

Research Article

Providing Databases for Different Indoor Positioning Technologies: Pros and Cons of Magnetic Field and Wi-Fi Based Positioning

Joaquín Torres-Sospedra, Raul Montoliu, Germán M. Mendoza-Silva, Oscar Belmonte, David Rambla, and Joaquín Huerta

Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellón, Spain

Correspondence should be addressed to Joaquín Torres-Sospedra; jtorres@uji.es

Received 4 February 2016; Accepted 7 April 2016

Academic Editor: Robert Piché

Copyright © 2016 Joaquín Torres-Sospedra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Localization is one of the main pillars for indoor services. However, it is still very difficult for the mobile sensing community to compare state-of-the-art indoor positioning systems due to the scarcity of publicly available databases. To make fair and meaningful comparisons between indoor positioning systems, they must be evaluated in the same situation, or in the same sets of situations. In this paper, two databases are introduced for studying the performance of magnetic field and Wi-Fi fingerprinting based positioning systems in the same environment (i.e., indoor area). The “magnetic” database contains more than 40,000 discrete captures (270 continuous samples), whereas the “Wi-Fi” one contains 1,140 ones. The environment and both databases are fully detailed in this paper. A set of experiments is also presented where two simple but effective baselines have been developed to test the suitability of the databases. Finally, the pros and cons of both types of positioning techniques are discussed in detail.

1. Introduction

Localization, with an expected market to grow to \$4.4 billion in 2019 [1, 2], is essential to support indoor services. Most of the newest applications need the user’s location to customize their services [3–5], monitor people [6], or track Internet-of-Things objects [7], among others. Moreover, location can also be used to detect the user’s activities and provide custom services based on them.

Many different approaches have been used to solve the problem of indoor positioning in the last years. They can be categorized, according to [8], as infrastructure-based and infrastructure-less technologies. The alternatives based on the former category require the deployment of custom beacons and instrumentation to sense the environment and improve indoor positioning accuracy, whereas the systems based on the latter category use “information” already present in the environment. Among all the possible technologies that can be used for positioning, this work focuses on those based on magnetic field and Wi-Fi fingerprinting. The two technologies for positioning are quite different. The former is based on the uniqueness of the disturbances in the magnetic

field produced by the structural elements of a building and tends to be used when the user is moving. The Wi-Fi one is based on the use of fingerprinting techniques and it tends to be more representative when the user is stationary.

Magnetic field based systems and Wi-Fi fingerprinting techniques belong to the infrastructure-less category and they have been attracting the attention of many researchers in the last years due to their low deployment costs. Wi-Fi fingerprinting techniques can only be considered infrastructure-less when they rely on existing Wi-Fi networks designed for communication purposes. When a Wi-Fi network is deployed on purpose to improve positioning, Wi-Fi based positioning should be considered an infrastructure-based system. In contrast, magnetic field based techniques are always considered infrastructure-less since they do not need any external device to support the indoor localization.

Although there are several works dealing with indoor localization, each one establishes its own evaluation procedure. Therefore, the available works cannot be directly compared, even when they use the same indoor positioning technologies, since they use different evaluation metrics, evaluation indoor areas (environments), mapping strategies,

and/or hardware elements, among many others. To perform meaningful comparisons, the evaluation of the indoor positioning systems has to be done using the same methodology.

The publication of databases [9, 10], the organization of open competitions [11–15], and the proliferation of new benchmarking initiatives [16, 17] and standards, such as the ISO/IEC DIS 18305 standard [18], are overcoming the main drawback in the indoor localization research field, which is the lack of a common set of databases and frameworks for meaningful evaluation and comparison of methods. Although all these measures are promoting the fair evaluation of indoor positioning systems, there is still a long way to go to fully cover all possible environments. The databases presented in [9, 10] were presented to address these shortcomings. They can be used to compare Wi-Fi fingerprinting and magnetic field, respectively, based on positioning methods. However, each database covers a different environment and, therefore, it is not possible to compare different technologies in the same environment.

In this paper, two different databases are introduced to allow for the comparison of two different technologies in the same environment. The first one, the *UJIIndoorLoc-Mag* database (<http://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc-Mag>), previously presented in the 2015 IPIN conference [10], is the first publicly available database that can be used to make comparisons among different magnetic field based methods. The second one consists of a set of Wi-Fi fingerprints captured in the same area as in the *UJIIndoorLoc-Mag* database.

This paper extends the work presented in [10] by including a new Wi-Fi fingerprint based dataset in the same environment, new baselines to evaluate the two technologies, a comprehensive comparison, and a discussion about the pros and cons of the two technologies. As far as we know, this is the first contribution where two different technologies for indoor positioning (magnetic field and Wi-Fi signals) are compared in the same environment.

Mobile devices (smartphones and tablets) have been used as positioning devices for all of the experiments presented in this paper. They are becoming an important alternative for positioning, which allow the implementation of pedestrian navigation systems. Moreover, the number of mobile applications using positioning for different purposes is rising. In particular, Android devices have been used since they allow full access to sensors and they dominate the worldwide smartphone OS market with approximately 80–85% of the market share.

The main contribution of this work can be summarized as follows:

- (1) We introduce two databases covering the same environment, one for magnetic field based positioning and the other one for Wi-Fi fingerprinting. Both databases have public access (<http://indoorloc.uji.es>).
- (2) We present a set of new baselines to test the suitability of magnetic field and Wi-Fi based positioning technologies in the same environment.

- (3) We present a comprehensive comparison of the two well-known technologies for indoor positioning, showing the pros and cons of each one.

The remainder of the paper is organized as follows. Section 2 presents the related work. Section 3 provides the analysis and design of the datasets. Section 4 introduces the material (databases) and methods (baselines). Section 5 shows the results and discussion. Finally, Section 6 presents the most important conclusions that have arisen from this work.

2. Related Work

This section introduces some previous works on both technologies for indoor positioning focusing on the kind of datasets that have been used to test the proposals.

2.1. Previous Works on Magnetic Field Based Positioning. There are many papers in the literature dealing with magnetic field based methods for indoor localization problems. Some of them are reviewed in this section [19–25]. We focus on the dataset used for testing the proposed algorithms and we also state whether they are publicly available.

Four experiments were done in [19] to demonstrate the feasibility of using the magnetic field for positioning. In the first one, data were collected at one specific location in six different environments. In the second one, data were collected at five overlapping corridors. In the third one, data were collected in the intersections of two different squared and regular grids. In the last one, magnetic field changes in the vertical direction were studied with 5 cm of resolution. Although the experiments and results were detailed, the details about the databases were not included. Finally, the authors stated that, in some cases, other technologies (such as Wi-Fi based ones) may be utilised to avoid severe errors and to constrain the localization area. The authors concluded that submeter accuracy, and even subdecimetre accuracy, was possible with magnetic-based positioning.

The experiment presented by authors of [21] took place on a rectangular-shaped, $67 \times 12 \text{ m}^2$, corridor whose surroundings included spaces such as lab, office, and library. So they considered an environment of 4 straight corridors, where the distance between parallel corridors was high, 12 m and 67 m. Moreover, data were statically collected with 45 cm intervals and 10 seconds spent in each location. Their training database consisted of 350 samples (approx.) with 5 features, including location (x, y) and magnetometer values in the three axes. However, information about collected data as well as their magnitudes was not described. They compared the nearest neighbour, the particle filter, and the modifier particle filter they proposed. Their system provided the best accuracy and they obtained a mean error of 0.95 m.

In [22], the authors demonstrated that geomagnetic localization performs reasonably well when the three components of the magnetic field— x -, y -, and z -axis—are considered. They tested their positioning system in three different environments: a suburban house, a city centered apartment, and a university lab. Data were collected as the magnetic flux

TABLE 1: Main characteristics of the data from selected papers. *Group* corresponds to the type of database we have identified, *# of 1D spaces* is the number of one-dimensional spaces considered in the database, and *total lineal m* is the length of the 1D spaces considered. *Total area m²* is the total number of 2D spaces considered in the database and the surface is the area they cover. *# of features* is the number of features stored in a single sample, where discrete ones have a single measure (or an average), whereas continuous ones have multiple measures. *# of samples* is the total number of samples included in the database.

Paper	Group	# of 1D spaces	Total lineal m	# of 2D spaces	Total area m ²	# of features	# of samples
[19] Exp. 1	3	—	—	6	—	3* + loc	N/A
[19] Exp. 2	1	5	N/A	—	—	3* × length + loc	5
[19] Exp. 3	2	16 + 12 = 28	34.2 + 3 = 37.2	—	—	3* + loc	64 + 36 = 100****
[19] Exp. 4	2	1	1	—	—	3* + loc	20
[21]	2	4	178	—	—	3* + 2 (x & y)	350
[22] Env. 1	3	—	—	1	14 × 16	3* + loc	14 × 16
[22] Env. 2	3	—	—	1	9 × 12	3* + loc	9 × 12
[22] Env. 3	3	—	—	1	6 × 19	3* + loc	6 × 19
[22] Env. 4	3	—	—	1	3.5 × 3.5	3* + loc	7 × 7
[23]	1	3	108	—	—	3* × length + loc	3
[24] Env. 1	2	1	187	—	—	3* + loc	37200
[24] Env. 2	3	—	—	1	13.8 × 9.9	3* + loc	40800
[24] Env. 3	2	2	80 (approx.)	—	—	3* + loc	12000
[25]	1	1	170 m	—	—	3* + gravity + loc	N/A
Our DB cont.	1	26 × 2	650 (approx.)	—	—	10** × length + m***	270 + 11
Our DB disc.	1	26 × 2	650 (approx.)	—	—	10** + m***	≈40000

In group 1, continuous samples, length corresponds to the number of individual measures taken in a single continuous sample.

*Three components of the magnetometer.

**Three components of the magnetometer + 3 components of the orientation + 3 components of the accelerometer + timestamp.

*** *m* stands for the features describing the coordinates for each corridor or sample.

For each corridor/segment in the trajectory we store the *xy* coordinates of initial and final points and the indexes of the initial and final samples.

****The authors commented that there were 100 datasets (not samples).

density at 1 m spacing. Moreover, they also conducted a magnetic fingerprint test in a $3.5 \times 3.5 \text{ m}^2$ bedroom. However, they did not detail the number of samples. Their experiments reported a global mean accuracy of 1.4 m when the three components of the magnetometer were used; this error decreased to 0.9 m in case of knowing the room where the user was located.

In [23], the authors selected a corridor of a multilevel building to evaluate the performance of using geomagnetic field information for positioning with four different devices. The corridor was about 36 m in length and 2 m wide. Samples were taken along the corridor at three different positions: (1) centered position, (2) 60 cm left to the corridor center, and (3) 60 cm right to the corridor center. Altogether 20 points were used for testing purposes. Their environment was narrow and realistic, because three different parallel paths in a 2 m wide corridor composed it. They obtained, in some cases, errors of 0.6 m.

An indoor location system based on a wearable device was successfully introduced in [24]. The system is tested in two very different environments, a 187 m corridor loop environment (37200 training samples and 310 test data points) and an atrium environment (40800 training samples and 408 test data points). They also examined the fingerprint difference between floors using a dataset with 60 points from each floor. They used a special device with four magnetometer sensors for sampling the magnetic fingerprints, so vectors consisted of 12 elements. They reported that the accuracy of

their system was 4.7 m, the median was 0.71 m, and the 90th percentile was 1.64 m. They also stated that they could achieve 0.45 m accuracy by combining magnetic and Wi-Fi received signal strength (RSS) methods (i.e., fingerprinting).

GIPSy, a positioning system that provided a median error of 2.1 m, was introduced in [25]. The main feature of this system is that the magnetic values are transformed to make them orientation independent using the gravity force values. The tests were done on a single floor of the Bahen Center for Information Technology (University of Toronto, Canada). Although a single path of about 170 m was mapped several times under different conditions, details about data were not provided.

