A Comparison of Multiple Algorithms for Fingerprinting using IEEE802.11

Carlos Serodio *Member, IAENG,* Luís Coutinho, Hugo Pinto, Pedro Mestre *Member, IAENG*

***Abstract*—The effectiveness of Location Based Systems depends on the correct location of users and mobile devices. While the outdoor location can be easily calculated, using technologies such as GPS (Global Positioning System), it is more difficult to obtain when the location scenario is an indoor environment. Several technologies and location techniques can be used in this field. One of these techniques is fingerprinting which consists in two different phases: the first phase is the calibration phase when data is collected and the Fingerprint Map is generated; the second phase is the on-line phase where data collected by the mobile device and the data collected in the calibration phase are used to estimate the location of the mobile node. From the several wireless communications technologies IEEE802.11 is probably the most used in Wireless Local Area Networks, so it was chosen for this work. In this paper a comparison between different type of Location Estimation Algorithms is presented: Nearest Neighbour, k-Nearest Neighbour, Weighted k-Nearest Neighbour and an algorithm based on Fuzzy Logic. For these tests three different mobile terminals where used, two mobile phones and one laptop computer. In these tests Fuzzy Logic had the best precision when the terminal used to do the location is not the same used to do the Fingerprint Map. It also had the lower standard deviation for all the test cases. The values of precision and standard deviation, as expected, are dependent on the type of terminal used to do the location.**

***Index Terms*—LBS, Location Estimation Algorithms, Fingerprinting, Fuzzy Logic.**

I. Introduction

From emergency location systems based on mobile networks, such as E911 (Enhanced 911), to the latest concept of applications that are adapted for the end-user and are dynamically delivery based on the user's location [1], all the Location Based Services (LBS) depend on the correct estimation of the users' location.

While in outdoor environments technologies such as GPS (Global Positioning System) can be successfully used, the same is not true when the operating scenario are indoor environments. In such scenarios alternative location technologies and methodologies must therefore be used, making this a very challenging research area where in the last years several different types of solutions have been developed.

Manuscript received March 05, 2011.

C. Serodio is with CITAB - Centre for the Research and Technology of Agro-Environment and Biological Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, email: cserodio@utad.pt

H. Pinto is with UTAD - University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, email: htpinto@gmail.com

L. Coutinho is with UTAD - University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, email: luis_ coutinho_86@hotmail.com

P. Mestre is with CITAB - Centre for the Research and Technology of Agro-Environment and Biological Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, email: pmestre@utad.pt

Some of the most used technologies used for indoor location include the use of infra-red [2], ultrasonic waves [3], [4], pressure sensors [5], RFID (Radio Frequency Identification) [6] and wireless communications networks [7], [8], [9]. In what concerns to the methodologies used to obtain the location, they can be divided into three main areas [10]: Triangulation, Proximity and Scene Analysis.

This paper is focused on a particular location technique, which uses wireless communications networks as location technology, and a methodology based on scene analysis: location using fingerprinting.

One of the objectives of this work is to study methods that can be used to locate standard mobile devices, such as PDAs (Personal Digital Assistant) and mobile phones. These devices usually have wireless communications based on Bluetooh, WiFi (IEEE802.11) and, in the case of mobile phones, have access to the GSM (Global System for Mobile Communications) or 3G communications networks. Bluetooth was not considered for this work mainly due to two reasons: Bluetooth fixed stations are not as ubiquitous as IEEE802.11, GSM or 3G technologies; Bluetooth as a long association time which makes very difficult to use it in the location of moving terminals. Since a better accuracy is obtained with the lower range technologies [11], [12], when choosing between mobile communications (GSM and 3G) or IEEE802.11, the last has been the chosen technology.

