Generation of Multi-Fingerprint Maps using Propagation Models

Pedro Mestre, Member, IAENG, Joao Cordeiro, Luis Reigoto, Luis Coutinho, Paulo Botelho, and Carlos Serodio, Member, IAENG

Abstract—One of the drawbacks of using Fingerprinting as a Location Estimation technique is the time needed to acquire the data to build the Fingerprint Map (FM). Initial data to generate the FM must be acquired in all points of the Spatial Domain that will be mapped into the Signals Domain. After this phase, any further modification that occur in the physical space implies acquisition of new data, at least in some of the points that belong to the Spatial Domain. Because many samples of the Radio-Frequency signal (e.g. the Received Signal Strength) must be acquired at each point, collecting data to build the FM can be a very time-consuming task. Furthermore, if the Location Estimation Algorithm (LEA) uses multiple Fingerprint Maps (e.g. North, South, East and West), this task will be even more time-consuming because more data must be collected. A solution to overcome this issue is to use propagation models to simulate data to build the FM. In this paper are presented some tests to assess the feasibility of using Fingerprint Maps generated using propagation models. Those tests were made both with algorithms that use only a single FM and algorithms that use multiple FM. Classic LEA (Nearest Neighbour, k-Nearest Neighbour) and Fuzzy Logic based algorithms and some probabilistic algorithms were used in those tests.

Index Terms—Fingerprinting, Location Estimation Algorithm, Indoor, Propagation Models.

I. INTRODUCTION

ONE of the Indoor Location Estimation Algorithms (LEA) that can use the already existing Wireless LAN (Local Area Network) infrastructure, such as WiFi, which as gained an increasing attention is Fingerprinting [1]. In fact, Fingerprinting has become one of the most used localization techniques for indoor environments [2].

Because it can be implemented using the existing wireless network infrastructure, and it is not needed to know the location of the network Access Points, are some of the the reasons why Fingerprinting is so popular. Other techniques such as those that are based on Multilateration require an exact knowledge of the location of some key elements of the wireless network infrastructure, e.g. the location of the Access Points.

Fingerprinting is a localization technique, that belongs to the scene analysis methodologies [3], which comprises two different phases [4],[5], [6]:
- Offline phase: during this phase data from the wireless network is acquired and stored in a database. This database is called the Fingerprint Map (FM) and it will be used by the LEA (during the online phase) to locate mobile terminals;
- Online phase: it is in this phase that the localization of the mobile terminal is made. Live data acquired from the wireless network interface card, of the node, is compared to those data previously stored in the FM. From this comparison it is expected that the Location Estimation Algorithm calculates the current location of the mobile node.

Although any property of the wireless signal can be used to build the Fingerprint Map, and be used by the Location Estimation Algorithm, usually it is used the RSS (Received Signal Strength) value. Several values of the RSS are acquired and their average is stored in the FM. Some algorithms, such as the probabilistic ones require that the standard deviation values, of the RSS samples, must also be recorded.

To build the Fingerprint Map, during the offline phase data must be collected at each point of the Spatial Domain. At each point are acquired and stored, in the database, RSS values related to the infrastructure Access Points detected at that point. These data represent the coordinates of that point in the Signals Domain, which is an N-dimensional space (where N is the number of references).

For the data to be representative of the wireless signals that are received at each point, several samples must be acquired at each location. Some Location Estimation Algorithms use data from two or more directions, such as the ones presented by the authors in [7]. This means that many data must be collected at each point that belongs to the map. As a consequence, the task of acquiring RSS data to build the first FM can be a very time-consuming task. Also any further modification that may occur in the physical space implies acquisition of new data, at least at some of the points of the Spatial Domain.

To cope with this issue authors presented in [8] a solution that uses Propagation Models to build the FM. Data about the scenario is imported into an application that generates the FM for the LEA. This application uses Indoor Propagation Models to simulate the FM.
In this paper are presented some tests to assess the feasibility of using propagation models to generate the Fingerprint Maps. These tests were made using: some of the classic Location Estimation Algorithms (Nearest Neighbour, k-Nearest Neighbour and Weighted k-Nearest Neighbour); Probabilistic Algorithms based on Bayesian probability with a Gaussian probability density function; Fuzzy Logic based algorithm. The probabilistic LEA include two algorithms inspired in k-Nearest Neighbour and Weighted k-Nearest Neighbour.