Table 1 summarizes the databases used in the previous reviewed works [19, 21–25]. We have identified three different types of databases (groups 1, 2, and 3 in the table) according to how samples were taken: (1) continuous samples taken in a lineal environment (such as a corridor), (2) discrete samples taken in a lineal environment, and (3) discrete samples taken in a two-dimensional space. Please note that a single continuous sample corresponds to a sequence of consecutive discrete samples taken in a lineal environment. We also include the data about our database with the continuous and discrete versions.

The soundness of results and conclusions presented in all of these contributions is high, but the databases employed were not totally detailed and their access was restricted (not public) in all studied cases. For instance, the information

about how locations are stored is not always provided. This information is described in some works (such as [21]), but it is omitted in the majority of contributions (such as in [19, 22–25]). To denote that this information was not provided, we used `loc` in Table 1.

Although the number of continuous samples used in the experiments seems to be low, 5 in [19] and 3 in [23], the length of the vectors was high enough to perform the experiments. However, our database contains more information than theirs and it includes 270 continuous samples (35,779 discrete samples) for training and 11 complex continuous samples (4,380 discrete samples) for testing. In our case we consider not only corridors but also combinations of two connected corridors (turns changing corridor). In some works, information about the gathered data is not described [25].

2.2. Previous Works on Wi-Fi Based Positioning. There are many papers in the literature dealing with Wi-Fi based methods for indoor localization problems. Some of them are reviewed in this section [26–30]. Similarly to Section 2.1, we focus on the dataset used for testing the proposed algorithms and we also state whether or not they are publicly available.

The experiments introduced in [26] introduced RADAR, which was the pioneer Wi-Fi RSS based indoor positioning system. Some important analyses were performed in [26]: the impact of the number of data points and number of samples, the importance of user orientation, and the problem of tracking a mobile user, among others. The experimental test bed was located on the second floor of a 3-storey building. Although this environment covered an area of 980 m² (43.5 m × 22.5 m) and included more than 50 rooms, the experiments were conducted only at the corridors. The experiments relied on an infrastructure-based topology where 3 base stations (WAPs) were deployed. The mobile host was a laptop computer. At least 20 different fingerprints were taken at 70 reference points using four different user orientations. The authors concluded that a median error between 2 and 3 m was feasible with this technology.

The experiments presented by authors of [27] introduced EZ, which was based on a genetic algorithm devoted to performing indoor positioning and avoiding explicit pre-deployment effort. Four experiments were conducted in two different environments, a small one (27 × 18 m, 486 m²) and a large one (140 × 90 m, 12600 m²). To create the reference data for EZ and two well-known indoor positioning systems (RADAR [29] and HORUS [31]), they collected the fingerprints at grid locations with 1.5 m separation for the small environment and 3 m separation for the large environment. They gathered 10,000 measurements per location during approximately 5 minutes. For the EZ system they introduced, they gathered data at 48 + 3 points (small environments) and 101 + 15 points (large environment). The user walked through the environment and stood for approximately 3 seconds at each location. They report that their system gave a median error of 2 m, for the small environment, and 7 m for the large environment. RADAR provided better results with median errors of 1.3 m and 5 m in both testing environments.

The experiments presented by authors of [28] were carried out in the Tietotalo building at Tampere University of Technology (10,000 m² approximately). The reference data were collected in 96 reference points and 30 RSS measurements were taken. To eliminate the fading effect, they computed the mean of the 30 measurements in their IPS. A total of 206 WAPs were detected in the training phase, where a Nokia N900 smartphone was the device used to collect data. In the operation phase, data was collected in 43 testing points with a Nokia N900 and, two weeks later, a laptop. Again 30 RSS measurements were taken with the Nokia N900, and 20 RSS measurements were taken with the laptop. Their system reported a median error of 4 m.

The experiments performed in [29] considered multi-building and multifloor positioning. The experiments were performed in two three-storey buildings at the University of Minho, Portugal. The data for the calibration and operational phases was collected using a laptop computer equipped with three network interfaces; so each observation collected multiple fingerprints at three different heights. For the training phase, three reference points were selected for each room. As a result, a total of 392 calibration points were established in the whole environment and 9,358 calibration samples were taken. For testing purposes, 472 uniformly distributed points were selected and 3 samples were collected. The system reported an average error of 3.35 m, a room detection rate of 74.1%, a floor detection rate of 99.5%, and a building detection rate of 100%.

The experiments done by authors of [30] followed the comprehensive benchmarking methodology developed in the EVARILOS Project [16, 17]. They deployed custom WAPs for localization and used a MacBook Pro notebook as a client's device. They performed the experiments in four different environments: a small office, a medium lab, a big office, and a big open space. In the experiments, each training point consisted of 40 RSSI scans. For the small office they collected the reference fingerprints at 41 locations, and 20 fingerprints were collected for testing the algorithm. For the medium lab, they collected 56 reference and 20 testing fingerprints. For the big sized office, they collected 123 reference and 60 testing fingerprints. For the last open space, they collected the reference fingerprints at 100 locations and testing fingerprints at 27 equally distributed locations. They reported an error of about 2 m in the small office, medium lab, and big sized office, whereas an error of about 7 m was reported for the challenging big open space.

Table 2 summarizes the databases used in the previous reviewed works [26–30]. Moreover, the *UJIIndoorLoc* database [10], which was used in the 2015 EvAAL-ETRI competition [15] at the IPIN conference, is also included in Table 2. We have identified two main groups according to the infrastructure used for positioning: infrastructure-based (they deployed custom elements to support positioning) and infrastructure-less (to support positioning, they used the already deployed antennas for connectivity). We also show the number of WAPs deployed/used for positioning, the covered area, the number of reference points, and the number

TABLE 2: Main characteristics of the data from selected papers. *Infrastructure* corresponds to the kind of infrastructure required to support the localization (free means no explicit infrastructure), *# of WAPs* is the number of Wireless Access Points detected/considered/required for positioning, *Area* is the surface covered for positioning in m^2 , *# of RefPoints* is the number of reference points where the training samples were collected, and *# of Fp/Place* is the number of fingerprints taken at each reference points. *Multibuilding?* and *Multifloor?* denote if the database includes samples from multiple buildings and multiple floors, respectively.

Paper	Infrastructure	# of WAPs	Area (approx.)	# of RefPoints	# of Fp/Place	Multibuilding?	Multifloor?
[9]	Free	520	110,000 m^2	933	At least 20	YES	YES
[26]	Based	3	22.5 \times 43.5	70 \times 4	At least 20	NO	NO
[27] Exp. 1	Free	48	27 \times 18	216*	10,000**	NO	NO
[27] Exp. 2	Free	4****	27 \times 18	48 + 3	3 seconds		
[27] Exp. 3	Free	156	140 \times 90	1400*	10,000**	NO	NO
[27] Exp. 4	Free	10/12****	140 \times 90	101+15	3 seconds	NO	NO
[28]	Free	206	10,000 m^2	96	30***	NO	NO
[29]	Free	101	2 buildings	392 \times 3	3 \times N	YES	YES
[30]	Based Wi-Fi router	4	450 m^2	41	40	NO	NO
[30]	Based Nodes	6	800 m^2	56	40	NO	NO
[30]	Based Wi-Fi router	3 \times 4	3 \times 450 m^2	3 \times 41	40	NO	YES
[30]	Based Zotac PCs	4	1000 m^2	100	40	NO	NO
Our DB	Free	93	200 m^2	34	20	NO	NO

* Approximately, they used grids at 1.5 and 3 m.

** They collected 10,000 beacons scanning data during 5 minutes approx.

*** They collect 30 scans, but then they apply the average function.

**** Selected with the iLoc algorithm.

of fingerprints per reference points, and we also identify whether the environment is multibuilding and/or multifloor.

Although the soundness of results and conclusions presented in all these contributions is also high, the databases employed were not completely detailed in all the cases and the environments were radically different. In order to fairly compare the different proposed algorithms, they have to be implemented and tested on the new environments as was done in [27]. In particular, the authors of [27] compared RADAR [26] and HORUS [31] in the same indoor area, and the results show the importance of using a common database to compare systems. For instance, RADAR provided a median error of about 1.3 m (small environment) and 5 m (large environment). The median resolution of the RADAR system was in the range of 2 to 3 m in [26].

3. Analysis of the Problem and Datasets Design

This section (1) presents the environment chosen for performing the comparison of the two technologies, (2) shows some basic tests to determine the feasibility of using the magnetic field, (3) presents the Wi-Fi signals for indoor positioning using mobile phones in the selected environment and the design of the datasets, and (4) introduces the design of the two databases.

3.1. Environment. All the experiments have been carried out at the Geospatial Technologies Research Group's office. It is located on the fifth floor of the Espaitec-2 building at the *Universitat Jaume I* university campus. This main office is about 260 m^2 and it has 21 bookcases and 18 desktops as shown in Figures 1 and 2.

3.2. Is It Feasible to Use the Magnetic Field for Location in the Proposed Environment? An experiment has been performed to study the feasibility of the use of the magnetic field, measured by a Google Nexus 4 mobile phone, in the corridors of the laboratory to provide indoor positioning. Two simple trajectories in the laboratory (see Figure 1) were selected. The first one consists of two segments; the user comes into the laboratory and goes straight on until arriving to the top side windows and then turns right and goes straight on until the right side windows. The second one is a simpler trajectory where the user goes straight on through a corridor. The experiment consisted of recording the values provided by the magnetometer of a mobile phone while walking along the two trajectories. It was repeated 5 times on different days and time slots; the last repetition was done 16 days and 6 hours after the first one so that different factors, including occupancy distribution, were considered. The sampling frequency for the magnetometer was set to 10 Hz to balance computational costs and energy consumption with time series resolution.

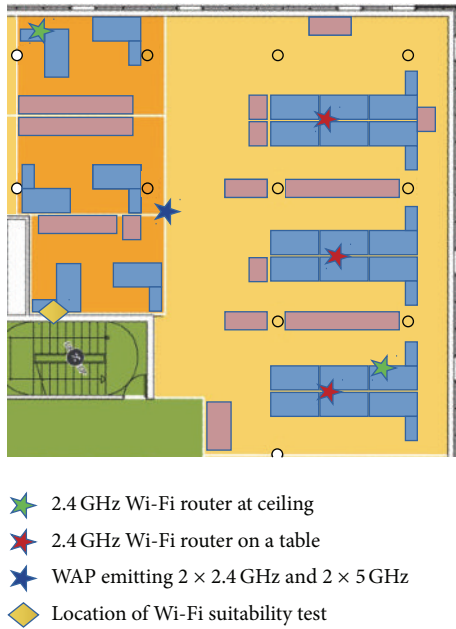


FIGURE 1: Distribution of the GEOTEC main laboratory (in yellow background) and professors' small offices (in orange background). The main corridor to access the laboratory is in green background. The bookcases are indicated with red boxes and the desktop tables are denoted with blue boxes. The locations of the WAPs supporting Internet connectivity are indicated with stars. The yellow box corresponds to the location of the mobile phone used for the feasibility study described in Section 3.3.

The magnetometer provides a vector that corresponds to the strength and direction of the magnetic field. This vector is relative to the mobile device as shown in Figure 3 and the values are measured in microteslas (μT). The example vector showed in Figure 3 means that there is a magnetic field of $46.669 \mu\text{T}$ strength in the direction of 45 degrees to the y -axis and z -axis of the device.

Figure 4 shows the recorded magnetometer values through both trajectories, where left plots ((a), (b), and (c)) correspond to the values measured by the magnetometer (in x -, y -, and z -axis, resp.) for the first trajectory (two corridors) and right plots ((d), (e), and (f)) correspond to the second trajectory (single corridor). Please note that the horizontal and vertical scales are different in the trajectories.

It can be observed that the magnetometer values are similar for the five runs according to the plots of the first trajectory. Although in the second trajectory the magnetic values are not exactly the same as in the fifth trajectory's run, their differences are low, about $5 \mu\text{T}$. In both cases, the form of the curve is very similar in the five runs, with small variations in the magnitude measured in each location.

From the results obtained by this test, it can be concluded that the magnetic field measured in the same location remains almost constant in time and, therefore, indoor positioning methods can be developed based on this fact.

