	11.18	42.10
14	11.99	35.58
16	10.52	35.22
18	9.56	28.59

Table 5.1 shows two main facts. First, the more points are used for fitting the regression, the better the estimates are. Second, and more importantly, the RSS estimation quality heavily depends on the distribution of the randomly chosen positions, as absolute differences between the maximum and minimum metric values are up to 30.92 dBm.

Selecting random points creates much uncertainty in the quality of the RSS estimations. An alternative is to manually define the set of positions. The distribution of positions should take into account the extent of the collection positions and the influence of the obstacles. In this work, six alternative sets were considered likely manual choices that could provide fine RSS estimations through regression. The process of determining the tentative sets of positions is time-consuming, and it is especially cumbersome due to the large number of alternatives for each set's size. Table 5.2 presents the value of the RSS error metric for each alternative set. The ID of each set indicates its amount of positions. Figure 5.2 shows the position distribution of each set.

The results presented in Table 5.2 reinforce the importance of the distribution of the collection positions. The estimation quality does not strictly improve as the

Table 5.2: Values for RSS error metric for manually defined sets of positions.

Set ID	Metric (dBm)
6A	21.39
8A	23.45
8B	19.63
12A	23.13
12B	11.90
14A	12.20

number of positions used for regression fit rises, as seen when comparing the set 6A with set 8A, set 8B with set 12A, and set 12B with set 14A. In each of those pairs of sets, the value of the RSS error metric is larger for the set with more points. The distribution of positions of each set, as shown in Figure 5.2, sheds some light on the previous fact. The convex hulls of sets 6A and 8B include more of the target area than those of sets 8A and 12A, respectively. Nevertheless, the convex hulls of sets 12B and 14A are equal, and the set 12B provide better estimations than set 14A despite having a smaller number of positions. Thus, even a well-designed set of positions may not be the best choice. Additionally, a set of positions that is optimal for a given environment may not be optimal for another one, a fact that that is left unproven because is beyond the focus of this section.

It is possible to search for fine positions for fitting the regression with the optimization strategy based on a genetic algorithm. Specifically, the genetic algorithm implementation provided in Burjorjee [4], which is in turn based on Mitchell [22] was used for the experiments. The population size was composed of 200 individuals. The genetic algorithm used the fitness function previously proposed and it was run for 100 generations. After testing several values, the numbers of 200 and 100 for population size and algorithm generations were the ones that provided higher stability (reproducibility) in the outputted solution. The obtained elite individual and the solution set using a higher frequency threshold of 0.9, are depicted in Figure 5.3. The value of the RSS error metric for the elite set (11 positions) was 8.85 dBm, which is lower than any of the values obtained using the previous two strategies. The metric value for the solution set (eight positions) is 11.18 dBm, which is still lower than most of the values obtained using the previous two strategies.

Figure 5.3 shows two important facts. First, all target positions are contained in the convex hull of both optimized sets, which avoids the usage of extrapolation. Second, and more importantly, the distribution of positions in the optimized sets do not resemble those of the manual choice strategies, nor they are intuitive. Therefore, the positions chosen for fitting a regression should not be random, and determining a

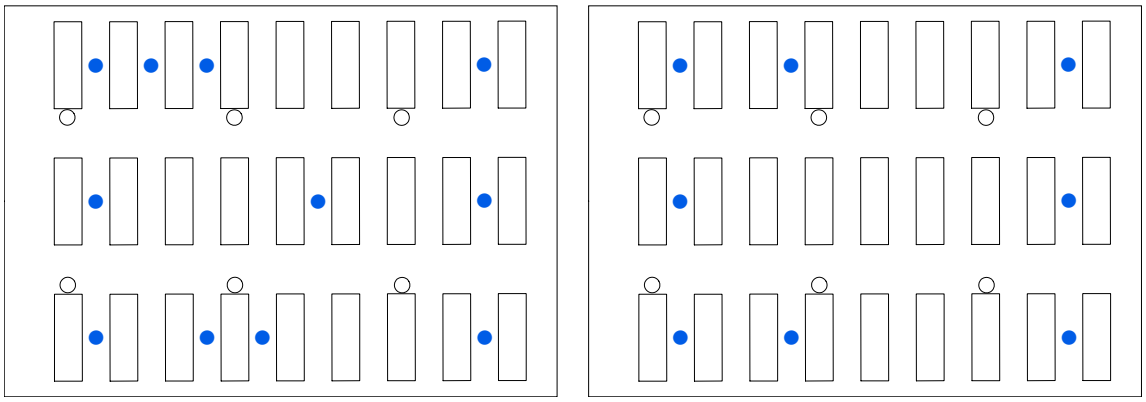
(a) 6A (b) 8A

(c) 8B (d) 12A

(e) 12B (f) 14A

Figure 5.2: Manually chosen sets of positions.

small set of positions that provides good estimations when used for fitting a regression is not a trivial problem.



(a) Positions represented by the elite individual

(b) Solution set (positions with higher frequency)

Figure 5.3: Set of positions obtained through optimization using a genetic algorithm.

Usage of the Solution Set for RSS Radio Map Update

The following experiments explored the goodness of the proposed position selection approach for WiFi radio map updates along 9 months (month 2 to 10) in the Library dataset. The goodness is evaluated in terms of regression residuals and the accuracy of a fingerprinting-based IPS.

Regarding regression residuals, the experiments tested three sets of positions and three RSS difference metrics. Table 5.3 shows the results of these experiments. Each table header indicates the usage of a particular set of positions and a specific RSS difference metric.

To explore the suitability of a set of positions for each month, a regression was fit using their associated measurements of the training set from the first month. Besides the solution set (GA), the sets 8A and 8B were also used for regression fitting. The sets 8A and 8B, previously introduced, are now used for baseline comparisons. As RSS difference metrics, the experiments used:

1. MRD: The MRD value previously introduced for the fitness function definition.
2. Mean: Its value is computed in a way similar to MRD, but the mean value is used instead of the maximum. This metric is included because it is frequently used in the literature for evaluation of WiFi RSS regression methods.
3. MeanP: It is calculated as: Compute the AP-wise regression residuals. Compute per each position the mean of those residuals. Take the maximum of those mean values. This metric is included because indicates how much the RSS difference may affect a RSS distance-based method like kNN.

Table 5.3: Values for RSS differences (dBm) according to metrics MRD, Mean and MeanP for sets 8A, 8B and GA.

		Month								
Metric	Set	2	3	4	5	6	7	8	9	10
MRD	8A	21.1	19.9	20.5	19.7	23.9	21.5	26.5	22.9	23.7
MRD	8B	21.2	20.8	21.0	23.8	21.8	24.9	30.0	25.4	22.0
MRD	GA	19.4	16.9	18.7	17.2	21.9	26.6	18.9	20.2	18.0
Mean	8A	1.8	1.6	1.7	1.5	1.6	1.5	1.6	1.5	1.8
Mean	8B	1.5	1.4	1.5	1.3	1.4	1.4	1.4	1.3	1.6
Mean	GA	1.7	1.5	1.5	1.3	1.5	1.5	1.4	1.2	1.6
MeanP	8A	5.2	5.3	4.6	4.9	4.7	4.8	5.0	4.5	5.3
MeanP	8B	3.7	3.7	3.8	4.6	3.6	4.2	4.3	3.4	4.3
MeanP	GA	3.8	3.0	3.0	2.9	3.7	3.3	3.5	3.3	4.1

The results presented in Table 5.3 indicate that the solution set obtained using proposed method is a better choice than the other two sets as a set of collection positions for periodical updates. Regarding the MRD metric, the solution set provides the best result for most months. It is noticeable that for month number seven, the value for the solution set is 5.1 dBm worse than the one for the set 8A. Some insights on that behavior will be later provided when analyzing the set effects on fingerprinting-based IPS accuracy. Regarding the Mean metric, the solution set is consistently better than the set 8A and slightly worse than the set 8B for some months. As for the MeanP metric, the solution set is much better than the 8A set. In comparison with the set 8B, the solution set is notably better for five of the months, and only slightly worse for two of them.

The experiments also explored the impact of the proposed update approach on the positioning accuracy of an IPS. A kNN fingerprinting approach was used for the experiments. Given a training set of fingerprints with known position labels, a query fingerprint, and two parameters specified by the value of k and a similarity metric on the fingerprint space, the kNN method finds the k fingerprints in the training set that are the most similar to the query fingerprint. The position label is estimated as the centroid of the position labels of the selected k closest fingerprints.

To measure the accuracy of an IPS, a test set of query fingerprints is usually used. The position labels are also known for the test set fingerprints, so that, for each fingerprint, the position estimation provided by the IPS and its original position label are used to compute a positioning error distance. In this section, the positioning distance has been calculated using the Euclidean distance and the positioning accuracy has been explored using the 75^{th} percentile of the computed distances for the test set.

The tested kNN used the RSS Euclidean distance as the fingerprint similarity

metric. The k parameter value was experimentally determined using the training and test sets of the first month of the WiFi Library dataset. Figure 5.4 shows the resulting positioning accuracies. The value of k that provides the best metric value is nine, and it is the one used for kNN in the remaining experiments. This value may appear large, but it is a reasonable value given that 12 fingerprints were taken at each position and no aggregation operation was performed for fingerprints with the same position label.

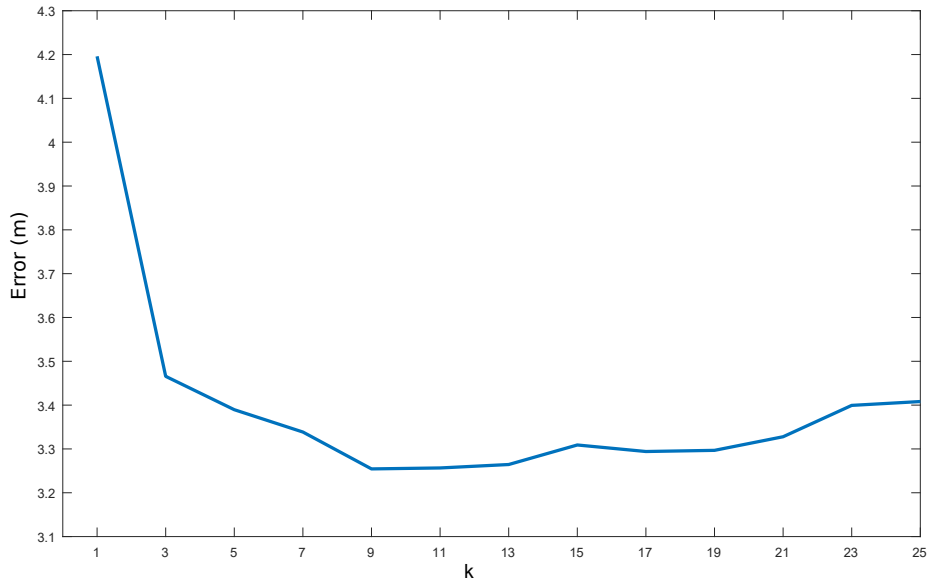


Figure 5.4: 75 percentile of positioning error using kNN for the first month.

For comparisons, the experiments included an evaluation of the radio map update at each month using all the training measurements collected at that month. Two updates strategies were tested: Replacement and addition. With the replacement strategy, all training fingerprints collected at one month replaced all fingerprints from the previous month in the WiFi radio map. With the addition strategy, the fingerprints of each month were added without any replacement or deletion from the previous months' fingerprints. The kNN method was used to estimate, for each month, the positions associated with the fingerprints of the test set of that month. Figure 5.5 shows the behavior of each update strategy along the time.

The strategy of addition provides values for the positioning error metric that are smaller and smoother than those provided by the replacement strategy. The metric values for the strategy of addition range from 2.84 m to 3.25 m. For the replacement strategy, however, the positioning error metric range from 4.10 m to 3.14 m. The

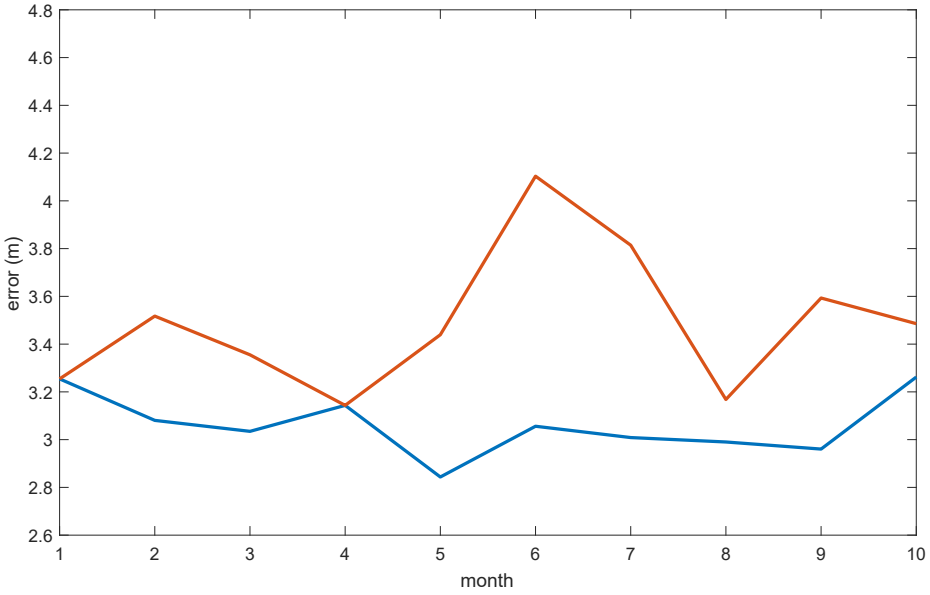


Figure 5.5: Comparison of the strategies of replacement (red) and addition (blue). Measurements for all positions are available.

months six and seven have the highest metric values, which may indicate that the training values for those months were not as good (representative) as they were for other months.

The evaluation of using the solution set for a radio map update was conducted as follows. For each month, the solution set was used to fit a regression, and the RSS values were estimated for the rest of the positions – i.e., the rest of the reference points. The estimation provides one fingerprint per position. However, the training and test sets in the database contain 12 fingerprints per position. Furthermore, the k value determined above for the kNN fingerprinting is the best under the assumption that there are 12 fingerprints per point. Therefore, 12 fingerprints per position were created using the one fingerprint per position obtained through regression estimation and adding a random value. In the training set from the first month, the AP-wise standard deviations values were less than 6 dBm in 80% of cases. The added random value is then uniformly chosen in the interval [-6;6]. The random value addition is specific to the evaluation presented in this study and will not be needed for an IPS radio map update, for which it may be desirable to collect only one fingerprint per point. The fingerprints newly estimated for each month were considered for radio map update following the strategies of replacement and addition described above. Figure 5.6 presents the positioning accuracy metric values for both strategies.