II. LOCATION ESTIMATION ALGORITHMS

Three different types of Location Estimation Algorithms are used in this paper to test the Fingerprint Maps generated using indoor Propagation Models. The key focus of this work are the algorithms, which are used only as a tool, but rather the technique to generate the Fingerprint Maps. A short description of each LEA that was used can be found below.

A. Nearest Neighbour Algorithms

When Nearest Neighbour Algorithms are used, the first step is to calculate the distance to a set of candidate points to verify which one(s) is(are) the nearest point(s) to the current location.

These algorithms use therefore the concept of distance. In this case it is not the distance on the Spatial Domain, measured in meters or centimetres, but the distance on the Signals Domain.

There are several ways to calculate the distance between the current point and the candidate points of the FM, the one that will be used is the Euclidean distance (Eq. 1):

$$d_j = \sqrt{\sum_{i=0}^{n} (P_{ri} - P_{FMj,i})^2} \quad (1)$$

were:
- $d_j$ is the distance to the point $j$;
- $n$ is the number of dimensions;
- $P_{ri}$ is the power received from reference $i$;
- $P_{FMj,i}$ the value of the power of reference $i$ registered in the FM for point $j$.

After calculating the distance (in dBm) between the current point and all the points of the Fingerprint Map that contains the reference $i$, the mapping between the coordinates of the current point in the Signal Domain to the coordinates in the Spatial Domain is made using one of the following classic algorithms:
- Nearest Neighbour (NN) – which considers that the current location is exactly at the coordinates (in the Spatial Domain) of the nearest neighbour, i.e., the neighbour point that is at the shortest distance (in the signal domain);
- k-Nearest Neighbour (kNN) – finds the $k$ nearest neighbours, in the signal domain, and assumes that the current location spatial coordinates are the average of the spatial coordinates of these $k$ neighbours;
- Weighted k-Nearest Neighbour (WkNN) – similar to the above algorithm, but it uses a weighted average of the $k$ nearest neighbours coordinates instead, to estimate the current location coordinates.

B. Probabilistic Algorithms

Based on Baysian theory it was built a Probabilist algorithm that uses a Gaussian kernel function [10] to estimate which is the probability of a point in the FM to be the solution.

This algorithm, for each point $P_i$ of the Fingerprint Map that contains a reference observed by the receiver, calculates which is the probability that $P_i$ is the solution, i.e., the actual point where the mobile terminal is located. After iteration every possible candidate point, the one with the highest probability value is considered as the current point.

To calculate the probability of a given RSS value to belong to certain point a Gaussian Probability Density function, Eq. 2, is used. There is a *sine qua non* condition for this kernel function to be used: besides the RSS values, the Fingerprint Map must also contain the standard deviation value of RSS.

$$Prob(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad (2)$$

where:
- $x$ is the Received Signal Strength value for a reference;
- $\mu$ is the Received Signal Strength value average, at a Signal Domain point;
- $\sigma$ is the Standard deviation of the Received Signal Strength, at a Signal Domain point.

For each candidate point, the probability for each reference is calculated. If we have two references, $i$ and $j$, then $Prob(x_i)$ and $Prob(x_j)$ are independent. Therefore the final probability for each point is the compound probability of all references probabilities.

The above described algorithm, that will be referred in the rest of this paper as Gaussian Probability (GP) algorithm, like Nearest Neighbour algorithm selects the point that is considered the "best candidate. However some other "almost as good" candidates (some of them even could be the real point) are not considered by the algorithm.

So, inspired in the Probabilistic Algorithm and the Nearest Neighbours Algorithms, besides the GP algorithm it was also implemented two other algorithms that take into account more than a single point. These algorithms use similar concepts to k-Nearest Neighbour and Weighted k-Nearest Neighbour, however instead of dealing with distances in the Signal Domain, probabilities are used instead. Those algorithms are:
- k-Gaussian Probability (kGP); it considers that the coordinates (in the Spatial Domain) of the current point are is the average of the coordinates of the $k$ points with higher probability values;
- Weighted k-Gaussian Probability (WkGP): similar to the above, but it uses a weighted average instead;

C. Fuzzy Logic based Algorithm

Based on the distance concept, in [6] authors presented a Fuzzy Logic based algorithm for use in Fingerprinting-based localization systems. This algorithm classifies the
point of the Fingerprint Map as 'Very Close', 'Near' and 'Far' according to their distance to the current point, and based on this classification it calculates the contribution of each point to the final point coordinates. It used the same principle as Weighted k-Nearest Neighbour, but the number of neighbours and their weight is not known beforehand - it is calculated by the algorithm.

