Hybrid technique for Fingerprinting using IEEE802.11 Wireless Networks

Pedro Mestre*, Luis Coutinho†, Luis Reigoto†, Joao Matias‡ and Carlos Serodio *
* CITAB-UTAD, University of UTAD, 5000-801 Vila Real, Portugal
e-mail: pmestre@utad.pt, cserodio@utad.pt
† University of UTAD, 5000-801 Vila Real, Portugal
e-mail: luis_coutinho_86@hotmail.com, luisreigoto@gmail.com
† CM-UTAD, University of UTAD, 5000-801 Vila Real, Portugal
e-mail: j_matias@utad.pt

***Abstract*—Location based on fingerprinting comprises two distinct phases: the first phase, called the off-line phase, is related with two tasks, a site survey to collect the data for the Fingerprint Map, and its generation based on the collected data; the second phase, called the on-line phase, is related with the location of a mobile terminal, by doing the comparison of the data acquired from the wireless transceiver with the data previously stored in the Fingerprint Map. Typically the first phase is very time consuming because multiple samples of the wireless signals, received from the multiple references, must be acquired and stored, for every spatial point to be considered in the Fingerprint Map. In this paper a hybrid approach using propagation models for indoor environments and fingerprinting is presented. In this approach the Fingerprint Map is generated using indoor propagation models and the map of the scenario for which it is being generated, instead of requiring a full site survey. This makes the process of generating the Fingerprint Map much faster when comparing with the traditional approach. Also when changes occur in the scenario, the Fingerprint Map can be easily recalculated. Several Location Estimation Algorithms were used to test the Fingerprint Map and a precision of $3.249m$ was achieved using k-Nearest Neighbour and a standard deviation of $1.928m$ was achieved using a Location Estimation Algorithm based on Fuzzy Logic.**

***Index Terms*—Fingerprinting, IEEE802.11, Hybrid Technique, Propagation Models, Indoor Location**

I. Introduction

The techniques used to determine the location of a mobile terminal can be divided into three main categories [1]: Triangulation, Proximity and Scene Analysis. When using wireless networks to do the location estimation, any property of the wireless signal can be used in the process, as long as it is possible to establish a relation between the property of the wireless signal used to determine the location and the actual coordinates of the mobile device.

Fingerprinting is a location technique that uses the scene analysis technique. It consists in observing and analysing the patterns of electromagnetic signals and their variations along time. A given property of the electromagnetic signal received from a set of references are read and then are compared with a set of previously stored values, called the Fingerprint Map (FM) [2], [3], [4]. From this comparison, the Location Estimation Algorithm (LEA) outputs the current location.

Two different concepts of domain exists in this technique: the signals domain and the spatial domain. The signals domain is an n-dimensional space, where the number of dimensions equals the number of references that exist in the scene under analysis. The second domain is related to the space where the mobile node to locate is. The number of dimensions of this space depends on the type of output of the Location estimation Algorithm (LEA). It can be a single label indicating the location of the user (e.g. classroom 1), can be a two dimension coordinates system (x, y), a three dimensional coordinates system (x, y, z), etc.

Fingerprinting comprises two distinct phases [5], one on which data is collected and the Fingerprint Map (FM) is generated and a second phase, also called on-line phase when data collected from the wireless communications transceiver are compared with information on the FM and the locationis made.

The first phase of fingerprinting consists in collecting several samples of the value of the property of the wireless signal to be used, e.g. the Received Signal Strength (RSS), at each point of the spatial domain, for all references detected at each point. Due to the number of points that make part of the FM and the number of samples that must be collected at each point, this can be a very time consuming task.

Also, whenever a change occurs in the scenario, e.g. due to the placement of new furniture or when the type of antenna used in the Access Points is changed, the FM must be recalculated. This means that the first phase must be repeated, at least for a part of the scenario.

To overcome the need for this data collection phase, either to build the FM or to update it, in this paper a hybrid solution for fingerprinting-based location, combining the traditional fingerprinting (at the on-line phase) with propagation models (at the off-line phase), is proposed. With the proposed approach, the FM is built using propagation models, not by doing a site survey. After knowing the number of references to be used and their correct placement, the expected signal strength at each point of the spatial domain considered in the FM is calculated. Whenever a change occurs in the scenario, it is only needed to recalculate the FM values, without the need to collect new data *in loco*.

In this paper some tests made at the authors' university are presented. The FM was built based on the map of the testing scenario, the location of the references and indoor propagation models. Two FMs were generated, one using the Multi-Wall Model (MWM) with the parameters defined in COST231[6], and a second FM calculated taking into account data for attenuation collected experimentally. To test the two FMs, the data was acquired at different points of the testing scenario using two mobile phones and one laptop computer with IEEE802.11 communications. Data was then processed using an application developed in Matlab, where the following Location Estimation Algorithms were implemented: Nearest Neighbour; k Nearest Neighbour; Weighted k Nearest Neighbour; an algorithm based on Fuzzy Logic.