Data to generate the Fingerprint Map and perform the location tests were collected using IEEE802.11g Access Points. These tests were made in classrooms of the University of Trás-os-Montes and Alto Douro, using three mobile terminals: a laptop computer and two Android-based mobile phones. To estimate the location of the mobile terminal, four Location Estimation Algorithms were used, and their performance was analysed. The following LEA were considered to do this analysis:

- Nearest Neighbour – which considers the coordinates of the nearest reference as coordinates of the actual location;
- k-Nearest Neighbour – which uses the average of the coordinates of the k nearest neighbours;
- Weighted k-Nearest Neighbour – which uses a weighted average of the coordinates of the k nearest neighbours;
- A Fuzzy Logic based algorithm – which selects which references should be included when determining the current location.

In the Fuzzy Logic based algorithm, the number of neighbours to be considered when calculating the current location and their correspondent weight are dynamically

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

determined by the algorithm, based on a set of rules and membership functions.

II. Location using Wireless Networks

Location using wireless networks is based on the properties of wireless signals. Any property of a wireless signal can be used in location systems, as long as there is a relation between it and the current location of the mobile terminal. The signal properties that usually are used in location systems are:

- Time-of-Flight – the time needed by the information to travel from the transmitter to the receiver;
- Received Signal Strength (RSS) – which indicates the power received by the wireless transceiver.

A. Location using Triangulation

Triangulation uses the geometric properties of triangles to determine the location of the mobile node [10]. It can be divided into Lateration and Angulation.

Laterarion uses the distances to determine the location. It takes into account the distances between the mobile node to be located and the references. It can be made using Circular Triangulation or Hyperbolic Triangulation, with time as independent variable. The distance between the wireless node and the references can be determined based on ToA (Time of Arrival), TDoA (Time Difference of Arrival), or the attenuation value.

To use angulation the angle of incidence of a signal must be known. By analysing the Angle of Arrival of a wireless signal relatively to a given reference, it is possible to determine the location of the mobile node. One example of the application of this method is VOR (VHF Omnidirectional Range).

B. Localization using Proximity

Location using this methodology consists in discovering the nearest reference to the mobile terminal, therefore its spatial resolution is dependent on the number of used references.

Cell of Origin is one of the techniques based on Proximity that can be implemented using wireless networks. It is based on the RF cell concept [13], [14]. To determine the location of a mobile node it is only needed to know the cell on which the node is (e.g. the WiFi access point to which the node is associated or the GSM cell in which the mobile phone is registered). If the location of the cell is known, then the location of the mobile node is also known. Mobile operators use the Cell of Origin method to determine the localization of their subscribers (as first approach).

Another technology that uses the Proximity concept to determine the location of objects is RFID. Two different concepts can be used with this technology: Readers are spread along the scenario and detect the presence of the tags transported by the users, when they are within the reading distance [6]; Tags are embedded in the environment and users transport small portable RFID readers that detect the presence of the tags [15]. In the first case the location of the user is the location of the reader while in the second is the location of the tag.

C. Location using Fingerprinting

Fingerprinting is a scene analysis technique. In scene analysis a scene is "observed" and its patterns and variations along the time are observed. The information about a scene in the case of fingerprinting is obtained from one or more properties of electromagnetic signals from the references.

This location methodology consists in reading a given parameter of an electromagnetic signal in real-time, typically the value of the Received Signal Strength (RSS), and compare it with a set of previously stored values, called the Fingerprinting Map (FM) [8], [16], [17].

Two different concepts of domain are used in fingerprinting: the spatial domain and the signal domain. The first is related with the physical space, i.e., this is the domain where the object to be located is. The second domain is a space of N dimensions (each reference is a dimension), that has several values of RSS for each reference. To do the location of a mobile node using fingerprinting, it is first made a search in the signal domain and then it is made the mapping of the information to the spatial domain.

Location using fingerprinting comprises two different phases:

- Calibration phase, which is an off-line phase, i.e., no location is made. It is in this phase that the FM is generated and the mapping between the spatial and the signal domains is made;
- On-line phase, on which the location of the mobile node is made. In this phase the value of the RSSS signal is acquired from the wireless interface, it is processed by the LEA and the coordinates, in the spatial domain, are calculated.

