UJIIndoorLoc-Mag: A New Database for Magnetic Field-Based Localization Problems

Joaquín Torres-Sospedra, David Rambla, Raul Montoliu, Oscar Belmonte, and Joaquín Huerta
Institute of New Imaging Technologies, Universitat Jaume I
Avda Vicente Sos Baynat S/N, Castellón, Spain
[jtorres , drambla , montoliu , belfern , huerta]@uji.es

***Abstract*—Indoor localization is a key topic for mobile computing. However, it is still very difficult for the mobile sensing community to compare state-of-art Indoor Positioning Systems due to the scarcity of publicly available databases. Magnetic field-based methods are becoming an important trend in this research field. Here, we present *UJIIndoorLoc-Mag* database, which can be used to compare magnetic field-based indoor localization methods. It consists of 270 continuous samples for training and 11 for testing. Each sample comprises a set of discrete captures taken along a corridor with a period of 0.1 seconds. In total, there are 40,159 discrete captures, where each one contains features obtained from the magnetometer, the accelerometer and the orientation sensor of the device. The accuracy results obtained using two baseline methods are also presented to show the suitability of the presented database for further comparisons.**

Keywords—Indoor Localization; Magnetic field; Database; Comparison of methods

I. INTRODUCTION

Many real world applications need to know the localization of a user in the world to provide their services. Automatic user localization consists of estimating the position of the user by using an electronic device, usually a mobile phone. Outdoor localization problem can be solved very accurately thanks to the inclusion of GPS sensors into mobile devices. However, GPS has severe problems in indoor environments. Many different approaches tried to solve the problem of indoor positioning in the last years. They can be categorized, according to [1], as infrastructure-based (RFID, infrared, ultrasound, Bluetooth) and infrastructure-less (Wi-Fi [5], FM radio frequencies [12], Magnetic field [3-4]) technologies.

The use of the Earth magnetic field for indoor localization is an interesting infrastructure-less method that is attracting the attention of many researchers in the last years. Indoor environments have some structures (ferrous structural materials, pipes, wires, etc.), which alter the Earth magnetic field. Even the presence of quotidian objects, such as metallic stoves or speakers, may alter the magnetic flux density in the surrounding areas [8]. Actually, the measured magnetic field can substantially vary between two points very close in the space. Thanks to that, sub-meter accuracy level can be theoretically achieved for indoor location. The variations in the magnetic field in indoor environments can be measured and recorded with available sensors inside smart phones [3, 7].

Although there are many papers in the literature trying to solve the indoor localization problem using a magnetic field-based method, there still exists one important drawback in this field, which is the lack of a common database for comparison of methods. Each approach presents its estimated results using its own database. Under these conditions, it is not possible to compare different methods since the particularities of each experiment are hardly reproducible. In the Pattern Recognition and Machine Learning research fields, the common practice is to test the results of each proposal either using a well-known dataset or providing the dataset used. In this way, researchers are able to fairly compare different methodologies in the literature. The UCI Machine Learning Repository [6] is a well-known example in this sense. In fact, there is an available database for comparing WLAN fingerprint-based indoor localization methods [2]. However, in the magnetic field-based indoor localization field does not exist such kind of database.

The main contribution of this work is the creation and the introduction of the *UJIIndoorLoc-Mag* database, which is the first publicly available database that could be used to make comparisons among different methods in this field. It has been published on the UCI Machine Learning Repository [6]: http://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc-Mag.

The database consists of 281 continuous samples (270 for training and 11 for testing) taken in our 260m^2 (15x20m approx.) laboratory. Each sample comprises a set of discrete captures taken along the 8 corridors (including intersections) of the laboratory with a period of 0.1 seconds. There are almost 40,000 discrete captures obtained from the magnetometer, accelerometer and orientation sensor of a mobile phone.

Two basic baselines are also presented to test the suitability of *UJIIndoorLoc-Mag* and they can be considered a simple starting point for further comparisons. We do not expect to obtain high accuracy with the baseline since we test the suitability of the database, we are not introducing a new indoor positioning system.

The rest of the paper is organized as follows. Section II presents the related work. Section III shows some prior tests we performed on magnetic field indoor positioning. Section IV introduces the main elements of the database and how it was made. Section V is devoted to explain the two baseline algorithms tested and the results obtained. Finally, Section VI presents the most important conclusions arisen from this work.

II. Related Work

As it has been commented before, 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 [3,4,7,8,9,10]. We focus on the dataset used for testing the proposed algorithms and we also show whether they are publicly available or not.

Four experiments were done in [3] 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, some basic information about the databases was not commented.

The experiment presented in [7] took place on a rectangular-shaped, 67x12 m^2, corridor where its surroundings included spaces such as lab, office, and library. So they considered and environment of 4 lineal 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 and their magnitudes were not described.

In [8], the authors demonstrated that geomagnetic localization performs reasonably well when the three components of the magnetic field - X, Y, and Z axes - 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 density at 1 m. spacing. Moreover, they also conducted a magnetic fingerprint test in a 3.5x3.5 m^2 bedroom. However, they did not detail the number of samples.

In [9], the authors selected a corridor of a multi-level 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, 2) 60 cm. left to the corridor the center and 3) 60 cm. right to the corridor center. A total of 20 points were used for testing purposes. Their scenario was narrow and realistic, because three different parallel paths in a 2 m. wide corridor composed it.

An indoor location system based on a wearable device was successfully introduced in [10]. The system is tested in two very different environments, a 187 m. corridor loop scenario (37200 training samples and 310 test data points), and an atrium scenario (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.

Table I summarizes the databases used in the previous reviewed works [3,7,8,9,10]. We have identified three different types of databases (groups 1, 2 and 3 in the table) according to how samples have been 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 some consecutive discrete samples taken in a lineal environment.