3.3. Is It Feasible to Use the Wi-Fi RSSI Values for Location in the Proposed Environment? Following the same aims,

a new experiment has been performed in order to know the feasibility of the use of the Wi-Fi RSSI values captured by using mobile phones to provide indoor positioning in the proposed environment.

In this experiment, a Samsung Galaxy S4 smartphone was set at a fixed position inside a small office (see Figure 1) and a Wi-Fi collector application was run to record the fingerprints during the experiment. The application gathered 4 fingerprints per minute, 24 hours a day, for a total of 8 (eight) consecutive days. So, $34,000$ fingerprints were collected (4 fingerprints/minute \cdot 60 minutes/hour \cdot 24 hours/day \cdot 8 days). The application stores all raw fingerprints without applying any restriction, so all detected WAPs in a Wi-Fi scan were registered, even those related to antennas outside the laboratory. Although we detected almost 40 different WAPs in this experiment, we have focused on the signals emitted by the six antennas located in our lab (see Figure 1), since the stability and robustness of distant antennas were too low and we do not have complete information about their location, usage, and availability.

Figure 5 shows the mean RSSI value at intervals of five minutes from 0:00 to 23:55 during two different days, a Saturday (nonworking day) and a Monday (fully working day) for 4 WAPs present in the laboratory used for Internet connectivity. Figures 5(a) and 5(b) correspond to two Wi-Fi routers placed near the office (the red star located at the top of Figure 1 and the yellow diamond located at the bottom of Figure 1). Figures 5(c) and 5(d) correspond to a unique enterprise WAP, which emitted 2 networks in the 2.4 GHz band and 2 networks in the 5 GHz band (blue star in Figure 1).

According to this experiment, the Wi-Fi signal is stable when the number of people is low (Saturday) or when the signal is emitted in the 5 GHz band. Although the Wi-Fi emitted in the 2.4 GHz band is affected by the presence of people [32] and wooden elements, fingerprinting seems to be valid for the proposed environment because the received signal strength indicator (RSSI) values tend to be similar during the day and during different days. This behaviour is also similar for the other detected WAPs in the office.

3.4. Databases Design. Developing an indoor positioning system (IPS) has a critical step, generating good training data. According to the literature, an existing method can provide better results and/or even worse results than the ones provided in the original reference depending on the environment and the strategy followed to gather the reference data (see RADAR results in [26, 27]). Since generating good training data is crucial, a protocol has been established to generate a training (or reference) dataset and a validation dataset for the two radically different technologies: magnetic field and Wi-Fi fingerprinting. In particular, training data should be collected at the eight corridors that compose the GEOTEC laboratory (see Section 3.1) in both directions. This capture process should be repeated 5 times. Both datasets also should include timing information, which may be useful for further spatiotemporal analysis.

Regarding the magnetic database, data from magnetometer, accelerometer, and orientation sensors were included in

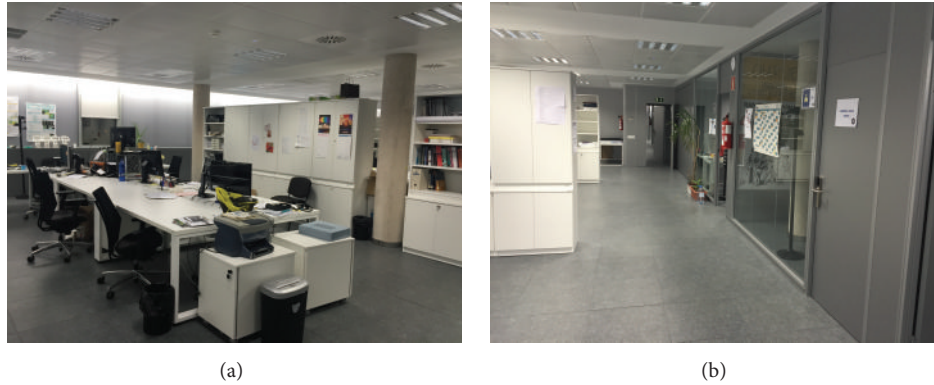


FIGURE 2: Pictures taken at the GEOTEC main laboratory. (a) shows the top zone of the laboratory, whereas (b) shows the left corridor. Both pictures show that the laboratory is full of wooden desktops with metal legs and bookshelves. Glass walls are also present.

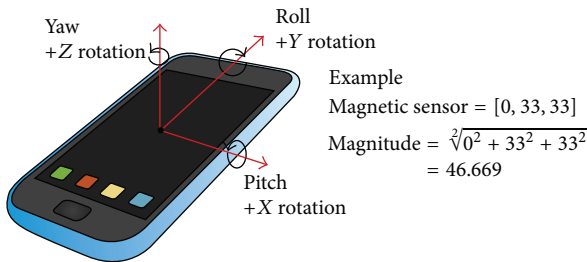


FIGURE 3: Meaning of the axis provided by the mobile phone sensory system.

the database to have better knowledge about the environment and context. Therefore, IPSs may be able to determine the user's speed, the user's turns, and other common situations when the user is navigating through the indoor facilities. Data should be collected in all the corridors as lineal segments in both directions. Generally, this is the natural way people walk. Moreover, data in the intersections between corridors should also be collected in order to have information about the user's turns. Most of the situations that may occur in an indoor environment (e.g., the presence of people and other obstacles in a corridor) should be considered while mapping it. Turns, including L-Turns and U-Turns, should be mapped to have a complete reference database, because the IPS's accuracy may depend on the situations recorded in the reference database.

Regarding the Wi-Fi based database, several consecutive fingerprints at each reference point should be taken. As shown in Figure 5, the RSSI values are not always constant. In fact, two consecutive (in time) RSSI readings can slightly vary, even for reading one second apart. To have a better reference dataset and consider the RSSI fluctuations, 5 consecutive fingerprints were collected for the training dataset to add diversity to the database. Several consecutive fingerprints should also be taken at each location in the validation dataset. So, real-time single fingerprint and off-line multiple-fingerprint methods can be evaluated with our database. Some IPSs found in the literature compute the average value of a set of fingerprints to estimate the user's position. Therefore, the proposed dataset should include this concern.

4. Materials and Methods

This section introduces the materials (databases) and methods (the baselines to assess the suitability of the datasets).

4.1. The Database for Magnetic Field Based Positioning. The database for the magnetic field is based on variations of the measured magnetic field produced by the structural elements present in a particular environment. The magnetic fingerprints were taken when the user was walking through the proposed environment (see Section 3.1). To allow this continuous mapping, the main routes through the GEOTEC laboratory were classified into 8 main corridors (see Figure 6).

As discussed previously in Section 3.4, the database contains mapping samples alongside the 8 corridors that compose the environment and all the intersections between two corridors (see Figure 7). Mapping "intersections" could make a more robust reference database, so the sensor values were also recorded when the user was turning to change the corridor where he/she was walking through. The 8 corridors and 19 intersections were mapped in two different directions with a Google Nexus 4 running Android 5.0.1. As a result, there were 54 different alternative paths. Sampling on every path was repeated 5 times, so the database designed for training purposes is composed of 270 different continuous samples. The samples of the eight corridors were collected on April 23rd, 2014, whereas the samples of the 19 intersections were mostly collected between September 26th and 29th, 2014, in order to have temporal diversity in the database. The five repetitions were consecutively taken.

The mapping process captured the data coming from three different sensor sources: magnetometer, accelerometer, and rotation sensor. The first source provided the raw data of the magnetometer sensor in the three axes [x , y , and z]. The second source was obtained from the raw data of the accelerometer also in the three axes minus the gravity force. The last one represented the orientation as the angle of rotation in the three axes. When capturing data, the user moved from a starting point to an ending point, and data were collected at every 0.1s. So continuous magnetic field fingerprints were stored. Each continuous sample contains

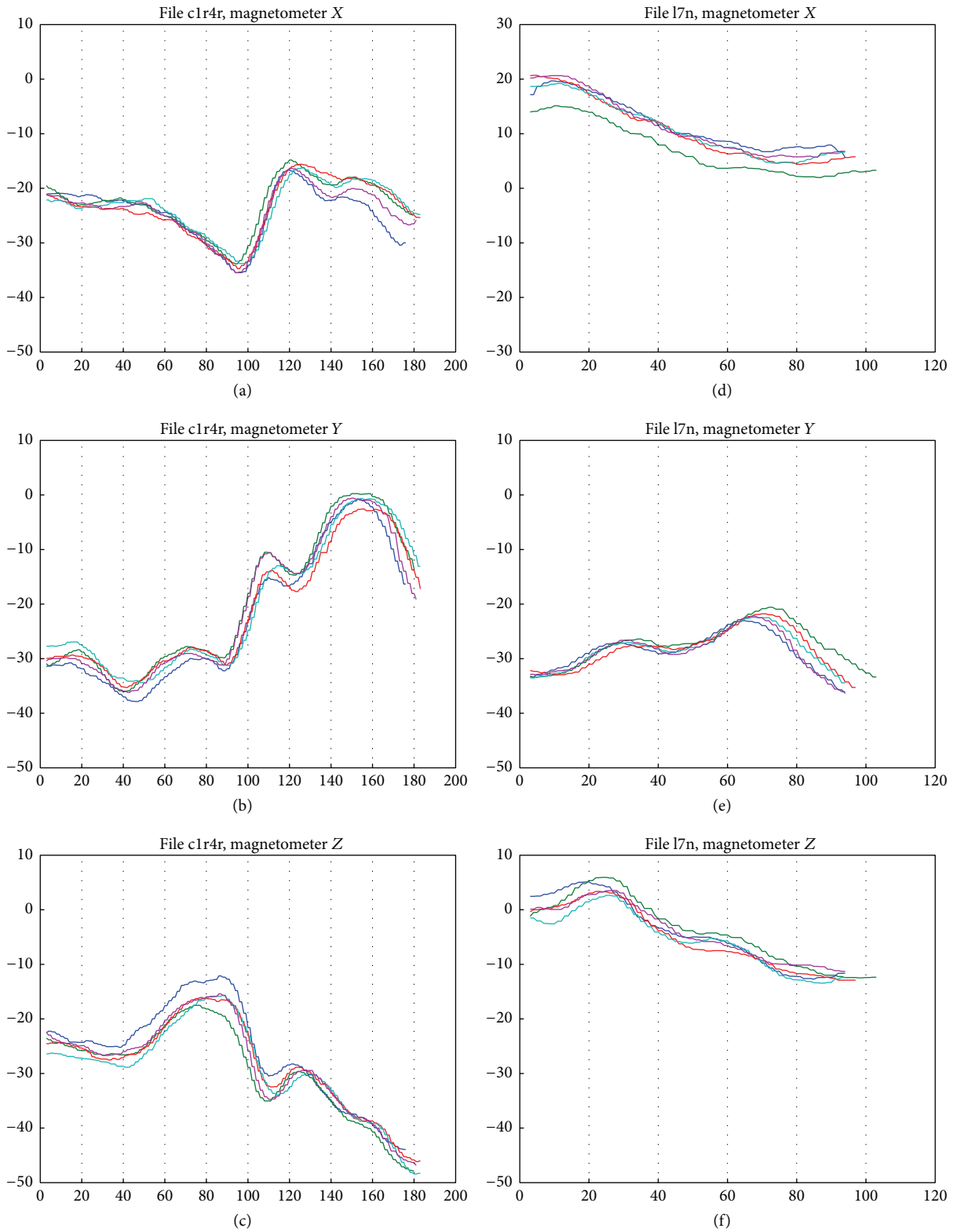


FIGURE 4: The magnetic field values (in the three axes) in two different trajectories. Note differences in x - and y -axis scales.