III. Location using Fingerprinting

In this section some details about the procedures used to determine the location of a mobile node using fingerprinting and the above mentioned LEA algorithms are presented. To do the location, the two phases required by fingerprinting must be made.

Although any property of the wireless signal can be used in this type of analysis, in this work the fingerprinting analysis will be made based on the RSS values.

A. Building the Fingerprint Map

Prior to the acquisition of the fingerprint map, it must be established the location of the points (spatial domain) where the RSS readings are going to be made. Data is then acquired for each one of the previously defined points.

Data collected and stored in this phase includes the value of the RSS for each reference found by the mobile terminal, for each point of the spatial domain. The number of references (and therefore dimensions) might be different for each point in the spatial domain.

After collecting all data, the FM is generated. This is made by calculating the average value of the RSS for every reference at each point. In the database it is then stored for each point and reference the corresponding values of the average RSS.

In this work, to do the fingerprint map, an application developed in Java, presented in [11], was used.

B. Location using the Nearest Neighbour based algorithms

When using the Nearest Neighbour, the k-Nearest Neighbour and Weighted k-Nearest Neighbour, the first step after acquiring the current value of the received power is to determine the distance between the current point (in the signal domain) and all the points that make part of the FM.

In this work to calculate the distance between the points it is used the Euclidean distance (Eq. 1):

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

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 and $P_{FMj,i}$ the value of the power of reference i registered in the FM for the point j.

This distance is calculated for all points that belong to the FM and that contains the reference i. The mapping of the current location between the signal and the spatial domain is then made using one of the following algorithms:

- Nearest Neighbour – The coordinates of the point in the spatial domain which has the shortest distance (in the signal domain) to the current point are considered as the coordinates of the current location;
- k-Nearest Neighbour – The k nearest points to the current point (in the signal domain) are selected and it is calculated the average of their coordinates in the spatial domain. This value is considered as the coordinates of the current location;
- Weighted k-Nearest Neighbour – In this case the procedure is similar to the k-Nearest Neighbour. The only difference is that the average of the coordinates is a weighted average.

C. Location using Fuzzy Logic

In the procedure here described the coordinates for the current location are also estimated by calculating the average of the coordinates of several points of the FM. Fuzzy Logic is used to select which points are the most important to calculate the final coordinates of the current location and to assess their corresponding weight in the average.

As for the other algorithms, the first step, after acquiring the current value of the received power, is to determine the distance in the signal domain between the current location and all the points that make part of the fingerprint map. The next step is to transform these distance values into grades of membership, i.e., it is made the fuzzification. For this phase a set of membership functions, such has the presented in Fig. 1, must be used.

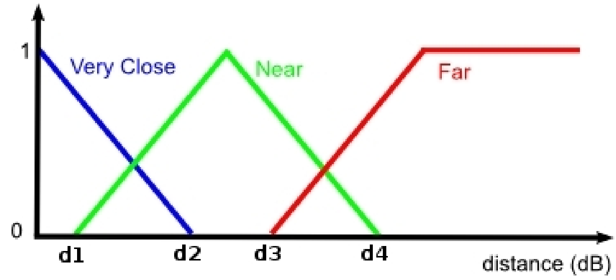

Figure 1. Example of membership functions.

In the above presented example the distance, d (in dB), between the current location and the point of the FM (in the signal domain) can be classified as 'Very Close' if $d < d_2$, 'Near' if $d_1 < d < d_4$ or 'Far' if $d > d_3$.

In this case the distance calculation cannot be made using the Euclidean distance (Eq. 1) because the number of dimensions might not be the same for all points under test. Since the number of dimensions influences the value of the distance, it would require the definition of several sets of membership functions, one per possible number of dimensions. The number of dimensions cannot be predicted beforehand because it will depend on the maximum number of references, the distance to the references and the sensitivity of the mobile receiver. To cope with this, the distance will be calculated using Eq. 2:

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

Based on the distance between the point under test and the reference, a weight for that reference will be chosen. This is done after calculating the grades of membership of the point of the FM under test using the fuzzy inference. In this stage the following simple IF THEN rules are used:

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

To the values for 'high', 'medium' and 'low' are assigned different weights, W_1, W_2 and W_3, such that $W_1 > W_2 > W_3$.