TABLE I: MAIN CHARACTERISTICS OF THE DATA FROM SELECTED PAPERS. GROUP CORRESPONDS TO THE TYPE OF DATABASE WE HAVE IDENTIFIED, # 1D SPACES IS THE NUMBER OF ONE DIMENSIONAL SPACES CONSIDERED IN THE DATABASE, METERS IS THE TOTAL LINEAL METERS OF THE 1D SPACES CONSIDERED. # 2D SPACES AND SURFACE CORRESPOND TO THE TOTAL NUMBER OF 2D SPACES CONSIDERED IN THE DATABASE AND THE SURFACE IS M2 THEY COVER (AREA). # FEATURES IS THE NUMBER OF FEATURES STORES IN A SINGLE SAMPLE, TAKE NOTE THAT DISCRETE ONES HAVE A SINGLE MEASURE (OR AN AVERAGE), WHEREAS CONTINUOUS ONES HAVE MULTIPLE MEASURES. # SAMPLES IS THE TOTAL NUMBER OF SAMPLES INCLUDED IN THE DATABASE.

Paper	*Group*	*# of 1D spaces*	*total lineal meters*	*# of 2D spaces*	*total area m^2*	*# of features*	*# of samples*
[3] Experiment 1	3	-	-	6	-	3* + loc	N/A
[3] Experiment 2	1	5	N/A	-	-	3* x lenght + loc	5
[3] Experiment 3	2	16 + 12 = 28	34.2+3 = 37.2	-	-	3* + loc	64+36 = 100****
[3] Experiment 4	2	1	1	-	-	3* + loc	20
[7]	2	4	178	-	-	3* + 2 (x & y)	350
[8] Scenario 1	3	-	-	1	14x16	3* + loc	14x16
[8] Scenario 2	3	-	-	1	9x12	3* + loc	9x12
[8] Scenario 3	3	-	-	1	6x19	3* + loc	6x19
[8] Scenario 4	3	-	-	1	3.5x3.5	3* + loc	7x7
[9]	1	3	108	-	-	3* x lenght + loc	3
[10] Scenario 1	2	1	187	-	-	3* + loc	37200
[10] Scenario 1	3	-	-	1	13.8x9.9	3* + loc	40800
[10] Scenario 1	2	2	80 (approx.)	-	-	3* + loc	12000
UJIIndoorLoc-Mag	1	26x2	650 (approx.)	-	-	10** x lenght + 6 x m***	270 + 11 ≅40.000 discrete

* In the group 1, continuous samples, length corresponds to the number of individual measures taken in a single continuous sample
* 3 components of the magnetometer
** 3 components of the magnetometer + 3 components of the orientation + 3 components of the accelerometer + timestamp
*** *m* stands for the number of corridors or samples.
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)

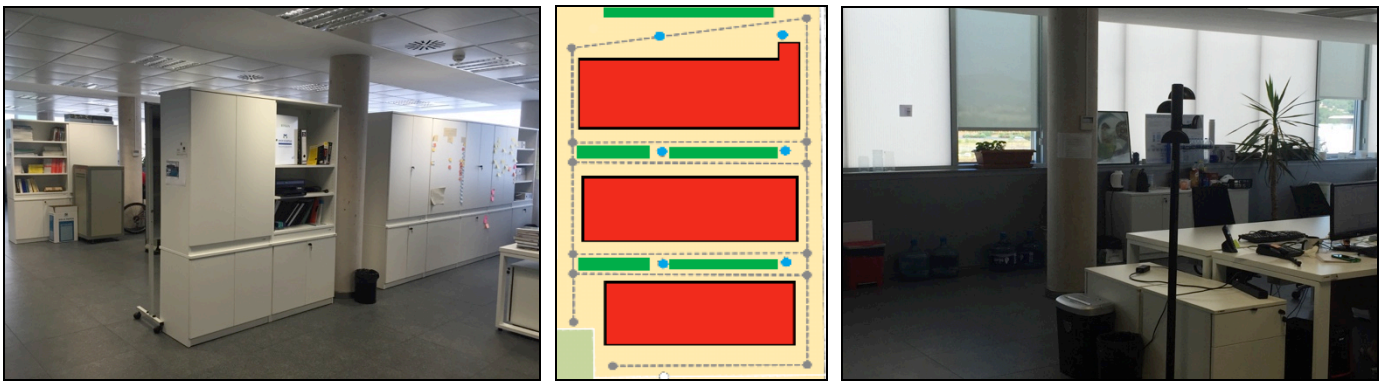


Fig. 1. Lab where samples were taken. In the middle: The lab's map where desktop tables are highlighted in red, bookcases in green, and Columns in blue. Left and right images are photos taken in the lab to show the real scenario. Corridors are numbered from left (1) to right (2) and from bottom (3) to top (8). The vertical corridors start at the bottom point, whereas horizontal ones start at right point.

Although the soundness of results and conclusions presented in all those contributions is high, 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 [7]), but it is omitted in the majority of contributions (such as in [3,8,9,10]). To denote that this information was not provided, we used *loc* to refer to location in Table I.

Although the number of continuous samples used in the experiments seems to be low, 5 in [3] and 3 in [9], the length of the vectors was high enough to perform the experiments. However, our database contains more information than their ones 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 do not only consider corridors, but also combinations of two connected corridors (turns changing corridor).

III. Prior Tests

Prior to generating the database, we performed some basic tests to determine the feasibility of using the Earth's Magnetic Field for indoor positioning using mobile phones. Moreover, we also gathered information about the features to be stored.

A. *It is feasible to use the magnetic field for location?*

First, we selected two simple trajectories in our laboratory (see Fig. 1). The first one consists of two segments; the user comes into the laboratory and goes straight on until the top side windows, then turns right and goes straight on until arriving the right side windows. The second one is a simpler scenario where the user goes straight on through a corridor.