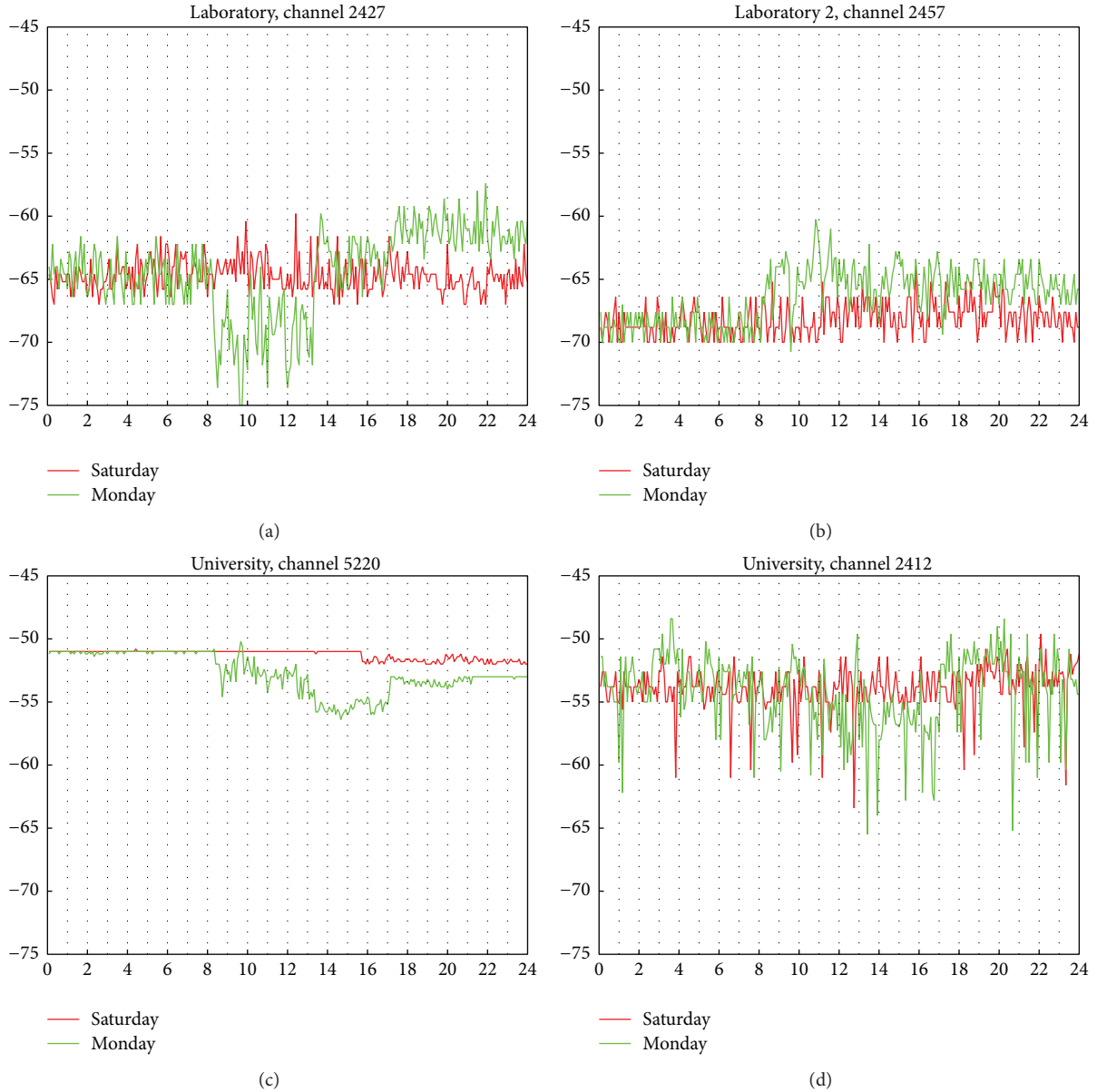


FIGURE 5: Mean RSSI values of four WAPs during two different days at five-minute intervals. The values from a full working day (8 am to 6 pm) are indicated in green. The values from a free day, when nobody was at the office, are shown in red.

the coordinates of initial and end points and also the coordinates of all turning points when capturing intersections. Moreover, it contains n discrete captures, each one with the 9 above-mentioned features plus the timestamp. With the initial/turning/end positions and the timestamps, it is possible to calculate the position of the discrete captures since the user was constrained to walk at constant speed while capturing the magnetic field values.

The mapping process was performed with an Android application that has direct access to sensors' data. The user's role in the application is to indicate in which zones the data capture process will be performed. Initially, the application shows a map centered in the proposed environment. Then,

the user draws the trajectory that he/she wants to follow to capture the data (see Figure 7(a)). This trajectory can consist of a path in a single corridor or in several ones. The user needs to be placed in the starting point of the route and, then, after clicking the "Start recording" button, the app starts to collect data until the user reaches the ending point and clicks the "Tap at end" button (see Figure 7(c)). In case of a multicorridor path, the user has to press the "Tap at turn" button to indicate that they are placed at the i th intersection (see Figure 7(b)).

For testing purposes, 9 complex routes (see Figure 8) along the laboratory were mapped. Each of these routes starts from different corridors and performs different trajectories.

The nine trajectories were mapped with the above-mentioned Google Nexus 4 smartphone. Two of them, routes 2 and 7, were also mapped with an LG G3 smartphone with Android 5.0 and stored as the 10th and 11th testing samples (files TT10 and TT11, resp.). So, a total of 11 complex continuous samples are available for testing purposes. Please note that the 11 testing trajectories are complex and were taken in more than one corridor. Although the 8th and 9th trajectories are placed in a single corridor environment, they may also be considered multicorridor since a U-Turn (180°) was done.

The data stored in each sample is proportional to the amount of time needed to complete an established path, due to sampling period of 0.1 seconds. So, the data provided by the accelerometer, magnetometer, and the orientation sensor of the device is stored 10 times per second. For example, if the user takes 12 seconds to map a corridor, the corresponding continuous sample will have 1200 values (12 s. \times 10 discrete captures \times 10 features).

All the paths, intersections, and turnings have been mapped with very high precision; since knowing that a person in normal conditions can cover a distance of 1.39 m per second, data have been captured approximately at every 0.139 m. The users walked through single corridor and multicorridor trajectories without any obstacle. Although the research group members and researchers were present in the office, nobody stood in the corridor.

4.1.1. Description of Database Files. The database consists of 281 continuous samples; 270 are for training and 11 are for testing. The corridors have been identified according to their numbering (see Figure 6) and orientation. The orientation for the vertical corridors is “normal” when the user was

ts_1	mx_1	my_1	mz_1	ax_1	ay_1	az_1	ox_1	oy_1	oz_1
\vdots									
ts_n	mx_n	my_n	mz_n	ax_n	ay_n	az_n	ox_n	oy_n	oz_n
$\langle m \rangle$									
lat_1	lon_1	lat_2	lon_2	FS_1	LS_1				
\vdots									
lat_m	lon_m	lat_{m+1}	lon_{m+1}	FS_m	LS_m				

Here, n is the number of samples collected in the trajectory at a 0.1-second frequency and m is the number of segments (corridors) in the trajectory. Each sample contains the timestamp, ts , and the values from magnetometer, accelerometer, and orientation sensors in the three axes, which are denoted with mx , my , mz , ax , ay , az , ox , oy , and oz . Finally, lat_i and lon_i correspond to the coordinates, latitude and longitude, of the initial, intermediate (intersections), and final points. To represent the coordinates, the WGS84 standard with the decimal degree representation has been selected. A trajectory with m corridors has $m + 1$ points. FS_i and LS_i stand for the i th trajectory’s first and last sample, respectively, in the full sequence of samples collected during the trajectory mapping.

walking from bottom to top (in the figure) and “reverse” when walking from top to bottom. Similarly, the orientation for the horizontal corridors is “normal” when the user was walking from left to right and “reverse” when walking from right to left. The samples have been stored as text files. The training files are grouped into two main categories “lines” and “curves”:

- (i) The “lines” group has 80 files and they stand for the single corridor case. The format for filename is “1X.Z.txt” where 1 stands for lowercase L (line), X stands for the number of corridors and orientations (n or r), and Z stands for repetition. For example, 13r_04.txt stands for the samples taken at the third corridor with normal orientation and the fourth repetition.
- (ii) The “curves” group has 190 files and they stand for all possible trajectories considering two connected corridors only. The format for that group’s filename is “cXXYY.ZZ.txt” where c stands for lowercase c (curve), XX and YY stand for the number of corridors and orientations for the first and second corridors in the two corridors’ trajectory, and ZZ stands for repetition. For example, c5n1r_06.txt stands for the samples taken at the fifth corridor with normal orientation and first corridor with reverse orientation and the sixth repetition.
- (iii) The testing files’ filename format is “ttPP.txt” where PP stands for the complex testing trajectory number (see Figure 8), for example, tt03.txt.

In each file, data have been stored as follows:

According to the previous structure, the text files are composed of two well-differentiated parts separated by the row indicating the number of segments in the trajectory: (1) the sequence of discrete samples taken during the trajectory mapping and (2) the configuration data.

The first part contains the timestamp (the UNIX time format in milliseconds) and the vector data from the magnetometer (Android’s TYPE_MAGNETIC_FIELD), accelerometer (TYPE_LINEAR_ACCELERATION), and orientation sensor (TYPE_ORIENTATION). The accelerometer’s values do not include the gravity force to have a better representation of a user’s real movement. Two consecutive samples (vertically represented here) from 6th testing trajectory are as

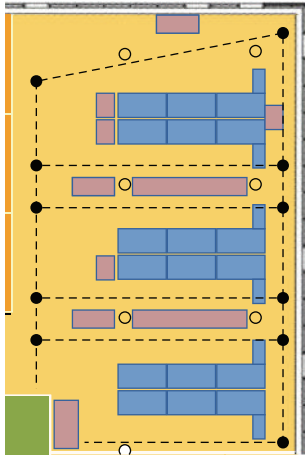


FIGURE 6: Distribution of the 8 corridors identified at the GEOTEC main laboratory. Corridors are numbered as follows: (1) the vertical corridor located at the left, (2) the vertical corridor located at the right, and (3–8) the horizontal corridors from bottom to top.

Lat _{start}	Lon _{start}	Lat _{start}
39.99389	-0.07375	39.99393
39.99393	-0.07384	39.99386
39.99386	-0.07389	39.99388

where latitude and longitude coordinates have been truncated, in this document, to 5 decimals for representation purposes. Three segments compose this particular example, so the number of intermediate points (intersections) is four. The mapped length of the first and second segments is similar, and the third segment’s length is slightly smaller.

4.2. *The Database for Wi-Fi Based Positioning.* The database for Wi-Fi fingerprinting is based on the received signal strength from a set of Wireless Access Points. Since this is an asynchronous task in Android devices and its frequency depends on the device, the Wi-Fi fingerprints were captured by the user being static at a known location inside the laboratory. To allow this discrete mapping, some reference points were identified for training (see Figure 9(a)) and testing (see Figure 9(b)) in the 8 main corridors at the GEOTEC laboratory.

As in the magnetic field based database, the Wi-Fi based database also contains mapping samples alongside the 8 corridors that compose the GEOTEC main laboratory. We consider that mapping between 4 and 5 reference points per corridor was enough due to the short length of the corridors. So, the 8 corridors were mapped in two different directions with a Samsung S3 (Android 4.3) and LG Spirit 4G LTE (Android 5.0.1). Sampling on every reference point was repeated 5 consecutive times, so that the database designed for training purposes is composed of a total of 680 different discrete samples. Similarly, the 8 corridors were mapped again at different reference points to generate the validation set, which is composed of 460 different discrete samples.

follows:

ts ₂₄	1417178330528	ts ₂₅	1417178330629
mx ₂₄	24.899292	mx ₂₅	24.719238
my ₂₄	-10.319519	my ₂₅	-11.219788
mz ₂₄	-49.55902	mz ₂₅	-49.319458
ax ₂₄	-0.12917818	ax ₂₅	-0.15856716
ay ₂₄	0.52311563	ay ₂₅	0.68318987
az ₂₄	-0.19135952	az ₂₅	-0.15023136
ox ₂₄	-64.537674	ox ₂₅	-62.273254
oy ₂₄	-21.03711	oy ₂₅	-21.420563
oz ₂₄	0.15363675	oz ₂₅	0.5122262

The second part contains the information about location of initial, intermediate, and ending points. Moreover, the samples can be associated with corridor segments and, furthermore, information about turnings is also provided in all the samples.

