3

Lessons Learned in Generating Ground Truth for Indoor Positioning Systems Based on Wi-Fi Fingerprinting

Joaquín Torres-Sospedra*, Óscar Belmonte-Fernández*, Germán M. Mendoza-Silva*, Raul Montoliu*, Adrián Puertas-Cabedo[†], Luis E. Rodríguez-Pupo*, Sergio Trilles*, Andrea Calia[‡], Mauri Benedito-Bordonau[§], Joaquín Huerta*

*INSTITUTE OF NEW IMAGING TECHNOLOGIES, JAUME I UNIVERSITY, CASTELLÓN, SPAIN [†]SOLUCIONES CUATROOCHENTA S.L., ESPAITEC2, CASTELLÓN, SPAIN [‡]DEPARTMENT BE-OP, EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH, CERN CEDEX, FRANCE [§]ESTUDIOS GIS, C/ALBERT EINSTEIN, VITORIA-GASTEIZ, SPAIN

1 Introduction

The development of applications and systems relying on localization-based services has boosted over the last few years. These service providers are eager to exploit customer information with the widespread adoption of smartphones, the proliferation of affordable mobile devices, and the ubiquity of Internet connections (4G and Wi-Fi). Automatic user localization can be considered a hot research and business topic with an expected \$3.96 billion market by 2019 (Markets&Markets, 2014). Indoor positioning and communications will be crucial elements in next-gen location-based services.

Applications based on users location need to know the position or localization for providing customized services, tracking assets and people, among others. Although outdoor localization is already solved due to the inclusion of GNSS support in localization devices, obtaining high precision in indoor positioning is still an unsolved problem for general-purpose applications. Two different approaches have been commonly applied to develop an indoor positioning system (IPS): infrastructure-based and infrastructure-less solutions. The former solutions need the deployment and installation of specific custom hardware in-site, while the later solutions only use a sensory system embedded in the location device to estimate the current location.

Geographical and Fingerprinting Data for Positioning and Navigation Systems. https://doi.org/10.1016/B978-0-12-813189-3.00003-4

For both infrastructure-based and infrastructure-less approaches, the fingerprinting technique has been widely used in indoor positioning. In fingerprinting, measurements (fingerprints) of a physical quantity are collected at known locations, in order to characterize the target environment. The environment characterization is then used in estimating the location where new measurements are taken. The measured quantities most often used in indoor positioning studies include wireless network signals strength (BLE (Čabarkapa et al., 2015), ZigBee (Marti et al., 2012), Wi-Fi (He and Chan, 2016)), which mostly use an infrastructure-based approach, and magnetic field directed strength, which uses a infrastructure-less approach. This chapter focuses on the development of IPS based on Wi-Fi fingerprinting to support heterogeneous mobile applications: indoor localization, pedestrian indoor navigation, and monitoring people at home.

Traditional Wi-Fi fingerprinting requires two steps or phases (Marques et al., 2012): the *training phase*, where the database with the ground truth or reference data are generated, and the *operational phase*, where the location algorithm estimates the location using previous knowledge. The training phase, or calibration phase, usually requires to measure the received signal strength indicator of the surrounding Wi-Fi access points at many (usually predefined) locations. After this phase, a reference database (or radio map) is available to estimate the positions of users at the operational phase. This simple approach is the base of many working IPS.

It is known that the process of site survey to generate the radio map is really time-consuming and labor intensive (Liu et al., 2014; Zhang et al., 2016a; Hossain and Soh, 2015). To overcome this issue, some approaches apply regression techniques to obtain a dense consistent radio map from less measurements (Hernández et al., 2017; Ezpeleta et al., 2015), whereas other authors artificially generate the radio map by using equations based on the radio signal (path loss) propagation (Chiou et al., 2010; Deasy and Scanlon, 2007) or ray tracing (Raspopoulos et al., 2012; Ayadi et al., 2015; El-Kafrawy et al., 2010). Although these advanced approaches to radio map generation provide interesting results, real measurements are still required in most of the works based on Wi-Fi fingerprinting.

Most of the scientific solutions found in the research literature are tested and evaluated on small- or medium-sized controlled environments (Zhang et al., 2016b; Li et al., 2016; Mizmizi and Reggiani, 2016), also known as laboratory environments, where the time required to generate the radio map might be affordable depending on the experimental setup. However, there is currently a raising interest in the development and deployment of realistic IPS in large realistic environments (Marques et al., 2012; Torres-Sospedra et al., 2014, 2015; Mathisen et al., 2016; Berkvens et al., 2016; Liu et al., 2016; Guimarães et al., 2016). In large scenarios, the calibration might be critical step since it requires the collection of many actual measurements. However, as far as we know, there is little information regarding how data are collected, the effort required to generate the radio map, and the problems that arise at this calibration stage in such large scenarios.

This chapter fills this gap and describes the lessons learned regarding collecting the necessary labeled measurements for real indoor positioning applications related to in-home monitoring and pedestrian navigation. Both contexts have been selected since they are uncontrolled environments (no control about AP placement and management) and they are representative enough of the current research and commercial trends. Moreover, the interest on these two main environments can be seen in the competitions held in related top conferences, where researchers have to deploy and evaluate their system on realistic large scenarios, including conference venues, university facilities, and living labs.

2 Lessons Learned at In-Home Scenarios

This section presents the application of indoor positioning to two study cases related to in-home monitoring. For both of them, details about each solution and experiences gathered with it are presented.

2.1 Calibration and Experimental Setup

The demand for higher comfort levels at home is increasing since some segments of the population, such as patients with mental disorders and senior citizens, spend most of their time at home. Remote monitoring, including in-home monitoring, and remote health care are valid alternatives for disease management in order to reduce frequent hospitalizations, including emergency visits, and to improve the patients quality of life. An IPS for remote monitoring operates over a very large environment, which is the aggregation of many “small” contexts (particular flats or houses) characterized by a specific user or device.

2.1.1 Smartphone-Based Patient Monitoring

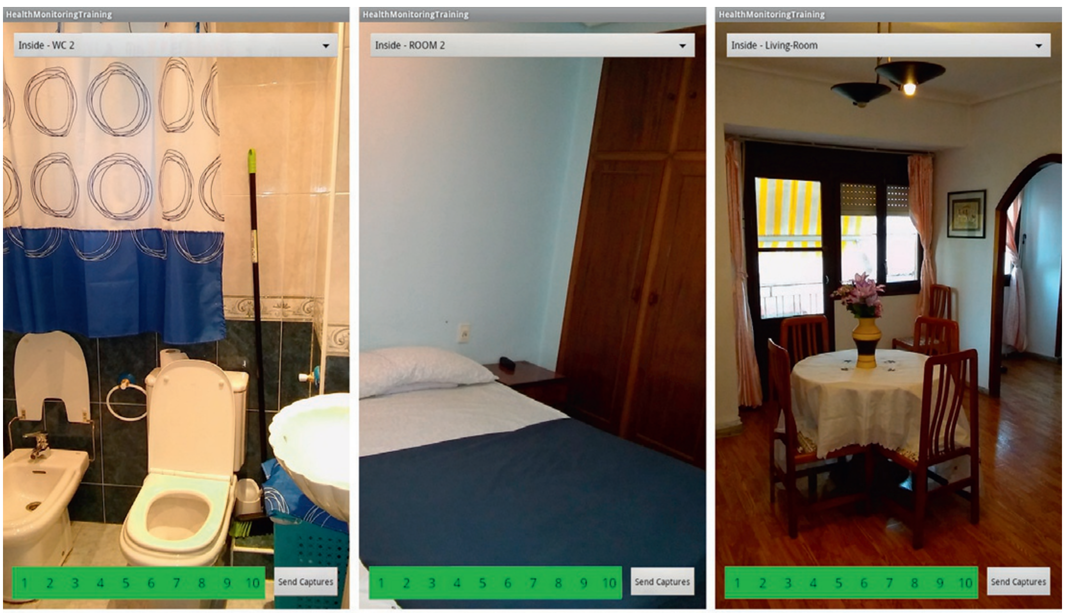
An Android application was developed in 2014 for monitoring patients at their homes, with a preliminary feasibility study being conducted to determine whether nonobtrusive Wi-Fi fingerprinting was suitable for in-home monitoring in Spanish flats. The indoor localization service developed for this application was a infrastructure-less solution since we relied on the already existing Wi-Fi network topology, that is, we did not deploy any extra device to support positioning, and we did not know the layout of each home. The IPS only processed the data gathered by smartphones to estimate the users’ position with room-level precision.