The first test consisted in recording the values provided by the magnetometer. This first test was repeated 5 times. It is important to mention that the sensor 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 Fig. 2 and the values are measured in microtesla (μT). The example vector shown in the figure means that there is a magnetic field of 46.669 μT strong in the direction of 45 degree to Y-axis and Z-axis of the device.

The sampling frequency has been set to 10 samples per second. It balanced the computational costs, energy consumption and time series resolution.

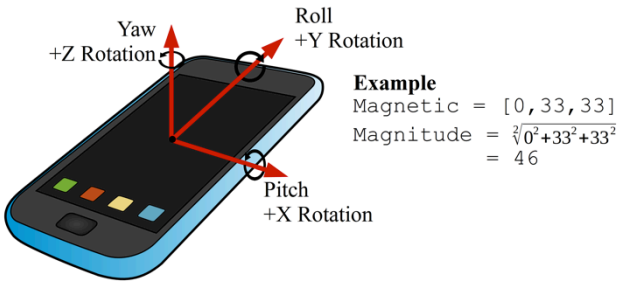


Fig. 2. Meaning of the Axis provided by the Mobile Phone sensory system.

Fig. 3 shows the recorded magnetometer values through both trajectories, where top plots correspond to the first trajectory (two corridors) and bottom ones correspond to the second one (single corridor). Please note that the horizontal scale is different in the two trajectories. The vertical scale and range values are also different in the second trajectory.

At first sight, it can be observed that the magnetometer values and the curves are similar for the five runs according to the plots of the first trajectory. But when we show the results with more detail (second trajectory), we can see that the magnetic values are not exactly the same in the five trajectory's runs, but their differences are low, about 5 μT.

However, it is not trivial to detect a user's orientation change (turn) with the information provided by the magnetometer. Therefore, we decided to record raw data from the orientation sensor too. The orientation sensor provides the direction vector and the values are measured in degrees. This vector is also relative to the mobile phone (see Fig. 2).

Fig. 4 shows the orientation of the device for the two trajectories. In this case, both plots also have different scales. Moreover, we show a simplified orientation instead the vector values for visualization purposes. There was a significant change of user's orientation in the first trajectory (≅90º) according to the Fig. 4 (left) because the user did a L-turn, whereas the changes of user's orientation in the second trajectory (see Fig. 4 right) should be considered insignificant (≅5º) and they may be due to user's movement.

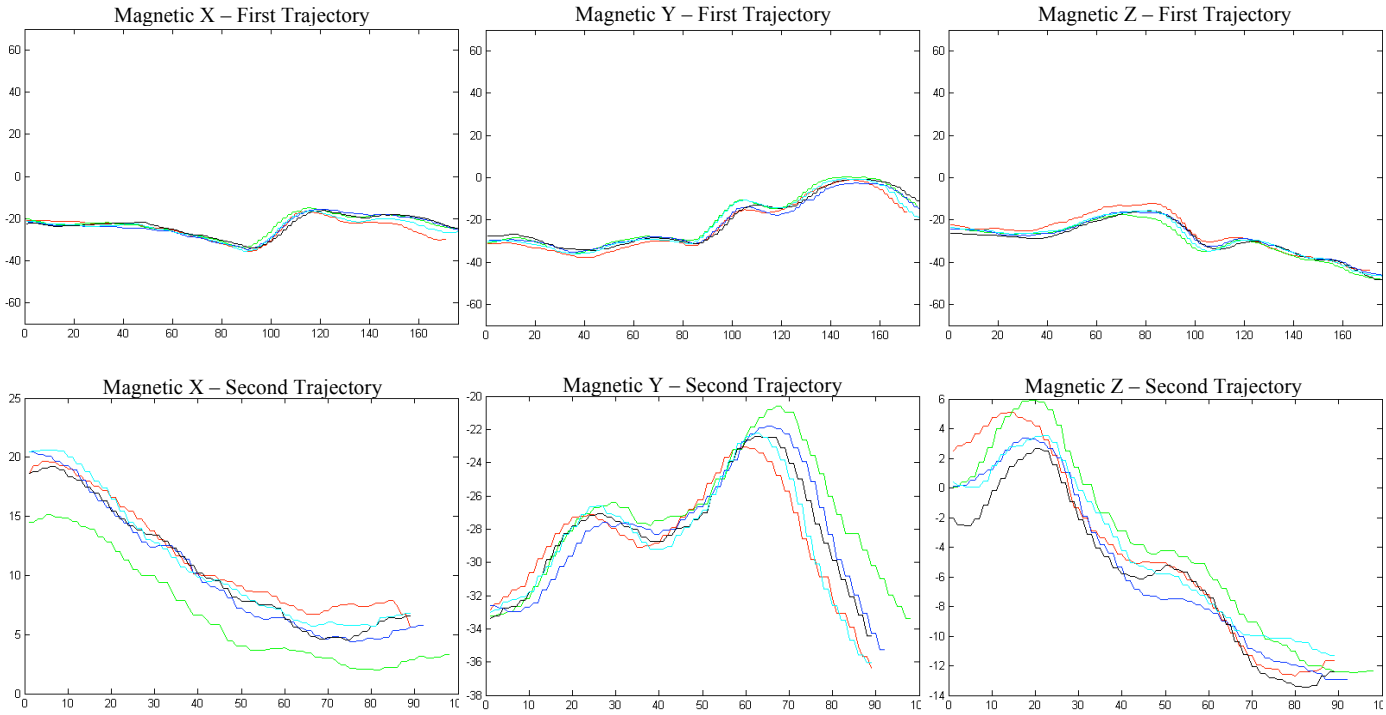


Fig. 3. The magnetic field values (in the three axes) in two different trajectories. Note differences in x-Axis scale.