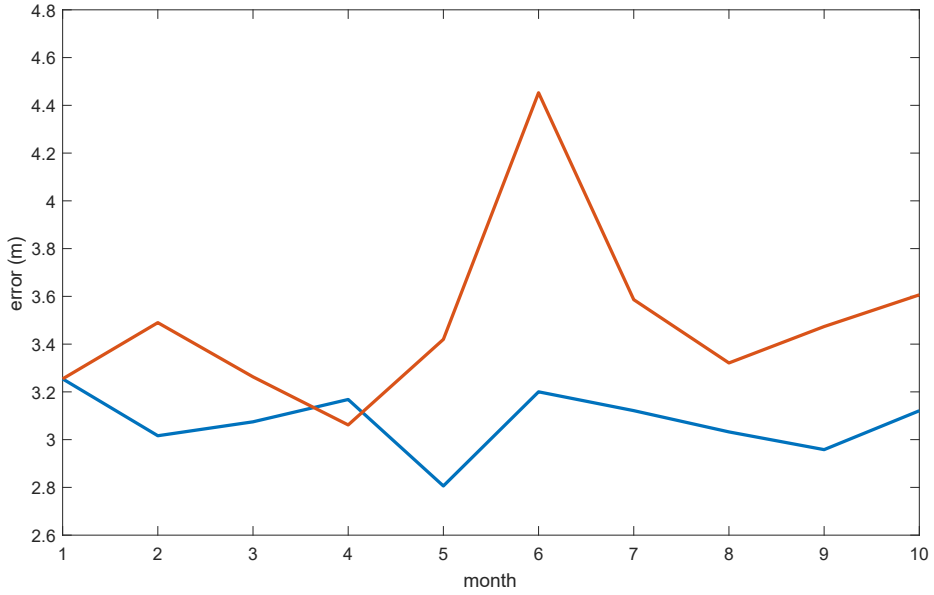


Figure 5.6: Comparison of the strategies of replacement (red) and addition (blue). Measurements are available only for the solution set.

The results obtained using the strategy of addition and the RSS estimations from the solution set are very similar to those using that strategy and the measurements available for all positions. The positioning error metric for the strategy of addition ranges from 3.25 m to 2.81 m, which is the same interval obtained when using the RSS measurements for all positions. The strategy of replacement showed larger metric values with higher variations than the addition strategy. When compared to using the same strategy and the measurements from all positions, the usage of the estimations from the solution set caused larger variability, with the metric values ranging from 4.45 m to 3.06 m, having a steeper variation for month six.

The above results suggest that the usage of the solution set as collection positions for WiFi radio map update is a reasonable choice for the tested environment. The approach of determining the solution set is automatic. Thus, the specialized and cumbersome task of manually determining a proper set of collections positions is avoided. The MeanP value, i.e, the maximum of the position-wise mean RSS differences between estimations and measurements, was lower than the detected AP-wise standard deviation for all months. Additionally, the accuracy of the tested kNN using measurements from all position and estimations had a behavior similar to the accuracy achieved when the estimates from a regression based on the solution set were used.

The previous results have been obtained for an initial radio map collection month

and across nine months of radio map updates. During the ten months, no drastic changes in the presence or power configuration of the APs were observed, apart from wireless networks with very low presence in the data. Therefore, the usage of this section's proposal is advised for environments that allow an initial relatively dense collection and as long as no drastic (non-nominal) change happens to the detected APs. The required initial collection is costly, but if the cost is assumed, the knowledge of the positions for performing periodic updates could translate into a better IPS performance. The following sections briefly addresses the nominality of changes.

5.2 Nominality of Environment Changes

A nominal situation for an IPS is one that corresponds with the initial setup of the IPS. Changes in the environment are to be considered nominal situations if they do not significantly decrease the positioning accuracy below the one that is expected for an IPS. A fingerprinting-based IPS is expected to keep its normal functioning if newly detected signal intensities are still similar to the signal characterization of the environment portrayed by the IPS' radio map. The boundaries for such similarity in WiFi fingerprinting, to the best of this work's knowledge, has not been established yet. It is known that WiFi signals are affected by fast fading [6]. It is common to model WiFi signals with a Gaussian variation of less than 5 dBm [2]. However, the signals may be subject to more drastic variations in real environments.

The experiments presented in this section used subsets of the Library dataset [21]. The Library dataset was also used in Section 5.1 and Chapter 3 provided more details about it. The subset used in the first experiment considered the data from Month 15 corresponding to samples from the 3^{rd} floor of the Test01 and Test05 sets. This subset contained 24 positions, 4 collection directions per position, and 6 samples per each position and collection direction. For the experiment, the first sample was dropped. As shown in Chapter 4, the first sample tends to an RSS value that is the most different of the six samples consecutively collected for a position and direction.

Figure 5.7 shows the counts for differences between maximum and minimum RSS values of the 20 samples of each position, considering the RSS values of AP 7. The AP 7 was strongly seen in the collection area (3^{rd} floor), with a mean detected intensity of -72.9 dBm. The histogram shows differences that are notably higher that the variations commonly used in WiFi models. Though variations above the 20 dBm were found only for two positions, 15 positions (more than half of the positions) had variations from 11 dBm to 18 dBm. The large variations are a result of the AP signal intensity in the area, the considered collection directions, and the environment characteristics. As explained in Chapter 4, signal variations are higher for strongly detected APs than for weak, far away APs. One of the collection directions faced the

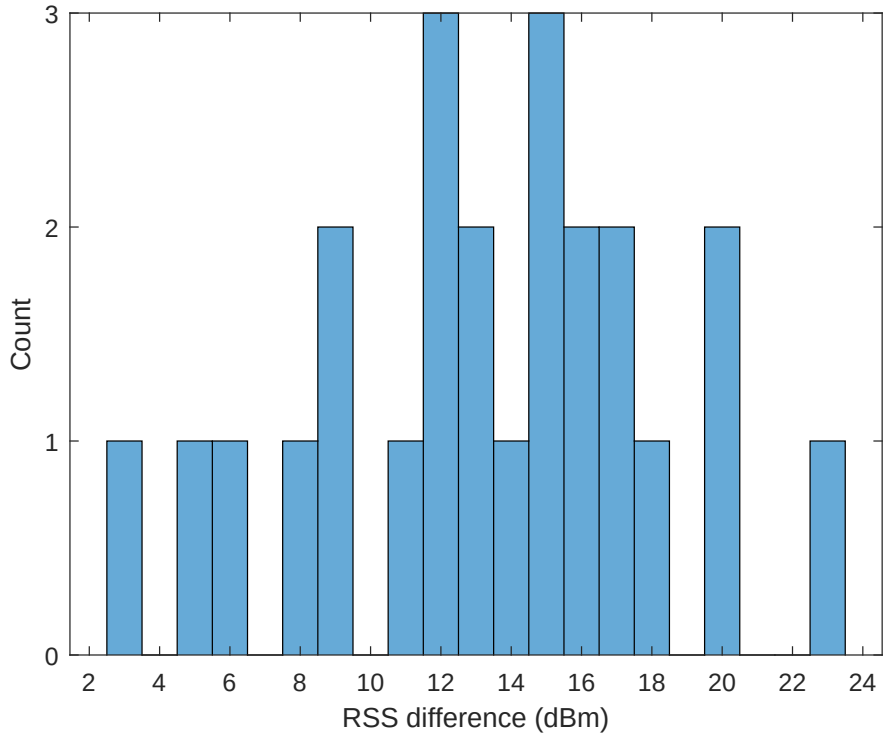


Figure 5.7: Histogram of differences between maximum and minimum RSS values per each position.

device emitting the AP, while other direction faced the opposite direction. Furthermore, the collection was performed at narrow lanes among bookshelves in the Library. If one or two Library user got close while the subject performed the collection, their bodies may have likely obstructed the AP signals. Thus, notable changes of over 20 dBm for the received signals from an AP can still be considered nominal. Drastic changes in the signal intensity of an AP measured in an area may also be a result of non-nominal situations. For example, the position where the device emitting an AP is deployed may change. It is also possible that a new fixed obstacle may appear in the environment. If the obstacle is very thick or includes metallic materials, the change in intensity may be significant.

An extreme situation of change in the detected signal from an AP is the non-detection case. The non-detection may occur because the signal strength is below the detection sensitivity of the receptor – usually a smartphone – or because the AP is no longer being broadcast. The former case is a nominal situation while the latter is not. The nominal non-detection cases are very common for APs that are already detected with weak intensities in the target environment. Figure 5.8 shows how the non-detection relates to the mean intensity value for an AP in the Library. The chart

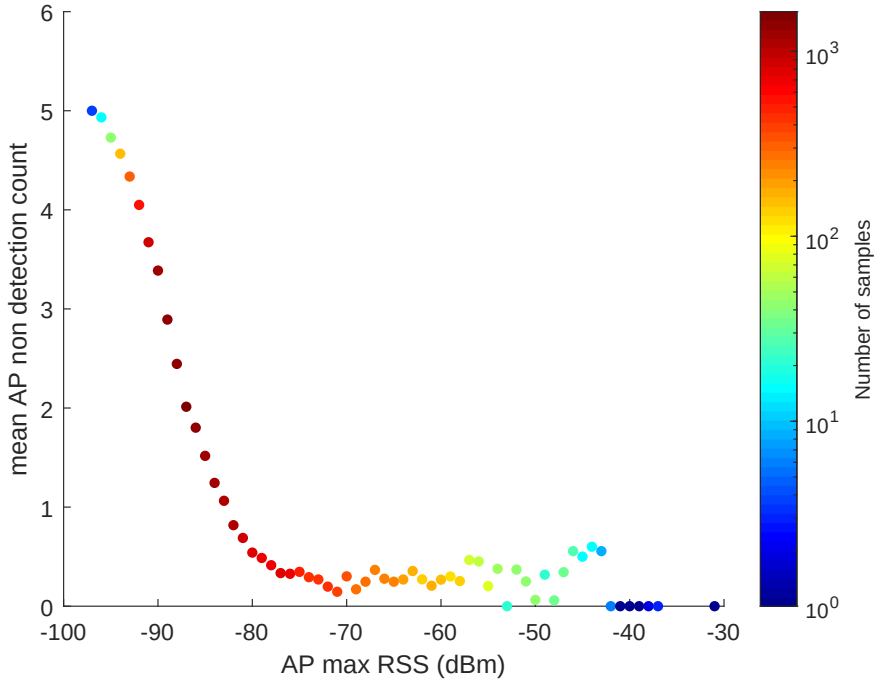


Figure 5.8: Mean number of times an AP was not detected for a position and direction.

results from an experiment that used a subset of the Library dataset. The subset included data from Training01 sets from the first 15 months. An AP filtering was performed to assure that the APs selected for the experiment had not been modified, i.e., their devices were not changed or switched off. The selected (48) APs were the same APs selected for experiments in the Library environment in Chapter 4.

The experiment harnesses the fact that there are six samples per each position and direction collection in the dataset. In the experiment, the maximum RSS value and the number of non-detection cases were computed for each month, AP, position and collection direction. Only cases where there was at least one detection of the AP in the five samples (the first one is removed) were considered. The numbers of non-detection cases were grouped by the maximum RSS values. The chart of Figure 5.8 shows the mean value of the numbers of non-detection cases associated with each maximum RSS value. As presented by the figure, the nominal non-detection situations occur mainly for APs weakly detected in an environment. The non-detection is very common for positions where an AP was detected with signals weaker than -90 dBm. However, non-detection cases also appeared in positions where an AP was detected with signals stronger than -50 dBm. Thus, the nominality of non-detection cases results not only from the number of those cases but also from whether the AP is strongly seen in the target area. Such APs are commonly close to the target area and very relevant for

positioning.

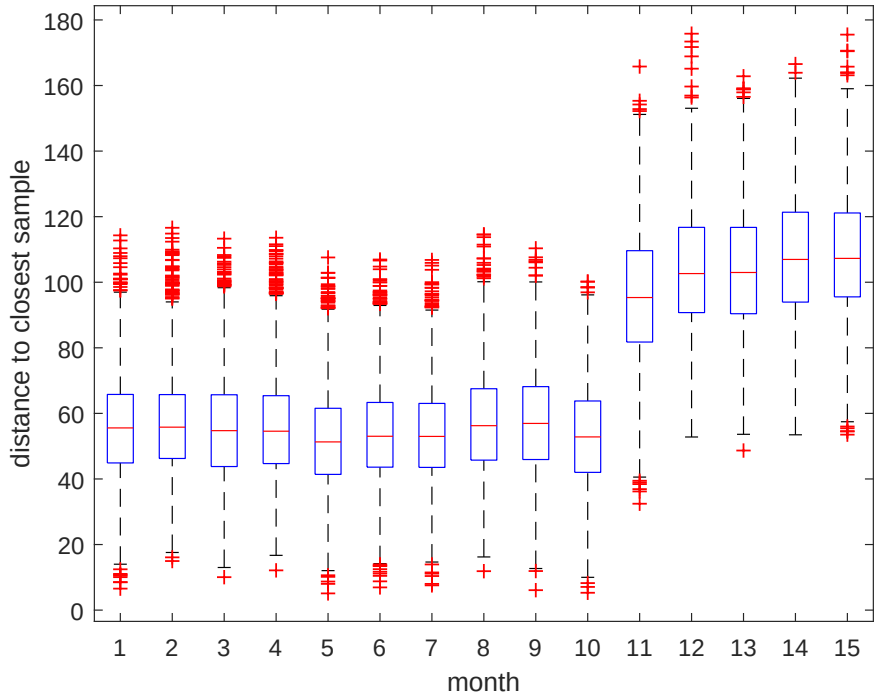


Figure 5.9: Distance from the operational fingerprint to the most similar fingerprint in the radio map along several months.

Non-detection cases that are non-nominal situations may drastically affect the positioning accuracy of an IPS. During the collection of samples for Month 11 in the Library dataset, some WiFi devices were changed. As presented in Chapter 3, some APs stopped being detected while others appeared, thus hinting at changes at the broadcasting parameters of the WiFi devices. Those changes modified the radio characteristics of the environment, thus affecting the similarity between fingerprints. Figure 5.9 depicts the changes that appeared in the fingerprint similarity as a result of the device modification in the Library environment.

The experiment for Figure 5.9 considered the set Training01 corresponding to the first month as training dataset and all test sets from each month. For each operational fingerprint of each test set, its most similar fingerprint in the training set was determined. The similarity was computed as the Euclidean distance in the fingerprint space. The changes in the WiFi devices occurred in Month 11 led to a drastic increase in the similarity distances. The increase is seen across the whole range of values, including the minimum, median and maximum values. The changes in fingerprint similarity can be used to detect non-nominal situations, as presented in Torres-Sospedra et. al. [27].

Changes are part of any environment and thus an IPS and its maintainers should be able to detect and adapt to them. Position estimations could be used to detect them [27]. Also, periodic updates to the radio map can be used to detect the changes if the positioning errors or the fingerprint similarities for the collected positions raise in drastic ways.

5.3 Conclusions

This chapter has addressed the WiFi radio map maintenance. The initial signal characterization of an environment is likely to need updates as a consequence of changes in obstacles from the environment or the devices that broadcast the APs. The chapter first addressed a method that guides the updates when the changes in the environment are not too drastic. The method chooses the positions in the environment where new samples are to be collected. The suggested positions are the set that is found to be the best compromise between a small number of collection positions and a good training set for a regression. The regression is used to estimate signal intensities for other positions in the environment. The chapter finally provided a brief characterization of the nominality of changes in an environment. The changes were explained through experiments that addressed changes in the RSS values and non-detection situations of an AP.

Some materials presented in this chapter are supported by one publication [20].

References

[1] Muhammad Usman Ali, Soojung Hur, and Yongwan Park. "LOCALI: Calibration-Free Systematic Localization Approach for Indoor Positioning". In: *Sensors* 17.6 (2017). ISSN: 1424-8220. DOI: 10.3390/s17061213.