Eight different volunteers accepted to participate in the evaluation of target in-home monitoring system. We met them to provide the monitoring application and gave them some basic instructions about the four stages of our evaluation: installation, configuration, mapping, and operation. The volunteers were untrained people who had not used any Wi-Fi mapping application before. The only requirement to participate in the evaluation was to own an Android smartphone.

The feasibility study had eight very distinct in-home scenarios scattered in three cities (province of Castellón, Spain). The exact location and distribution of five of them were not provided to us due to privacy concerns. However, the detected APs indicated that they were not close to each other. Each AP was seen only at one of the eight scenarios. All volunteers installed the application when we met them and then they were distributed in four groups depending on the number of weeks (one or two) spent on the mapping and operation stages. Using such approach, we could get some additional feedback from the interactive stages.

In the configuration stage, the user introduced some important data about the flat: number of different rooms, room labels, room pictures, and rooms distribution, among others. To avoid mapping issues and improve the mapping application usability, supplying the room labels and taking pictures were compulsory. The configuration data and pictures remained in the volunteer phone and they were not provided to us. This stage was used to facilitate the mapping stage for the user and it took, an average, less than 6 min.

In the mapping stage, the user had to collect Wi-Fi fingerprints at each room within a period of 1 or 2 weeks (see Fig. 1). The application interface was as simple as possible, and they only had to select the room where they were using a combo box and click on a button in order to collect up to 10 consecutive Wi-Fi fingerprints (see Fig. 1). To avoid human errors while selecting the room, the picture attached to the room was shown before letting the user click on the *Collect* button. This approach considerably reduced the chances of selecting a wrong room label during the mapping procedure, thus avoiding entering errors into the IPS reference database. After this stage, a basic model based on 1-NN was created in each volunteer's phone and the reference database (Wi-Fi fingerprints and location labels) was sent to a centralized server. Only one volunteer quit the evaluation at this stage due to external factors.



(A) (B) (C)

FIG. 1 Example of the Health Monitoring application at the mapping stage (smartphone) in the guest bath room (A), guest bed room (B) and living room (C).

Once the mapping stage finished and the in-home monitoring mode was activated on the phone, the system collected fingerprints every 10 min. At the operation stage, the fingerprints were processed on the phone to estimate the user's position. Once a day, the collected fingerprints and the predicted locations were sent to the centralized server. Moreover, we also allowed an active localization with feedback where the user could ask the system for the actual position and correct it in case of a wrong classification. These "active" fingerprints, the predicted location and the actual location, were also sent to the centralized server. This last stage also lasted 1 or 2 weeks.

After the 2–4 weeks evaluation, the users provided their comments about configuring and using this in-home monitoring application. We will focus only on the experiences they reported about mapping and collecting the active location fingerprints with feedback.

2.1.2 Smartwatch-Based Patient Monitoring

Similarly, we developed an Android Wear application for monitoring people at their homes and we also performed a preliminary feasibility study to determine whether nonobtrusive Wi-Fi fingerprinting with smartwatches was suitable for in-home monitoring in Spanish flats.

All scenarios used during the experimentation phase were urban flats, three of them located in the city of Castellón de la Plana and the other two in the city of Valencia. A total of five people participated in this study. The flats located in the same city are far enough to each other not to share any common AP.

In the calibration stage, the smartwatch asked the user to go to every place and, once there, to start the collecting fingerprints for that place (see Fig. 2) by just pressing a button on the screen and waiting during, more or less, 1 min.

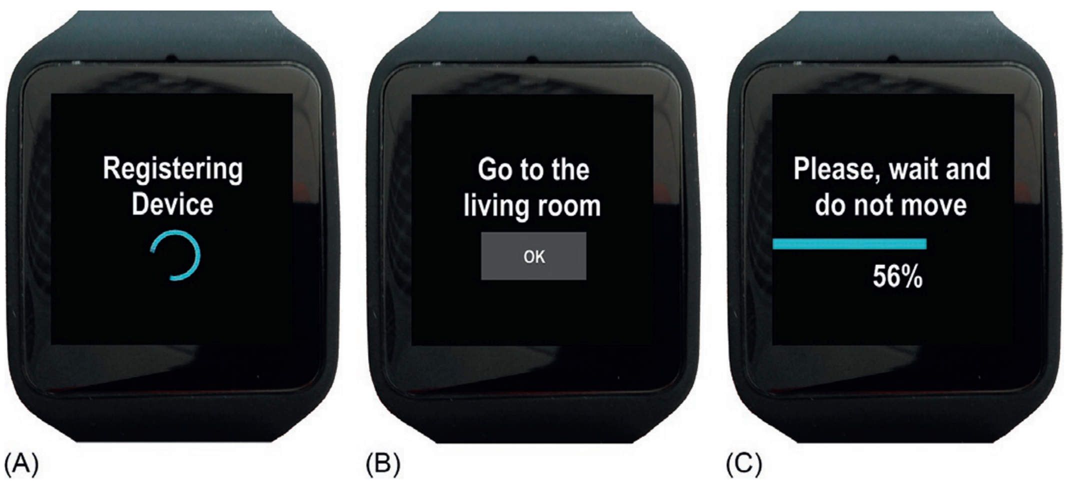


FIG. 2 Example of the Health Monitoring application at the mapping stage (smartwatch) when registering device (A), indicating a place to collect fingerprints (B) and collecting fingerprints at the selected place (C).

In the operational stage, a background process running in the wearable device collected one fingerprint per minute. Those fingerprints were stored in an internal database and sent to the server periodically. Moreover, two additional operations were allowed at this stage: (1) send new samples for recalibration and (2) force active localization to obtain the current position (feedback).

In contrast to the smartphone-based alternative presented before, the proposed monitoring system was less intrusive. After the 3 weeks evaluation, the users provided their comments about configuring and using this in-home monitoring application. As done before, we will focus only on the experiences they reported on mapping and collecting the active localization fingerprints with feedback.

2.2 Experiences and Lessons Learned

In-home data were collected to assess the feasibility of two different Wi-Fi-based monitoring systems. In most applications, like e-health and Aging-Assisted Living (AAL), only the location of the user at room level is needed.

2.2.1 Smartphone-Based Patient Monitoring

To perform the study with smartphones, the volunteers collected data in two stages. The data collected in the first and the second stages were used to train and test the solution, respectively. These data were collected in seven distinct scenarios (the volunteers' homes) and the stages' collection times were at least 1 week apart for all cases. Volunteers were asked to collect data at different hours throughout the day, to avoid that the usage patterns of neighbors' wireless networks to bias the measurements. This was not always possible due to work restrictions, so most data were taken at weekends (Table 1).

The first difference arose from the layout of the departments, whose area ranged from 76 to 114 m^2. Although all of them were one-floor flats, some of them had a square-like layout, and others had a rectangle-like layout.

The size and layout had a huge impact on the volunteers. In fact, the volunteer from scenario 1 reported that capturing one round of measurements in all the rooms was

Table 1 Main Characteristics of the Scenarios

Scenario	Area (m^2)	No. of APs	$\overline{APs}$	Time (s)	Training	Validation
1	114	127	15.78	12	5808	4313
2	76	103	14.90	5	1018	740
3	91	81	5.37	7.5	897	804
4	95	108	17.65	11	2019	1873
5	89	107	13.88	11	1900	1890
6	99	104	27.96	3.5	3627	2929
7	97	123	15.77	44	1960	1723
				Total	17,229	14,272

exhausting, whereas the user from scenario 3 collected multiple rounds of measurements without any complaint. Moreover, one complaint from volunteers was that the number of rounds of measurements was not defined. The previous facts explain the difference in captured fingerprints in the scenarios (more than 10,000 fingerprints in scenario 1 and ≈1800 in scenarios 2 and 3). According to the volunteers, they did not have any feedback about the quality of the collected fingerprints, so they could not determine when to finalize the training stage.