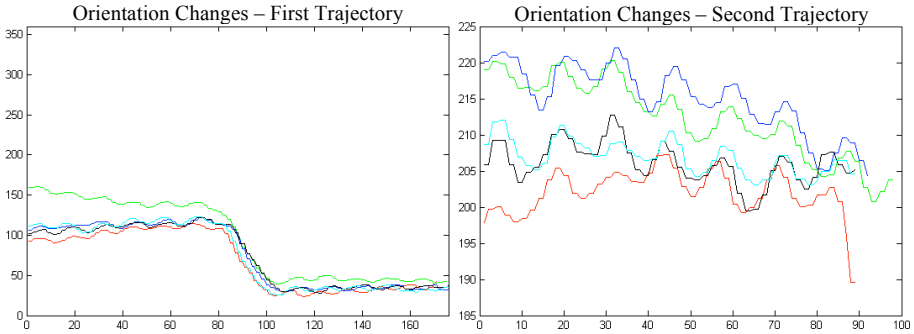


Fig. 4. Simplified orientation sensor values for the two trajectories. Note differences in x-Axis scale.

B. It is necessary to include accelerometer values?

The second test consisted in storing the magnetic field values provided by the magnetometer along with the values provided by the accelerometer. The later sensor provides a vector with the accelerometer values expressed in m/s^2. Those values have been processed to remove the gravity forces and therefore to have an estimation of user's real movement.

In particular, we recorded the magnetometer and processed accelerometer values through a corridor. We repeated this test with three different speed conditions. In the first one, the user was walking slower than usual. In the second one, the user was walking at a normal speed. Finally, the user was walking faster than usual, without getting running speed. Fig. 5 shows the combination of magnetic values and the processed accelerometer values on the Y-Axis. We found that this axis was representative enough to detect the user's steps, and therefore estimate the speed.

First of all, the shape of the magnetic curves in the three axes may be considered similar for the three different cases. However, the horizontal scale (time) varies significantly in the three configurations. In the first case, slow speed, the time required to capture values through the trajectory was 12 seconds approximately (121 samples), whereas time was reduced to a half in the third case with the fastest speed.

We consider that there may be two alternatives to deal with user's speed in indoor positioning. The first one consists of resampling the training or the operational samples to allow its comparison. Resampling is the procedure to dilate or compress the sequence of discrete captures to have the same spatio-temporal resolution. The other alternative consists of mapping the scenario under some different speed configuration, and using an advanced method to determine the speed configuration at operational stage. Therefore, the appropriate training samples from the full training/reference set could be selected depending on user's speed.

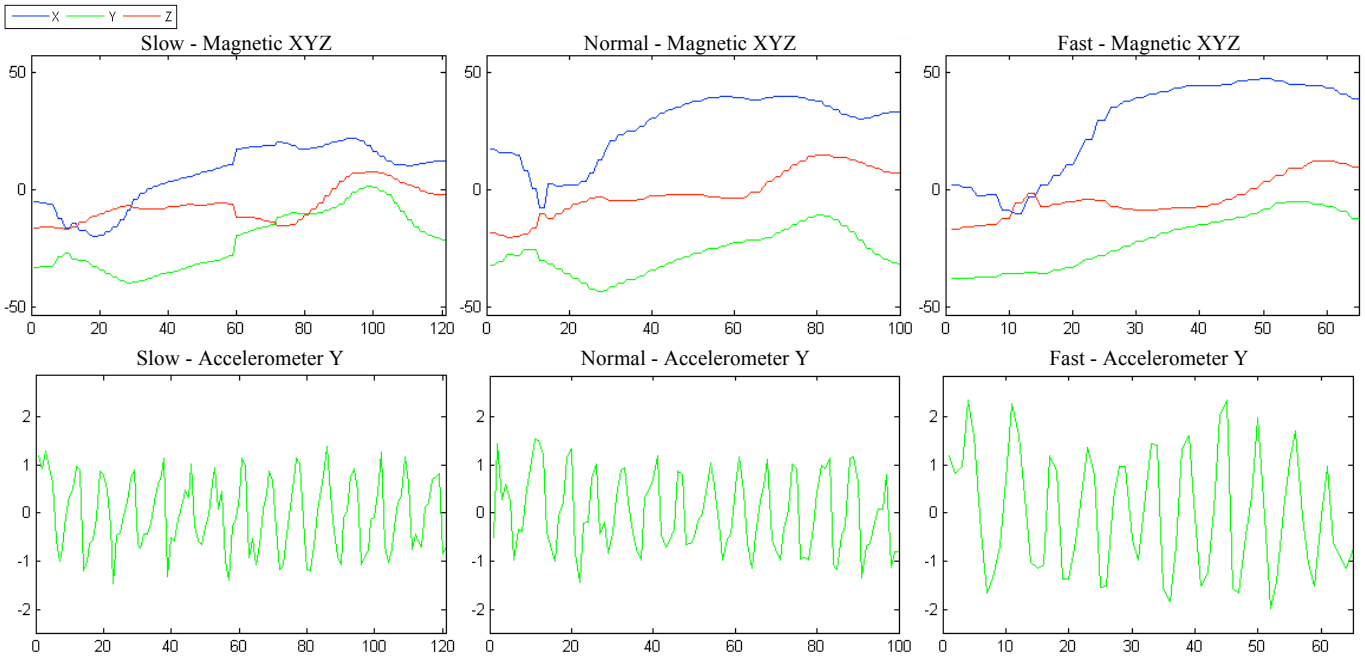


Fig. 5. The magnetic field values in XYZ axes (top) and the Accelerometer values in the Y-axis (bottom) for different speed conditions