[2] Rafael Berkvens, Herbert Peremans, and Maarten Weyn. "Conditional entropy and location error in indoor localization using probabilistic Wi-Fi fingerprinting". In: *Sensors* 16.10 (2016), p. 1636.

[3] Wonsun Bong and Yong Cheol Kim. "Fingerprint Wi-Fi radio map interpolated by discontinuity preserving smoothing". In: *International Conference on Hybrid Information Technology*. Springer. 2012, pp. 138–145.

[4] Keki M. Burjorjee. *SpeedyGA: A Fast Simple Genetic Algorithm*. 2009. URL: https://es.mathworks.com/matlabcentral/fileexchange/15164-speedyga--a-fast-simple-genetic-algorithm (visited on).

[5] Santiago Ezpeleta et al. "RF-based location using interpolation functions to reduce fingerprint mapping". In: *Sensors* 15.10 (2015), pp. 27322–27340. ISSN: 1424-8220.

[6] Ramsey M. Faragher and Robert K. Harle. "Towards an Efficient, Intelligent, Opportunistic Smartphone Indoor Positioning System". In: *Navigation* 62.1 (Mar. 2015), pp. 55–72. DOI: 10.1002/navi.76. URL: https://doi.org/10.1002/navi.76.

[7] Zhuan Gu et al. "Reducing fingerprint collection for indoor localization". In: *Computer Communications*. Vol. 83. 2016, pp. 56–63. DOI: 10.1016/j.comcom.2015.09.022.

[8] Noelia Hernández et al. "Continuous Space Estimation: Increasing WiFi-Based Indoor Localization Resolution without Increasing the Site-Survey Effort". In: *Sensors* 17.1 (2017).

[9] Shau-Shiun Jan, Shuo-Ju Yeh, and Ya-Wen Liu. "Received Signal Strength Database Interpolation by Kriging for a Wi-Fi Indoor Positioning System". In: *Sensors* 15.9 (2015), pp. 21377–21393. ISSN: 1424-8220. DOI: 10.3390/s150921377.

[10] Siddharth Joshi and Stephen Boyd. "Sensor selection via convex optimization". In: *IEEE Transactions on Signal Processing* 57.2 (2009), pp. 451–462.

[11] Loizos Kanaris et al. "Sample Size Determination Algorithm for fingerprint-based indoor localization systems". In: *Computer Networks* 101 (2016), pp. 169–177. ISSN: 1389-1286. DOI: http://dx.doi.org/10.1016/j.comnet.2015.12.015.

[12] John Krumm and John Platt. "Minimizing calibration effort for an indoor 802.11 device location measurement system". In: *Microsoft Research, November* (2003).

[13] M. Lee and D. Han. "Voronoi Tessellation Based Interpolation Method for Wi-Fi Radio Map Construction". In: *IEEE Communications Letters* 16.3 (Mar. 2012), pp. 404–407. ISSN: 1089-7798. DOI: 10.1109/LCOMM.2012.020212.111992.

[14] B. Li et al. "Method for yielding a database of location fingerprints in WLAN". In: *IEE Proceedings - Communications* 152.5 (Oct. 2005), pp. 580–586. ISSN: 1350-2425. DOI: 10.1049/ip-com:20050078.

[15] Liqun Li et al. "Experiencing and Handling the Diversity in Data Density and Environmental Locality in an Indoor Positioning Service". In: *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*. MobiCom '14. Maui, Hawaii, USA: ACM, 2014, pp. 459–470. ISBN: 978-1-4503-2783-1. DOI: 10.1145/2639108.2639118. URL: http://doi.acm.org/10.1145/2639108.2639118.

[16] C. Liu et al. "A Kriging algorithm for location fingerprinting based on received signal strength". In: *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. Oct. 2015, pp. 1–6. DOI: 10.1109/SDF.2015.7347695.

[17] R. Macho-Pedroso et al. "Optimal microphone placement for indoor acoustic localization using evolutionary optimization". In: *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Oct. 2016, pp. 1–8. DOI: 10.1109/IPIN.2016.7743609.

[18] MathWorks®. *Extrapolating Scattered Data, in MATLAB® R2017b*. 2017. URL: https://es.mathworks.com/help/matlab/math/scattered-data-extrapolation.html (visited on).

[19] MathWorks®. *Support Vector Machine Regression, in MATLAB® R2017b and Statistics and Machine Learning Toolbox™*. 2017. URL: https://es.mathworks.com/help/stats/support-vector-machine-regression.html (visited on).

[20] Germán M Mendoza-Silva, Joaquín Torres-Sospedra, and Joaquín Huerta. "Locations selection for periodic radio map update in wifi fingerprinting". In: *LBS 2018: 14th International Conference on Location Based Services*. Springer. 2018, pp. 3–24.

[21] Germán Martín Mendoza-Silva et al. *Long-Term Wi-Fi fingerprinting dataset and supporting material*. Nov. 2017. DOI: 10.5281/zenodo.1066041. URL: https://doi.org/10.5281/zenodo.1066041.

[22] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.

[23] Juri Ranieri, Amina Chebira, and Martin Vetterli. "Near-optimal sensor placement for linear inverse problems". In: *IEEE Transactions on signal processing* 62.5 (2014), pp. 1135–1146.

[24] Hosam Rowaihy et al. "A survey of sensor selection schemes in wireless sensor networks". In: *Proc. SPIE*. Vol. 6562. 2007, A1–A13.

[25] Venkat Roy, Andrea Simonetto, and Geert Leus. "Spatio-temporal sensor management for environmental field estimation". In: *Signal Processing* 128 (2016), pp. 369–381.

[26] J. Talvitie, M. Renfors, and E. S. Lohan. "Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization With Indoor Wireless Signals". In: *IEEE Transactions on Vehicular Technology* 64.4 (Apr. 2015), pp. 1340–1353. ISSN: 0018-9545. DOI: 10.1109/TVT.2015.2397598.

[27] J. Torres-Sospedra et al. "Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates". In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sept. 2018, pp. 1–8. DOI: 10.1109/IPIN.2018.8533791.

[28] Leehter Yao, William A Sethares, and Daniel C Kammer. "Sensor placement for on-orbit modal identification via a genetic algorithm". In: *AIAA journal* 31.10 (1993), pp. 1922–1928.

[29] Simon Yiu et al. "Wireless RSSI fingerprinting localization". In: *Signal Processing* 131 (2017), pp. 235–244.

Chapter 6

Fingerprinting-based IPS Evaluation

Errors in position estimations are common to (indoor) positioning systems. Despite those errors may range from within only a few centimeters for technologies like UWB or Ultrasound, the most often used IPS have error ranges of several meters. The scenario where the positioning is performed may determine the significance of the magnitude of errors. Also, the scenario usually influences the behavior of the IPS, as they usually rely on measurements of a signal – e.g., WiFi or BLE RSS values – that are heavily affected by the characteristics of the indoor environment. Moreover, the scenario is highly relevant given that, commonly, the position is determined for a person using a smartphone. Therefore, IPS evaluation is not only important for performing straightforward comparisons to determine which solution is the most adequate for a given application, but also because IPS evaluation is an integral part of the development of an IPS. As acknowledged in Chapter 2, the proper evaluation of IPS proposals is one of the main challenges for indoor positioning, and it is the subject of this chapter.

This chapter first addresses the evaluation within the context of IPS competition challenges and reproducible research, i.e., the chapter first provides insights into how to make the offline evaluation of IPS possible. Later, the chapter focuses on the measurement procedure of positioning errors, describing the proposal of using the pedestrian path length as a magnitude of the positioning error that effectively deals with obstacles, building, and floor miss-identifications. Finally, the chapter describes a local-level analysis of positioning errors, which study the variance of position estimates provided for the same test point and its behavior across an environment.

6.1 Offline Evaluation

Offline platforms for IPS evaluation [26] and IPS competitions [37, 36] have been among the proposed solutions for the IPS evaluation challenge. The offline evaluation is attractive because it does not incur in hardware deployment costs and provides IPS testers or challenge competitors with significant time to configure their positioning methods. Offline evaluation is particularly suitable for the evaluation of smartphone-based IPS, given that for collecting the data one just needs a smartphone with the required collection software. In the case of evaluation platforms, despite they may start providing services with some evaluation datasets in their stock, they mainly rely on the wiliness of their users to provided more datasets. Although the general target of all datasets may be IPS evaluation, their specific goals and related characteristics may vary significantly. Regardless of the goal of a dataset, and following approaches typically used in machine learning, the dataset should contain enough data to fit the model for an IPS to a specific environment – training set – and to perform the unbiased evaluation of a model fit and the hyper-parameter tuning – validation set. Also, a set is required for the unbiased evaluation of a final fit – test set – and its data should preferably be distinct from those used for training and validation.

Competitions are normally held at a specific date, and the data collection and curation should be completed well in advance to that date. For example, for the 2016 and 2017 IPIN conferences, the call for competition for the Track 3 – the track corresponding to smartphone-based offline evaluation – was published a few months before the competition date [37, 36]. Those periods are required to grant competitors enough time to calibrate and evaluate their proposals. For those competitions, collection was performed using several modern smartphones and a dedicated Android application named *GetSensorData* [16]. The data consisted of WiFi RSS, inertial data, magnetic field, GPS, and pressure measures, among others. The competition of 2017 also included BLE RSS measures. In the case of the offline evaluation in IPS competitions, the data provided to competitors and used to evaluate them try to reflect an IPS usage by people as realistic as possible. Therefore, the collection is performed by a person moving as natural as possible across a predefined set of waypoints. Also, the person holds the phone in natural ways, either at chest level simulating common interactions with apps or with the arm relaxed and stretched down performing natural swings. Apart from the signal measurement data and the position tags, additional information like floor plan maps, map-based reference trajectories, and videos were also provided in the hope to help the competitors achieve better position estimations.

For the offline evaluation in IPS competitions, other challenges exist apart from depicting IPS conditions as realistic as possible. It is not trivial to choose the collection reference points or waypoints that properly characterize the environment but still keep a challenging character for the evaluation set. For example, in the 2017

IPIN Track 3 competition, the training and validation routes for the *UJIUB* building alone included 460 waypoints across five floors, and a large number – not to be disclosed – of waypoints for test routes. As consecutive waypoints were separated by a few meters, training and validation routes described much of the environment. Thus, test routes had to be chosen carefully to avoid repeating large patches of other routes. Performing the position – and time – tagging as accurate as the available means allow requires a committed and careful work. The subject performing the collection task has to become familiar with the environment, has to have clear means to follow the plan, and landmarks or waypoints have to be made easily identifiable. Furthermore, the collected data should be carefully reviewed to find any inconsistency. Finally, providing a final ranking of competitors as agreeable – "fair" – as possible by all people involved in the competition requires tact and compromise. The competition rules should be properly described well in advance to the date when it is held.

Offline evaluation is not the ultimate solution for the challenge of IPS evaluation. There are IPS with hardware interactions unable to be tested offline. Furthermore, the offline evaluation allows conveniences like computationally heavy procedures, setting up completely different IPS for different scenarios, manual reviews of position estimates or several estimation attempts that are not normally applicable to real deployed systems or their online evaluation. Nevertheless, offline evaluation is acknowledged as very valuable [26, 31] and it will be even more valuable as new, relevant evaluation datasets become publicly available.

6.2 Error Measurement as Pedestrian Path Distances

Errors in IPS are commonly measured as the Euclidean distance between the true position, which is also called ground truth in some machine learning contexts, and the position estimated by an IPS. The Euclidean distance is easy and fast to compute, and the scenario should not matter if error ranges are very small – within centimeters – or the targeted scenario is free of relevant obstacles. However, the error ranges are not small for the most often used IPS. Also, the walls that divide indoor spaces and the floors that separate stories in a building, among others, are relevant obstacles for people.

The *ISO/IEC 18305:2016* International Standard [21] describes the methodology suggested for generic IPS evaluation. It defines the tests for an IPS under several evaluation scenarios and types of entities to be positioned. The standard provides suggestions about the density, disposition, and surveying accuracy of the test points. It also defines that the positioning error should be computed as the Euclidean distance

(l_2 norm) between the actual and the estimated positions. The resulting distances are later resumed using several statistics. Despite the existence of the standard, the evaluation of IPS is still an open challenge, as stated by Potortì et al. [32]. In Potortì et al. [32], the authors provide a critical reading of the standard, refining which statistics should be used – percentile values – and highlighting the complexity of error measurement in environments distinct from single-floor open spaces. They suggest three alternatives for error measurement:

1. 3D Euclidean distance, as defined in the standard. It is widely applied in IPS academic works and it is very easy to compute. However, it over- or underestimates the importance of floor or building estimation errors.

2. 2D Euclidean distance with floor and building penalties, as used in several IPS competitions [37, 36]. It is also very easy to compute, and the over- or underestimation of the importance of floor or building estimation error is less significant.

3. Real distance, as proposed by a preliminary work of this section [25]. It is not easy to compute as it requires the application of pathfinding methods based on map information. However, it provides an approximate distance of the path that a pedestrian would walk.

The precise map information may not exist for a given scenario. However, its confection would be valuable not only for IPS evaluation but also for other stages of the IPS realization. The map information may be employed during radio-map creation for collection campaigns and radio-map enrichment [19] or IPS estimates corrections using map matching [43]. Furthermore, the real distance computation requires pathfinding in the indoor scenario, which could be reuse to provide navigation services, i.e., directions, to the IPS users.

The following two subsections address the error measurement procedures. In them, the position estimates from an IPS and their corresponding ground truth positions are considered the endpoints of paths that a pedestrian would follow. The lengths of such paths are regarded as the magnitudes of the measured positioning errors. Such error magnitudes provide a more realistic estimation of distances from the perspective of a person, or from the perspective of navigable paths for robot-based positioning. First, Section 6.2.1 chooses and describes three 2D pathfinding methods. Then, Section 6.2.2 describes the integration of the 2D pathfinding methods into multi-floor and multi-building path determination procedures. Also, in Section 6.2.2, the procedures are tested and compared using position estimations from several IPS for a large environment.

6.2.1 2D Pathfinding

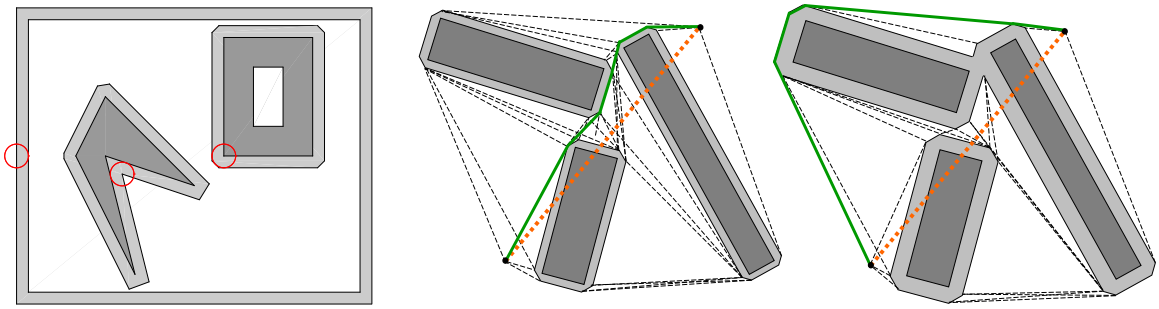
Pathfinding is an important topic for computer games [15, 7, 1] and robotics [24], and thus it has been intensively studied. This section describes the three 2D pathfinding methods chosen for demonstrating the IPS error measurement procedures and briefly mentions other alternatives. The first two methods are Visibility Graphs (VG) and Navigation meshes (NM). Both of them are vector-based methods, i.e., they operate on the polygonal representation of the environment, and create network (graph) representations of the input environment. The VG method assures finding the shortest paths while the NM method creates networks that are simpler than those created by the VG while also finding paths similar to those found by VG. The third method is Fast Marching (FM). FM operates on raster (image) representations and provides paths smoother than those determined by the VG or NM methods.