A reported issue related with the quality of the device used during the acquisition stage was the time each measure lasted. The time was 3.5 s on average in the best case for a high-end mobile phone, and 44 s for the worst device used in the acquisition stage. If the worst case is taken apart, the mean time for acquisition is about 10 s, which was a quite reasonable time for all users. The volunteers, specially those having a problematic device, suggested to add a timer to avoid being stuck at a position for a long time.

Finally, the battery drainage was a problem highlighted by a volunteer that took many measurements in weekdays after work. The battery status of the device was poor (a 2-year-old device) and a few mapping rounds could not be finished because the battery drained. The volunteer asked to implement a safe-energy mode or to avoid capturing fingerprints when the battery was below a safety threshold.

2.2.2 Smartwatch-Based Patient Monitoring

The experiments for the smartwatch-based monitoring were performed with the training and re-calibration modes (normal data collection and repeated collection, respectively, presented in Section 2.1.2) of the smartwatch application. Five volunteers participated in this study (see Table 2).

Four databases were created for each scenario. A total of 50 fingerprints per location (room) were considered in any of the 4 datasets, with each dataset differing from the others to cover different situations:

- Set 1: The user was standing up while collecting training fingerprints.
- Set 2: The user was moving around the room while collecting training fingerprints.
- Set 3: The user was standing up while collecting re-calibration fingerprints.
- Set 4: The user was moving around the room while collecting re-calibration fingerprints.

Table 2 Scenario characteristics

Scenario	Area (m^2)	No. of APs	Locations Mapped
1	120	33	Kitchen, office, living-room, bedroom
2	80	36	Kitchen, office, living-room, bathroom
3	90	27	Kitchen, office, living-room, bedroom
4	80	43	Kitchen, office, living-room, bedroom
5	62	23	Kitchen, office, living-room, bedroom

Sets 1 and 2 were collected during the first week of the experimental setup (calibration stage), whereas sets 3 and 4 were collected 2 weeks later (re-calibration at the operational stage).

A few users who participated in the experiments with smartphones also participated in the experiments with smartwatches, just to track the evolution of the procedure to collect fingerprints with a less invasive device.

First, the time required to capture the fingerprints at one location was not excessive according to the feedback provided by the volunteers than also participated in the previous experiment (with smartphones). Moreover, the fact of being guided was considered a better approach. However, some volunteers suggested to add a “skip location” button to avoid capturing fingerprint in a nonaccessible location (e.g., the bathroom when someone else is inside).

In general, the procedure to collect the data for this experiment was faster than for the previous one because: (1) people where guided about how, where, and how much measurements were to be captured; (2) the total number of fingerprints required was defined before starting the experiments (we avoid relying on the user to stop data capture) and (3) the smartwatch was an additional device provided for this experiment (the volunteers did not have to use their personal mobile device).

3 Lessons Learned at Very Large Scenarios

This section presents several efforts iteratively performed to apply an indoor positioning solution to the large scenario of a university campus. In a way similar to the previous section, details about each effort and experiences gathered with them are presented.

3.1 Calibration and Experimental Setup

There is a raising interest in providing applications and custom services inside shopping centers, airports and railway stations, and public institutions. Commercial IPS would be helpful for reaching a place of interest indoors, for example, a classroom in a university campus or the boarding gate in an airport. However, they can also be used for safe evacuation in case of an emergency disaster, or for monitoring and tracking patients in a hospital. These kinds of scenarios have common features: they are very large and heterogeneous scenarios.

We began to develop an Indoor and Outdoor Positioning System in 2013 (Torres–Sospedra et al., 2015). The main objectives were: improving mobility throughout the campus, obtaining high location accuracy indoors and outdoors, minimizing costs, being as less intrusive as possible, and developing as a smartphone-based application. Thus, we decided to develop a Wi-Fi fingerprinting technique for indoor positioning. We consider that our navigation application was a infrastructure-less solution since we did not install any Wi-Fi antenna for localization purposes and we only use the information gathered by smartphones to provide indoor location services.

Wi-Fi fingerprinting mapping at the university campus was done in six very differentiated stages, where different strategies and applications were carried out. In the first, second, and third stages, we collected the data that yielded the *UJIIndoorLoc* database as a result, including the public and private datasets. In the fourth stage, we mapped the building where the 17th AGILE Conference on Geographic Information Science was held. The fifth stage was done in parallel to an official Wi-Fi quality and coverage study at the university facilities. The last stage collected a year-long measurements at the university's library.

- The first stage consisted of collecting Wi-Fi fingerprints for the UJIIndoorLoc training/reference dataset. The fingerprints were taken at 933 reference points located in 3 different multifloor buildings (Universitat Jaume I). Although 18 people (mappers) participated in this process and installed the label-based mapping application on their phones, we divided the whole scenario into smaller areas. Each area was mapped by, at least, two different mappers with different Android devices (smartphone or tablet). The application interface was as simple as possible, and they only had to select their reference point using four combo boxes and click on a button in order to collect up to 10 consecutive Wi-Fi fingerprints (see Fig. 3A). Almost 20,000 valid fingerprints were collected.
- The second stage consisted of collecting Wi-Fi fingerprints for the UJIIndoorLoc public evaluation/testing dataset. The fingerprints were taken at any place inside the three previously mentioned buildings, that is, there was not any reference point. The scenario was divided into smaller areas as it was done for the first stage. The map-based mapping application collected one fingerprint and estimated the current position. Then, the user had to pin the actual current position on a map and click on a button to store the fingerprint, the estimated position, and the actual position (see Fig. 3B). This mapping procedure was simpler and more effective. More than 1000 valid fingerprints were collected at this stage by 12 people with different devices.
- The third stage consisted of collecting Wi-Fi fingerprints for the UJIIndoorLoc private evaluation/testing dataset. The fingerprints were also taken in any place inside the three previously mentioned buildings with another map-based application. Only six volunteers participated in this stage and installed a map-based mapping application on their phones. The interface was simpler and they only had to pin their real position and click on a button in order to collect up to 5 or 10 (depending on the user) consecutive Wi-Fi fingerprints (see Fig. 3C). More than 5100 fingerprints were collected at this stage.
- The fourth stage consisted of collecting samples in the Faculty of Law and Economics main building. We developed some applications for the 17th AGILE Conference on Geographic Information Science, which required indoor positioning techniques, so we mapped this building with precision in order to support the conference applications. At this stage, 18 people participated in collecting the Wi-Fi fingerprints. They installed another map-based mapping application on their phones, so they only had to pin