After the defuzzification, the weigh of the point of the FM under test (W_{FMj}) is known, and the current coordinates (C_p) can be calculated using Eq. 3:

$$C_p = \frac{\sum_{j=0}^{n} (W_{FMj} \times C_i)}{\sum_{i=0}^{n} (W_{FMi})} \tag{3}$$

where W_{FMj} is the weight of point j of the FM and C_j represents the point coordinates in the spatial domain.

IV. Tests and Results

In this section are presented the results for the tests made in two classrooms in the University of Trás-os-Montes and Alto Douro. The map of the scenario where the tests where made is presented in Fig. 2.

In the map are marked all the points that were considered in this test. These are the points where the data to generate the FM and to do the the LEA tests were collected. The distance between the points is $2.5m$. In Fig. 2 also the location of the Access Points in the testing scenario is presented.

In this test five IEEE802.11g Cisco Aironet 1200 Access Points, with $6.5dBi$ patch antennas, where used (Fig. 3).

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

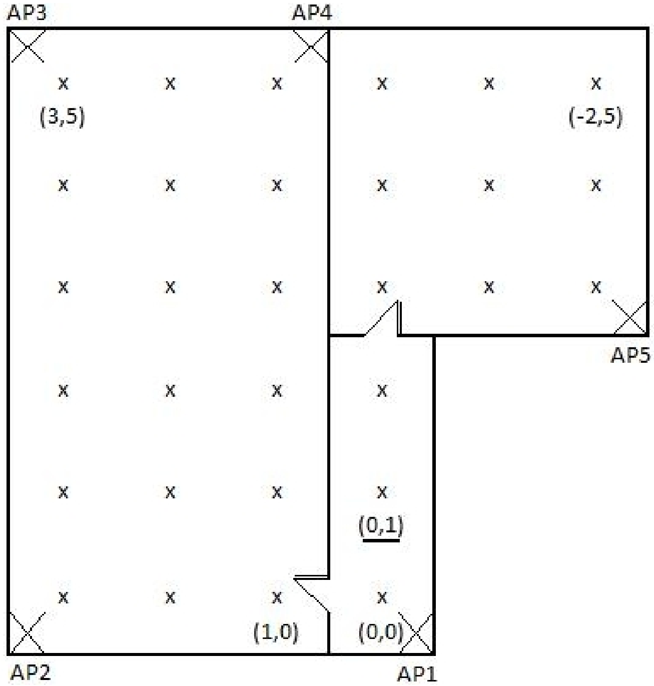

Figure 2. Map of the area where the location tests were made, the marks represent the location of the points considered for the fingerprint map.

These are the same Access Points and antennas used in our University in the 'eduroam' wireless network. In these tests the Acces Points where not under service, they were configured only for our testing scenario. The objective was to use a fully controlled scenario without external influences.

The first step was the acquisition of the data to build the FM. It was made using an application developed using the Java-based Framework (Fig. 4) presented in [11], using a laptop computer (ASUS Notebook K50IN with an Atheros AR9285 Wireless Network Adapter) running Ubuntu Linux. This data collection was made without the presence of the user near the laptop. For each sampling point, 20 values of the RSS for each reference was stored in the database.