II. Experimental Procedure

When fingerprinting is used, one of the first steps is to select the references that will be used to built the FM. It must be chosen references which location will not change, preferably for a long time. Although in the traditional approach to fingerprinting the exact location of the references does not need to be known, in the approach presented in this work it is a *sine qua non* condition. Also the exact location of the points of the spatial domain to be included in the FM, as in the traditional approach to fingerprinting, must be known in beforehand.

Once the exact location of the references and the points that will be considered are known, the FM can then be built using the propagation models. In the FM are included the expected RSS values for each reference at each point. In this step, the expected RSS value for each point of the Fingerprint Map is calculated using two different approaches: using the COST 231 model for indoor propagation [6] and using attenuation data collected from the testing scenario.

A. Using the COST 231 model

The Multi-Wall Model considers that the indoor attenuation between two points is given by Eq. 1:

$$L = L_{FS} + L_C + \sum_{i=0}^{I} k_{wi} L_{wi} + K_f^{\frac{k_f+2}{k_f+1}-b} \qquad (1)$$

where L_C is a constant loss in dB, k_{wi} is the number of penetrated walls of type i, k_f is the number of penetrated floors, L_{wi} is the loss of wall type i in dB, L_f is the loss between adjacent floors, b is an empirical parameter, I the number of wall types and L_{FS} is the free space loss given by Eq. 2:

$$L = L_0 + 10Nlog(d) \qquad (2)$$

where L_0 is the path loss at a distance of $1m$, N is the power decay index and d the distance between the transmitter and the receiver in m.

The parameters considered for the above presented equations were as presented in [6].

Obstacle	Att.(dB)
Brick wall	9.00 dB
Wood door	3.04 dB

TABLE I
Attenuation values for the obstacles in the testing scenario.

B. Using data acquired from the testing scenario

In the second approach, the values for the attenuation are calculated using Eq. 3:

$$L = PL_{FS} + \sum_{i=0}^{n} L_i \qquad (3)$$

where PL_{FS} is the Path Loss in free space in db an L_i is the attenuation of obstacle i, also in db, and n is the total number of obstacles between the transmitter and the receiver.

The term for the Path Loss in free space is the same as presented in Eq. 2, using $N = 2.57$. This value of N was obtained by measuring the pass loss in the working scenario and then calculated using the least squares fitting procedure. In Table I are presented the values for the attenuation (in dB) for the various obstacles found in the testing scenario. These values were obtained measuring the actual value for the attenuation for each of the elements in the table.

C. Location Estimation Algorithms

To test the location system, values of the RSS for each point of the spatial domain are then acquired and stored. Data collected in this step are processed using an application developed in Matlab, which implements the following LEA:

- *Nearest Neighbour* – which considers the coordinates, in the spatial domain, of the nearest point in the FM as the coordinates for the actual location;
- *k-Nearest Neighbour* – where the average of the coordinates, in the spatial domain, of the k nearest points belonging to the FM is calculated. This average is then considered as the coordinates of the current location;
- *Weighted k-Nearest Neighbour* – similar to the previous method, however it is used a weighted average;
- *An algorithm based on Fuzzy Logic* – where it is also made a weighted average of the nearest points.

While in the first three methods the number of point to be considered in the location estimation and their weights are known in beforehand, in the last method the number of points for which the average is calculated and their weights are unknown. These parameters are dependent on the output of the Fuzzy Logic inference engine.

D. Calibration Procedure

For each type of mobile terminal used to determine the location of the user, the expected RSS values are different because the received power depends, among others, of the gain of the receiving antenna (Eq. 4).

$$P_r = P_t - PL + G_t + G_r \qquad (4)$$

where P_r is the received power in dBm, PL is the total Path Loss, in free space and due to obstacles (db), G_t the gain of the transmitting antenna (dB or dBi) and G_r the gain of the receiving antenna (dB or dBi).

This means that for each mobile terminal, used to do the location estimation, the expected values for RSS are different. It is also expected that these values will be different from those calculated using the propagation models.

To overcome these differences, before using the location system, a calibration procedure must be made. This calibration is made by acquiring several values for the RSS at a calibration point, for each reference detected at that point with the mobile terminal under calibration. The average of the differences between the value of the RSS in the Fingerprint Map and the actual value of the RSS reported by the mobile terminal is used as calibration value, Eq. 5:

$$c = \frac{1}{m}\sum_{i=0}^{m}\left(\frac{1}{n}\sum_{j=0}^{n} RSS_i - PFM_i\right) \quad (5)$$

where m is the number of samples and n the number of references.