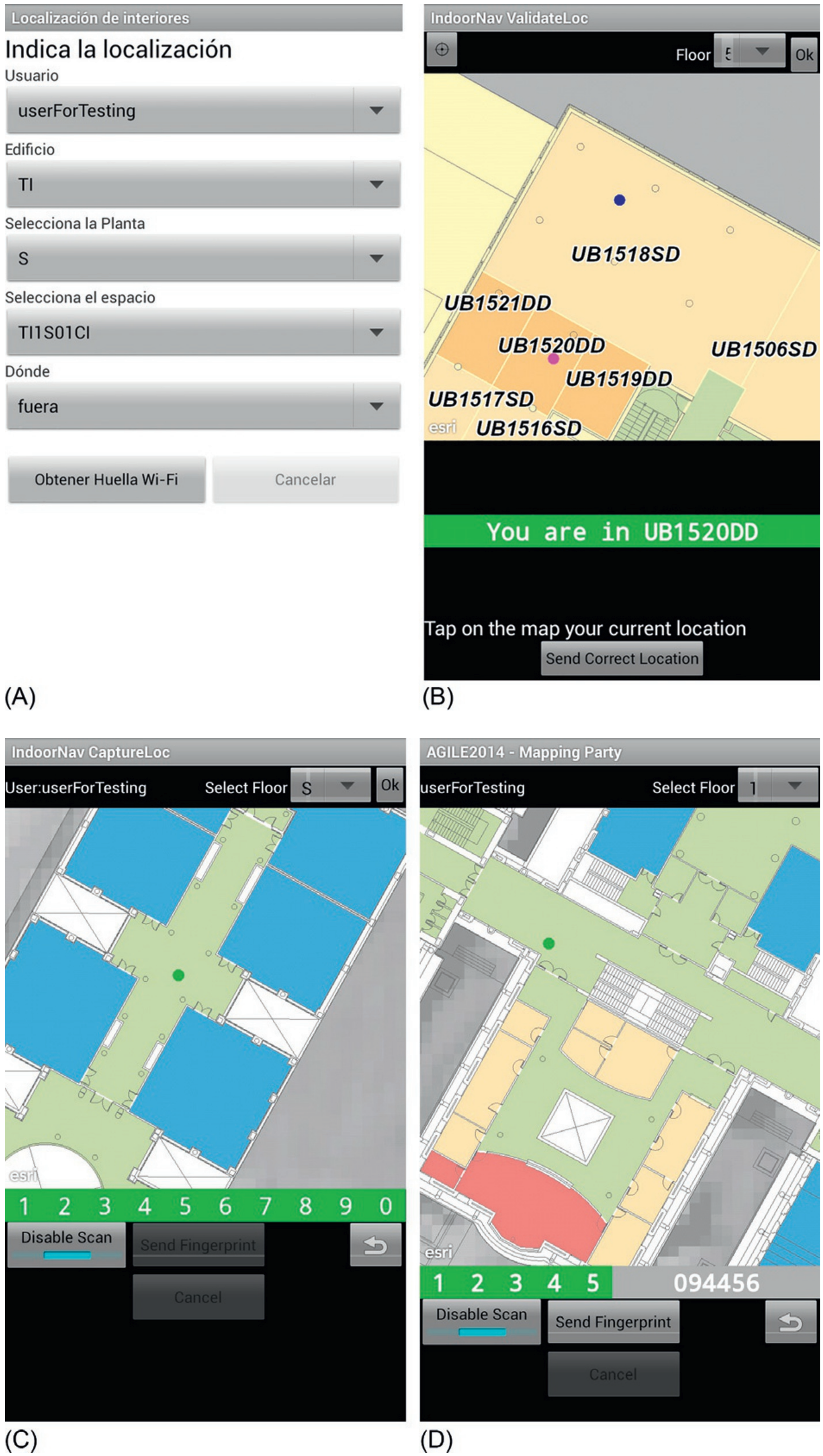


FIG. 3 Example of the UJI Monitoring applications for the first (A), second (B), third (C) and fourth (D) stages.

their real position and click on a button in order to collect up to five consecutive Wi-Fi fingerprints. In this application, we showed a timestamp code and we also introduced some warning messages to avoid errors (see Fig. 3D). In this case, we did not divide the whole scenario in smaller areas, but we divided it into routes. We calculated the most common routes among all the conference rooms and places and each route was mapped by, at least, two different people with different devices. Moreover, two users mapped the whole building (including stairs and the conference rooms) with higher precision. More than 12,000 fingerprints were collected by the 16 route-based volunteers and more than 9000 fingerprints were collected by 2 researchers with a background in Wi-Fi fingerprinting mapping.

- The fifth stage consisted in a full mapping of all the buildings located at the UJI's campus. We developed a Wi-Fi quality application that simultaneously run some Internet connectivity tests and recorded Wi-Fi fingerprints. This application was used once at every single facility (office, classroom, laboratory, among others) in the university to gather the information required to create Wi-Fi coverage and quality maps, and to collect up to 10 consecutive fingerprints (see Fig. 4). Only one

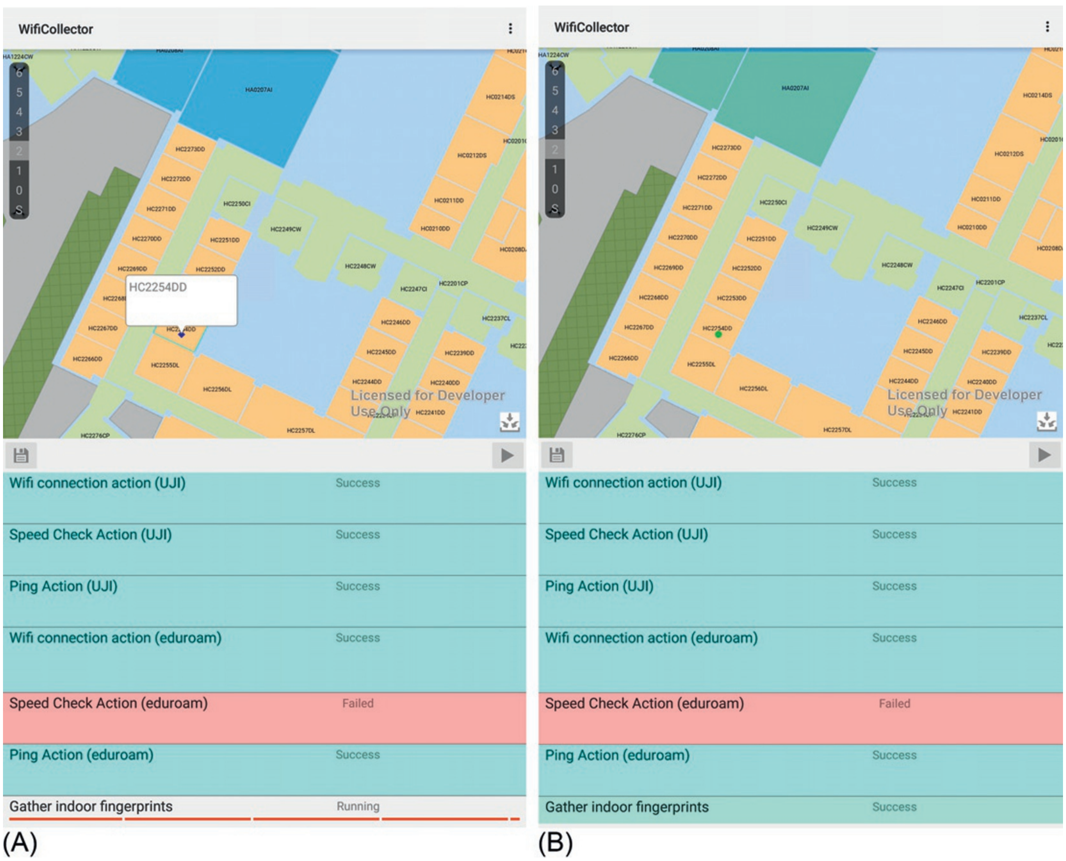
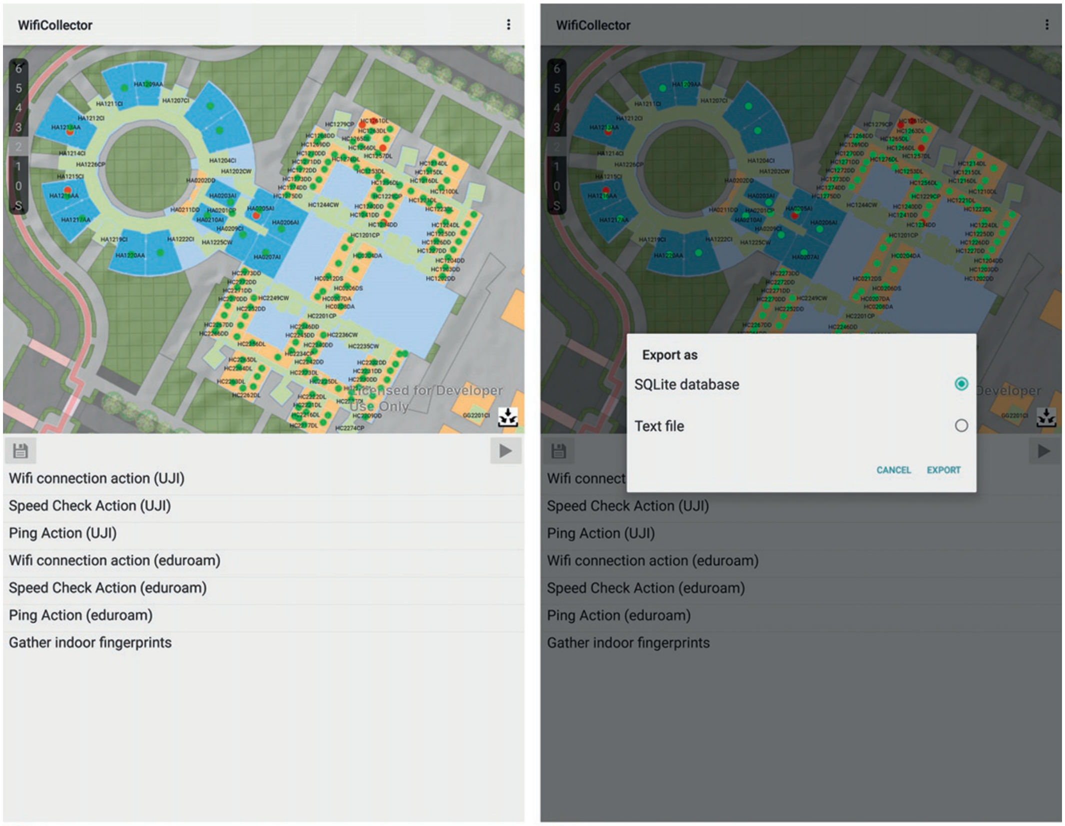