After collecting the data to build the FM, it was collected the data to test the chosen LEA. These data were collected using two Android mobile phones, a XPERIA X10 mini

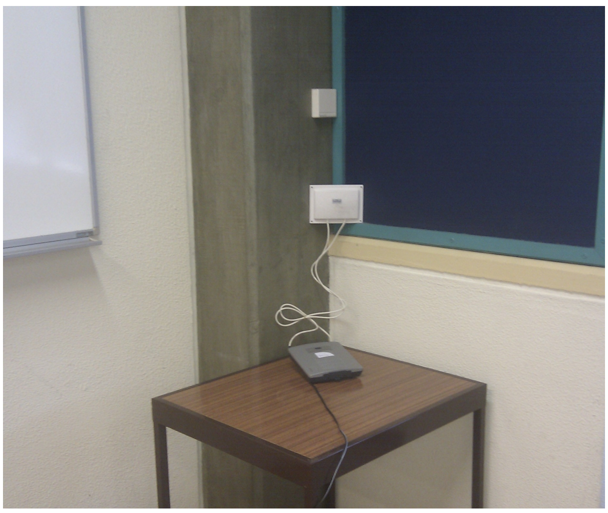

Figure 3. One of the Access Points used in the tests.

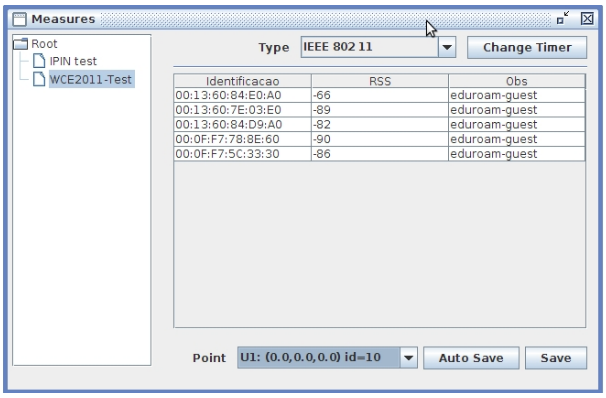

Figure 4. Detail of the Java application used in the laptop computer to acquire data to generate the Fingerprint Map and to test the Location Estimation Algorithms.

(Sony Ericsson) and a HTC Desire.

To collect data with the mobile phones it was developed an application for the Android platform (Fig. 5). Since the objective of this application is to collect data to test the different LEA and to fine tune the weights of Weighted k-Nearest Neighbour and the membership functions of the fuzzy algorithm, this application does not make any data processing. It stores in a file the values of RSS for each Access Point in range, for each sampling point. Data is then uploaded to a computer where it is analysed. To collect data using the laptop computer, it was used same application that was used to collect the data to generate the FM.

To do the data collection using the mobile phones, in the case of the XPERIA X10 mini it was held by the user during the test, with the HTC two tests were made, one with the phone on a table and other with the user holding it.

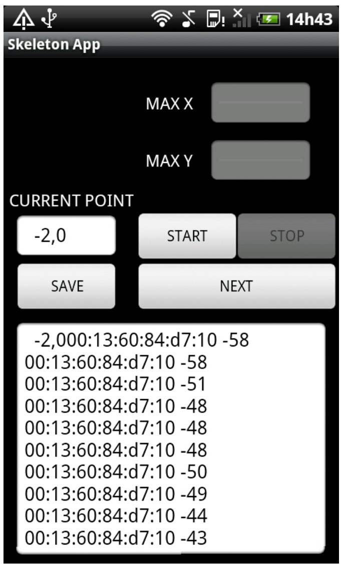

Figure 5. The data collection application developed for the Android platform which was used to collect data using the mobile phones.

Data collected using both applications were analysed using an application developed using Matlab. A detail of this Application is depicted in Fig. 6, where the membership functions used in this work are presented. The values used for the distances were chosen empirically. The marks on the plot represent the different membership degrees for the different

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table I
COMPARISON OF THE VARIOUS METHODS USING THREE DIFFERENT MOBILE TERMINALS.