For instance, the configuration part for the 6th testing trajectory is as follows:

Lon _{start}	Sample _{start}	Sample _{end}
-0.07384	0	71
-0.07389	72	159
-0.07394	160	223

The mapping process consisted of scanning the environment for Wi-Fi networks and recording this information, the fingerprint, at the preestablished reference points located in the laboratory. Any Wi-Fi fingerprint contains the RSSI values of all the detected WAPs in the Wi-Fi scan. Moreover, any fingerprint also contains the coordinates of the reference point and information about the user and device and, finally, a timestamp. Although our environment has 5 Wi-Fi routers and a special enterprise WAP that emits four networks (two in the 2.4 GHz band and two in the 5 GHz band), several external WAPs were detected inside the laboratory. It is worth mentioning that we have not applied any restriction when the RSSI values were recorded, so the RSSI values belonging to external WAPs have also been stored.

The mapping process was performed with an Android application that has direct access to the sensors’ data. The user’s role in the application is to indicate in which zones the data capture is going to be performed. Initially the application shows a map of the proposed environment. Then, the user taps its current position on the map and the application starts collecting 5 consecutive Wi-Fi fingerprints (see Figure 10).

4.2.1. *Description of Database Files.* The database consists of 1,140 Wi-Fi fingerprints; 680 are for training and 460 are for testing. After processing all the individual fingerprints, a total of 97 different WAPs (9 of them located in the environment and the rest, 88, located outside) were detected. Due to privacy issues, the MAC address for each WAP has been anonymised, so the database uses the virtual identifiers WAPXX instead of the MAC addresses.

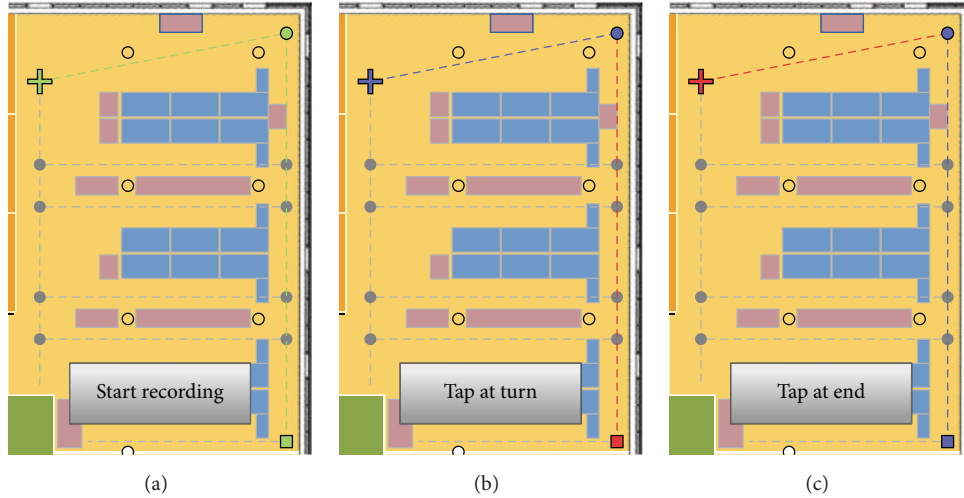


FIGURE 7: Three screenshots of the Android application used to capture the data. (a) shows the path (green) in which the data capture is going to be done and the “Start recording” button. ((b) and (c)) The current segment where the user is walking is highlighted in blue; the user has to press the button when she/he arrives to the 1st intersection (b) or the final destination (c).

The database includes the RSSI values from 9 known WAPs (the five routers plus the four networks provided by the special enterprise WAP) and 88 “unknown” WAPs installed outside the laboratory, which may mainly be located in the nearby facilities. The correspondence between the known WAPs (see Figure 1) and the virtual identifiers is as follows:

- (i) WAP67: green star located at the top-left of the image.
- (ii) WAP68: green star located at the bottom-right of the image.
- (iii) WAP70 and WAP80: blue star, the two networks emitted in the 5.2 GHz band.

- (iv) WAP74 and WAP84: blue star, the two networks emitted in the 2.4 GHz band.
- (v) WAP95: red star located at the center desktop-zone.
- (vi) WAP96: red star located at the bottom desktop-zone.
- (vii) WAP97: red star located at the top desktop-zone.

The fingerprints have been stored as two independent text files, where each line contains the following data:

```
RSSI_WAP01      RSSI_WAP02      ...      RSSI_WAP97      lon      lat      Phone_ID      Timestamp
```

Here $RSSI_{WAPXX}$ is the RSSI value for the WAP whose anonymised identifier is WAPXX, lat and lon are the coordinates (WGS84 in decimal degree format), $Phone_ID$ identifies the device used for mapping (1 for Samsung Galaxy S3; 2 for LG Spirit 4G LTE), and $timestamp$ is the timestamp. The nonrealistic RSSI value +100 has been used

to denote that the WAP was not detected in the Wi-Fi scan because it was switched off or the signal was too weak for the smartphone to detect it. For instance, the first training sample has the following data (without left headings):

```
RSSI_WAP01 to RSSI_WAP10      100,100,100,100,100,100,100,100,-88,100,100,
RSSI_WAP11 to RSSI_WAP20      -92,100,100,100,100,100,100,-77,100,100,
RSSI_WAP21 to RSSI_WAP30      100,100,100,100,100,100,100,100,100,100,100,
RSSI_WAP31 to RSSI_WAP40      100,100,100,100,100,100,100,100,100,100,100,
RSSI_WAP41 to RSSI_WAP50      100,100,100,100,100,100,100,100,100,100,100,
RSSI_WAP51 to RSSI_WAP60      100,100,100,100,100,-89,100,100,100,100,
RSSI_WAP61 to RSSI_WAP70      100,100,100,-73,-87,-81,-66,-67,100,-54,
RSSI_WAP71 to RSSI_WAP80      -83,100,100,-47,-60,100,100,100,100,-55,
RSSI_WAP81 to RSSI_WAP90      -80,100,-77,-45,-61,100,100,100,100,100,
RSSI_WAP91 to RSSI_WAP97      100,100,100,-87,-49,-57,-53,
Lon and Lat                    -0.07384642645414766,39.99384618213595,
User, Phone and ts             2,1,1450452282
```

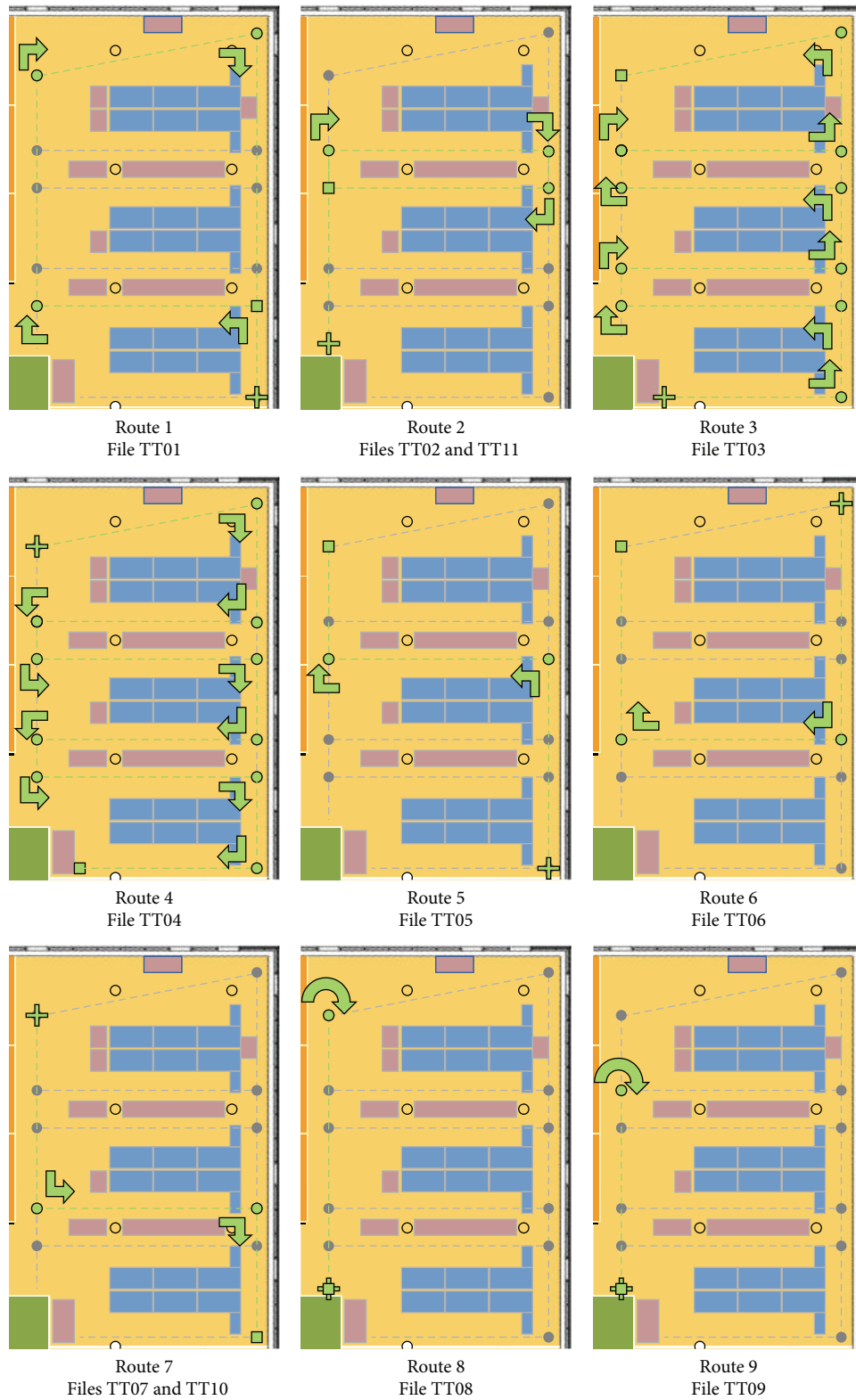



FIGURE 8: The 11 testing trajectories. The starting point is denoted with the green cross, green bullets are the intersections and user pose changes, and the ending points are denoted by green squares. The green arrows are introduced to clarify the user's pose change.

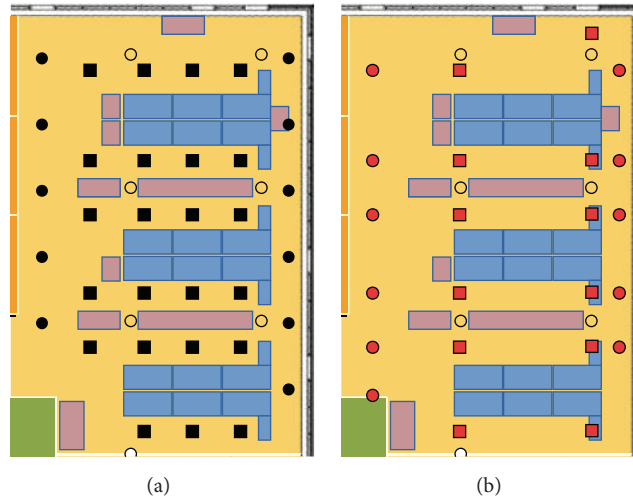


FIGURE 9: Distribution of the training (a) and validation samples (b) through the 8 corridors identified at the GEOTEC main laboratory. The circles represent the reference points taken at the vertical corridors (1 and 2) and squares represent the reference points taken at the horizontal corridors (3 to 8).