Visibility Graphs and Navigation Meshes

The VG and NM methods require the representation of the free space of the environment for performing 2D pathfinding. The free space refers to all positions where the subject for whom the path is computed can be without colliding with, or trespassing, the boundaries of the environment or the obstacles it contains. The free space may be obtained from floor plans of the target buildings by doing some processing using the appropriate tools, as Section 6.2.2 will later describe. The free space is finally represented using polygons with holes, neither of which are necessarily convex. Vector-based methods like VG and NM effectively avoid that the computed paths may trespass boundaries. However, the free space representation is not enough for avoiding that a real-world subject – which occupies an area larger than zero – may touch the boundaries while following such paths. To take into consideration the size of the subject, a concept known as *forbidden space* is used.

The determination of the forbidden space is as presented by Figure 6.1a. In plain terms, the limits of the environment – polygon boundaries – are grown inward and the inner obstacles – polygon holes – are grown outward. The growth should be related to the size of the subject. In Figure 6.1a, the size of the subject is represented by red circles. The size of the subject is no longer a concern after computing the forbidden space, i.e., the new environment representation. The new environment representation precisely identifies the areas that are valid for path computation. The CAD/CAM technique known as polygon offsetting [6] has been used in this work. It is adequate for outer offsets and inner offsets (insets). The magnitude of the growth should be carefully considered, given that some offsetting operations may block ways that would otherwise be available for path determination, as presented in Figure 6.1b.

The new environment representation is then directly used to compute VG or NM.



(a) Offsetting determination. (b) Effect of growth size on path determination.

Figure 6.1: Offsetting and VG path computation examples. Obstacles are represented in dark gray color and their corresponding offsets in light gray color. The red circles represent the size of the subject. Black dots indicate the paths' endpoints. The black dashed lines, the solid green line and the dashed orange line show arcs of the visibility graph, the shortest path, and the straight-line (Euclidean) distance, respectively.

Given a set S of disjoint polygonal obstacles, the *Visibility Graph* of S has as nodes the vertices from S and an arc connects two nodes v and w if it does not collide with any edge from an obstacle or the environment's limits. Algorithms for the efficient construction of visibility graphs already exist [24, 28]. The VG can be used to find the shortest path between two points among obstacles. The shortest path determination requires to set the arc weights to the actual distance between each pair of vertices and the usage of a graph searching algorithm like e.g., Dijkstra's [9] or A^* [17]. The shortest collision-free path between two points is then composed by arcs of the visibility graph and by the path endpoints [28].

The paths found with visibility graphs usually have hard turns on the edge's vertices. The hard turns are not ideal for describing people movement. Also, the VG typically have a large number of arcs, which increases the computational cost of pathfinding. Figure 6.1b presents examples of VG and the determination of the paths between two endpoints. Notice that graph arcs are included for every pair of vertices visible to each other, including the endpoints. For the simple environment of Figure 6.1b, the number of visibility arcs is higher than 30. For the experiments presented in this chapter, a MATLAB©implementation for VG construction that performs brute-force testing of arc eligibility was created. The determination of VG requires several geometrical operations with varying degrees of complexity, both in terms of computational cost or implementation difficulty. The most used operations were point-in-polygon and ray- or segment-to-environment intersection.

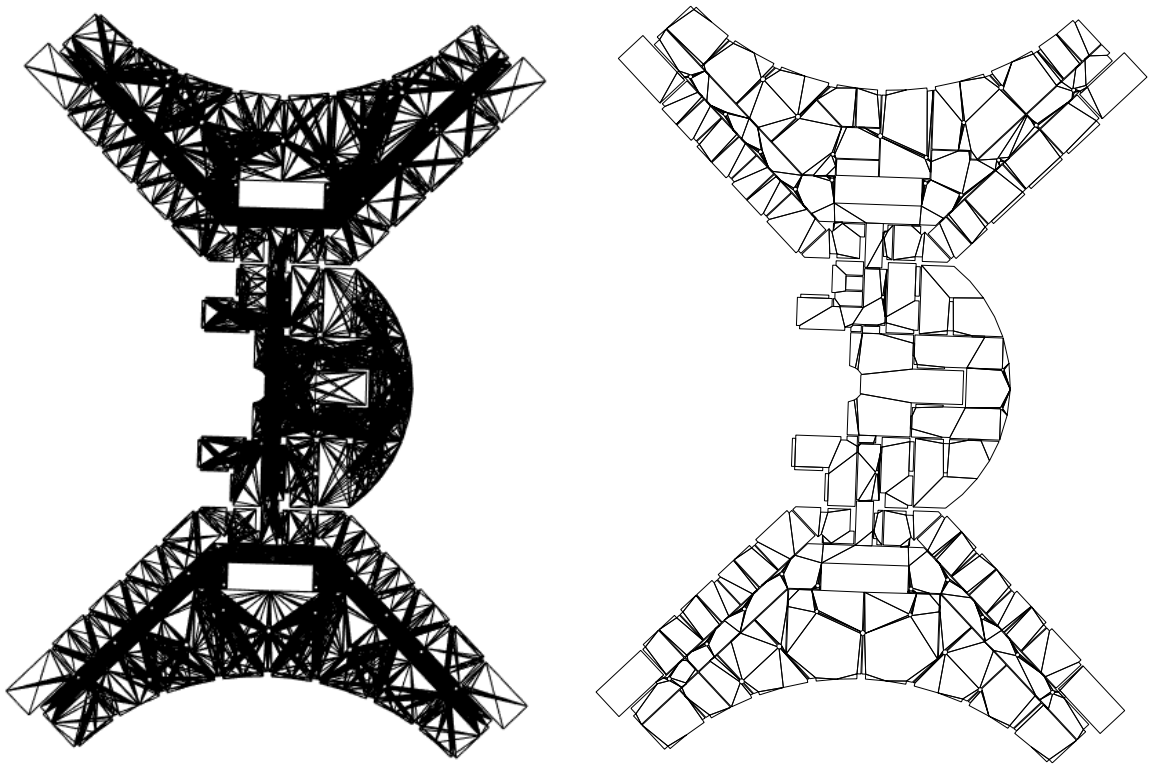
A *Navigation Mesh* [34, 39] subdivide the free space of the environment, or its representation after offsetting, in as a set of convex polygons. In robotics, the same concept is called meadow maps [39]. The convex polygons represent spaces where

movement between two points of the polygon's boundary is possible without colliding with an obstacle. A constrained triangulation could be used as an NM. For example, a navigation mesh can be produced by creating a Delaunay triangulation [8] from the environment vertices and then removing the triangles that do not belong to the environment.

Triangles are the simplest convex polygons. Thus, an NM based solely on a triangulation normally has too many unnecessary divisions of the space. An NM construction can start from a triangulation. Then, the construction iteratively combines adjacent polygons to remove nonessential edges, thus obtaining new convex polygons [39]. Several approaches exist for NM construction [41]. The pathfinding algorithm applied over an NM may consider only the vertices of the NM's polygons, only the middle points of the edges of the NM's polygons, or both vertices and middle points. Additionally, the resulting path may be smoothed using line-of-sight testing [34]. A MATLAB©implementation for NM was created for the experiments presented in this chapter. The implementation used the construction steps previously mentioned. It first builds a triangulation and then combines triangles into more complex convex polygons. The path is computed using both vertices and middle points and the resulting path is smoothed.

In summary, the VG and NM methods are used to compute graphs that represent an environment. Figure 6.2 presents an example of the resulting graph for a building (TI) that belongs to the environment where the experiments later presented in this chapter were performed. Notice that the graph obtained by VG (see Figure 6.2a) has much more nodes and arcs than the one produced by NM (see Figure 6.2b). The more complex the resulting graph is, then more time it takes to compute paths on the graph. Path determination in VG is considerably slower than in NM for complex environments, despite the NM method performs path smoothing. In both VG and NM, the resulting graphs have to be temporally modified to account for the path endpoints. In the case of VG, the pathfinding method first has to add new edges that connect the endpoints to the environment vertices that are visible from the endpoints. In the case of NM, for each endpoint, the pathfinding method first has to add new edges that connect the endpoint to the vertices of the polygon where that endpoint lies.

The VG and NM methods allow the computation of paths between endpoints that lie at valid areas, i.e., inside the environment limits but outside the obstacles. If an endpoint lies outside the valid areas, a correction step is required. The correction step translates the endpoint to the closest point that is in a valid area. Such point is the closest point in an edge of the environment translated by a very small distance inward or outward.



(a) Visibility Graph (b) Navigation Mesh

Figure 6.2: Pathways representations for the TI building, lowest floor.

Fast Marching

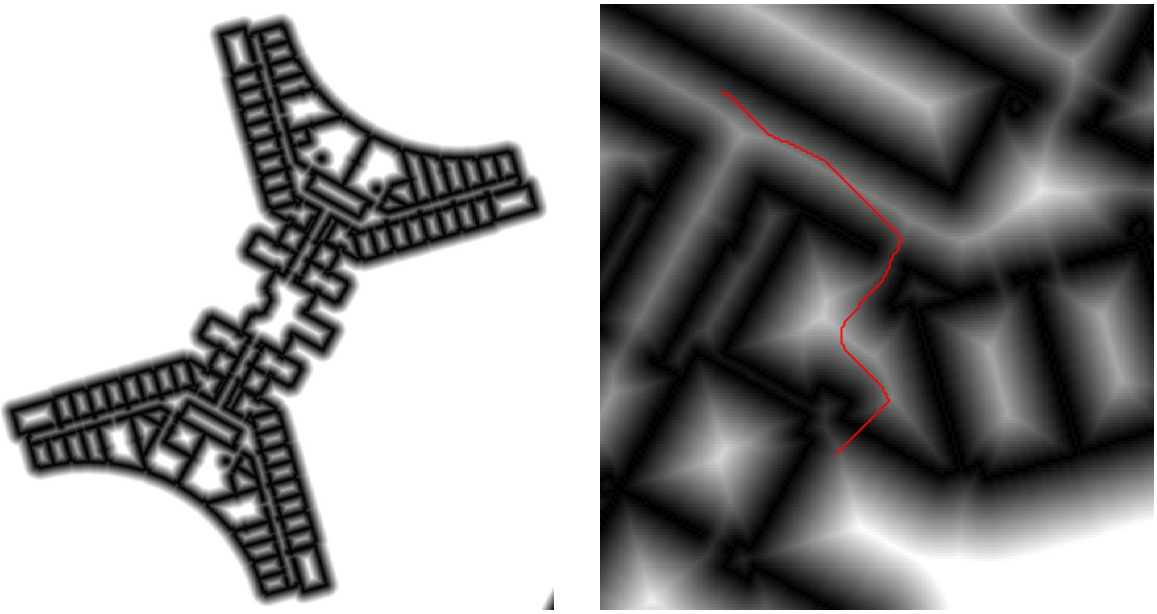
Fast marching [33] is a method that, for a given domain and propagation speeds, finds the time that a wavefront would reach each point in the domain from given source points. In the case of images, the domain is the grid cells and the cell values may define the propagation speeds. The algorithm is similar to Dijkstra's [9] and uses the known discrete solution to the Eikonal equation [5] to update cell values and thus find the time at which the front crosses a point. For pathfinding, the speed in cells representing obstacles can be set zero. As the algorithm produces no local minimum [40], a path between source and destination points is determined following the maximum gradient direction from the initial point. Further improvements to the original algorithm, like the MultiStencils Fast Marching that solves the Eikonal equation through several stencils [18], produce smoother paths. The implementation used in the experiments presented in this chapter is based on MultiStencils [22].

Unlike the VG and NM methods, FM does not create a graph representation of the environment. Instead, the input image is transformed into a map of speeds. In

this work, the map of speed was created from a distance transformation applied to the input image, in which new cell values represent the distance to the nearest cell that corresponds to a boundary. The distance value of each cell is then transformed into a speed value using the following formula:

$$v = d. * (d < 30) + 30 * (d >= 30) + 0.01, \tag{6.1}$$

where d is the cell distance value and v is the resulting speed. Figure 6.3a shows the resulting transformation for the image of a building. Thus, the computation time for FM to determine a path depends on the size of the input image, in pixels. The size of the input image is directly related to the size of the environment and its complexity, given that low-definition images commonly lose details that may be relevant for path computations. Also, the endpoint corrections performed in VG and NM is managed differently in FM. Given that its path determination avoids crossing the environment boundaries, and areas outside the environment are also considered with the lowest speed, the correction is intrinsically incorporated into the computed path. Figure 6.3b presents examples of speed transformation and path determination. In Figure 6.3, areas outside the building were given white color for better visualization.



(a) Speed transformation. (b) No correction example

Figure 6.3: Fast Marching examples.

Other alternatives

A network of predefined waypoints, or points of visibility, is commonly used for pathfinding [15]. Usually, the nodes and edges are set manually or generated using a grid automatically generated by a GIS tool like ArcGIS© indoor [10]. The resulting network of points is usually simple enough for the path computations using Dijkstra [9] or A^* [17] to be fast. However, it is not guaranteed to find the shortest path between two endpoints on a network of waypoints. Also, the manual recreation of the network if the scenario changes is a very time-consuming task.

Algorithms like Dijkstra or A^* can be directly applied to raster environment representations [15]. The raster's discretization affects the paths that may be found, which, besides, have unrealistic looks. The $Theta^*$ algorithm [27], which does not constrain paths to grid edges, produce better paths than Dijkstra or A^* in terms of realistic look and path length.

The map representation can be transformed before the application of a pathfinding algorithm. Quadtrees [11] are used to reduce the number of nodes that represent the environments. With them, the whole space is iteratively divided into four new ones until a subspace does not contain any obstacles or has a minimum size [20]. Skeletons are also used as a reduced representation of the environment [3, 13]. The skeletons commonly represent a set of points that have more than one closest point on the boundary of the polygon. They can be computed either for vector [12, 14] or rater [23] data.

Finally, it is common to subdivide large environments in map hierarchies. A path computation with map hierarchies first requires the determination of the involved scenarios. Then, sub-paths are computed for each of them. Ultimately, the sub-paths are concatenated to form the output, whole path. If the proper division and composition techniques are applied, the pathfinding is sped up in map hierarchies with respect to considering the whole map at once for large scenarios [4, 30].