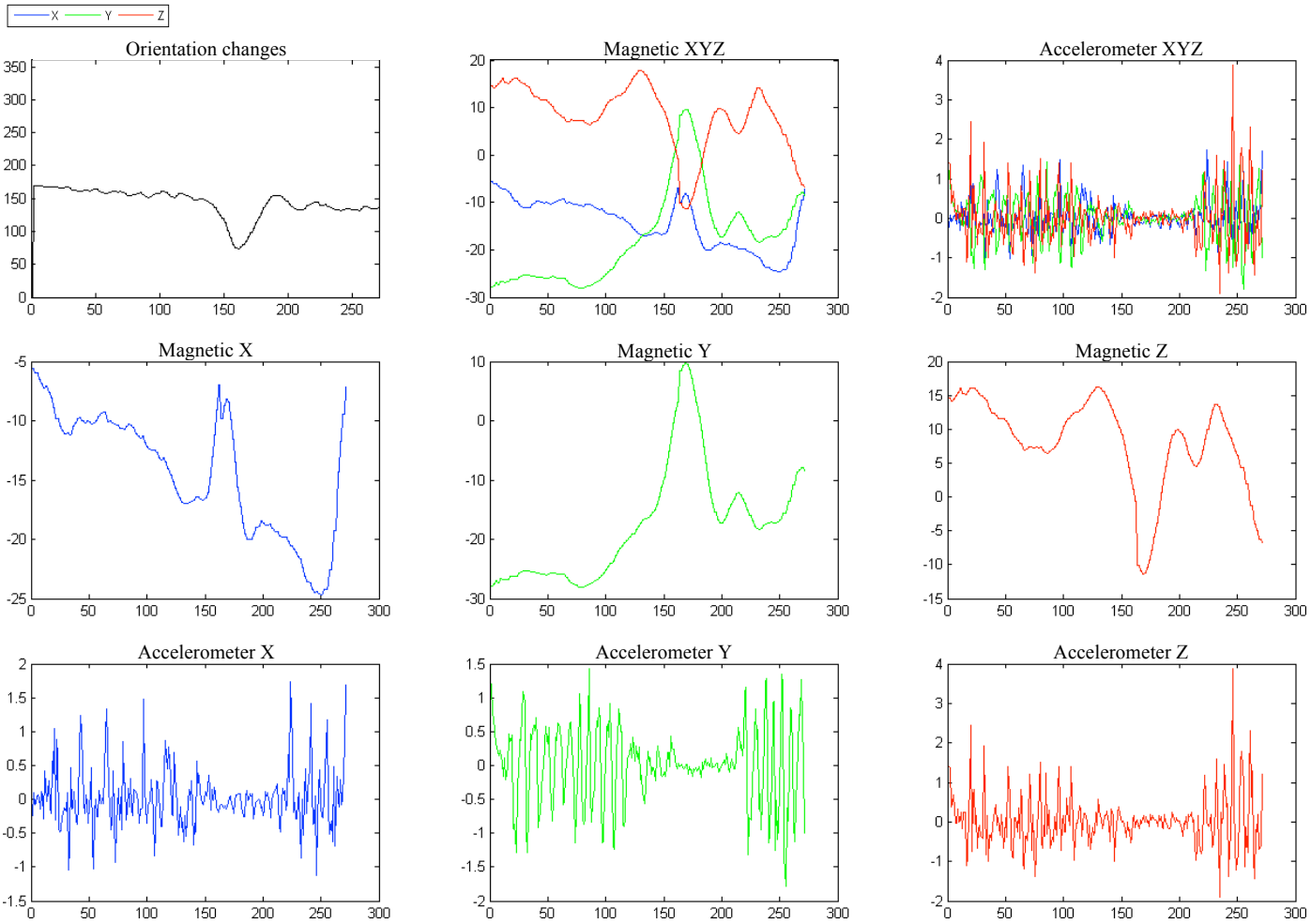


Fig. 6. Data values collected for a trajectory. Simplified orientation is shown for visualization purposed. Moreover the values of magnetic and accelerometer are also shown in separate plots.

C. Lessons learnt from prior tests

After performing some prior tests, including the ones shown in this section, we decided that data from magnetometer, accelerometer and orientation sensors should be included in the proposed public database. Researchers may combine all this information in order to improve the indoor positioning systems. For example, the user's speed, turns, and other common situations could be estimated, and this new information could benefit Indoor Positioning Systems' (IPS) accuracy.

Moreover, we also considered important to record the exact moment in which each discrete sample was taken. We detected that some minor delays could be introduced between two consecutive samples. Moreover, this timing information may be useful for further spatio-temporal analysis such as 'is the time a factor to consider for magnetic field based indoor location?'. However, this kind of questions is out-of-scope.

Fig. 6 graphically shows the information gathered when a trajectory inside an indoor environment was mapped. The first row shows the simplified orientation, the magnetometer values and the accelerometer values. For clarification purposes, the values of the magnetometer are shown in the second row in different plots. Similarly, the third row shows the processed values of the accelerometer (gravity force has been removed).

Although the analysis of this information is complex, some information can be extracted by interpreting the different plots. For instance, the user turned to the left and then turned to the right, such as in the testing trajectories number 5, 7 & 10 (see Fig. 8). The user reduced the speed between the first and second turn because she/he was, maybe, avoiding and "obstacle" because there were some people in the middle of the corridor. Moreover, the two consecutive turns produced an abrupt change in the magnetic field for the three axes.

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. If turns were not considered in the mapping procedure (generation of the reference databases), we would be unable to detect them only with the magnetic data at the operational stage.

Thus, the most important lesson, which we learned from the prior testing experiments, was that having a good reference dataset was essential to develop an accurate Indoor Positioning System based on Earth's Magnetic Field. Therefore we planned to map our laboratory considering all possible natural turns (see Section IV).

Here we publish a dataset in which values from time, magnetometer, accelerometer, and orientation sensor have been recorded. The procedure to map took some time to plan and develop it. So, our principle while collecting data was to record as maximum information as possible according to our current knowledge. The unuseful data can be removed or omitted by the location algorithm.