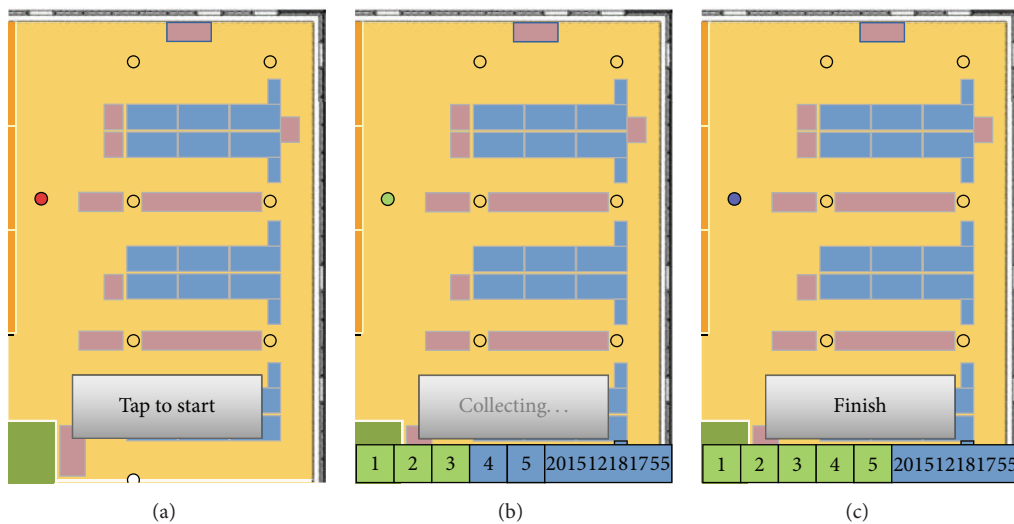


FIGURE 10: Three screenshots of the Android application used to capture the data. (a) shows the selected reference point (red) where the data capture is going to be performed and the “Tap to start” button. (b) shows the reference point in green while the 5 fingerprints are captured. The application indicates that the fingerprints are being collected by turning the corresponding panel located at the bottom of the screen to green. (c) shows that captures have been taken and the user finishes the mapping procedure.

In this example fingerprint, the Samsung Galaxy S3 detected 22 WAPs at the first training reference point (bottom-left circle in Figure 9(a)). The strongest signal was provided by WAP84, which corresponds to the special enterprise WAP. It can be observed that the RSSI provided by WAP74 and WAP84 are not the same, even though they are provided by the same device. This is the same for WAP70 and WAP80. Some external WAPs report strong-moderate signal strength: WAP67, WAP68, WAP75, and WAP85. We consider that all the signals may be important for indoor location and, therefore, we did not remove or filter any value.

4.3. Comparison of the Two Databases. Two independent databases gathered in the corridors of the GEOTEC laboratory are presented in this paper, one for magnetic field based positioning and the other for Wi-Fi based fingerprinting. The samples were gathered while the user was walking for the magnetic field based database, whereas the user had to remain for some seconds at a reference point for the Wi-Fi based database.

As commonly done in the literature, we applied different mapping strategies for each technology. The refreshment update frequency is much higher for the magnetometer than

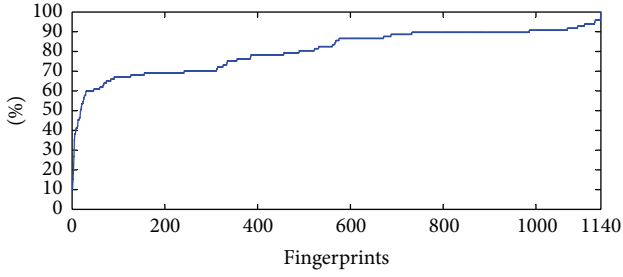


FIGURE 11: Cumulative distribution of the WAPs and fingerprints.

for the Wi-Fi chipset. When the user is walking it is possible to assign the magnetic strength to any intermediate position. However, the Wi-Fi fingerprint cannot be assigned to a particular position unless the user is static at a reference point. If the user is walking when the Wi-Fi scan is performed, the fingerprint could only be assigned to a long path segment (not a particular reference point) because the Wi-Fi scan process lasts between 4 and 6 seconds.

Despite the use of different mapping strategies, the databases have some common features: (1) samples were collected in the same corridors, (2) both datasets include timing information, (3) there are separate training and testing datasets, (4) the samples were taken 5 times per location (corridor in the magnetic field based and reference point in the Wi-Fi based), and (5) we have the same coordinate system in both datasets (WGS84 in decimal degree format) using our own cartography (<http://smart.uji.es>) (<http://indoorloc.uji.es>).

Regarding the number of features of each discrete capture, each discrete element of the magnetic database includes 9 meaningful values from 3 different sensors, whereas the Wi-Fi database includes the RSSI values from 97 WAPs. In fact, an average of 19.03 WAPs per fingerprint has been computed from the database records. Including 97 RSSI values does not mean that the 97 WAPs had been detected in all the fingerprints; it means that 97 different WAPs were detected at least once in the database. To explain more clearly this effect, Figure 11 shows the relation between number of fingerprints (horizontal axis) and number of WAPs (vertical axis). The plot shows that 58 WAPs (60%) have been detected in less than 30 fingerprints. Moreover, 74 WAPs (76.25%) are detected in less than a third of the fingerprints included in the database, so we consider that a significant number of the 97 WAPs have a minor presence in the database. A total of 8 WAPs are detected in at least 1026 fingerprints (90% of the total), with only 4 WAPs detected in all the fingerprints. Those 4 WAPs are all in the laboratory: WAP74, WAP84, WAP95, and WAP97.

4.4. A Realistic Baseline for Magnetic Field Based Positioning.

Two very simple baseline methods have been developed and tested to provide a starting point that any more sophisticated indoor localization algorithm should be able to overcome. The first method uses a discrete method to obtain the position of the discrete test points obtained from the continuous test samples. In this case, the experiment tries to obtain an

answer to the following question: *Is it possible to obtain precise location using only the data obtained from the magnetometer for a discrete point?*

The second baseline method is a continuous method that obtains the position of the user taking into account several seconds of data instead of single discrete samples. In this second case, the research question to answer can be formulated as follows: *Is it needed to take into account several consecutive captures to obtain accurate locations?*

In both cases, the methods have only used the training samples taken in the 8 corridors and from the magnetometer.

4.4.1. Discrete Method. For each continuous sample, the location of each discrete capture can be easily estimated since the coordinates of the initial and final points of the path are known, the timestamps were recorded, and the user velocity was almost constant.

All the discrete captures extracted from the continuous training samples of the corridors are used as the training dataset, where each element consists of 5 features: the location where the capture was taken [lat, lon] and the measurement obtained by the magnetometer in this location [m_x, m_y, m_z]. The same procedure was performed to extract the discrete captures from the test paths. In total, there are 8,943 samples for training and 4,380 for testing.

The 1-NN algorithm [33] was used to estimate the location of each test sample, so the test current location would correspond to the most similar training sample. The location of the most similar sample in the training set is the one assigned to the test sample. Although other distances or similarity metrics could have been used [34, 35], the distance between two samples, $m_1 = [m_{1,x}, m_{1,y}, m_{1,z}]$ and $m_2 = [m_{2,x}, m_{2,y}, m_{2,z}]$, corresponds to the Euclidean distance and it is estimated as follows:

$$d(m_1, m_2) = \sqrt{\sum_{i=\{X,Y,Z\}} (m_{1,i} - m_{2,i})^2}. \quad (1)$$

In this case, the error in positioning for each test path has been estimated as the mean distance between actual position and predicted position for all test captures along the path. This distance between two points does not correspond to the Euclidean distance between them since the points correspond to the latitude (lat) and longitude (lon) coordinates in decimal degrees; they are not expressed in linear meters. So, the *haversine formula* (2) is used instead. The standard error of the mean is also shown in the table:

$$d_{\text{haversine}} = R \cdot c, \quad (2)$$

where R is the radius of Earth, 6373 km approximately, and

$$c = 2 \cdot \arctan 2 \left(\sqrt{a}, \sqrt{1-a} \right),$$

$$a = \sin^2 \left(\frac{\Delta \text{lat}}{2} \right) + \cos(\text{lat}_1) \cdot \cos(\text{lat}_2) \cdot \sin^2 \left(\frac{\Delta \text{lon}}{2} \right). \quad (3)$$

4.4.2. Continuous Method. For the continuous case, each continuous training sample is divided into several subsamples of 5 seconds each. For instance, if a sample is 10 seconds long and has 100 discrete samples, then it is divided into 6 continuous subsamples, [1–50], [11–60], . . . , [51–100]. Each overlapping subsample includes information about the location of the initial and final point of the subpath and the 50 captures of the three components of the magnetic field measured.

All the subsamples extracted from the training samples of the corridors are used as the training dataset. The test samples are also divided into subsamples of 5 seconds. All the subsamples extracted from the test paths are used as the test dataset. In total, there are 540 subsamples for training and 231 for testing. For each test subsample, NN-based method (similar to the one introduced for the discrete case) is performed to look for the most similar training subsample.

The distance between two continuous subsamples $vm_1 = [vm_{X,1}, vm_{Y,1}, vm_{Z,1}]$ and $vm_2 = [vm_{X,2}, vm_{Y,2}, vm_{Z,2}]$ is also based on the Euclidean distance, and it is given by the following equation:

$$d(vm_1, vm_2) = \frac{1}{N} \cdot \sum_{i=1}^N d(vm_1[i], vm_2[i]), \quad (4)$$

where $vm[i]$ is the i th element of the vector vm , d corresponds to Euclidean distance (see (1)), and N is the number of discrete captures of each continuous subsample. In our case, $N = 50$ since each continuous subsample contains 50 discrete captures.

4.5. A Realistic Baseline for Wi-Fi Based Positioning. As in the case of the “magnetic field” database, a baseline method has been developed and tested using the Wi-Fi database. In this case, the RSSI values of the fingerprints are negative values that express the signal strength in dBm. The artificial value +100 was used to denote those WAPs which were not detected in a fingerprint scan. Equation (5) has been used to modify data and store it in a more convenient format for computing:

$$\text{NewValue} = \begin{cases} 0 & \text{if RSSI} = +100 \\ \text{RSSI} - \min + 1 & \text{otherwise} \end{cases} \quad (5)$$

The values have been made positive by subtracting the lowest possible value (−97 dBm in this particular case), so the weakest signal value corresponds to 1 and 0 denotes those WAPs which were not detected.

The 1-NN algorithm [33] was applied as baseline to estimate the location of each test fingerprint, so the test current location would correspond to the most similar training sample. The location of the most similar sample in the training set is the one assigned to the test sample. Although other distance or similarity metrics could have been used [34, 35], the distance between two fingerprints corresponds to the Euclidean distance and it is estimated with

$$d(fp_1, fp_2) = \sqrt{\sum_{i=1}^{N_{\text{WAPS}}} (fp_{1,i} - fp_{2,i})^2}, \quad (6)$$

where $fp_{i,j}$ stands for the j th WAP RSSI value for the i th fingerprint and N_{WAPS} is the number of existing WAPs in the database.

Apart from the *normal configuration* where the training set is used as reference dataset and the IPS evaluation (test) is done with the validation set, we have also implemented some interesting variations for the baseline:

- (i) The *continuous* configuration: the original training set includes fingerprints taken at different reference points, 34 in this case. A procedure has been applied to the training set in order to emulate a continuous mapping along all the corridors similarly to the magnetic field based database. So, new intermediate reference points have been generated between two original and consecutive training reference points, each new point at regular distances between the two original points. In particular, 2 alternatives, *complete* and *simple*, have been tested. Depending on the alternative used, 25 (complete) or 5 (simple) new artificial fingerprints have been created for each new intermediate point with linear interpolation. So, this procedure increases the size of the training set depending on the number of artificial points created between two real consecutive reference points and the interpolation used.
- (ii) The *average* configuration: a total of 5 consecutive fingerprints have been captured at each reference point. Averaging the 5 consecutive fingerprints has been done to test if this procedure can generate more representative fingerprints, for training and validation, as some authors have done in the literature (e.g., [28]). This procedure reduces the number of fingerprints by a factor of 5.
- (iii) The *threshold* configuration: according to the generated datasets, the RSSI values range is [−98 dBm, . . . , −28 dBm]. Some previous works state that the RSSI values below a threshold, −90 dBm [30], −85 dBm [36], or −80 dBm [30], should be computed as non-detected WAPs. A thresholding procedure has been applied to the datasets to test whether or not thresholding improves the accuracy in our database.
- (iv) The *known MACs* configuration: some works only use the WAPs whose location is known [27] or the WAPs they have deployed [30]. The accuracy of a reduced version of 1-NN classifier has been tested using the 6 Wi-Fi antennas [WAP67, WAP68, WAP70, WAP74, WAP80, WAP84, WAP95, WAP96, and WAP97] deployed in the GEOTEC laboratory for connectivity purposes (see Figure 1).