The first step of this algorithm is to calculate the distance between the current point and those stored in the FM that are relevant. Because the number of dimensions (in the Signals Domain) is equal to the number of references and it can change from point to point, Eq. 3 is used to calculate the distance.

\[
d_j = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (P_{ri} - P_{FMj,i})^2}
\]

where \( n \) is the number of dimensions.

After having a classification for the relevant points in the Fingerprint Map, according to their distance to the current point, to each one it is assigned a weight, using fuzzy inference.

To assign weights to the points, the following simple IF THEN rules are used:

- IF the distance is 'Very Close' THEN the point weight is set to 'high';
- IF the distance is 'Near' THEN the point weight is set to 'medium';
- IF the distance is 'Far' THEN the point weight is set to 'low';

To the values for 'high', 'medium' and 'low' are assigned weight values, \( W_1, W_2 \) and \( W_3 \), such that \( W_1 > W_2 > W_3 \), and after the defuzzification process the weight of each relevant point of the FM \( W_{FMJ} \) is now known. The final value for the coordinates \( (C_p) \) can then be calculated using Eq. 4:

\[
C_p = \frac{\sum_{i=0}^{n} (W_{FMJ} \times C_i)}{\sum_{i=0}^{n} (W_{FMJ})}
\]

where \( W_{FMJ} \) is the weight of point \( j \) of the FM and \( C_j \) represents the point coordinates in the Spatial Domain.

### III. FINGERPRINT MAPS

With the objective of increasing the performance of Location Estimation Algorithms, the authors have presented in [7] a solution for Fingerprinting-based localization that uses multiple Fingerprint Maps. During the offline phase, instead of collecting only RSS related data, it must also be collected information about the user direction.

Based on the user direction it is then possible to build multiple Fingerprint Maps, and allow the LEA to choose which the best map to be used to locate the user.

Two strategies to build these maps were presented in [7], one that chooses a single map and another that weights the contribution of several maps. In this paper the latter approach was used.

Therefore two sets of Fingerprint Maps were built to do the tests tests presented in this paper: one for use with algorithms that require a single FM; another for algorithms that can use the user direction information and select the most suitable map(s).

Below are presented the concepts and models used to generate the Fingerprint Maps using indoor Propagation Models. These concepts are applied both for single and multiple Fingerprint Maps (FM with user orientation information), Fingerprint Maps for those algorithms that require a single FM were built by averaging the values for all directions in the multi Fingerprint Maps (as it would be done with real data).

#### A. Generating the Fingerprint Maps using Propagation Models

Generating Fingerprint Maps using propagation models implies that we need to have in beforehand some information about the location where the localisation process will run (wall, doors, Access Point location, etc). This can be pointed out as an disadvantage of this method because it requires a survey. However the "traditional" method also requires a survey that can take several hours or even days, depending on the complexity of the scenario. Furthermore many information needed to build the FM can be obtained using the blueprints of the location.

Also, if there are changes in the scenario, the "traditional" method implies to go again to the location and acquire new data. Using the proposed method, it is only needed to change the scenario parameters and in a very short time we have new Fingerprint Maps simulated.

This method requires the knowledge of the exact location of the Access Points, which might seem a contradiction. But, the Location Estimation Algorithms do not use this information, they are only used to calculate the FM. The algorithms "cannot" distinguish between an FM generated using Propagation Models and those that are built using the "traditional" method.

To build the Fingerprint Maps, we need to estimate the value of the expected Received Signal Strength for reference, at the points that will be added to the database. It is required that the model used to estimate these values take into account the following parameters:

- Attenuation in free-space;
- Attenuation because of obstructions that we can find in indoor environments such as walls, floors, etc;
- Absorption of electromagnetic waves caused by the user (that will be holding the mobile terminal).