6.2.2 Position Estimation datasets

This section shows the error measurement procedures using the 2D pathfinding methods introduced in Section 6.2.1. The measurement procedures were applied to position estimates provided by IPS trained and evaluated on data from the UJIIndoorLoc dataset [38]. The dataset allows the evaluation of WiFi fingerprinting solutions on a complex environment. The environment encompasses three buildings at the UJI, each having 4-5 floors (see Figure 6.4). The positioning based on WiFi fingerprinting is challenging. An AP seen with the strongest intensity is not necessarily the closest one as a result of intrinsic signal fluctuations and environment-induced variations, including absorption by people. The UJIIndoorLoc includes training and validation sets

that are publicly available and a test set that is kept secret by the dataset curators. The test set was used in the IPIN 2015 Track 3 competition [35] to evaluate WiFi fingerprinting solutions. Track 3 of the IPIN 2015 competition allowed the competing teams up to five sets of position estimates to be submitted. Only the best according to the mean accuracy was selected for competing in the final ranking. The competing teams were identified by the names "RTLSUM", "HFTS", "ICSL", and "MOSAIC". In total, the teams provided 14 sets of position estimates.

Figure 6.4: 3D areal view, from Google Earth, of the three buildings considered in this section's experiments. From left to right, the buildings are identified as TI, TD and TC.

The intention of this section is only demonstrative, and it does not suggest that the results of any competition were unfair nor the proposed error measurement procedure is the definitive, fair way to evaluate competing IPS. Fairness is a very subjective concept that is not within the focus of this dissertation.

Environment Information Preparation

The pathfinding methods require the compilation of the target environment information. The environment is composed of the three university buildings where the samples of UJIIndoorLoc dataset were collected. The environment information included a representation of each building's floor that depicted the building's boundaries and the inner (structural) obstacles. The representation of a floor was either a set of polygons or an image. Also, the environment information included a representation of the communications pathways between floors inside a building and between buildings. In the target environment, floors adjacent to each other communicate through stairs or elevators. Buildings communicate through outdoor paths between their entrance doors.

The environment information was obtained from transformations to CAD and GIS data. The data were carefully created for previous projects in the university [2]. The positions for environment representation and the position tags of the UJIIndoorLoc datasets are represented using the same coordinate system. That coordinate system allows straight-line distance computation directly using the Euclidean distance.

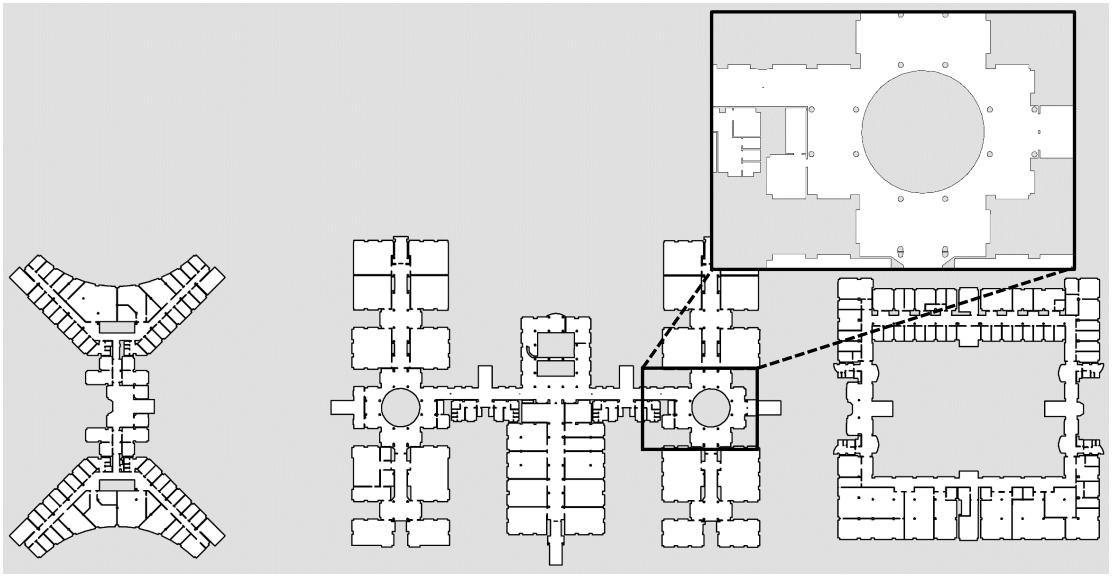


Figure 6.5: Geometric data of a building representing the free space (white spaces).

To create the representation of each floor from each building, room's doors and similar barriers were removed to obtain a continuous free space upon which pedestrian path could be computed – after offsetting, if needed. Figure 6.5 shows the resulting free space representation for the first floor of each of the three buildings. In such representation, a path may span from any two point where a subject is allowed to roam.

Polygon simplification was performed to reduce the number of vertices and edges. In the case of polygonal representations, the simplification is highly relevant if the geometries include curves. For example, in the case of circular pillars, the simplification reduced the number of vertices by a factor of 5, from about 40 to 8 vertices. Also, the polygonal representation required the offsetting process described in Section 6.2.1 before the application of the VG and NM methods. The images created for raster representations used a map scale and a pixel density that allowed the direct transformation from pixel distance to real-world distance. Besides, the symbology used to represent lines allowed thin but continuous edges representations.

The inter-floor (vertical) and inter-building communication ways were represented as static information in the form of triplets. The triplets contained two endpoint tags

Table 6.1: General characteristics of environment's buildings.

Building	Area (m^2)	Floors	Vert. Comm.	Ext. Doors
TI	$\approx 15,600$	4	18	7
TD	$\approx 37,150$	4	37	8
TC	$\approx 27,100$	5	24	4

and the distance between the two endpoints. The endpoint tags contained a 2D position, the building and floor that the endpoint belonged to, and an endpoint ID. The distances were measured using a GIS tool as fixed pedestrian paths. Such paths went through stairs and elevators for inter-floor ways and through pedestrian routes that communicated building doors for inter-building ways. Table 6.1 presents the numbers of vertical and inter-building ways along with other building information. The compiled vertical and inter-building ways data considered any vertical way between two adjacent floors from the same building and one route for any pair of doors from two buildings. The integration of vertical and inter-building links information to the 2D pathfinding is explained in the following section. The ways that connect floors and buildings were manually compiled for this work. Despite automatic methods for topology extraction or features identification from floor plans or other indoor models exist [42, 29], this work favored the manual extraction to assure the accuracy of the data.

IPS Error Measurement Procedures

Let a and b be endpoints for which a path is to be determined. Without loss of generality, let a be the origin and b the destination of the path. If a and b are located on distinct floors or buildings, determining the path that connects them requires finding inter-floor (vertical) and inter-building (outdoor) path segments. Those segments, when orderly concatenated together, transits from the floor-building of the origin to the floor-building of the destination. Let $I_{a,b} = \{i_1, ..., i_n\}$ be the result of concatenating such segments, where i_j is a segment endpoint. Each i_j is either an endpoint of a vertical communication way or an endpoint of an inter-building route. For the test environments, the full path is the concatenation of the 2D path between a and i_1, the path represented by the $I_{a,b}$, and 2D path between i_n and b. In Figure 6.6, $I_{a,b}$ is represented by the green and blue segments. Notice that the green segments are known in advance. A blue segment represents a connection between an endpoint of a vertical way and an endpoint of an inter-building route. The blue segments are 2D paths that have to be dynamically determined. The 2D path determination, i.e., inside the floor of a building, is performed using either of the VG, NM or FM 2D

pathfinding methods. If a and b belong to the same floor and building, then $I_{a,b} = \emptyset$.

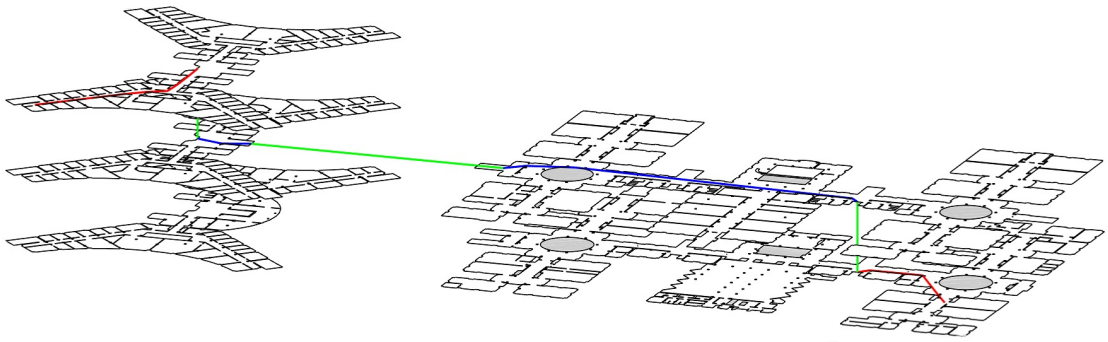


Figure 6.6: Shortest distance path between two endpoints located in distinct buildings and floors.

It is difficult to determine in advance the set $I_{a,b}$ that would provide the shortest path between a and b. The usage of heuristics like, e.g., choosing the two doors from the origin and destination buildings which are the closest among them, does not guarantee to obtain the shortest path. Therefore, to assure to find the shortest path and to enable the exploration of longer paths, the basic path determination algorithm in this work follows the approach of determining all possible sets $I_{a,b}$ that are always-forward alternatives. Always-forward alternatives are those where floor changes are made only if they represent a floor-wise or building-wise advancement towards the destination. Paths longer than the shortest distance ones are possible choices for a pedestrian unfamiliar with the environment.

The previous approach is directly applicable to any of the VG, NM or FM methods. However, the number of all possible always-forward sets $I_{a,b}$ is large and exponentially related to the number of vertical and inter-building links in the environment. Given that the VG and NM methods produce representations of the environments as undirected graphs, it was also tested a variant of the basic algorithm in which the individual 2D graphs produced for each floor by VG and NM are connected to represent vertical and inter-building communication ways. This variant produces a single graph and avoids the determination of the sets $I_{a,b}$. Such 3D graph is the classical representation used for navigation services in indoor environments [10].

Algorithm 1 presents the basic path determination algorithm, which is hereinafter called the EE variant. Algorithm 2 presents the variant that creates a single 3D graph, which is hereinafter called the SM variant. Apart from the environment representation, the algorithms require endpoints as an input. The endpoints are organized into separated datasets, which in turn are composed of sets of endpoints. Recall that the endpoints of a path represent a position estimate and its corresponding ground truth position. The position estimates were provided by teams that participated in

```
   Input: Environment Data Representation, Endpoints
   Output: Paths between endpoints
1  Floor model creation
2  for each endpoints dataset do
3  |  for each endpoints set do
4  |  |  Endpoints expansion
5  |  |  Endpoints correction
6  |  |  Paths computation
7  |  |  Endpoints contraction
8  |  end
9  end
```

Algorithm 1: Path determination steps. EE (Endpoints Expansion) variant.

```
   Input: Environment Data Representation, Endpoints
   Output: Paths between endpoints
1  Floor model creation
2  Single model integration
3  for each endpoints dataset do
4  |  for each endpoints set do
5  |  |  Endpoints correction
6  |  |  Paths computation
7  |  end
8  end
```

Algorithm 2: Path determination steps. SM (Single Model) variant.

the Track 3 of the 2015 IPIN competition. Each team was able to provide up to 5 sets of estimations. Thus, an endpoint dataset is associated with all the estimations that a team provided, and a set of endpoints is associated with a particular set of estimations of a team. The explanation of the steps for Algorithm 1 is as follows:

- *Floor model creation*: This step loads the environment data, i.e., the polygonal representations or images of each floor plan and the vertical and inter-building connection information. For the VG and NM methods, the step computes the 2D graph representation. For the FM method, it performs an image preprocessing - e.g., image binarization and speed transformation.

- *Endpoints expansion*: This step transforms the endpoint data - formed by pairs (a, b) - to account for paths between different floors or buildings. After determining the all always-forward sets $I_{a,b}$, the endpoint data is transformed by

removing the endpoint pair (a, b), and by adding (a, i_1) and (i_n, b) for every $I_{a,b}$. If a segment $(i_k, i_{k+1}) \subset I_{a,b} \wedge k \leq |I_{a,b}|$ is such that the (i_k and i_{k+1}) endpoints are one an endpoint of a vertical way and the other an endpoint of an inter-building route, the endpoint pair (i_k, i_{k+1}) is also added. The performed transformations are tracked for the *Endpoints contraction* step.

- *Endpoints correction*: For VG and NM methods, it corrects endpoints lying outside the environment to the closest point inside the building. The FM method does not requires this step.

- *Path computation*: For VG and NM methods, this step computes shortest path among endpoints pairs using the graph representation of the environment and the A^* algorithm. In the case of FM, the MultiStencils Fast Marching method is applied upon the pre-processed floor images for the endpoint pairs.

- *Endpoints contraction*: The track of the endpoint transformations performed in the *Endpoints expansion* step is used for path concatenation, i.e., to construct paths connecting the original endpoint pairs (a, b).

Algorithm 2 performs the *Floor model creation* and *Path computation* steps explained for Algorithm 1. Algorithm 2 perform the *Single model integration* step, which avoids the *Endpoints expansion* and *Endpoints contraction* steps found in Algorithm 1. The *Single Model Integration* step links together into a single graph the graphs created for each floor of each building by the NM and VG methods. The graphs are linked attending to the vertical and inter-building connections. Each connection represents a new arc that is added to the graph. Each endpoint of a newly added arc is joined to the graph of its corresponding floor by adding new connecting arcs to the graph. The new connecting arcs are determined using the same approach used in the path computation for attaching the origin and destination points to the environment. For VG, the connecting arcs are the visibility arcs that irradiate from the endpoint. For NM, the connecting arcs are segments that connect the endpoint to the vertices of the (convex) polygon that contains the endpoint.

The output of Algorithm 1 and Algorithm 2 are the paths that connect their inputted endpoints. The application of Algorithm 1 or Algorithm 2 with any of the proposed 2D pathfinding methods is hereinafter called measurement procedure. Five measurement procedures are addressed:

- *VG-EE*: Applies Algorithm 1 using VG for 2D pathfinding.

- *VG-SM*: Applies Algorithm 2 using VG for 2D pathfinding.

- *NM-EE*: Applies Algorithm 1 using NM for 2D pathfinding.

- *NM-SM*: Applies Algorithm 2 using NM for 2D pathfinding.
- *FM*: Applies Algorithm 1 using FM for 2D pathfinding.

The procedures VG-SM and NM-SM apply A^* over the entire (3D) graph and produce only the shortest distance paths. The procedures VG-EE, NM-EE, and FM, however, produce several alternative paths for each pair of endpoints lying on different floors or buildings. The shortest distance path has to be sought among all those alternatives. Other measurement alternatives could have been used instead of the shortest distance. One example is the path that includes the nearest floor connection or the nearest building door to the origin endpoint. The measurement alternatives to the shortest distance are useful for pedestrian navigation. However, they are not addressed here because they are unlikely to be chosen by non-human subjects (robots) or people who are familiar with the environment.

The outputs of the measurement procedures vary not only on the paths they produce but also on the time they take to first prepare their models and then to compute the paths. All the results presented in this section were obtained from experiments carried out in a Desktop Computer with Intel(R) Core(TM) i7-8700 CPU @ 3.2 GHz, 16 GB of RAM memory, running MATLAB ©R2019a on a MX Linux 18.2 Continuum. The implementation of the procedures was performed in the MATLAB©language, favoring correctness over efficiency. Also, the measured times were not the result of many repetitions. Therefore, the measured times are only presented to grasp a general idea of computation time requirements.

Table 6.2: Measured computation times for steps from Algorithm 1. The steps are *Floor model creation* (FMS), *Endpoints expansion* (EES), *Endpoints correction* (ECS) and *Paths computation* (PCS). FMC reports total time, EES reports mean time per endpoint set, ECS reports mean time per endpoint, and PCS reports mean time per computed path.