IV. THE UJIINDOORLOC-MAG DATABASE

This section introduced the *UJIIndoorLoc-Mag* database main features. All the samples were taken in our 260 m^2 laboratory, which is composed by 8 corridors.

In this office, bookcases and desktop tables are the elements that separate the corridors as shown in Fig. 1. The laboratory is located in the fifth floor of the Espaitec-2 building at *Universitat Jaume I* university campus.

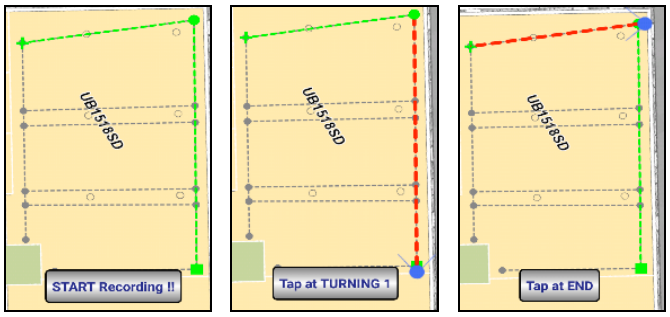
A. General description

The database contains mapping samples alongside the 8 corridors and all the intersections between two corridors. We consider that mapping "intersections" could make a more robust reference database, so we recorded the sensors values 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's Nexus 4 and 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 by a total of 270 different continuous samples.

We used Android devices since they allow full access to sensors and they dominate the mobile phones market with, approximately, 78% of share.

Our 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 came 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. User was moving when capturing data from a starting point to an ending point, and data were collected at every 0.1s. So continuous magnetic fingerprints were stored. Each continuous sample contains 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 samples since the user's speed was almost constant 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 is going to be performed the data capture process. Initially the application shows a map centered into the users current approximately location provided by the GPS sensor. Then, the user draws the trajectory that wants to follow to capture the data (see Fig. 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 Fig. 7-C). In case of a multi-corridor path, the user has to press the *"Tap at Turning i"* button to indicate that they are placed at the *i*-th intersection (see Fig. 7-B).



A B C

Fig. 7. Map of the Lab where samples were taken and Three screenshots of the Android application used to capture the data. A: Shows the path in where the data capture is going to be done and the "Start Recording" button. B&C: The current segment where the user is walking is highlighted in red, the user has to press the button (Turning in B or End in C) when she/he arrives to the 1-st intersection (B) or the final destination (C).

For testing purposes, 9 complex routes (see Fig. 8) along the laboratory were mapped. Each of these routes goes from different corridors and performs different trajectories. Two of them were mapped with two different mobile devices, the above-mentioned Nexus 4 and a LG G3 Smartphone with Android 5.0. So, a total of 11 complex continuous samples are available for testing purposes.

Our approach provides a geo-magnetic database, which contains information about continuous recordings from one or two corridors (training) and multiple corridors (testing). 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 of the device is stored 10 times per second. E.g., if it takes 12 seconds to map a corridor, the corresponding continuous sample will have 1200 values (12 s. x 10 discrete captures x 10 features).

Please note that the 11 testing trajectories are complex and were taken in more than one corridor. Although the 8-th and 9-th trajectories are placed in a single corridor scenario, they may also be considered multi-corridor since a U-turn (180º) is done on them. In those two trajectories, two different directions in the corridor are considered, so they cannot be considered pure single-corridor trajectories.

Due to the complexity of data recorded, each training and testing continuous sample has been stored as an independent text file, whose description is detailed in Section IV-B.

The continuous mapping we have performed may provide an accurate positioning. All the paths, intersections and turnings have been mapped with very high precision. Moreover, knowing that a person in normal conditions can cover a distance of 1.39 m. per second, our approach captures data approximately at every 0.139 m. that means that the accuracy over the path is very high. In *UJIIndoorLoc-Mag*, the users are walking at a normal speed through single and multi-corridor trajectories without any obstacle. Although the research group members and researchers were present in the office, nobody stood in the corridor.

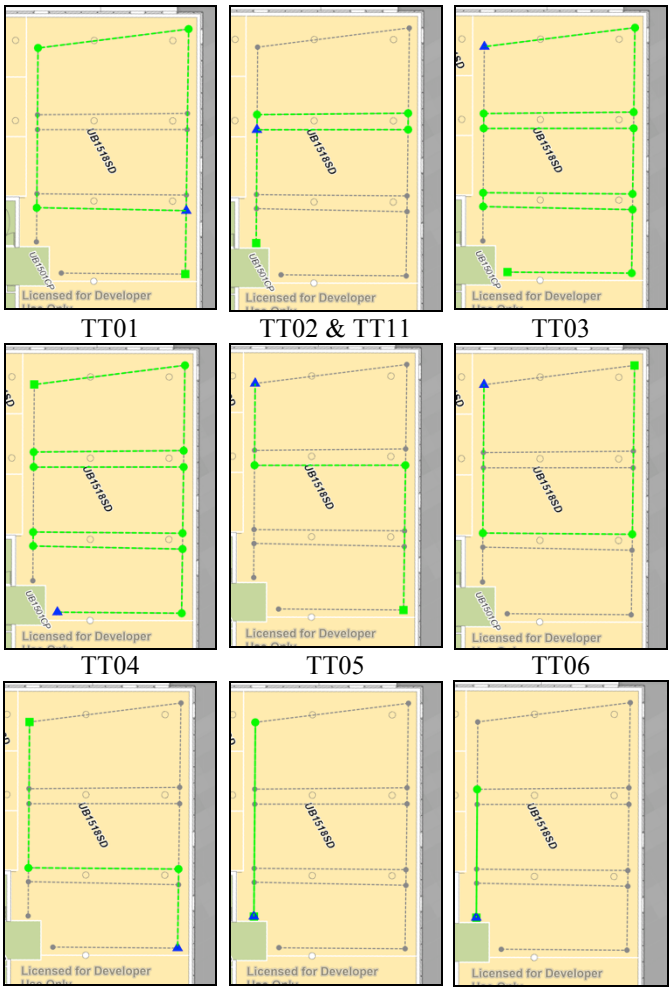


Fig. 8. The 11 testing trajectories (TT). Green squares are the starting point, blue triangle is the ending point and green circles.