Taking into consideration the results obtained in [11], Eq. 5, which was also used in [9], will be the base to build the Fingerprint Map. It is based on Motley-Keenan model [12].

\[
PL(d) = PL(d_0) + U_n + 10nlog \left( \frac{d}{d_0} \right) + \sum_{i=1}^{N} k_i L_{0i} 2^{\log_2 \left( \frac{d}{d_0} \right)}
\]

where:

---

**Engineering Letters, 25:1, EL_25_1_11**

(Advance online publication: 22 February 2017)
• $PL(d)$ is the total Path Loss as a function of the distance between the transmitter and the receiver ($d$);
• $PL(d_0)$ is the Path Loss at a reference distance $d_0$;
• $U_a$ is the User Attenuation, according to [5] and [13] a single human body can cause an attenuation in the range of $3.5\, dB$ to $5.0\, dB$;
• $n$ is the Path Loss exponent, which may vary according to the structure of the building [14];
• $L_{0i}$: is the attenuation of a reference wall with thickness $\epsilon_0$;
• $k_i$: is the number of type $i$ walls that have thinness $\epsilon_i$.

As discussed in [9] both single Fingerprint Maps and multiple Fingerprint Maps can be generated using Eq.5. However, "real life" Fingerprint Maps have some randomness because of the intrinsic properties of RF signal propagation.

To mimic this randomness, in [9] Fingerprint Maps were also built by simulating Radio-Frequency (RF) signals. To simulate RF signals, instead of using Eq.5 as it is, we must add to it a random variable $X_\sigma$, that denotes a Gaussian variable with zero mean and standard deviation $\sigma$ [15], Eq. 6:

$$PL(d) = PL(d_0) + U_a + 10nlog\left(\frac{d}{d_0}\right) + \sum_{i=1}^{N} k_i L_{0i} 2^{log_{10}\left(\frac{d}{d_0}\right)} + X_\sigma \quad (6)$$

A set of simulated RF signals is generated, using Eq. 6, and then those values are averaged to build the FM. This is the same process as if the FM was generated using real values. Obviously that we cannot have too many samples, otherwise the average of $X_\sigma$ will be zero and its contribution to the FM will disappear.

Adding $X_\sigma$ to the model to generate the FM presented better results, therefore it was the one selected to generate the Fingerprint Maps used to do the tests whose results are presented in section V.

IV. TESTING SCENARIO AND CONDITIONS

Tests made to assess the feasibility of the proposed method (to generate Fingerprint Maps) consisted in generating the FM using Propagation Models, and then feed the LEA with real data to calculate the Precision, Standard Deviation, Maximum Error and Minimum Error for each algorithm.

To do these tests data was acquired in a real scenario, located at the University of Trás-os-Montes and Alto Douro in Portugal. In Fig. 1 it is depicted the map of the testing scenario, where the location of each Spatial Domain point that was mapped into the FM is represented. At each one of these points, 20 samples of the RSS and azimuth values were acquired, with the user facing each of the four directions that were considered in the FM (North, South, East and West). This is the same scenario that was use in [9].

To generate the Fingerprint Maps it was considered that $X_\sigma$ has a standard deviation of $4dBm$ and the user attenuation is $5dB$. For all the tests the value of $k$ is 3, and the weights are 0.7, 0.2 and 0.1.

V. NUMERICAL RESULTS AND DISCUSSION

In this section are presented the results that were achieved in the tests made using both single Fingerprint Maps and multiple Fingerprint Maps. Values presented in this Section were obtained using the following Location Estimation Algorithms:

- Nearest Neighbour;
- k-Nearest Neighbour;
- Weighted k-Nearest Neighbour;
- Gaussian Probability;
- k-Gaussian Probability;
- Weighted k-Gaussian Probability;
- Fuzzy Logic based Algorithm.

The objective of these tests is not to test the algorithms themselves, but to check if values obtained using simulated Fingerprint Maps are similar to those that are obtained using real data, and therefore if it is feasible to use such a technique to generate the FM.

A. Reference Values

In Table I are presented the normalized values for the Precision (Pec.), Minimum Error (Min. Err.), Maximum Error (Max. Err.) and Standard Deviation (St. Dev.) obtained in the testing scenario and using real values (both for the Fingerprint Map and data used to feed the LEA).

These data were acquired using an Android Smartphone and the Application shown in Fig. 2, which is the same that was used in [7].

Values shown in Table I are reference values, to be used for performance comparison and assessment of the several Location Estimation Algorithms using propagation models based Fingerprint Maps. These are the same values presented in [9] and were obtained using the three classic Location Estimation Algorithms based on the Euclidean Distance. As stated in [9] these are not optimal values, but are "real life" values that are here presented for comparison purposes only.

B. Multiple Fingerprint Maps

One of the set of tests, made to assess the performance of the proposed Fingerprint Map generation strategy, consisted in using all the above presented Location Estimation
Algorithms with Multiple Fingerprint Maps (generated using propagation models) and real data acquired in the testing scenario.