Step	VG-EE	NM-EE	FM
FMC Σ (s)	53.33	123.47	1.30
EES $\mu \pm \varphi$(s)	22.48 ± 55.09	26.68 ± 65.69	22.58 ± 55.43
ECS μ (s)	1.5×10^{-3}	1.6×10^{-3}	NA
PCS $\mu \pm \varphi$ (s)	0.21 ± 0.04	0.03 ± 0.01	2.51 ± 0.12

Table 6.2 presents the measured times for some steps of Algorithm 1. The *Floor Model Creation* step is performed by all procedures. Though the time this step takes is not negligible, it is not significant given that it takes less than 3 min to complete for all approaches. Furthermore, the step is to be executed only once and the created

graphs or processed images can be stored for their later usage. The stored information can be later used indistinctly by procedures based on Algorithm 1 or Algorithm 2 to perform as many path computations as required. The times for the *Endpoints expansion* step are small for all procedures when compared to the time employed in the *Paths computation* step. However, they are not negligible and only one order of magnitude lower than the time of the *Paths computation* step for the FM procedure. Their φ values were not reported as their were close to zero.

The φ values are larger than μ for the *Endpoints expansion* step because the time measurements for teams RTLSUM, HFTS, and ICSL were small – taking 0.84 ± 0.32 – and the two corresponding the MOSAIC team were large – 183.17 and 180.28. In the EE variants, several paths are computed for some endpoint pairs. The number of such alternative paths may be significantly large depending on the amount of floor and building miss-identifications in the estimates of the evaluated IPS. As a mean, the number of path computations processed using this modality was 5.83 times larger than the number of position estimates. Thus, the total path processing time in for NM and VG methods in the EE variants was almost 6 times larger than in the SM variant counterparts. The largest increase in the number of computed paths was 26.70 and corresponded to one of the estimation sets of the team MOSAIC. Furthermore, the EE variants demand a lot of memory resources, mainly because of the path information is kept for later analysis.

The times for the *Paths computation* step are the most important measures presented in Table 6.2. The times for the NM-EE procedure are small and thus affordable for evaluation purposes – which is typically an offline procedure – using most modern desktop computers, even for large IPS evaluation sets. The times for the VG-EE procedure are about seven times larger than those for NM-EE. Nevertheless, time affordability should not be a hard concern in most scenarios for VG-EE. The FM procedure, however, uses more than 2 s as a mean to run a single path computation. Such large time implies that it could take weeks to complete the paths computation for large evaluation sets. Large evaluation sets are indeed possible: The one used in the experiments of this section accounted for more than 70,000 evaluation pairs considering all teams' sets. The FM procedure can be applied for small evaluation sets using modern computers if it is desirable to obtain paths with fewer hard turns than those provided by the VG-EE and NM-EE procedures.

Table 6.3 presents the measured times for some steps of Algorithm 2. The times for the *Floor model creation* step are omitted in Table 6.3 as they are equal to those shown in Table 6.2. The *Single model integration* step is performed only once. The NM-SM runs significantly faster than the VG-SM in this step because the floor graphs produced by NM are significantly less complex than those produced by VG. The times for the *Endpoints correction* step are smaller in Table 6.3 than those presented in Table 6.2. Endpoints requiring correction may get repeated in the *Endpoints*

Table 6.3: Measured computation times for steps from Algorithm 2. The steps are *Single model integration* (SMS), *Endpoints correction* (ECS) and *Paths computation* (PCS). SMS reports total time, ECS reports mean time per endpoint, and PCS reports mean time per computed path.

Step	VG-SM	NM-SM
SMS Σ (s)	32.21	0.94
ECS μ (s)	1.0×10^{-4}	1.2×10^{-4}
PCS $\mu \pm \varphi$ (s)	0.22 ± 0.01	0.02 ± 0.00

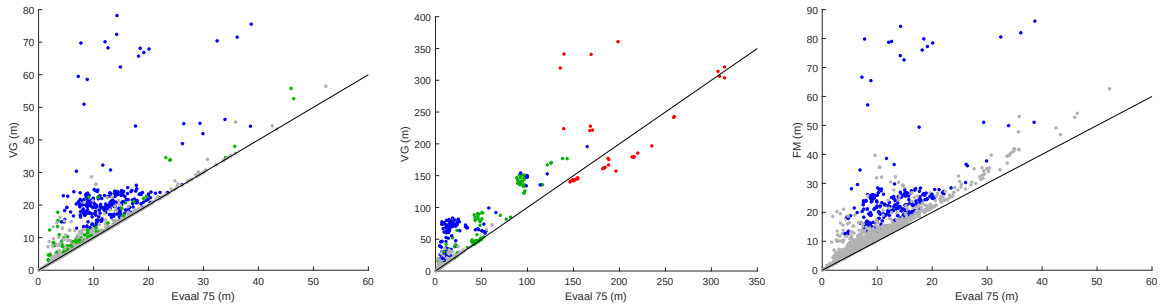
expansion step of Algorithm 1, thus driving up the mean times. The times of *Paths computation* step are similar between Table 6.2 and Table 6.3.

Table 6.4: Complexity of the environment representation. The numbers of nodes and arcs are the totals resulting from summing up the respective numbers from the graph of each floor from each building.

VG	2.3×10^4 nodes – 3.2×10^5 arcs
NM	2.3×10^4 nodes – 7.7×10^4 arcs
FM	13 floor images of $2,971 \times 2,101$ pixels

The times presented for the *Paths computation* steps both in Table 6.3 and Table 6.2 can be directly related to the complexity of the environment upon which the 2D pathfinding methods is applied. As seen in Section 6.2.1, VG produces graph representations that are more complex than those obtained using NM. Table 6.4 shows the complexity of the graphs produced by VG and NM. Despite the number of nodes is roughly the same, the number of arcs for VG is four times larger than for NM. The FM method computes the paths directly from image transformation whose sizes match those of the original images of the environment. The images used in the experiments are large. Each of them contained over 6 million pixels, which is a result of large and complex environment representations.

The evaluation time, while important, matters only for time affordability, and thus it is a secondary aspect of the proposed procedures. The main aspect is the error magnitudes as measured by the procedures. To set those magnitudes in the context of IPS evaluation, Figure 6.7 presents comparisons between the errors as measured by the VG or FM procedures and as measured by the EvAAL procedure. The EvAAL procedure determined error magnitudes as the 2D Euclidean distance between the positions of the ground truth and estimated positions, with penalties added for floor error and building error of 4 m and 50 m, respectively:



(a) RTLSUM - VG (b) MOSAIC - VG (c) RTLSUM - FM

Figure 6.7: Examples of error measurement differences. Green and gray dots represent cases for which ground truths and estimations were in the same floor, with or without applying corrections, respectively. Blue and red dots represent cases of floor and building miss-identification, respectively.

$$\begin{aligned} error = & distance_{lat,lon}(estimated, real) \\ & + 4 * abs(estimated_{floor} - real_{floor}) \\ & + 50 * (estimated_{building} != real_{building}) \end{aligned} \tag{6.2}$$

The figure provides color codes that allow the identification of specific types of paths used for error measurement. Positions correctly estimated in the target floor and building by an IPS are identified by gray and green dots and tend to have EvAAL error magnitudes similar to those from the proposed procedures. Notice that gray and green dots lie near the diagonal reference lines in the plots. The green dots additionally identify cases for which the position estimates had to be corrected because it lied outside the evaluation area, i.e., outside a building or inside an obstacle. The green dots never appear in the plots lying over the reference lines. Notice that Figure 6.7c does not have green dots, which is a result of the fact that the FM procedure does not make estimate corrections before computing the paths. The dots that represent larger differences in the error magnitudes between VG or FM procedures and the EvAAL procedure are those in blue or red colors. The blue dots represent paths that accounted for a floor, but not building, miss-identification, with or without applied corrections.

The different-floor same-building cases always reflect an error underestimation for the EvAAL procedure. The error underestimation is large for some samples. The 4 m floor penalty of the EvAAL procedure approximately reflected the height difference between two adjacent floors. However, 4 m is a value too small to consider the displacements required for an actual subject to change floors. The red dots, only

shown in Figure 6.7b, represent building miss-identifications with or without floor miss-identification or estimation correction. The position estimates with building miss-identifications have the largest errors, and they were only provided by the MOSAIC team. The EvAAL procedure heavily underestimated the error magnitudes for some estimations, while for others it overestimated their magnitudes. The 50 m building penalty was too large for cases where the estimations and the ground truth lied in nearby positions at the ground floors of adjacent buildings. Also, such penalty was far too small for other cases that required changing floors at both buildings, as previously shown by Figure 6.6. In general, when compared to pathfinding-based alternatives, the EvAAL procedure mostly underestimated the magnitude of the error for the evaluated environment and estimation sets.

Table 6.5: Team evaluation results, using best set of estimations according to the mean value of EvAAL error.

Team	EvAAL Mean (m)	VG Mean (m)	NM Mean (m)	FM Mean (m)
RTLSUM	6.20	7.24	7.25	8.28
ICSL	7.67	9.61	9.63	10.96
HFTS	8.49	9.48	9.49	10.80
MOSAIC	11.64	13.25	13.28	15.16

The magnitude of the error is the main evaluation aspect of an IPS. A metric like the mean or the 75^{th} percentile is commonly applied over the set of determined magnitudes to create a rank of the competing IPS. Recall that, in the case of the Track 3 of 2015 IPIN competition, each competing team provided several estimation sets. For each set, its mean error value was computed. The lowest mean value of a team was then used to rank the team. The values in Table 6.5 were computed using only the estimation set of each team that provided the best mean value of the EvAAL errors. Thus, values from distinct procedures can be compared. Notice that the table does not make the distinction between the variants EE or SM given that they produce the same measurement results either for VG or NM procedures, which is a result of using the best distance as error magnitude quantity.

According to Table 6.5, the team HFTS consistently performs better than the ICSL in the VG, NM and FM procedures, thus ranking 2nd if a new ranking was possible. The team HFTS had a floor detection rate notably better than the ICSL team [35]. As previously seen, the EvAAL procedure mostly underestimated the error in comparison to the proposed procedures. Furthermore, EvAAL error underestimation was notable for many estimates that miss-identified the floor. The MOSAIC and RTLSUM teams would keep ranking last and first, respectively, using any of the addressed procedures. For all teams, the value obtained using the EvAAL procedure

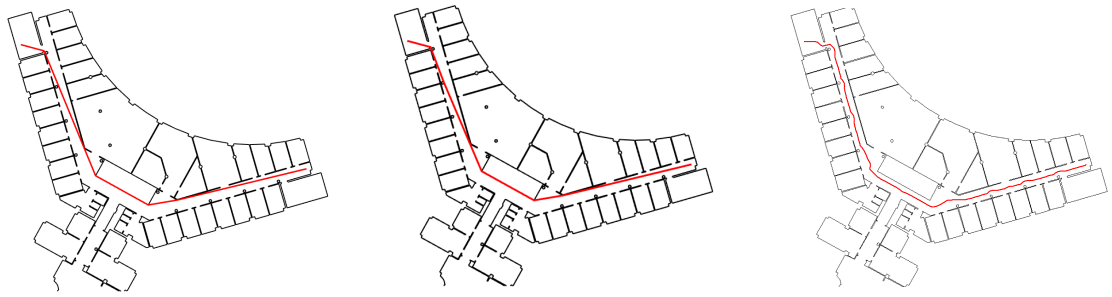
was lower than the one obtained using the VG procedure. In turn, the value obtained using the VG procedure was slightly lower than the one obtained using NM. The procedure FM produced the largest values in all cases.

Table 6.6: Team evaluation results, using best set of estimations according to the 75^{th} percentile of the EvAAL error.

Team	EvAAL 75^{th} (m)	VG 75^{th} (m)	NM 75^{th} (m)	FM 75^{th} (m)
RTLSUM	8.34	8.96	8.96	10.46
ICSL	10.87	13.26	13.26	14.50
HFTS	11.60	12.57	12.57	14.08
MOSAIC	10.65	12.16	12.18	14.25

To assess the teams' ranking using a new metric, the values in Table 6.6 were computed using only the estimation set of each team that provided the best 75^{th} percentile value of the EvAAL errors. The 75^{th} percentile has been used in competitions after the 2015 IPIN competition [37, 36]. The usage of 75^{th} percentile drastically changed the ranking of teams, even for the EvAAL procedure. Notice that the 75^{th} percentile values are lower than the mean values for the MOSAIC team. The team MOSAIC has a few very-large positioning errors, as a consequence of building miss-identification, which severely affected the mean positioning error. It was the only team to have such errors. According to the 75^{th} percentile, the team MOSAIC would rank second for the EvAAL, VG and NM procedures. Table 6.6 supports the order that Table 6.5 showed for teams HFTS and ICSL and the proposed procedures. The team HFTS performed better than the team ICSL according to the 75^{th} percentile and the VG, NM, and FM procedures. The RTLSUM would keep in any case the top place in a ranking. The team provided remarkably good position estimates. The closer the position estimations to the ground truth, the lesser the chances that the path that joins them has to avoid obstacles.

Apart from the change in the ranking induced by using a particular error measurement procedure or metric shown in Table 6.5 and Table 6.6, it is notable the similarity between the metric values provided by the VG and NM procedures. Figure 6.8 exemplifies the reason between this similarity. Notice that the path provided by the VG procedure is almost identical to that provided by the NM procedure. The path initially found by the NM method is later smoothed using line-of-sight testing. The smoothing removes non-essential vertices, often creating visibility edges as those used in the VG method. The FM method, instead, actively avoids boundaries, which create smooth but sinuous paths that divert from the shortest distance paths found by the VG method or their approximation provided by the NM method.



(a) Visibility Graphs (b) Navigation Meshes (c) Fast Marching

Figure 6.8: Examples of computed paths using the three proposed 2D pathfinding methods.

Having the building and floor correctly predicted is of utmost importance for an IPS. The proposed measurement procedures can correctly penalize floor and building miss-identifications, and thus avoid over or underestimations. The NM method produces path estimations that are a good approximation to the shortest distance paths while also keeping a moderate consumption of the computing resources.

The proposed measurement procedures focused on error measurement at indoors. The indoor character influenced aspects of procedures, like the endpoint correction. In general, a positioning system that has building (map) information should validate the position estimations to avoid non-accessible areas like, for example, obstacles. A position estimate should be corrected by, e.g., computing the position in the valid areas which is the closest to the original estimate. If the estimates should strictly lie indoors, then outside areas can be considered non-accessible and the correction procedure should be applied. However, there are local positioning systems that work smoothly between indoor and outdoor areas, and no correction is required for them.

6.3 Local Effects of Position Estimations

Section 6.2.1 showed that the 75^{th} percentile and the mean values may provide distinct rankings of competing IPS. Thus, the selection of the accuracy metric for evaluation is very important. One popular option for characterizing the accuracy of an IPS is a CDF plot. CDF plots are very useful and recommended [32] for determining the variability of the positioning errors of an IPS. Also, they allow an easy determination of the number, or percentage, of position measurements that are below a given error. However, the CDF and related metrics like the mean, percentiles and ranges are broad descriptions for a large environment. It is difficult to directly translate that broad description to a smaller area given that IPS are commonly affected by the (local)

environment characteristics.