FIG. 4, CONT'D See legend on next page



(C) (D)

FIG. 4 Example of the UJI Wi-Fi quality test application while gathering Wi-Fi fingerprints (A), after collecting all fingerprints (B), showing locations already mapped (C), exporting the database (D).

professional user mapper participated in this stage, using a Samsung Galaxy Tablet. More than 22,000 fingerprints were collected at this stage.

- The last stage created a long-term, dense mapping of mid-sized areas among the bookshelves of two floors at the UJI's library building. A trained professional performed the mapping using a Samsung Galaxy S3 smartphone and a new application. The application organized the signals measurement process around campaigns defined by an organizer and composed of ordered capture locations. Each capture location defined guidance aids for the mapper, that is, map's zoom and orientation, the current capture location, the direction that the mapper should face when collecting the batch of fingerprints, the locations already measured and those pending for collection (Fig. 5). The application let the mapper choose which available campaign to work with. The batch of fingerprints to be taken for each location contained six (the first one is later discarded) fingerprints. The mapping process did

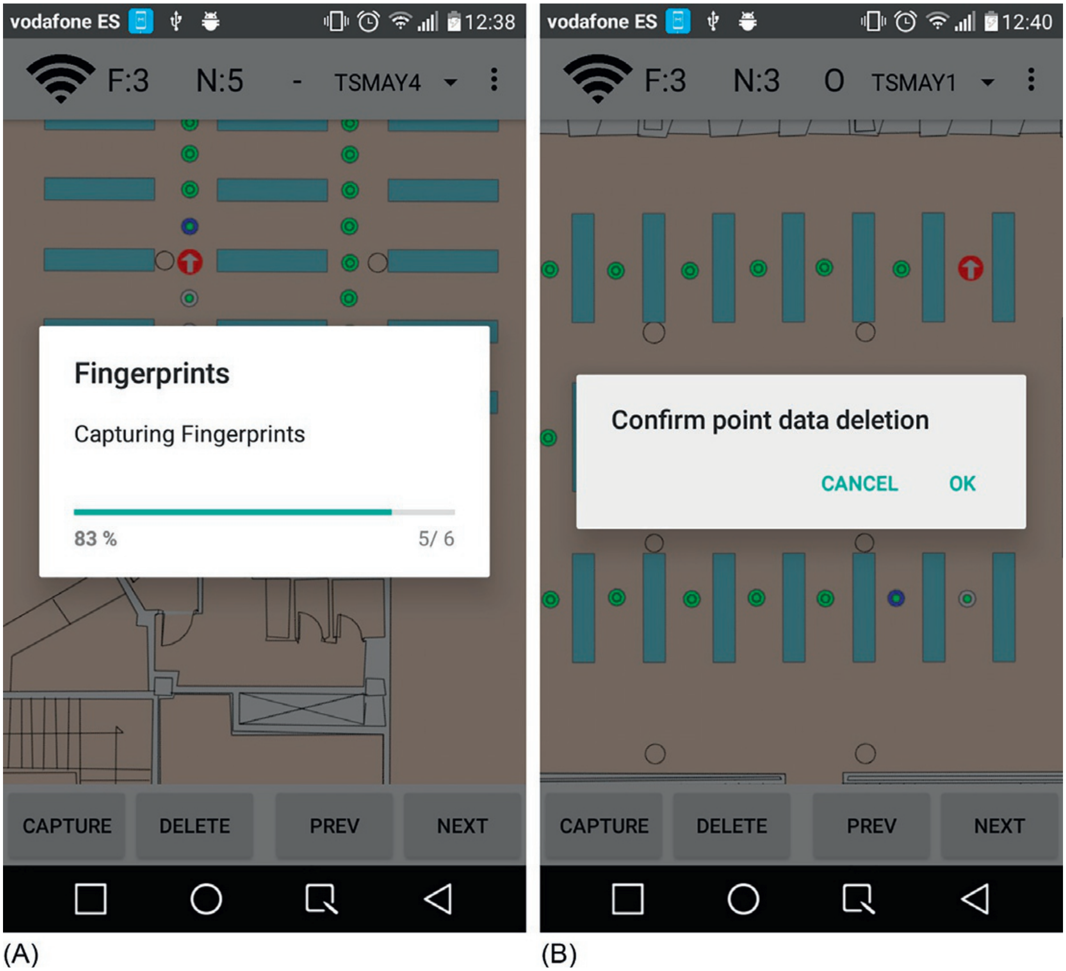


FIG. 5 Example of the latest UJI Wi-Fi campaign-based application while capturing Wi-Fi fingerprints (A) and while deleting a database record (B).

not require an Internet connection. The number of fingerprints captured at this stage was 46,656.

After finishing each stage, the users reported the issues they had during the mapping procedure and they also reported their experiences about using the different mapping applications. We will focus on the experiences reported by a total of 28 people that participated in such mapping stages with different mapping applications. As far as we know, there are not any previous studies about mapping in a large scenario including such quantity of people, devices, and mapping strategies. Some of those 28 volunteers participated in more than one stage, and they were able to compare the procedures and applications used in one stage to the previous one, and provide their valuable feedback to us. The volunteers who had no previous experiences with Wi-Fi mapping also provided valuable information, as they were not biased by previous mapping approaches.

3.2 Experiences and Lessons Learned

Mapping the UJI's university campus was done in six well-differentiated stages. The first three stages corresponded to the generation of the public and private subsets of the *UJIIndoorLoc* database, and their fingerprints were collected in three different buildings at UJI's campus. The fourth stage was done to support the applications for a conference, whose venue was the main building of the Faculty of Law. The fifth stage was collected in parallel with a Wi-Fi quality evaluation in all the buildings of the UJI's Campus. The last stage collected measurements during a year at areas among bookshelves from the university's library.