B. *Description of database files*

The database consists of 281 continuous samples, 270 are for training and 11 for testing. They have been stored as independent text files. The training ones are grouped into two main categories "lines" and "curves".

- The "lines" group has 80 files and they stand for the single corridor case. The format for filename is "`l`*`XX`*`_`*`ZZ`*`.txt`" where *XX* stands for the number of corridor & orientation (`n` or `r`) and *ZZ* stands for repetition. Example: `l3r_03.txt`
- 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 "`c`*`XXYY`*`_`*`ZZ`*`.txt`" where *XX* and *YY* stand for the number of corridor & orientation for the first and second corridors in the two corridors trajectory, and *ZZ* stands for repetition. Example: `c5n1r_05.txt`
- The testing files' filename format is "`tt`*`PP`*`.txt`" where *PP* stands for the complex testing trajectory number (see Fig. 8). Example: `tt03.txt`

Data has been stored as a simple text file as follows:

```
ts1   mx1   my1   mz1   ax1   ay1   az1   ox1   oy1   oz1
…
tsn   mxn   myn   mzn   axn   ayn   azn   oxn   oyn   ozn
<m>
lat1  lon1  lat2  lon2  FS1   LS1
…
latm  lonm  latm+1 lonm+1 FSm   LSm
```

Where n is the number of samples collected in the trajectory at a 0.1 seconds 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 corresponds to the coordinates (latitude & longitude in decimal degrees) of the initial, intermediate (intersections) and final points. A trajectory with m corridors has $m+1$ points. FS_i and LS_i state for the i-th trajectory's first and last sample respectively in the full sequence of samples collected during the trajectory mapping.

According to the previous structure, the text files are composed by 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 magnetometer (Android's `TYPE_MAGNETIC_FIELD`), accelerometer (`TYPE_LINEAR_ACCELERATION`) and orientation (`TYPE_ORIENTATION`) sensors. The accelerometer's values do not include the gravity force to have a better representation of user's real movement. Two consecutive samples (vertically represented here) from 6-th testing trajectory are:

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 to corridor segments and, moreover, information about turnings is also provided in all the samples.

For instance the configuration part for the 6-th testing trajectory is:

```
39.99389  -0.07375   39.99393   -0.07384    0     71
39.99393  -0.07384   39.99386   -0.07389   72    159
39.99386  -0.07389   39.99388   -0.07394  160    223
```

Where latitude and longitude coordinates have been truncated 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 lower.

V. BASELINE

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 one uses a discrete method to obtain the position of the discrete test points obtained from the continuous test samples. The second one uses a continuous method that obtains the position of the user taking into account 5 seconds of data instead of simple discrete samples. Both algorithms only use the training samples taken on the 8 corridors and from the magnetometer. The 190 two-corridor continuous samples were not used for training purposes in the baselines.

A. Discrete method

For each continuous sample, the localization 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 has been performed to extract the discrete captures from the test paths. In total, there are 8943 samples for training and 4380 for testing.

The k-NN algorithm [11] with `k = 1` has been used to estimate the location of each test sample, so the test current location would correspond to the most similar train 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 [12,13], the distance between two samples, $m_1 = [m_{X,1}, m_{Y,1}, m_{Z,1}]$ and $m_2 = [m_{X,2}, m_{Y,2}, m_{Z,2}]$, corresponds to the Euclidean's distance and it is estimated as follows:

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

Table II shows the baseline results for the discrete method. The error in positioning corresponds to the mean distance between actual position and predicted position. This distance between two points does not correspond to the Euclidean's distance between them since the points corresponds to the latitude (lat) & longitude (lon) coordinates in decimal degrees, they are not expressed in linear meters. So, the *haversine formula*, eq.2, is used instead. The standard error of the mean is also shown in the table.

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

Where R is the radius of Earth, 6373 km approximately, and:

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

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1) \cdot \cos(lat_2) + \sin^2\left(\frac{\Delta lon}{2}\right)$$

In general, 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 of Table I.

B. Continuous method

For the continuous case, each continuous training sample is divided in several subsamples of 5 seconds each one. For instance, if a sample is 10 seconds long and has 100 discrete samples, then it is divided in 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 sub-path, 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 in 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, a 1NN-based method (similar to the one introduced for the discrete case) is performed to look for the more 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's distance*, and it is given by the following equation:

$$d_{vec}(vm_1, vm_2) = \frac{1}{N}\sum_{i=1}^{N} d(vm_1[i], vm_2[i]) \quad (3)$$

where $vm[i]$ is the i-th element of the vector vm, d corresponds to *Euclidean's distance* (see eq.(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 samples.

Table II also shows the baseline for the continuous method similarly than for the discrete method. In this case, the mean error in positioning (considering the 11 different testing paths) is lower: 6.05 ± 0.43 m.

TABLE II: MEAN POSITIONING ERROR FOR DISCRETE AND CONTINUOUS METHODS IN THE 11 TESTING PATHS.

	Discrete Method		***Continuous Method***	
Path	***#samples***	***Error***	***#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*

C. General Discusion

Note that the error provided by both baselines is not low, 7.23 ± 0.38 m. and 6.05 ± 0.43 m. respectively. The principal objective of this contribution is to introduce the *UJIIndoorLoc-Mag* database to the Scientific Community and describe how it has been created.