The *ISO/IEC 18305:2016* International Standard [21] defines two metrics to measure IPS accuracy at a more local level than that provided by the CDF plot: the *circular error 95* and the *coverage*. The circular error 95 indicates the radius of the circle that, given a center position, contains 95% of position estimates. Instead of considering a whole environment, it can be applied individually to separate areas or individually to each test point if they contain several samples. The coverage metric indicates the ratio of evaluation points that meet a minimum performance requirement. The performance requirement is met for a test point having several samples if the maximum value of the error measurements for that point is below or equal to some non-negative value defined for the IPS.

This section addresses local analyses of errors measured for IPS. The IPS are those also used in Section 6.2.2. For the analyses, fingerprints and position estimates corresponding to some target reference points in the test set were selected. The targeted reference points are those where 5 or more fingerprints were collected. The targeted reference points accounted for 654 out of the 723 unique positions found in the private test set of the UJIIndoorLoc dataset. The circular error 95 was not computed given that the amount of samples per reference point may be as low as 5 samples, i.e., a selection for 95% of estimates would mean a selection for the 100% of estimates. The computation of the coverage metric followed an adaptation from the definition of the *ISO/IEC 18305:2016* to consider the black box, off-line evaluation procedure followed in Section 6.2.2.

The number of reference points that have exactly five samples account for less than half of the total reference points, as presented in Figure 6.9. Thus, instead of using exactly five samples per point, five or more were used to increase the number of reference points available for the analysis and to avoid sampling randomness. The metric and measurement procedure used here are the 75^{th} percentile and the EvAAL, respectively. The EvAAL was chosen because from the measurement procedures addressed in Section 6.2.2 it is the one that provides the most optimistic perspective. The minimum performance requirement was chosen to be 10 m because it is the round number closest to the 75^{th} percentile of the EvAAL-measured errors for all teams. Only one of the estimation sets that each team submitted was used. The selected set was the one used for ranking in the Track 3 of the 2015 IPIN competition.

Table 6.7 shows the results of computing the coverage metric for each team. In general, the coverage indicates the presence of large errors in about half of the reference points for most teams. The team RTLSUM is appreciatively the best performing according to coverage. The team HFTS performed better than the ICSL, which is in line with the analyses presented in Section 6.2.2. The visualization of the coverage metric results can provide insights into the areas which insufficient IPS performance.

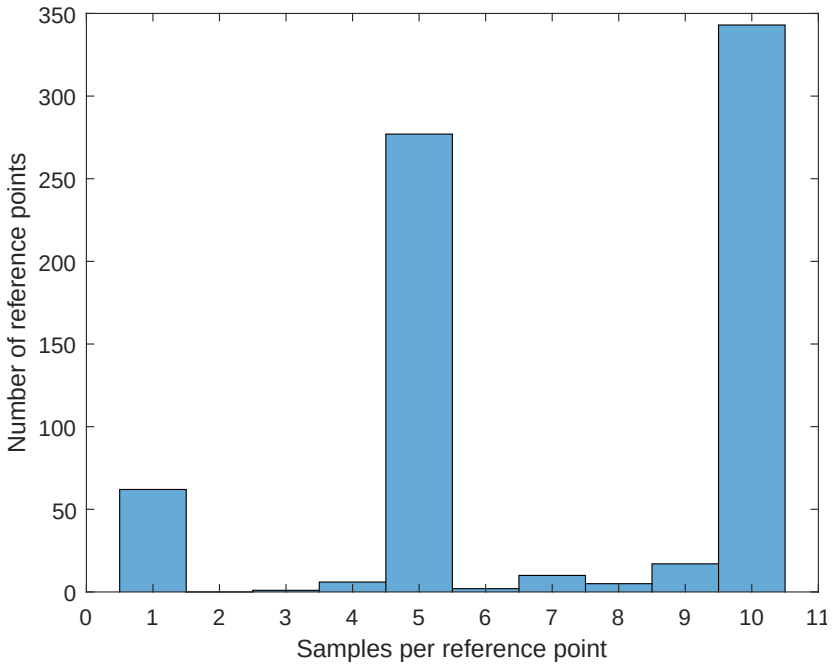


Figure 6.9: Histogram of samples per reference point in the private test set of the UJIIndoorLoc dataset.

Such visualization, despite it is recommended, is not provided here to maintain the private character of the UJIIndoorLoc's test set.

Table 6.7: Coverage metric adapted to the black box, off-line evaluation, considering 10 m as the minimum performance requirement.

RTLSUM	ICSL	HFTS	MOSAIC
0.62	0.50	0.58	0.46

The coverage metric exposed that a notable number of reference points have large errors. Such fact is significant given that almost 75% of errors should be lower than the chosen value for the minimum performance requirement. Thus, uncertainty in positioning accuracy should exist for many reference points. The uncertainty in measurements is sometimes addressed using sensibility analysis. WiFi fingerprinting models have a large number of input variables. Also, for large environments, some of the variables may strongly influence the position measurement in an area while being unimportant for other areas. Thus, the sensibility analysis should be done at a local scale, i.e., individually at small areas or each reference point. However, the black box, offline evaluation approach addressed in this section imposes hard constraints

upon the IPS responses for the input variables. The addressed evaluation approach rules out the possibility of getting new position estimates for selected variations in the fingerprints.

Thus, instead of performing a sensibility analysis, the relation between the variation of error magnitudes and the variation of the RSS was studied. The variation of error magnitudes was computed as the standard deviation of errors at each reference point. The variation of the RSS (input signals) was computed as the maximum value of the distances among the fingerprints from each reference point. The chosen distance between two fingerprints was the Euclidean distance in the fingerprint space, which is a measure that is commonly applied as a similarity metric for WiFi or BLE deterministic fingerprinting. The maximum fingerprint distance at each reference point ranged from 4.24 dBm to 163.8 dBm. The latter value is indicative of large variability in the input signals. The standard deviation of errors at each reference point ranged from 0 to 224.57 meters. The former value suggest a consistent estimation while the latter is indicative of very large and very low errors provided for the same reference point. Table 6.8 shows the *rho* values of the Pearson correlation test applied to the previous two measures. The p-values were not shown as they were significant ($p < 0.05$) for the team RTLSUM and very significant ($p \ll 0.05$) for the other teams.

Table 6.8: Pearson correlation rho values between standard deviation of errors and maximum fingerprint distance.

RTLSUM	ICSL	HFTS	MOSAIC
0.16	0.37	0.24	0.16

The correlation between the standard deviation of errors and maximum fingerprint distance is always positive, i.e., the larger the fingerprint distances, the larger the error standard deviations. Also, despite the correlation is weak for the RTLSUM and MOSAIC teams, it is notable for the ICSL team. The ICSL team has a low floor detection rate [35]. A large correlation may hint at a large susceptibility of a given IPS to variations in the input signals.

To further study the effect of variations of fingerprint distances on the position estimates, the standard deviation of error magnitudes was computed and later presented in Table 6.9 for two cases. The first case was called “All RPs”. In this case, the standard deviation of error magnitudes of all reference points was considered. Table 6.9 presents the median of those values. The second case was called “Sel RPs”. In this case, only the values from selected references points are considered. The selection is performed attending to maximum fingerprint distance computed for each reference point. A reference point was selected if its maximum fingerprint distance

Table 6.9: Standard deviation of error magnitudes when considering all reference points (All RP) and when considering only those with large – above the median – standard deviation values of fingerprint distance (Sel RP).

Selection	RTLSUM	ICSL	HFTS	MOSAIC
All RPs	1.46	1.95	0.97	0.78
Sel RPs	1.67	2.43	1.19	1.87

was above a threshold value. The threshold was set as the median of the maximum fingerprint distances of all reference points. The values presented for the case "Sel RPs" are consistently larger than those for the case "All RPs". The increase is particularly notable for the MOSAIC team, having increases larger than twice of those of the "All RPs" case. A likely reason for the MOSAIC team's notable increase is that the selected reference points should be those where building miss-identification occurred, which are the ones that have the largest estimation errors. The second most notable increase corresponds to the ICSL team, which is the team that showed a larger correlation in Table 6.8. The team with the least increase is RTLSUM, which had the smallest correlation in Table 6.8 and which was the one that provided best overall results across the error accuracy and coverage metrics.

The results from Table 6.8 and Table 6.9 suggest that the IPS from RTLSUM team may be more robust to nominal changes in the input signals, i.e., changes related to short-term signal variations resulting from multipath, device movement, or collection orientation, for example. Also, the results support a notion that is known for WiFi or BLE fingerprinting positioning: estimates corresponding to the same ground truth usually jump around its positions. Also, the results indicate that the instability of those jumps – random error and not bias – is related to distances among the operational fingerprints.

Reducing the magnitude of errors leads an IPS to perform better in any of the analyses explored in this section, as showed for the RTLSUM team. However, general metrics, like percentiles or their related CDF plot may hide local behaviors and are insufficient for analysing the robustness of a system. The *ISO/IEC 18305:2016* defines several conditions on the collection of test points for IPS evaluation, including for example the number or reference points. However, none of those recommendations can assure to cover the whole spectrum of nominal cases of signal readings, at least for fingerprinting-based systems. Therefore, further research efforts on IPS evaluation may dig into conditions on the evaluation data like, e.g., signal variations bounds applicable for a given environment and application. A dataset meeting such conditions may be used for evaluation instead of lots of other test data.

6.4 Conclusions

This chapter addressed the evaluation of IPS. The evaluation was first explained through insights gained from competition challenges. Furthermore, the chapter described a measurement procedure of positioning errors. The procedure proposed the measurement of errors as the length of pedestrian paths. The application of the procedure to several IPS in a large environment was also provided. Finally, the chapter presented local-level analysis of positioning errors that studied the variance of position estimates provided for the same test point and its behavior across an environment.

Some materials presented in this chapter are supported by one publication [25]. Other materials from this chapter were used as contributions in other publications [31, 37, 36].

References

[1] Zeyad Abd Algfoor, Mohd Shahrizal Sunar, and Hoshang Kolivand. "A comprehensive study on pathfinding techniques for robotics and video games". In: *International Journal of Computer Games Technology* 2015 (2015), p. 7.

[2] Mauri Benedito-Bordonau et al. "UJI Smart Campus: Un ejemplo de integracion de recursos en la Universitat Jaume I de Castello". In: *IV JIIDE*. Toledo, Spain, 2013, pp. 417–426.

[3] Harry Blum et al. "A transformation for extracting new descriptors of shape". In: *Models for the perception of speech and visual form* 19.5 (1967), pp. 362–380.

[4] Adi Botea, Martin Muller, and Jonathan Schaeffer. "Near optimal hierarchical path-finding". In: *Journal of game development* 1.1 (2004), pp. 7–28.

[5] Heinrich Bruns. *Das eikonal*. Vol. 35. S. Hirzel, 1895.

[6] Xiaorui Chen and Sara McMains. "Polygon offsetting by computing winding numbers". In: *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers. 2005, pp. 565–575.

[7] Xiao Cui and Hao Shi. "A*-based pathfinding in modern computer games". In: *International Journal of Computer Science and Network Security* 11.1 (2011), pp. 125–130.

[8] Boris Delaunay et al. "Sur la sphere vide". In: *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* 7.793-800 (1934), pp. 1–2.

[9] E. W. Dijkstra. "A note on two problems in connexion with graphs". In: *Numerische Mathematik* 1.1 (1959), pp. 269–271. ISSN: 0945-3245. DOI: 10.1007/BF01386390.

[10] Esri. *ArcGIS Indoors*. https://pro.arcgis.com/en/pro-app/help/data/indoors/get-started-with-arcgis-indoors.htm. Accessed on 6 August 2019.

[11] Raphael A. Finkel and Jon Louis Bentley. "Quad trees a data structure for retrieval on composite keys". In: *Acta informatica* 4.1 (1974), pp. 1–9.

[12] Roland Geraerts. "Planning short paths with clearance using explicit corridors". In: *2010 IEEE International Conference on Robotics and Automation*. IEEE. 2010, pp. 1997–2004.

[13] Joachim Giesen et al. "The scale axis transform". In: *Proceedings of the twenty-fifth annual symposium on Computational geometry*. ACM. 2009, pp. 106–115.

[14] Daniel Felipe Gonzalez Obando et al. "Vector-Based Morphological Operations on Polygons Using Straight Skeletons for Digital Pathology". In: *Discrete Geometry for Computer Imagery*. Ed. by Michel Couprie et al. Cham: Springer International Publishing, 2019, pp. 249–261. ISBN: 978-3-030-14085-4.

[15] Ross Graham, Hugh McCabe, and Stephen Sheridan. "Pathfinding in computer games". In: *The ITB Journal* 4.2 (2003), p. 6.

[16] LOPSI Research Group. *The GetSensorData Android App for Registering All Sensor Stream in a Smartphone*. https://lopsi.weebly.com/downloads.html. Accessed on 6 August 2019.

[17] Peter E Hart, Nils J Nilsson, and Bertram Raphael. "A formal basis for the heuristic determination of minimum cost paths". In: *IEEE transactions on Systems Science and Cybernetics* 4.2 (1968), pp. 100–107.

[18] M. S. Hassouna and A. A. Farag. "MultiStencils Fast Marching Methods: A Highly Accurate Solution to the Eikonal Equation on Cartesian Domains". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.9 (Sept. 2007), pp. 1563–1574. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1154.

[19] Noelia Hernández et al. "Continuous Space Estimation: Increasing WiFi-Based Indoor Localization Resolution without Increasing the Site-Survey Effort". In: *Sensors* 17.1 (2017). ISSN: 1424-8220. DOI: 10.3390/s17010147. URL: https://www.mdpi.com/1424-8220/17/1/147.

[20] Julian Hirt et al. "Using quadtrees for realtime pathfinding in indoor environments". In: *International Conference on Research and Education in Robotics*. Springer. 2010, pp. 72–78.

[21] ISO Central Secretary. *Information technology – Real time locating systems – Test and evaluation of localization and tracking systems*. en. Standard ISO/IEC 18305:2016. Geneva, CH: International Organization for Standardization, 2016. URL: https://www.iso.org/standard/62090.html.

[22] Dirk-Jan Kroon. *Accurate Fast Marching: Multistencils second order Fast Marching 2D and 3D*. https://www.mathworks.com/matlabcentral/fileexchange/24531-accurate-fast-marching. Accessed on 6 August 2019. 2009.

[23] Maximilian Langer, Aysylu Gabdulkhakova, and Walter G. Kropatsch. "Non-centered Voronoi Skeletons". In: *Discrete Geometry for Computer Imagery*. Ed. by Michel Couprie et al. Cham: Springer International Publishing, 2019, pp. 355–366. ISBN: 978-3-030-14085-4.

[24] J.C. Latombe. *Robot Motion Planning*. The Springer International Series in Engineering and Computer Science. Springer US, 1991. ISBN: 9780792391296.

[25] G. M. Mendoza-Silva, J. Torres-Sospedra, and J. Huerta. "A more realistic error distance calculation for indoor positioning systems accuracy evaluation". In: *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sept. 2017, pp. 1–8. DOI: 10.1109/IPIN.2017.8115950.

[26] Raul Montoliu et al. "IndoorLoc platform: A public repository for comparing and evaluating indoor positioning systems". In: *2017 International conference on indoor positioning and indoor navigation (IPIN)*. IEEE. 2017, pp. 1–8.