When estimating the location, the value of the received power to feed to the LEA is not directly the value acquired from the wireless transceiver. The value of RSS considered by the LEA is given by (Eq.6):

$$RSS_{LEA} = RSS - c \quad (6)$$

After calculationg the value of the RSS to feed to the LEA, for comparison purposes, the location will be made using all the above mentioned algorithms.

III. Tests and Results

The chosen scenario (Fig. 1) consists of two classrooms and a corridor connecting them, at the author's university. The small cross marks on the map represent the points where data was collected and for which the Fingerprint Map was generated. Also the location of the five Access Points used as references are presented in the map.

For these tests, five CISCO Aironet 1200 Access Points, equipped with $6.5dBi$ patch antennas were used. Although these are the same equipments used for the wireless network of our University, they were not in service at the moment of the tests. This is to keep the testing environment as stable as possible with the least number of external interferences.

Data to test the FM was collected using a laptop computer (ASUS Notebook K50IN with an Atheros AR9285 Wireless Network Adapter) running Ubuntu Linux, and two Android based mobile terminals, a HTC Desire (Mobile Phone 1) and a Sony Ericsson XPERIA X10 mini pro (Mobile Phone 2). To collect data with these two phones, an application for the Android platform was developed. Also a data acquisition application, presented in [7], was developed in Java for the laptop computer.These applications do not do any data processing. They store the collected data into files that are then downloaded to a computer for processing.

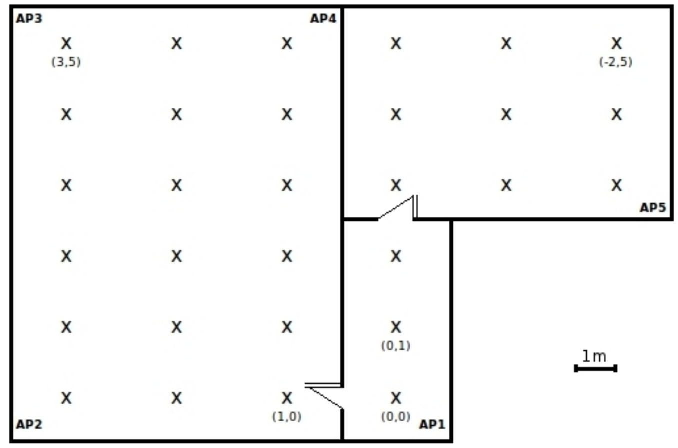

Fig. 1. Map of the area where the location tests were made, the marks represent the location of the points considered for the fingerprint map.

An application, developed using Matlab, was used to determine the location of the mobile node, using the above mentioned LEA. This applications does not make the real-time location of the mobile node, since it uses the data previously collected. It is only intended to assess the performance of the different LEA with the different FM, and to fine tune the parameters for the location algorithms.

Data to test the LEA was gathered at the coordinates signalled in Fig. 1. The coordinates of the map are spaced $2.5m$ from each other, and for each coordinate 20 samples of the RSS values, per available reference, for the three mobile terminals, were stored.

For the first test (Table II) the FM was calculated for the exact same coordinates that were used to acquire the data with the mobile nodes. FM1 was generated by calculating the attenuation between the references and the spatial domain coordinate using the COS231 model, and FM2 was generated using the second propagation method.

Analysing the data it can be concluded that the method with the best accuracy ($3.249m$) was the *k-Nearest Neighbour*, for the laptop computer, and the method with the best standard deviation was the LEA based on Fuzzy Logic ($1.964m$), also using the laptop computer. Comparing the performance of the LEA using both FMs, it can be concluded that the FM that had more best results was FM2. It had 8 best results out of 12, i.e., all the tests made with the mobile phones.

To assess the influence of the grid size on the performance of the chosen LEA, a second version of the fingerprint maps was built using a smaller grid. Each coordinate of the new FM corresponds to $1.25m$, i.e. half of the previous FM. In table III, data of the location estimation using these two new FM are presented.

It is noticeable that the reduction on the grid size had a positive impact on the performance of the HTC mobile phone when using the Fuzzy Logic LEA (with both FM) and for the laptop computer using FM2 using all LEA. In all other LEA for the HTC mobile phone and for all LEA using the second mobile phone the grid change had a negative impact.