Normalized values for the results obtained with the above mentioned tests are presented in Tables II, III and IV. The first presents results of tests made using the Classic Algorithms, the second of those made using Probabilistic Algorithms and the third table presents the results obtained using Fuzzy Logic.

Analyzing the results, from the Location Estimation Algorithm performance comparison point of view, it can be concluded that probabilistic algorithms have a slightly better performance that those based on the Euclidean Distance. If we compare the Fuzzy Logic based algorithms with the other base algorithms (NN and GP), it has a slightly better performance in terms of precision and a much better performance in terms of Maximum Error.

However the emphasis of this paper is not on the Location Estimation Algorithms but on the feasibility of using Propagation Models to generate the Fingerprint Maps. Comparing Tables I, II and IV it is obvious that if we use real data, the performance will be better. Nevertheless, results obtained using simulated Fingerprint Maps are not very different from those obtained with real data, also, it must be taken into consideration that the process of generating Fingerprint Map by simulation is much faster. This feature might be very important specially in a changing environment or in very large and complex areas.

C. Single Fingerprint Map

Another set of tests was made using a single Fingerprint Map. In this case the Fingerprint Map was generated by averaging all the values that were simulated for the four directions. Real data used to test the LEA was acquired facing all the four possible directions, and all these data, without any preprocessing, were fed to the Location Estimation Algorithm under test.

Table V presents the results that were obtained using the Classic Algorithms based on the Euclidean Distance, in Table VI are presented the results that were obtained using the Probabilistic Algorithms, and Table VII represents those values obtained using the Fuzzy Logic based algorithm.

As expected, according to the results presented in [7], when a single Fingerprint Map is used the results are worse, in comparison to those obtained with multiple Fingerprint Maps.

(Advance online publication: 22 February 2017)
TABLE VI
VALUES FOR THE PROBABILISTIC ALGORITHMS, USING A SINGLE FINGERPRINT MAP

<table>
<thead>
<tr>
<th></th>
<th>GP</th>
<th>kGP</th>
<th>WkGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec.</td>
<td>1.55</td>
<td>1.42</td>
<td>1.45</td>
</tr>
<tr>
<td>Max. Err.</td>
<td>5.83</td>
<td>5.00</td>
<td>5.44</td>
</tr>
<tr>
<td>Min. Err.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>0.95</td>
<td>0.78</td>
<td>0.82</td>
</tr>
</tbody>
</table>

TABLE VII
VALUES FOR THE FUZZY LOGIC BASED ALGORITHM, A SINGLE FINGERPRINT MAP

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec.</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>Max. Err.</td>
<td>3.73</td>
<td></td>
</tr>
<tr>
<td>Min. Err.</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>St. Dev.</td>
<td>0.77</td>
<td></td>
</tr>
</tbody>
</table>

Regarding to the feasibility of using Fingerprint Maps generated using propagation models, as it was already concluded for Multiple Fingerprint Maps (see subsection V-B), this is a feasible method.

VI. CONCLUSION

In [8] the authors presented a method to generate Fingerprint Maps, using information of the location scenario such as the blueprints, the location and type of obstacles and location of Access Points. This method is based in indoor Propagation Models and it can generate the Fingerprint Maps, without the need to collect RSS values in the real scenario. In the case that the scenario changes, the new FM can be generated very quickly.

As an improvement to this methodology to generate Fingerprint Maps, authors presented in [9] an extension to this method, that was used to generate Fingerprint Maps that include the user direction information. Such maps are used by algorithms that choose the most suitable map(s) based on the user direction (as in [7]).

In this paper, in addition to the tests presented in [9] using some of the classic LEA (NN, kNN and WkNN), it is presented a new set of tests that have as objective to assess the feasibility of FM generated by Propagation Models using other types of algorithms: Probabilistic Algorithms and a Fuzzy Logic based Algorithm.

The results obtained using propagation models are not as good as the ones achieved using real data, as it was already expected. However these results are very promising. Also, even though the results are worse, the obtained values are not that much different.

Having in mind some future work centred on obtaining a better model for the RF signals (that almost mimics real RF signals) we can use propagation models for the proposed purpose. That better model will also help to use this approach to simulate data to test new Location Estimation Algorithms, not only generating maps for the online phase.

REFERENCES