5. Results and Discussion

This section explains the results for the baselines using the magnetic field and Wi-Fi based databases and discusses the obtained results.

TABLE 3: Mean positioning error for magnetic field based discrete and continuous methods in the 11 testing paths.

Path	Discrete method		Continuous method	
	# of samples	Error	# of samples	Error
1	540	8.8 ± 0.18	35	8.74 ± 0.68
2	356	7.1 ± 0.21	21	7.44 ± 0.89
3	876	7.8 ± 0.14	44	7.89 ± 0.62
4	859	7.81 ± 0.14	41	7.21 ± 0.69
5	362	6.11 ± 0.19	23	5.14 ± 0.69
6	224	7.5 ± 0.21	9	6.05 ± 1.52
7	211	7.72 ± 0.29	8	6.24 ± 1.51
8	246	9.26 ± 0.22	16	6.58 ± 0.88
9	196	3.33 ± 0.21	11	1.25 ± 0.34
10	223	7.46 ± 0.3	10	5.33 ± 1.36
11	287	6.65 ± 0.18	13	4.7 ± 0.81
Mean		7.23 ± 0.38		6.05 ± 0.43

5.1. Results of the Baseline for Magnetic Field Based Positioning. Table 3 shows the baseline results for the discrete (see Section 4.4.1) and continuous methods (see Section 4.4.2). The mean error in positioning using the discrete method is 7.23 ± 0.38 m. This general error has been calculated considering the mean results in the 11 testing paths. In the continuous case, the mean error in positioning (considering the 11 different testing paths) is 6.05 ± 0.43 m. According to the results, the continuous method provides more accurate results, thereby improving the accuracy with respect to the discrete method by more than 1 meter.

Regarding the first research question: “Is it possible to obtain precise location using only the raw data obtained from the magnetometer in a discrete point?” according to the obtained results it seems that using only the three measurements of the magnetic field ($[m_x, m_y, m_z]$) in each point as fingerprint is not enough to provide accurate positioning. The use of only three features as fingerprint reduces the probability of having unique fingerprints in different positions.

Regarding the second research question: “Is it needed to take into account several consecutive samples to obtain accurate location?” according to the obtained results the use of data captured during several seconds (5 in our experiments) improves the accuracy, but there are several factors that should be studied in more detail to obtain more accurate results.

Therefore, 4 more experiments have been performed to improve the accuracy of the continuous baseline. The first two deal with the problem showed in Figure 4. This figure shows that not all training samples of the same path have the same values. The form of the curve is similar, but the absolute values in the same location differ from some samples to others. Two variations of the continuous method have been explored to deal with this problem. The first one (called E_1) applies a normalization procedure by subtracting from each training sample the mean of the 5 training samples of the same corridor and direction. The second one (called E_2) uses the mean sample as the unique training sample of a corridor and orientation. In order to obtain the mean of the 5 training

TABLE 4: Mean positioning error for the variation of magnetic field based continuous method in the 11 testing paths.

Path	E_1	E_2	E_3	E_4
1	5.72 ± 0.66	8.79 ± 0.67	6.48 ± 0.65	8.69 ± 0.44
2	6.07 ± 0.34	7.46 ± 0.89	5.28 ± 0.38	2.88 ± 0.49
3	4.71 ± 0.38	7.89 ± 0.62	4.70 ± 0.47	6.78 ± 0.62
4	6.55 ± 0.67	7.16 ± 0.70	6.97 ± 0.78	5.59 ± 0.67
5	3.30 ± 0.55	5.34 ± 0.67	3.31 ± 0.52	1.92 ± 0.57
6	6.31 ± 1.00	5.99 ± 1.51	7.31 ± 1.18	1.39 ± 0.34
7	4.30 ± 0.65	6.15 ± 1.54	5.28 ± 1.26	3.08 ± 1.48
8	7.16 ± 0.57	6.18 ± 0.78	7.42 ± 0.64	5.69 ± 1.00
9	5.59 ± 1.13	1.65 ± 0.31	5.95 ± 1.19	2.75 ± 1.06
10	3.51 ± 0.69	5.04 ± 1.22	3.04 ± 1.28	2.03 ± 0.58
11	7.96 ± 0.62	4.80 ± 0.80	6.45 ± 0.76	1.64 ± 0.51
Mean	5.56 ± 0.37	6.04 ± 0.42	5.65 ± 0.37	3.77 ± 0.46

samples from the same corridor and orientation, longer samples have been trimmed so that all training samples have the same length.

The second and third columns of Table 4 show the results obtained for two variations. The second one, E_1 , obtains the best results by reducing the mean error across the 11 test paths by a half meter with respect to the original continuous baseline. The second variation does not improve the accuracy as although the values measured from the magnetic field are quite similar in the same location, in two different time moments to allow the estimation of the user’s position, there are small variations, and therefore it seems that having several training samples adds diversity to the training database and this can help the classifier when looking for the closest sample. In other words, the samples from the same class (location in this case) fall in different places in the feature space.

Since the length of the continuous sample depends on the walking velocity of the user capturing the data, a new experiment (called E_3) has been performed to study if the user’s velocity affects the accuracy of the continuous method. In this experiment, after applying the normalization procedure presented at Experiment E_1 , all the training samples have been modified to obtain new ones as if the user had walked at the same speed. For this purpose, first, the average speed of the user in all training samples has been estimated using the data captured from the accelerometer. Then, each training sample is resampled using the estimated average velocity by using interpolation. For the test samples, first, a variable number of captures are taken depending on the velocity of the user to obtain a sample as if the user had walked at the average speed previously obtained. Second, the sample has been resampled by using interpolation to obtain a sample length of 50. The fourth column of Table 4 shows the results obtained. In this case, no improvements are obtained, since the user who captured the data of the database tried to walk by always using the same velocity. This fact can be confirmed since the variance of the velocity estimated for each training sample is very small.

TABLE 5: Baseline results for the Wi-Fi based database and different configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Normal	—	4.73	680	460	4.5s	9.78 ms	1

Time 1: time required to compute all testing fingerprints in seconds.

Time 2: time required to compute a single fingerprint in milliseconds.

Time rel.: relative time per fingerprint with respect to the normal configuration.

TABLE 6: Baseline results for the Wi-Fi based database and different configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Continuous complete	1 new point	4.51	3280	460	21.01	45.67	4.67
	3 new points	4.45	8480	460	55.01	119.59	12.23
	5 new points	4.49	13680	460	89.47	194.49	19.89
	7 new points	4.46	18880	460	124	269.56	27.56
Continuous simple	1 new point	4.61	1200	460	7.7	16.74	1.71
	3 new points	4.54	2240	460	14.12	30.7	3.14
	5 new points	4.60	3280	460	21.19	46.07	4.71
	7 new points	4.60	4320	460	27.22	59.17	6.05

The last variation of the continuous baseline (called E_4) consists of including in the training set the samples that can be extracted from all the intersections between two corridors also included in the database and that were not included in the training set in the previous experiments. The normalization procedure presented in Experiment E_1 has also been performed for the training samples. The last column of Table 4 shows the results obtained. In this case, the best results are obtained and the mean error across the 11 test paths is 3.77 m, more than 2 meters better than the original continuous baseline. This experiment confirms the conclusion previously extracted from Experiment E_2 that having more training samples adds diversity to the training database and improves accuracy.

5.2. Results of the Baseline for Wi-Fi Based Positioning. Table 5 shows the results for *normal configuration* baseline (see Section 4.5). With this configuration, which does not require any further parameter, the accuracy is 4.73 m and the time required to compute a single fingerprint is about 10 ms (NN implemented in Matlab and run in a 2nd-generation Intel i7 @ 3400 GHz and 8GB RAM). This time is used as baseline to compare the computational costs with respect to the other configurations.

Table 6 shows the results for the *continuous configuration*. In the *complete* configuration, 25 new artificial fingerprints have been generated for each new intermediate point. In the *simple* configuration, 5 new artificial fingerprints have been generated for each new intermediate point. The best result, 4.45 m, is provided by the *continuous complete configuration* with 3 new intermediate points between two consecutive training reference points. This experiment shows that the positioning accuracy can be improved using a *continuous configuration* of training samples, at the expense of having a training set 12 times larger than the original one.

Table 7 shows the results for the *average configuration*. The average of the five consecutive fingerprints has been applied

for three cases: only training set, only validation set, and both, training and validation sets. The best result, 4.49 m, is obtained when averaging is applied to the validation set but the training set remains unchanged. Although the time required to process a fingerprint is not altered (with respect to the *normal configuration*) when averaging is only applied to the validation set, the algorithm still requires having the five consecutive fingerprints. This experiment shows that positioning accuracy can be improved when the user is static by computing the averaged fingerprint in the validation set. However, the accuracy is not improved when averaging in the training set is applied since, as in the case of the magnetic field experiment (E_2), having diversity in the training set benefits the localization task.

Table 8 shows the results for the *threshold configuration*. Four different threshold values have been tested, and the best results (4.46 m) are obtained when the RSSI values below -80 dBm are removed. Also, the computational cost per fingerprint is the same as that for the normal configuration. This experiment shows that there was some noise present in the environment, which was removed by applying thresholding techniques to remove distant WAPs from the dataset.

Table 9 shows the results for the *known MACs configuration*. Only the WAPs close to or inside the GEOTEC laboratory are considered and the rest are removed. Although the computational cost per fingerprint with respect to the normal configuration is slightly lower, the accuracy is also slightly worse. This experiment shows that Wi-Fi fingerprinting has to be supported not only by the WAPs present in the environment but also by nearby ones because their presence has improved the positioning accuracy (see Table 5).

Tables 6–9 have shown the results for different ideas to improve the accuracy in indoor positioning techniques. In general, except for the *known MACs configuration*, all these ideas have separately improved the positioning accuracy with respect to the normal configuration (Table 5). Having the baseline results and access to the database, one may combine

TABLE 7: Baseline results for the Wi-Fi based database and different configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Average	T	4.84	136	460	1.05	2.27	0.23
	V	4.50	680	92	0.88	9.6	0.98
	T & V	4.86	136	92	0.24	2.61	0.27

TABLE 8: Baseline results for the Wi-Fi based database and different configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Threshold	-75 dBm	4.97	680	460	4.41	9.58	0.98
	-80 dBm	4.46	680	460	4.4	9.57	0.98
	-85 dBm	4.72	680	460	4.42	9.6	0.98
	-90 dBm	4.73	680	460	4.47	9.71	0.99

TABLE 9: Baseline results for the Wi-Fi based database and different configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Known MACs		4.90	680	460	3.9	8.48	0.87

different approaches in a single method. As an example, Table 10 shows the results of an advanced configuration which combines the “*continuous*” (complete and simple), the “*average*” (only on the validation set), and the “*threshold*” configuration (filtering RSSI values lower than -80 dBm) obtaining better results than any of the independent solutions shown in Tables 5–9.

5.3. *Discussion.* Magnetic fields and Wi-Fi technologies are radically different. Magnetic field positioning relies on the disturbances in the magnetic field in the environment whereas Wi-Fi based positioning uses the received signal strength indicator from multiple Wireless Access Points previously deployed for Internet connectivity in a particular environment. The WAP identifiers are unique so Wi-Fi fingerprints are highly attached to the place they were taken. In contrast, two distant places may have similar magnetic field strengths when using only the three values provided by the magnetometer in a particular location as fingerprint.

The mapping procedure is different for the two technologies. The magnetic field based database was collected while the user was moving, since the magnetic field strength can be sampled with a frequency of 10 Hz using smartphones. In contrast, reference points were statically sampled for the Wi-Fi based database, because getting the RSSI values from the environment lasted about 3 seconds with the LG device and 5 seconds with the Samsung device. Therefore, mapping the whole environment was faster for the magnetic field based database. Despite the differences, samples in both datasets have been captured using the same protocol, which led to data captured in the same corridors of the laboratory.