The two initial basic baselines have been performed to test the suitability of the database, and to establish a starting point for further comparisons that any more sophisticated indoor localization algorithm should overcome.

According to the results shown in Table II, the continuous method provides better positioning results (lower error) in 9 of 11 paths since more information (50 discrete samples) is used to predict the location. We consider that better results could be achieved if training samples taken at intersections were considered in the algorithm.

VI. CONCLUSIONS

This paper introduces a new database for indoor localization, *UJIIndoorLoc-Mag*, on the basis of variations on the magnetic field. The database description has been fully detailed, including the features used in the database. The procedure and the applications used to generate the database have also been described.

Two baseline methods have been introduced using the proposed database in order to show the viability of the usage of the magnetic database and also to encourage future researchers to use the database to compare their different approaches.

We consider that this contribution is useful for the research community. Researchers can use the presented database for testing their own indoor localization proposals based on Magnetic Field or performing data mining analysis.

Our further work will be focused on increasing the amount of samples of the database. Moreover, a more robust indoor positioning method will also be presented in order to show the viability of the indoor magnetic positioning approaches for navigation purposes.

ACKNOWLEGMENTS

This work was supported by *Ministerio de Economía y Competitividad* under the project "Smart Ways" (Convocatoría Retos-Colaboración, RTC-2014-1466-4).

We 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 the supporting services that allowed integrating the existing GIS services in the applications developed to create the *UJIIndoorLoc-Mag* database.

We also thank Javier Fernandez, Ángel Ramos, Álvaro Arranz and Guillermo Amat for their collaboration and comments, as members of project "Percepción" (*Ministerio de Industria, Energía y Comercio*, Programa Avanza2, TSI-020601–2012-50).

REFERENCES

[1] 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.

[2] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J.P. Avariento, T. Arnau, M. Benedito-Bordonau, J. Huerta, UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problem. 5th International Conference on Indoor Positioning and Indoor Navigation, (IPIN 2014).

[3] B. Li, T. Gallagher, A.G. Dempster, and C. Rizos, "How feasible is the use of magnetic field alone for indoor positioning?", 3th International conference on Indoor Positioning and Indoor Navigation (IPIN 2012).

[4] W. Storms, J. Shockley and J. Raquet, "Magnetic field navigation in an indoor environment", International Conference and Exhibition on Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS 2010)

[5] N. Marques, F. Meneses, and A. Moreira, "Combining similarity functions and majority rules for multi-building, multi-floor, wifi positioning," in Proceedings of the 3th the International Conference on Indoor Positioning and Indoor Navigation (IPIN'2012), 2012.

[6] K. Bache and M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[7] J. Song, H. Jeong, S. Hur and Y. Park; Improved indoor position estimation algorithm based on geo-magnetism intensity. 5th International Conference on Indoor Positioning and Indoor Navigation, (IPIN 2014)

[8] D. Vandermeulen, C. Vercauteren and M. Weyn. Indoor localization Using a Magnetic Flux Density Map of a Building: Feasibility study of geomagnetic indoor localization. The Third International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2013), pp 42-49. 2013.

[9] B. Li and T. Gallagher and Chris Rizos and Andrew G. Dempster. Using Geomagnetic Field for Indoor Positioning. International Global Navigation Satellite Systems Society IGNSS Symposium 2013. 2013.

[10] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, M. Wiseman. Indoor Location Sensing Using Geo-Magnetism. The International Conference on Mobile Systems, Applications, and Services, pp. 141-154, 2011.

[11] T.M. Cover, P.E. Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–37, 1967.

[12] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, J. Huerta. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. Expert Systems with Applications vol. 42, no. 23, pp. 9263–9278, 2015.

[13] 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.

Dr. Joaquín Torres-Sospedra is a Researcher at the Institute of New Imaging Technologies. He holds a Bachelor's Degree in Computer Science (2003) and a PhD with distinction (2011), both from Universitat Jaume I. His main research interests are in the areas of Indoor Positioning and Navigation, Wi-Fi fingerprinting, Machine Learning, Pattern Recognition, Multiple Classifier Systems, Sensor Fusion, Knowledge-based Systems and Interoperability.

David Rambla received his M.Sc in Computer Science at the University Jaume I (2015). In June, 2013 he joined the research team at the Institute of New Imaging Technologies (INIT) as a member of the GEOTEC group at the UJI. His main interests are mobile applications development and Augmented Reality integration with GIS.

Dr. Raúl Montoliu is currently an Assistant Lecturer at the Department of Computer Science and Engineering and Senior Researcher at the Institute of New Imaging Technologies (INIT), at Universiat Jaume I. He holds a Bachelor's Degree in Computer Science (1998) and a PhD with distinction (2008), both from Universitat Jaume I. His current research interests include indoor localization and navigation, human and social behavior from sensor data, sport video analysis and surveillance applications.

Dr. Óscar Belmonte is an associate professor in the Department of Computer Languages and Systems at Universitat Jaume I, where he teaches Programming and Graphics, particularly in the Master's Degree in Geospatial Technologies. He holds a Bachelor in Physics (1992) and a PhD in Physics (2002) both from the University of Valencia. His current research interests are Real-Time Computer Graphics, Geospatial Web technology, and Web-of-Things. He has led some research projects in the regional, national and European ares.

Dr. Joaquín Huerta is associate professor in the Department of Information Systems at UJI, where he teaches GIS and Internet Technologies. He holds a PhD in Computer Science from Universitat Jaume I. He is co-Director of the Master in Geospatial Technologies (Erasmus Mundus program) and co-Director of the Joint Doctorate in Geoinformatics (Marie Skłodowska-Curie Actions, International Training Networks and European Joint Doctorates). His main research activities are Smart Cities & Smart Campus, Context Aware Systems, Internet of Things and Sensor Networks.