[27] Alex Nash et al. "Theta^*: Any-angle path planning on grids". In: *AAAI*. Vol. 7. 2007, pp. 1177–1183.

[28] K. J. Obermeyer and Contributors. *The VisiLibity Library*. http://www.VisiLibity.org. 2008.

[29] Yueyong Pang et al. "Extracting Indoor Space Information in Complex Building Environments". In: *ISPRS International Journal of Geo-Information* 7.8 (2018), p. 321.

[30] Nuria Pelechano and Carlos Fuentes. "Hierarchical path-finding for Navigation Meshes (HNA*)". In: *Computers & Graphics* 59 (2016), pp. 68–78. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2016.05.023. URL: http://www.sciencedirect.com/science/article/pii/S0097849316300668.

[31] Francesco Potortì et al. "Comparing the Performance of Indoor Localization Systems through the EvAAL Framework". In: *Sensors* 17.10 (2017). ISSN: 1424-8220. DOI: 10.3390/s17102327. URL: https://www.mdpi.com/1424-8220/17/10/2327.

[32] F. Potortì et al. "Evaluation of Indoor Localisation Systems: Comments on the ISO/IEC 18305 Standard". In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sept. 2018, pp. 1–7. DOI: 10.1109/IPIN.2018.8533710.

[33] James A Sethian. "A fast marching level set method for monotonically advancing fronts". In: *Proceedings of the National Academy of Sciences* 93.4 (1996), pp. 1591–1595.

[34] Greg Snook. "Simplified 3D movement and pathfinding using navigation meshes". In: *Game programming gems* 1.1 (2000), pp. 288–304.

[35] Joaquín Torres-Sospedra et al. "A realistic evaluation of indoor positioning systems based on Wi-Fi fingerprinting: The 2015 EvAAL–ETRI competition". In: *Journal of Ambient Intelligence and Smart Environments* 9.2 (2017), pp. 263–279.

[36] Joaquín Torres-Sospedra et al. "Off-Line Evaluation of Mobile-Centric Indoor Positioning Systems: The Experiences from the 2017 IPIN Competition". In: *Sensors* 18.2 (2018). ISSN: 1424-8220. DOI: 10.3390/s18020487. URL: https://www.mdpi.com/1424-8220/18/2/487.

[37] Joaquín Torres-Sospedra et al. "The Smartphone-Based Offline Indoor Location Competition at IPIN 2016: Analysis and Future Work". In: *Sensors* 17.3 (2017). ISSN: 1424-8220. DOI: 10.3390/s17030557. URL: https://www.mdpi.com/1424-8220/17/3/557.

[38] J Torres-Sospedra et al. "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems". In: *International Conference on Indoor Positioning and Indoor Navigation*. 2014, pp. 261–270. DOI: 10.1109/IPIN.2014.7275492.

[39] Paul Tozour and IS Austin. "Building a near-optimal navigation mesh". In: *AI game programming wisdom* 1 (2002), pp. 298–304.

[40] A. Valero-Gomez et al. "The Path to Efficiency: Fast Marching Method for Safer, More Efficient Mobile Robot Trajectories". In: *IEEE Robotics Automation Magazine* 20.4 (Dec. 2013), pp. 111–120. ISSN: 1070-9932. DOI: 10.1109/MRA.2013.2248309.

[41] Wouter Van Toll et al. "A comparative study of navigation meshes". In: *Proceedings of the 9th International Conference on Motion in Games*. ACM. 2016, pp. 91–100.

[42] Qing Xiong et al. "Free multi-floor indoor space extraction from complex 3D building models". In: *Earth Science Informatics* 10.1 (2017), pp. 69–83.

[43] Francisco Zampella, Antonio Ramon Jimenez Ruiz, and Fernando Seco Granja. "Indoor positioning using efficient map matching, RSS measurements, and an improved motion model". In: *IEEE Transactions on Vehicular Technology* 64.4 (2015), pp. 1304–1317.

Chapter 7

Conclusions

A reliable and accurate positioning at indoor environments is a necessity yet to be satisfied. The necessity is created by the success of smartphones and a pervasive desire of as many as possible services provided through them. In many cases, the quality of the provided service is improved if the position of the device, i.e., the position of the user is used to customize the content that the service provides.

The technologies available today to support indoor positioning are insufficient to cover all requirements and environments. Thus, several IPS proposals have mainly focus on one of those technologies to provide solutions that fit some applications or scenarios. The number of proposal for most of the technologies is remarkable. One of the contributions of this dissertation is a meta-review that took IPS surveys published within the last 5 years to explore the relevance of published works. The meta-review was presented in Chapter 2, along with an introduction to key IPS concepts, technologies, techniques, methods and the most-recognized challenges in indoor positioning. The introduction and the meta-review allows a reader unfamiliar with IPS to learn the basics of indoor positioning and have a curated collection to browse from general to more specialized surveys.

Among the findings presented in Chapter 2, it is notable that most of the works referenced in surveys are only cited once, which suggested that only a few works have a significant and acknowledged relevance. A total of 3,901 works (with DOI number) were cited in the context of 61 reviews but only 75 works were cited more than 5 times in the selected surveys, being considered well-known or disruptive works. In some cases, the surveys cite recent papers that do not attract the indoor positioning experts. Finally, the meta-review also shows that a few papers published in early 2000 were disruptive and have had a huge impact on further developments. Most of the relevant works in the surveys are in the period 2010-2016, which match the birth of smartphones as we currently know them and the proliferation of new conferences

related to the topic, like the International Conference on Indoor Positioning and Indoor Navigation (IPIN).

The contribution presented in Chapter 3 is two open datasets intended to foster experimentation and reproducibility of indoor positioning solutions. The first dataset contains 103,584 WiFi fingerprints collected during 25 months in the library building. For the second dataset, three subjects collected 4,700 BLE fingerprints in two environments taking into account three BLE beacons intensity levels and three smartphone models. Chapter 3 presented experiments performed with the datasets to describe how each of them can be used to test specific traits of an IPS. The WiFi dataset allowed testing an IPS robustness to short and long term variations in the environment, while the BLE dataset allowed exploring an IPS robustness to low detection rates of BLE beacons and the effect of beacons positions on the IPS accuracy. The WiFi collected dataset was intensively used for the analyses presented in Chapter 4 and Chapter 5.

Chapter 2 deepened into fingerprinting, one of the star methods used for IPS. Fingerprinting is mainly used for WiFi signals, although it is used also for magnetic and BLE signals. One of the most acknowledges challenges in fingerprinting is the collection of samples to create the radio map, i.e., the signal characterization of the environment. The collection of samples is a process that is not trivial and requires a significant effort. Another contribution of this dissertation is a set of collection methodologies, experiences, and methods that reduce the collection effort. The process of collection involves obtaining or developing the collection tools, deciding how, when and where to conduct the survey, performing the revision of samples, and transforming the collected data. Chapter 4 of this dissertation addressed all the previous steps.

The contributions presented in Chapter 4 included insights gathered thought the development and usage of tools that assisted the collection of WiFi and BLE samples. The tools included smartphone and web applications that allowed a user to plan a collection as a point-based campaign. The experiences showed that proper collection tools help improve the collection time and reduce tagging errors. Also, concerns regarding the reported WiFi or BLE scans that affect the collection were described. Besides the suggestions on how to collect the samples, Chapter 4 contributed with suggestions on where to collect samples for the case of WiFi signals. The suggestions are based on experiments analyzed the effect that the distribution of collection points and the intensity of the APs in the environment have in (1) the accuracy of an IPS and (2) in the quality of a regression that could be applied to enrich the radio map. The suggestions highlighted the importance of situating collection points around the boundaries of the target environment. Also, zones that are close to APs require more points than other zones. The position of an AP was shown to be an important piece of information for the determination of collection point.

Another of the contributions presented in Chapter 4 is one method for determining the position of an AP and another method for determining if an AP is inside the area where WiFi samples were collected. The first method combines interpolation and the Weighted Centroid method to find the position of an AP. The method provided better estimates when compared with others found in the IPS literature. The other method uses the distribution of the AP's intensities seen in the collected samples to determine whether the position of AP can be precisely determined using the Weighted Centroid method. The final contribution presented in Chapter 4 uses the knowledge gained from previous contributions shown in the chapter to propose a regression model that improves the regression quality for APs that are close, and thus strongly seen, to the target environment. The method uses a raw reference position of the AP and raw map information of the obstacles in the environment.

Changes eventually occur for any environment where an IPS is deployed. Those changes may affect the behavior of an IPS, altering its accuracy. Thus, periodic radio map updates are commonly performed to allow the IPS to adapt to those changes. A method that suggests where to perform WiFi radio map updates was another contribution of this dissertation, presented in Chapter 5. The set of positions is chosen so that the newly collected samples are used to train a regression. The regression is then applied to estimate signal measurement for the remaining positions targeted in the environment. The method optimizes the set of collection positions based on a genetic algorithm, finding a fine compromise between a small number of points in the selected set and a good regression quality. Chapter 5 also contributes with a brief characterization of nominality of changes in an environment. The characterization is supported by experimental analyses that showed that large variations in the RSS values are to be expected under normal functioning conditions for an IPS. Also, the non-detection situations were analyzed, showing that they are common for AP weakly detected in an environment and that when APs important for the positioning have drastic changes the similarity between fingerprint is significantly affected.

The contributions of this dissertation presented in Chapter 6 were related to the evaluation of IPS, mainly those based on fingerprinting. The first contribution is the insights gained through experiences from competition challenges based on the off-line evaluation of fingerprinting based IPS. The insights explained the challenges of preparing the data that later allows the evaluation and the challenges in performing a fair assessment and comparison of distinct IPS.

Another contribution presented in Chapter 6 is a measurement procedure of positioning errors. The procedure estimates the error magnitude as the length of a pedestrian path that would connect an estimated position with its associated ground truth position. The procedure uses 2D pathfinding methods to compute paths inside building floors. In case of paths that traverse several floors or buildings, the floor paths are joined using pre-computed vertical and inter-building paths. Chapter 6

presented experiments that tested three 2D pathfinding methods: Visibility Graphs, Navigation Meshes, and Fast Marching. The measurement procedure was applied to position estimates provided by several IPS on a large environment. The results of the experiments showed that usual ways of measuring the positioning error may overestimate and frequently underestimate the magnitude of the error as could be perceived by the IPS users. Also, the results showed that the Navigation Meshes method is a fair way to determine the shortest pedestrian path without incurring in large computational overloads. The final contribution presented by Chapter 6 analyzed the locality of positioning errors associated with an IPS. The experiments show that the usage of the positioning error metric commonly used to compare IPS may mask low accuracy cases of an IPS that are not manifesting across the whole environment. The experiments also hinted the detection of the susceptibility of an IPS to variations in the signals.

7.1 Future Work

The contribution presented in this dissertation can be improved or continued in several ways. The collection tools and methodologies can be extended to continuous – route-based collection. The route-based collection is becoming increasingly popular and has been used to collect the data for IPS evaluation in recent competitions.

Also, the regression model presented in Chapter 4 could be considered a first step towards the definition of more general regression models or method. To the best of this work's knowledge, there is not any interpolation method, regression method, or tool that allow direct modeling of the environment influence of the measured phenomenon. The environment characteristics could be incorporated into the existent methods that consider the spatial relation between measurements.

The path-based error measurement presented in Chapter 6 used an accurate representation of the evaluation environment. Continuation lines of this work may try to obtain such representation in ways more automatic than those used in this dissertation, as well as include their implementations into GIS tools that enable easy usage of them.

7.2 Supporting Publications

The list of publication supporting with this dissertation is the following:

- Mendoza-Silva, G.M., Matey-Sanz, M., Torres-Sospedra, J. and Huerta, J.,

2019. BLE RSS Measurements Dataset for Research on Accurate Indoor Positioning. Data, 4(1), p.12.

- Mendoza-Silva, G., Richter, P., Torres-Sospedra, J., Lohan, E. and Huerta, J., 2018. Long-term WiFi fingerprinting dataset for research on robust indoor positioning. Data, 3(1), p.3.

- Mendoza-Silva, G.M., Torres-Sospedra, J. and Huerta, J., 2018, January. Locations selection for periodic radio map update in wifi fingerprinting. In LBS 2018: 14th International Conference on Location Based Services (pp. 3-24). Springer, Cham.

- Mendoza-Silva, G.M., Torres-Sospedra, J. and Huerta, J., 2017, September. A more realistic error distance calculation for indoor positioning systems accuracy evaluation. In 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 1-8). IEEE.

- Mendoza-Silva, G.M., Torres-Sospedra, J., Huerta, J., Montoliu, R., Benítez, F. and Belmonte, O., 2017. Situation Goodness Method for Weighted Centroid-Based Wi-Fi APs Localization. In Progress in Location-Based Services 2016 (pp. 27-47). Springer, Cham.

- Mendoza-Silva, G. M.; Rodriguez-Pupo, L. E.; Torres-Sospedra, J.; Huerta, J., 2016. Solutions for signal mapping campaigns of Wi-Fi networks. In Proceedings of the VII Jornadas Ibericas de Infraestructuras de Datos Espaciales.

The list of other publications that resulted from collaborations related to this dissertation:

- Torres-Sospedra, J., Belmonte-Fernández, Ó., Mendoza-Silva, G.M., Montoliu, R., Puertas-Cabedo, A., Rodríguez-Pupo, L.E., Trilles, S., Calia, A., Benedito-Bordonau, M. and Huerta, J., 2019. Lessons Learned in Generating Ground Truth for Indoor Positioning Systems Based on Wi-Fi Fingerprinting. In Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation (pp. 45-67). Academic Press.

- Torres-Sospedra, J., Richter, P., Mendoza-Silva, G., Lohan, E.S. and Huerta, J., 2018, September. Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates. In 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 1-8). IEEE.

- Torres-Sospedra, J., Jiménez, A., Moreira, A., Lungenstrass, T., Lu, W.C., Knauth, S., Mendoza-Silva, G., Seco, F., Pérez-Navarro, A., Nicolau, M. and Costa, A., 2018. Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition. Sensors, 18(2), p.487.

- Potortì, F., Park, S., Jiménez Ruiz, A., Barsocchi, P., Girolami, M., Crivello, A., Lee, S., Lim, J., Torres-Sospedra, J., Seco, F. and Montoliu, R., 2017. Comparing the performance of indoor localization systems through the EvAAL framework. Sensors, 17(10), p.2327.

- Torres-Sospedra, J., Jiménez, A., Knauth, S., Moreira, A., Beer, Y., Fetzer, T., Ta, V.C., Montoliu, R., Seco, F., Mendoza-Silva, G. and Belmonte, O., 2017. The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work. Sensors, 17(3), p.557.

- Torres-Sospedra, J., Mendoza-Silva, G.M., Montoliu, R., Belmonte, O., Benitez, F. and Huerta, J., 2016, October. Ensembles of indoor positioning systems based on fingerprinting: Simplifying parameter selection and obtaining robust systems. In 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 1-8). IEEE.

- Torres-Sospedra, J., Montoliu, R., Mendoza-Silva, G.M., Belmonte, O., Rambla, D. and Huerta, J., 2016. Providing databases for different indoor positioning technologies: Pros and cons of magnetic field and Wi-Fi based positioning. Mobile Information Systems, 2016.