Prior to the first stage, the *Indoor Positioning System* architect, the *Database* designer, and the *APP* developers tested the alpha and beta versions of the whole system. After some tests, they decided to release a label-based application to collect the fingerprints (see Fig. 3A). Due to the variability of Wi-Fi signals, 10 consecutive fingerprints were collected at each reference point.

3.2.1 First Stage

In the first stage, we provided a label-based application (Fig. 3A) to 18 volunteers in order to collect the training samples for the *UJIIndoorLoc* database. Moreover, we provided them with: (1) a list of the areas and reference points where they had to collect the Wi-Fi fingerprints and (2) a notebook to manually log this process and annotate the detected issues. In June 20, 2013, they collected almost 15,000 valid fingerprints, the 75% of the UJIIndoorLoc training samples. With this process, we collected most of the fingerprints with high diversity in approximately two-and-a-half hours.

After collecting all the training fingerprints, we revised all the collected fingerprints that the volunteers marked as issue. We manually fixed all the issues when possible, but we had to discard 750 fingerprints. The most common issues detected by the volunteers were:

- Wrong user. *The application crashed and I forgot to set my user. The fingerprints I collected with the default user are valid.*
- Wrong label about reference point. *I selected the XYZ label instead of YXZ at 12:34 a.m.*
- Wrong floor. *I selected the X floor instead of Y at 12:34 a.m.*
- Unknown reason. *The fingerprints I took from 12:34 a.m. to 12:45 a.m. are not valid.*

After the data correction, we matched the information provided by fingerprints with the manual logs provided by the volunteers. We detected that two volunteers did not report a few wrong label and wrong floor issues. So, we manually removed 125 more fingerprints from the database.

The volunteers reported that the office identifier was not appropriate to denote the reference points and, in some cases, it induced errors. So they suggested to use a map to improve the mapping task experiences.

At the end of this first stage, we realized that a label-based application was not appropriate for this kind of large-environment mapping because the volunteers felt uncomfortable during all the mapping procedure. Mapping 10 consecutive fingerprints at each reference point was very tedious according to the experiences reported by the volunteers. Despite the simplicity of the application interface, they had to stand about 1 or 2 min at each reference point in order to collect the fingerprints and log them on the notebook.

On the management side, a significant amount of time was spent inspecting fingerprints to assure the collection's quality. The fingerprints inspection included manually looking for inconsistencies in the temporal marks order, as well as considering the issues explicitly reported by the volunteers. We manually fixed about 21% of collected fingerprints, with 15% of them successfully edited and the remaining 6% removed. Afterward, a basic IPS was implemented using the 19,937 fingerprints collected at the first stage as reference data.

3.2.2 Second Stage

In the second stage, we developed a map-based mapping application (see Fig. 3B). This application collected one fingerprint, the IPS returned the estimated position and the user only had to select his/her current position if the estimation was not accurate. We introduced a map to avoid using labels in this large environment and it only collected 1 fingerprint (instead of 10) to perform a faster mapping. We provided this application to 12 volunteers, 2 of whom had not participated in the previous stage, to collect the fingerprints for the validation dataset (public test). Moreover, we provided them with: (1) a map with the areas (without explicit reference points) where they had to collect the Wi-Fi fingerprints and (2) a notebook to manually log this process and annotate the detected issues. On September 20, 2013, the volunteers collected almost 700 valid fingerprints. The 62.5% of the UJIIndoorLoc validation samples were taken in approximately 70 min.

The volunteers performed an initial mapping in which they simultaneously mapped the same area for a few minutes, and they were supported by the application developers. This initial mapping served as a training step on the application and the fingerprints were marked to be removed. According to the feedback provided by the volunteers, this initial fake mapping was useful to become familiar with the application.

After the initial mapping, the volunteers moved to the target collection areas. Although we improved the application interface, some issues were detected by users. We manually fixed all the issues when possible, but we had to discard 17 fingerprints. In this stage, the volunteers detected two issues:

- Wrong floor. *I did not realize that the system detected me in the wrong floor and I submitted my position with the wrong floor at 12:34.*
- Wrong location. *I wrongly established my position on the map at 12:34* or *I established my position on the map but I accidentally tapped another position when uploading the information at 12:34.*

Afterward, we matched the information provided by fingerprints with the manual logs provided by the volunteers. We only detected that two volunteers did not report one wrong location issue at this stage, so we removed the related fingerprints from the database. In fact, we detected more nonreported errors but they corresponded to the initial mapping, whose fingerprints had been previously marked as invalid and had not been considered for the final database.

At the end of the second stage we realized that using a map-based application is very useful for mapping since the visual information helps reducing errors. However, the negative point about this mapping is the density of fingerprints collected. We only collected an average of 0.90 fingerprints per minute and volunteer in this second stage, in contrast to the 5.55 fingerprints per minute and volunteer gathered at the first stage. Although the time required per capture has been the same in the two stages, 1 min approximately, the mapping procedure in the second stage has been more interactive and the volunteer felt more comfortable.

3.2.3 Third Stage

In the third stage, we combined some features from the first and second stages. This stage was done in two different periods of time, one in November 2013 and the other in March 2015. We improved the map-based application to collect 10 consecutive fingerprints for the first mapping period (see Fig. 3C), and we introduced two minor changes in the application for the second mapping period (see Fig. 3D). We provided the application to six volunteers, three of whom had not participated in the previous stages, to collect the fingerprints for the private test dataset. This dataset was used in an EvAAL-ETRI competition.

Each volunteer had to map, at least, one of the three buildings. Moreover, we only provided them with a notebook to manually log this process and annotate the detected issues. In November 2013 and March 2015, the volunteers collected (3779 + 1395) valid fingerprints for the private test dataset. Prior to each period, we trained the volunteers to use the application.

Although the mapping was successfully done without any important issues, the users suggested some minor changes to improve the mapping applications:

- The map size was not optimized for new versions of Android and it was too small under some configurations.
- Some devices, specially those supporting the 5 GHz Wi-Fi band, collected the fingerprints significantly slower than the others.

3.2.4 Fourth Stage

In the fourth stage, the scenario was the main building of the Faculty of Law and Economics, Universitat Jaume I. The main aim of this mapping stage was to collect the Wi-Fi fingerprints to support indoor positioning in the applications for the 17th AGILE Conference on Geographic Information Science. In this mapping stage, 2 experienced people performed a comprehensive mapping, whereas 16 volunteers focused in short common routes and in the main areas where the conference took place. The mapping was done during 4 different days in May 2014. The mapping application (see Fig. 3D) introduced minor changes with respect to the mapping application used in the third stage (first period), being the most remarkable ones: (1) the map fitted better on any display size, (2) the number of consecutive fingerprints was reduced to five, and (3) we introduced a capture identifier, so people could more easily log the fingerprints. These minor changes were also applied to the application used for the second mapping period of the third stage (March 2015). In essence, the mapping applications used for the third and fourth stages were very similar.

Although the route mapping was successfully done, the users who performed the comprehensive mapping reported that it was not possible to send the fingerprints to the server in a few areas out of 3G/4G and Wi-Fi stable coverage. Moreover, all volunteers had to use the 3G/4G Internet connection to send the fingerprints to the centralized server instead of the Wi-Fi. Due to the high density of Wi-Fi antennas and high presence of people in classrooms, some users reported issues with Wi-Fi connectivity in some areas. The problem was that the devices connected to the UJI's Wi-Fi networks tended to constantly disconnect from one AP and connect to another one with the strongest signal. This caused severe delays while collecting and sending the fingerprints to the server, so we decided to announce volunteers to quit using Wi-Fi and switch to 3G/4G connections.

The volunteers who performed the route-based mapping felt comfortable with the collection approach. The comprehensive mapping took approximately 7 h, with an average of 16 fingerprints per minute, for the first experienced mapper and about 9 h, with an average of 7.5 fingerprints per minute, for the second experienced mapper. In some places, the first experienced mapper collected more than 20 fingerprints per minute. Both experienced users reported that the comprehensive mapping was very demanding due to the size of the environment and the comprehensive mapping strategy.