Regarding the magnetic field based indoor positioning baseline, the use of the continuous samples instead of the discrete ones improves the accuracy. In addition, the normalization of the continuous training samples and the inclusion in the training set of the data captured in the intersection have been crucial to obtain good results, obtaining as the best result an error of 3.77 m.

In the case of the Wi-Fi based indoor positioning, some independent experiments (configurations in the baselines) show that there are different ways to improve the positioning accuracy: adding artificial fingerprints to increase the training database density, applying fingerprint averaging when the user is static in the operational stage, and removing noisy distant WAPs in fingerprints by applying thresholding. The mean positioning error is reduced to approximately 4.4 meters by applying these independent configurations. When all the configurations are considered in the same algorithm, the error is reduced to 4.26 meters. This indicates that combining diverse approaches can improve positioning accuracy.

Both technologies have advantages and disadvantages for indoor positioning. Some of them are summarized in Table 11. In general, the main advantages of the magnetic field based technique are as follows: (1) no infrastructure needs to be installed, (2) a good level of accuracy can be reached, and (3) the possibility of having high sampling frequency allows continuous mapping. Therefore, the mapping process can be very fast. In contrast, the main disadvantages are as follows: (1) the fact that each discrete sample has only three features, (2) the dependence on the user velocity, and (3) the dependence on the device orientation. The main advantages of the Wi-Fi fingerprinting based method include the following: (1) the high number of features of each fingerprint and (2) the good accuracy even using simple and fast algorithm such as NN. Nevertheless, it has some disadvantages, such as (1) the low refreshment frequency and (2) the fact that the RSSI values can be affected by the presence of people in the environment.

6. Conclusions

Research on indoor positioning requires extensive efforts in developing the required software and generating datasets, which often can be challenge with a limited budget for all the

TABLE 10: Results for the Wi-Fi based database and different mixed configurations.

Configuration	Parameters	Accuracy	# of training FP	# of test FP	Time 1	Time 2	Time rel.
Mixed configuration	3 NPC; AV; -80 dBm	4.27	8480	92	10.32	112.17	11.47
Mixed configuration	3 NPS; AV; -80 dBm	4.49	2240	92	2.74	29.78	3.04

Three NPC and 3 NPS stand for 3 new intermediate points (continuous complete configuration and continuous simple configuration), AV stands for applying averaging to the validation set, and -80 dBm stands for thresholding RSSI values below it.

TABLE 11: Pros and cons for magnetic field and Wi-Fi fingerprinting positioning.

	Pros	Cons
Magnetic field	Being fully infrastructure-less	Vector only contains 3 values
	Good accuracy	Requires user movement
	Continuous mapping	High dependence on user's velocity
	Continuous positioning	High dependence on user's orientation
	Fast mapping procedure	Large reference database
	Refreshment frequency @ 10 Hz	
Wi-Fi fingerprinting	Being partially infrastructure-less	Affected by people presence
	Good accuracy	Affected by user movement
	Average of 17 WAPs per FP	5 GHz band not being in all devices
	"Small" reference database	Refreshment frequency @ 0.2 to 0.3 Hz
	Discrete database	Required feature selection
	Allows precise spatiotemporal analysis	Presence of mobile hotspots
	1-NN valid and fast solution	

underlying investments. The researchers tend to use nearby facilities as a test bed to evaluate new indoor positioning systems so that the results published in the literature using private databases cannot be directly compared.

This paper has introduced two databases for indoor localization, one being magnetic field based and the other being Wi-Fi fingerprinting based database, on the same indoor area. The description, procedures, strategies, and applications used to generate the databases have been fully described. Several baselines have been introduced using the proposed databases in order to show the viability of the proposed databases and also to encourage researchers to use them in order to compare their different indoor positioning approaches.

Our further work will be focused on collecting and publishing new databases with different technologies, even hybrid ones. Moreover, our intention is to provide some tools for the visual representation of the datasets and the visual analysis of the different indoor positioning technologies and systems.

Competing Interests

The authors declare that the grant, scholarship, and/or funding mentioned in Acknowledgments do not lead to any conflict of interests. Additionally, the authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement no. 642332, <http://www.geo-c.eu/>). The authors also gratefully acknowledge funding from the Spanish Ministry of Economy and Competitiveness through the "Metodologías avanzadas para el diseño, desarrollo, evaluación e integración de algoritmos de localización en interiores" project (Proyectos I+D Excelencia, código TIN2015-70202-P) and the "Red de Posicionamiento y Navegación en Interiores" network (Redes de Excelencia, código TEC2015-71426-REDT). The authors would like to thank all the current and past members of the *Geospatial Technologies Research Group* and *Ubik Geospatial Solutions S.L.* for their valuable help in creating the *SmartUJI* platform and providing us with the supporting services that allowed integrating the existing GIS services in the applications developed to create both databases.

References

- [1] 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), 2014.
- [2] Markets & Markets, *Indoor Location Market by Solution (Tag-Based, RF-Based, Sensor-Based), by Application (Indoor Maps & Navigation, Indoor Location-Based Analytics, Tracking &*

- Tracing, Monitoring & Emergency Management*), by Service, by Vertical, & by Region—Global Forecast up to 2019, Markets & Markets, 2015.
- [3] A. G. Estevez and N. Carlsson, “Geo-location-aware emulations for performance evaluation of mobile applications,” in *Proceedings of the 11th Annual Conference on Wireless on-demand Network Systems and Services (WONS '14)*, pp. 73–76, Obergurgl, Austria, April 2014.
 - [4] A. R. de M Neves, Á. M. G. Carvalho, and C. G. Ralha, “Agent-based architecture for context-aware and personalized event recommendation,” *Expert Systems with Applications*, vol. 41, no. 2, pp. 563–573, 2014.
 - [5] J. Torres-Sospedra, J. Avariento, D. Rambla et al., “Enhancing integrated indoor/outdoor mobility in a smart campus,” *International Journal of Geographical Information Science*, vol. 29, no. 11, pp. 1955–1968, 2015.
 - [6] L. Calderoni, M. Ferrara, A. Franco, and D. Maio, “Indoor localization in a hospital environment using random forest classifiers,” *Expert Systems with Applications*, vol. 42, no. 1, pp. 125–134, 2015.
 - [7] W. Le, Z. Wang, J. Wang, G. Zhao, and H. Miao, “A novel wifi indoor positioning method based on genetic algorithm and twin support vector regression,” in *Proceedings of the 26th Chinese Control and Decision Conference (CCDC '14)*, pp. 4859–4862, IEEE, Changsha, China, June 2014.
 - [8] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, “Indoor localization using FM signals,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 8, pp. 1502–1517, 2013.
 - [9] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó et al., “UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '14)*, pp. 261–270, Busan, Republic of Korea, October 2014.
 - [10] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte, and J. Huerta, “UJIIndoorLoc-Mag: a new database for magnetic field-based localization problems,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation*, October 2015.
 - [11] P. Barsocchi, S. Chessa, F. Furfari, and F. Potorti, “Evaluating ambient assisted living solutions: the localization competition,” *IEEE Pervasive Computing*, vol. 12, no. 4, pp. 72–79, 2013.
 - [12] D. Salvi, P. Barsocchi, M. T. Arredondo, and J. P. L. Ramos, “EvAAL, evaluating AAL systems through competitive benchmarking, the experience of the 1st competition,” in *Evaluating AAL Systems through Competitive Benchmarking. Indoor Localization and Tracking*, S. Chessa and S. Knauth, Eds., vol. 309 of *Communications in Computer and Information Science*, pp. 14–25, Springer, Berlin, Germany, 2012.
 - [13] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, “A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned,” in *Proceedings of the 14th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN '15)*, pp. 178–189, ACM, 2015.
 - [14] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, S. Sen, and V. Handzinski, “Microsoft indoor localization competition: experiences and lessons learned,” *SIGMOBILE Mobile Computation and Communication Review (MC2R)*, vol. 18, no. 4, pp. 24–31, 2014.
 - [15] F. Potorti, P. Barsocchi, M. Girolami, J. Torres-Sospedra, and R. Montoliu, “Evaluating indoor localization solutions in large environments through competitive benchmarking: the EvAAL-ETRI competition,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '15)*, pp. 1–10, Alberta, Canada, October 2015.
 - [16] T. V. Haute, E. De Poorter, J. Rossey et al., *D2.1 Initial Version of the EVARILOS Benchmarking Handbook*, 2013.
 - [17] T. Van Haute, E. De Poorter, F. Lemic et al., “Platform for benchmarking of RF-based indoor localization solutions,” *IEEE Communications Magazine*, vol. 53, no. 9, pp. 126–133, 2015.
 - [18] ISO, “Information technology—real time locating systems—test and evaluation of localization and tracking systems,” Standard ISO/IEC DIS 18305, 2015, http://www.iso.org/iso/catalogue_detail.htm?csnumber=62090.
 - [19] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, “How feasible is the use of magnetic field alone for indoor positioning?” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–9, IEEE, Sydney, Australia, November 2012.
 - [20] W. Storms, J. Shockley, and J. Raquet, “Magnetic field navigation in an indoor environment,” in *Proceedings of the International Conference and Exhibition on Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS '10)*, pp. 1–10, Kirkkonummi, Finland, October 2010.
 - [21] J. Song, H. Jeong, S. Hur, and Y. Park, “Improved indoor position estimation algorithm based on geo-magnetism intensity,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '14)*, pp. 741–744, Busan, South Korea, October 2014.
 - [22] D. Vandermeulen, C. Vercauteren, and M. Weyn, “Indoor localization using a magnetic flux density map of a building: feasibility study of geomagnetic indoor localization,” in *Proceedings of the International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT '13)*, pp. 42–49, Porto, Portugal, 2013.
 - [23] B. Li, T. Gallagher, C. Rizos, and A. G. Dempster, “Using geomagnetic field for indoor positioning,” *Journal of Applied Geodesy*, vol. 7, no. 4, pp. 229–238, 2013.
 - [24] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, and M. Wiseman, “Indoor location sensing using geo-magnetism,” in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*, pp. 141–154, ACM, July 2011.
 - [25] S. Shahidi and S. Valaee, “GIPSY: geomagnetic indoor positioning system for smartphones,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '15)*, pp. 1–7, IEEE, Banff, Canada, 2015.
 - [26] P. Bahl and V. N. Padmanabhan, “RADAR: an in-building rf-based user location and tracking system,” in *Proceedings of the Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 775–784, 2000.
 - [27] K. Chintalapudi, A. P. Iyer, and V. N. Padmanabhan, “Indoor localization without the pain,” in *Proceedings of the 16th Annual Conference on Mobile Computing and Networking (MobiCom '10)*, pp. 173–184, September 2010.
 - [28] J. Machaj, P. Brida, and R. Piché, “Rank based fingerprinting algorithm for indoor positioning,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11)*, pp. 1–6, Guimaraes, Portugal, September 2011.
 - [29] N. Marques, F. Meneses, and A. Moreira, “Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning,” in *Proceedings of the International Conference*

- on *Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–9, IEEE, Sydney, Australia, November 2012.
- [30] F. Lemic, A. Behboodi, V. Handziski, and A. Wolisz, “Experimental decomposition of the performance of fingerprinting-based localization algorithms,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '14)*, pp. 355–364, Busan, Republic of Korea, October 2014.
- [31] M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*, pp. 205–218, Seattle, Wash, USA, June 2005.
- [32] S. Garcia-Villalonga and A. Perez-Navarro, “Influence of human absorption of Wi-Fi signal in indoor positioning with Wi-Fi fingerprinting,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '15)*, pp. 1–10, IEEE, Banff, Canada, October 2015.
- [33] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [34] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, and J. Huerta, “Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems,” *Expert Systems with Applications*, vol. 42, no. 23, Article ID 10227, pp. 9263–9278, 2015.
- [35] S.-H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, pp. 300–307, 2007.
- [36] Y. Zhuang, Z. Syed, J. Georgy, and N. El-Sheimy, “Autonomous smartphone-based WiFi positioning system by using access points localization and crowdsourcing,” *Pervasive and Mobile Computing*, vol. 18, pp. 118–136, 2015.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