LEA	Laptop Computer				Mobile Phone 1				Mobile Phone 2			
	FM1		FM2		FM1		FM2		FM1		FM2	
	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)
Nearest Neighbour	3,625	2,662	4,258	2,513	4,418	2,727	3,936	2,176	6,277	2,998	6,179	2,997
k Nearest Neighbour	3,249	2,224	4,221	2,288	3,845	2,167	3,563	1,969	6,143	2,884	6,109	3,034
W. k Nearest Neighbour	3,362	2,353	4,137	2,321	4,107	2,353	3,703	1,985	6,170	2,893	6,119	2,955
Fuzzy Logic	3,526	1,964	4,487	2,215	4,416	2,187	4,022	2,039	5,824	2,651	5,793	2,714

TABLE II
COMPARISON OF THE PERFORMANCE OF THE VARIOUS LEA, USING THE THREE MOBILE MOBILE NODES, FOR BOTH FM, WITH A GRID SIZE OF $2.5m$

LEA	Laptop Computer				Mobile Phone 1				Mobile Phone 2			
	FM1		FM2		FM1		FM2		FM1		FM2	
	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)	Prec(m)	σ(m)
Nearest Neighbour	3,934	2,717	3,731	2,466	5,219	2,452	4,390	2,347	6,869	3,450	7,195	3,562
k Nearest Neighbour	3,505	2,527	3,691	2,333	4,395	2,266	3,960	2,112	6,776	3,341	7,108	3,311
W. k Nearest Neighbour	3,647	2,632	3,675	2,372	4,760	2,344	4,120	2,219	6,811	3,385	7,130	3,403
Fuzzy Logic	3,421	1,928	4,184	2,188	4,178	1,991	3,793	1,903	6,222	2,796	6,375	2,868

TABLE III
COMPARISON OF THE PERFORMANCE OF THE VARIOUS LEA, USING THE THREE MOBILE MOBILE NODES, FOR BOTH FM, WITH A GRID SIZE OF $1.25m$

IV. CONCLUSION AND FUTURE WORK

In this paper a hybrid approach to fingerprinting was presented. It combines the use of propagation models with fingerprinting, replacing the need for a site survey by generating the FM using indoor propagation models. Using this approach, whenever a change occurs the the scene it is not needed to do another site survey to generate the correct FM, it is only needed to recalculate it based on the propagation models.

Two propagation models were used, one based on measurements made in the testing scenario and the other was the the MWM (COST231). Comparing all the obtained values the best results were obtained using the *k-Nearest Neighbour*.

In the presented tests, the best results were obtained with the laptop computer, using the FM based on COST231. Although the best performance was achieved using the *k-Nearest Neighbour*, the other methods also had a very interesting behaviour, specially Fuzzy Logic which had a best value for the standard deviation. For the mobile phones, using the $1,25m$ grid, the best results were achieved using Fuzzy Logic.

In other tests made by the authors [8], using the same scenario and equipments, on which the FM was generated with data collected from the scenario, the best values for the accuracy for the laptop computer, the HTC mobile phone and for the Sony Ericsson mobile phone were $1.686m$ (Nearest Neighbour), $5.507m$ (Fuzzy Logic) and $3.719m$ (Fuzzy Logic), respectively. It is noticeable an improvement on the performance when using the HTC mobile phone. To be noticed that the FM used in those tests was generated using data collected with the laptop computer. Analysing the results for the other nodes, it can be observed that from the first mobile phone there was a boost on the performance while for the second mobile phone results are worse.

Differences in the overall performance of the used mobile terminal can be related with factors such as the sensitivity of the wireless transceiver, resolution of the RSS indicator and the size of the terminal and consequent influence of the user in the values of the received power.

Although some modification must be made to the location system to improve the accuracy, by all the above presented results it can be concluded that the proposed hybrid technique for fingerprinting can be used. Some of the improvements will be related with the calibration procedure for each mobile terminal, and the use of other type of LEA, such as based on pattern search.

REFERENCES

[1] J. Hightower and G. Borriello, “Location sensing techniques,” University of Washington, Department of Computer Science and Engineering, Seattle, Tech. Rep., July 2001.

[2] P. Bahl and V. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system,” *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 775–784 vol.2, 2000.

[3] C. Komar and C. Ersoy, “Location tracking and location based service using IEEE 802.11 WLAN infrastructure,” in *European Wireless*, Barcelona Spain, February 2004.

[4] A. Taheri, A. Singh, and E. Agu, “Location fingerprinting on infrastructure 802.11 wireless local area networks,” in *LCN ’04: Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 676–683.

[5] H. Liu, H. Darabi, P. Banerjee, and J. Liu, “Survey of wireless indoor positioning techniques and systems,” *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1067 –1080, 2007.

[6] *Digital Mobile Radio Towards Future Generation Systems - COST 231 Final Report*, 1999.

[7] P. Mestre, H. Pinto, C. Serodio, J. Monteito, and C. Couto, “A multi-technology framework for lbs using fingerprinting,” in *Industrial Electronics, 2009. IECON ’09. 35th Annual Conference of IEEE*, 2009, pp. 2693 –2698.

[8] C. Serodio, L. Coutinho, H. Pinto, and P. Mestre, “A Comparison of Multiple Algorithms for Fingerprinting using IEEE802.11,” in *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2011, WCE 2011, 6 -8 July, 2011, London, U.K.*, 2011.