Only one volunteer reported severe errors and his fingerprints were not considered in the fourth stage. He received a phone call and after 5 min he got lost in the building. He did not know how many fingerprints had collected and the location of the last one, and he finally decided to quit the mapping process. We realized that we should have included in the application information about the fingerprints that the user, and other users, had collected to avoid these kind of situations.

3.2.5 Fifth Stage

With the experience gained in the fourth stage, we completely rebuilt the mapping application in order to be a multipurpose application (see Fig. 4). Although it allowed Wi-Fi tests and gathering Wi-Fi fingerprints, it could easily be extended to perform any test and collect data from any sensor embedded in the device. We highly improved the user interface to make it adaptable to a high diversity of devices and screen sizes; we added user configuration parameters and new functionalities:

- Maps were embedded in the application and they were simplified to improve the user experience in mobile devices. We realized that a simple but effective interface reduces the issues in the mapping process and improves the quality and size of the reference databases (see Fig. 4A and B).
- The number of consecutive Wi-Fi fingerprints or a timeout could be set to avoid time issues in those devices whose Wi-Fi refresh rate is very low.
- We showed the information of the previous gathered data in the map as red and green bullets (see the gray bullets in the black and white screenshot showed in Fig. 4C). The user knew where mapping had been done, as well as the status, success (green/light-gray bullets) or failure (red/dark-gray bullets), of each individual capture.
- The mapping application stored all gathered data in an internal database: quality tests, Wi-Fi fingerprints, and the information provided by other sensors. The user decided when to send this information to a centralized server (see Fig. 4D). Therefore, we avoided those cases in which severe communication problems, mainly due to low or null coverage, did not allow to register the fingerprints in the centralized server.

In the fifth stage, all the buildings of the UJI's Campus were mapped by a single professional person, who was explicitly hired to perform this task. Due to the coverage problems detected in previous stages, the new application collected the Wi-Fi fingerprints offline and they were registered in the centralized server when all the Campus was already mapped. The application was set to capture between 5 to 10 fingerprints per observation. This fifth mapping lasted 24 working days, starting on April 20, 2015 and ending on May 20, 2015. A total of 2700 observations (Wi-Fi tests) were made and 25,000 fingerprints were collected at 2300 different locations during this stage. Performing the quality tests and gathering the Wi-Fi fingerprints took, approximately, 2 min per each location (office, classroom, laboratory, among other spaces). In particular, 4.5 fingerprints per minute were taken with this professional mapping strategy.

Although the mapping was successfully done without any important issues, the professional user suggested some minor changes to improve the mapping applications:

- More detailed maps were required. He collected the fingerprints approximately at the center of each classroom or office. However, it was not easy to select his position with high precision in the map since the maps did not included the furniture in classrooms. The professional user suggested the inclusion of an additional layer with furniture found in the classrooms and the library.
- Include predefined reference points instead of manually selecting them in the map, since most of the measurements were collected at the geometric center of the surveyed classrooms.
- Add the feature to remove a set of measurements. Although the systems warn about the user's position, he had to triple check his position before taking the measurements, since there was no option to remove the captured measurements.

Moreover, the professional mapper suggested us to include a functionality to add predefined reference points with detailed instructions in the application. According to his experience, nonexpert mappers (without knowledge of radio signal propagation) may suffer stress while taking the measurements because they do not know if they are in the appropriate place to collect the fingerprint. In such case, detailed instructions (reference points, orientation, among other features) must be provided to them. In other words, we might not rely the responsibility of selecting the reference points on nonexperts.

3.2.6 Sixth Stage

This stage demonstrated that incorporating the mapping instructions to the application makes the mapping process less cumbersome, more accurate, and faster. The users no longer have to carry hard copy instructions to follow a well-planned mapping campaign. They just need to focus on the map and instructions from the application. They may finish one campaign and then switch to another one, and upload the campaigns' data at the time of their choosing.

In the mapping process with a map-based application, mappers may make two types of errors: (1) place themselves at wrong locations, and (2) wrongly indicate their current location on the application's map. In this stage, the second type of error is no longer present because users are not required to indicate their current floor or tap their location on the application's map. The application's design reduces the likeliness of the first type of error by:

- map's zoom and orientation adjustments at each location;
- indications of the facing direction at each location;
- actions for traversing the list of locations and deleting the captures that the user may consider erroneous; and
- indication of the point that follows the current one in the capture process.

The collection in this stage was performed by a professional mapper. With the new approach, almost the mapping time employed by the mapper was devoted to fingerprint capture. For example, for a campaign totalizing 96 fingerprint batches, each batch composed of 6 fingerprints, the mean time to complete it was 50 min. As the mean time for capturing one batch was 30 s, a mean of only 2 min was used for displacements, orientation, and other tasks. Among the factors influencing the huge time utilization are follows:

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

The new mapping approach provided notable results in terms of error avoidance and time utilization. The application improvements were accompanied by the automation of other tasks performed by the campaign organizers, like a tool for campaign definition, captured data validation (amount, order, and direction of collected data points), and data transformation. The new approach has consolidated the idea of providing a platform for the smartphone-based citizen-science data collection.

In general, mapping in large scenarios has been a complex task. Several factors have to be properly balanced:

- size of the environment;
- number and kind of people involved in mapping;
- density of fingerprints;
- distance between reference points; and
- distribution of reference points.

Other large environments may add new challenges to those we have faced in our university, for example, difficulties for mapping at very crowded places (e.g., in shopping centers during sales season), and the need of performing the same mapping at several times of the day due to environment variations (e.g., check-in areas in airports may be very crowded at some times and completely empty at others, which influences the measured Wi-Fi signal intensities). In general, the mapping process requires previous careful considerations of the target environment.

4 General Experiences

This chapter has shown the lessons learned from mapping large environments for Wi-Fi fingerprinting. Despite two very different contexts have been considered (in-home monitoring and pedestrian navigation inside a campus), the experiences have been positive in both of them. On the one hand, the in-home monitoring is interesting since many homes have been considered, and thus it can be seen as a large area composed by many independent small areas. On the other hand, the campus' size resembles a small city or a neighborhood of a big city, which is of interest for real deployments of indoor positioning in smart cities. In conclusion, the experience gained from both environments was positive and allowed us to prepare increasingly better mapping campaigns.

For in-home monitoring with smartphones, using different devices means different Wi-Fi hardware that can bias the feedback from users. The time required to collect a single fingerprint varied depending on the device. Moreover, giving freedom to users was seen positive by some volunteers, which collected the fingerprints at their own will without any intrusion. However, a few volunteers stated that they could not realize when additional an fingerprint capture was not required since they did not have a metric to know the quality of collected data. Moreover, battery drainage was highlighted as the major concern of the users.

The second strategy guided mapping with smartwatches was better suited for volunteers since they were totally guided and the number of fingerprints required to have an effective system was much lower. Minor details were pointed out by the volunteers to be improved in successive versions of the smartwatch in-home monitoring application.

The mapping strategy for collecting the *UJIIndoorLoc* database (public and private datasets) was very demanding and positive. We collected a realistic database for large multibuilding multifloor environments, which has supported all our developments in Indoor Positioning and Indoor Navigation. Moreover, it has public access through the *UCI Machine Learning Repository (University of California, Irvine, USA)*. Furthermore, this database has been used to evaluate the IPSs that participated in the EvAAL-ETRI Competition, which was organized in conjunction with the 2015 Indoor Positioning and Indoor Navigation Conference (October 13–16, Banff, Canada) (Potortì et al., 2015). Competitors and attendees reported that meaningful comparisons are possible with this huge database. Feedback provided by external undergrad students, external researchers, and, even, companies suggest that this database, and others, may serve as de facto standard to fairly compare different IPSs.

The strategy with 2 comprehensive and 16 route-based mapping was also very positive. Although reviewing this mapping was less demanding because severe issues were not reported, two people had to map the environment for some hours. The database was used to support the official applications for the 17th AGILE conference on Geographic Information Science. The feedback provided by the attendees who used the applications was positive because the applications were useful to attendees and they supported wayfinding to the rooms and other points of interest where the conference events took place.

The professional mapping experiences have shown us that mapping a large scenario, such as university campus, and making a dense mapping at a mid-sized scenario over a long time, are very demanding tasks. A basic mapping covering all the campus required 1 month. The dense, long-term mapping required at least 6 h a month during 12 months. Mapping for just one purpose may not compensate the efforts if the cost of a professional mapping is considered. However, professional mapping also sheds light on the interest of integrating different tests and observations about the environment, and a reliable mapping can provide insights into (long-term) signals variability that would be otherwise untrustworthy. Developing a multipurpose application that collects information from diverse sensors and runs different test may be valuable. The campaign-driven application's value lies not only in improving the collection experience and reducing the errors, but also in its potentials for crowdsourced collection usage.

Assisting the mapper when collecting samples is also useful, since the mapper has only to go the places shown in the map. According to the data we have collected, having the responsibility of selecting the references points might require additional time and might produce stress in the people who collects the data.

Finally, we can state that the process of generating a reference database for Wi-Fi-based fingerprinting is very hard and demanding. Moreover, it requires constant refinement in order to avoid errors. Label-based mapping applications are appropriate for label-based positioning, such as in-home monitoring, when the number of labels is low. However, it introduces some errors during the ground truth generation if the environment is large and the number of labels is also large (a building with multiple offices or reference points). For big environments, such as a university campus, offline mapping applications are a suitable solution to generate the ground truth. Moreover, the users reported better feedback when an extensive mapping process was complemented with route-based mapping. In general, it seems that crowdsourcing will be well-established in the future, so future work will also be focused on these kinds of methodologies to generate and keep up-to-date the reference data for fingerprinting.

Acknowledgments

Parts of this work were funded in the frame of the Spanish Ministry of Economy and Competitiveness through projects TIN2015-70202-P and TEC2015-71426-REDT. Germán M. Mendoza-Silva gratefully acknowledges funding from grant PREDOC/2016/55 by Universitat Jaume I.

References

Ayadi, M., Torjemen, N., Tabbane, S., 2015. Two-dimensional deterministic propagation models approach and comparison with calibrated empirical models. IEEE Trans. Wirel. Commun. 14 (10), 5714–5722. ISSN 1536-1276. https://doi.org/10.1109/TWC.2015.2442572.

Berkvens, R., Weyn, M., Peremans, H., 2016. Position error and entropy of probabilistic Wi-Fi fingerprinting in the UJIIndoorLoc dataset. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6.

Čabarkapa, D., Grujić, I., Pavlović, P., 2015. Comparative analysis of the Bluetooth low-energy indoor positioning systems. In: 2015 12th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), pp. 76–79.

Chiou, Y.S., Wang, C.L., Yeh, S.C., 2010. An adaptive location estimator using tracking algorithms for indoor WLANs. Wirel. Netw. 16 (7), 1987–2012. ISSN 1572-8196. https://doi.org/10.1007/s11276-010-0240-8.

Deasy, T.P., Scanlon, W.G., 2007. Simulation or measurement: the effect of radio map creation on indoor WLAN-based localisation accuracy. Wirel. Pers. Commun. 42 (4), 563–573. ISSN 1572-834X. https://doi.org/10.1007/s11277-006-9211-x.

El-Kafrawy, K., Youssef, M., El-Keyi, A., Naguib, A., 2010. Propagation modeling for accurate indoor WLAN RSS-based localization. In: 2010 IEEE 72nd Vehicular Technology Conference, Fall, pp. 1–5.

Ezpeleta, S., Claver, J.M., Pérez-Solano, J.J., Martí, J.V., 2015. RF-based location using interpolation functions to reduce fingerprint mapping. Sensors 15 (10), 27322–27340. ISSN 1424-8220. https://doi.org/10.3390/s151027322. Available from: http://www.mdpi.com/1424-8220/15/10/27322.

Guimarães, V.G., Castro, L., Carneiro, S., Monteiro, M., Rocha, T., Barandas, M., Machado, J., Vasconcelos, M., Gamboa, H., Elias, D., 2016. A motion tracking solution for indoor localization using smartphones. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8.

He, S., Chan, S.H.G., 2016. Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. IEEE Commun. Surv. Tutorials 18 (1), 466–490. ISSN 1553-877X. https://doi.org/10.1109/COMST.2015.2464084.

Hernández, N., Ocaña, M., Alonso, J.M., Kim, E., 2017. Continuous space estimation: increasing WiFi-based indoor localization resolution without increasing the site-survey effort. Sensors 17 (1). ISSN 1424-8220. https://doi.org/10.3390/s17010147. Available from: http://www.mdpi.com/1424-8220/17/1/147.

Hossain, A.K.M.M., Soh, W.-S., 2015. A survey of calibration-free indoor positioning systems. Comput. Commun. 66, 1–13. ISSN 0140-3664. https://doi.org/10.1016/j.comcom.2015.03.001. Available from: http://www.sciencedirect.com/science/article/pii/S0140366415001115.

Li, W., Wei, D., Yuan, H., Ouyang, G., 2016. A novel method of WiFi fingerprint positioning using spatial multi-points matching. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8.

Liu, W., Chen, Y., Xiong, Y., Sun, L., Zhu, H., 2014. Optimization of sampling cell size for fingerprint positioning. Int. J. Distrib. Sens. Netw. 10 (9), 273801. https://doi.org/10.1155/2014/273801.

Liu, W., Fu, X., Deng, Z., Xu, L., Jiao, J., 2016. Smallest enclosing circle-based fingerprint clustering and modified-WKNN matching algorithm for indoor positioning. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6.

Markets&Markets, 2014. 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). Available from: http://www.researchandmarkets.com/reports/2570920.

Marques, N., Meneses, F., Moreira, A., 2012. Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning. In: 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–9.

Marti, J.V., Sales, J., Marin, R., Jimenez-Ruiz, E., 2012. Localization of mobile sensors and actuators for intervention in low-visibility conditions: the zigbee fingerprinting approach. Int. J. Distrib. Sens. Netw. 8 (8), 951213. https://doi.org/10.1155/2012/951213.

Mathisen, A., Sørensen, S.K., Stisen, A., Blunck, H., Grønbæk, K., 2016. A comparative analysis of indoor WiFi positioning at a large building complex. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8.

Mizmizi, M., Reggiani, L., 2016. Design of RSSI based fingerprinting with reduced quantization measures. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6.

Potortì, F., Barsocchi, P., Girolami, M., Torres-Sospedra, J., Montoliu, R., 2015. Evaluating indoor localization solutions in large environments through competitive benchmarking: the EvAAL-ETRI competition. In: Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation.

Raspopoulos, M., Laoudias, C., Kanaris, L., Kokkinis, A., Panayiotou, C.G., Stavrou, S., 2012. 3D ray tracing for device-independent fingerprint-based positioning in WLANs. In: 2012 9th Workshop on Positioning, Navigation and Communication, pp. 109–113.

Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J.P., Arnau, T.J., Benedito-Bordonau, M., Huerta, J., 2014. UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In: Proceedings of the Fifth Conference on Indoor Positioning and Indoor Navigation, pp. 261–270.

Torres-Sospedra, J., Avariento, J., Rambla, D., Montoliu, R., Casteleyn, S., Benedito-Bordonau, M., Gould, M., Huerta, J., 2015. Enhancing integrated indoor/outdoor mobility in a smart campus. Int. J. Geogr. Inf. Sci. 29 (11), 1955–1968. https://doi.org/10.1080/13658816.2015.1049541.

Zhang, M., Pei, L., Deng, X., 2016a. GraphSLAM-based crowdsourcing framework for indoor Wi-Fi fingerprinting. In: 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), pp. 61–67.

Zhang, W., Hua, X., Yu, K., Qiu, W., Zhang, S., 2016b. Domain clustering based WiFi indoor positioning algorithm. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–5.