	Laptop		Mobile 1 (on table)		Mobile 1 (hand)		Mobile 2 (hand)	
LEA	Prec(m)	StDev(m)	Prec(m)	StDev(m)	Prec(m)	Stdevc(m)	Prec(m)	StDev(m)
Nearest Neighbour	1.686	2.190	6.185	1.492	7.327	1.625	5.001	2.036
k-Nearest Neighbour	2.805	1.295	4.857	0.920	6.287	0.952	3.814	1.473
W. k-Nearest Neighbour	2.126	1.495	5.561	1.024	6.850	1.193	4.318	1.638
Fuzzy Logic	3.373	0.883	4.481	0.620	5.507	0.674	3.719	0.810

points of the scene, for the current location. Also the values for 'high', 'medium' and 'low' ($W_1 = 100$, $W_2 = 1$ and $W_3 = 0.01$) were chosen empirically.

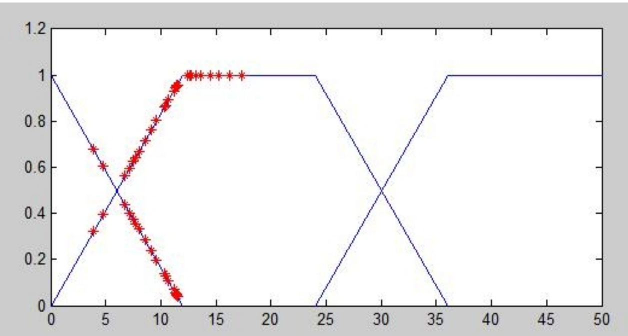

Figure 6. Detail of the output window of the application developed in Matlab where the used membership functions are presented.

In Table I it is presented a comparison of the results obtained with the used Location Estimation Algorithms using the data collected with the three mobile terminals. To be noticed that for both the laptop computer and the HTC mobile phone all five references were detected and for the other mobile phone only four references were detected at each sampling point. In this test a value of $k = 3$ was used and the weights used for the Weighted k-Nearest Neighbour were 0.7, 0.2 and 0.1, as in [11].

When the set of data used to do the location is the same that was used to do the Fingerprint Map, which is the case for the location using the laptop PC, the best method, considering the precision, was the Nearest Neighbour. However, the lower standard deviation was obtained using Fuzzy Logic.

For the mobile phones the best method was always the method based on Fuzzy Logic, either considering the standard deviation or the precision. The best value obtained for the precision was $3.719m$ for the Sony Ericsson mobile phone and the lower standard deviation, $0.620m$, for the HTC mobile phone.

Also to be noticed that, as it was expected, the worse values for the precision was obtained when the mobile device was being held by the user (comparing the values for the same mobile phone).

Observing the histogram for the distances (in the signal domain) off all data samples collected with the three terminals (Fig. 7, Fig. 8, Fig. 9 and Fig. 10), it was already expected the best performance of the Sonny Ericsson in what concerns to the precision, since it has more points that will be classified as 'Very Close'.

To test the influence of the distribution of the distances,

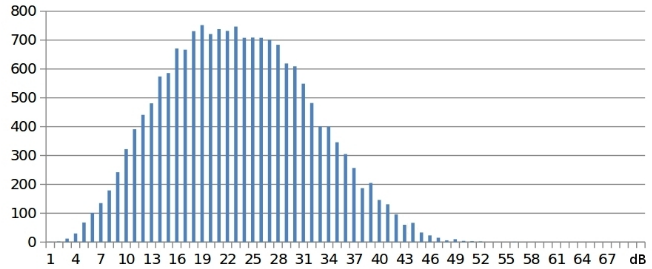

Figure 7. Histogram of the distances in the Signal Domain for all data samples using the PC.

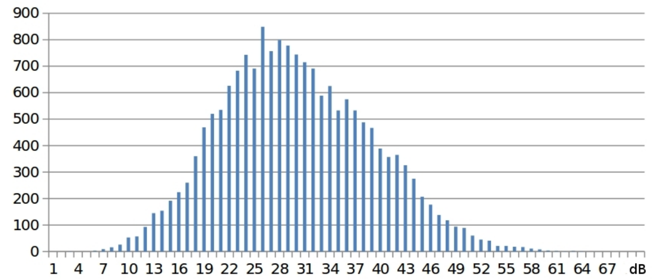

Figure 8. Histogram of the distances in the Signal Domain for all data samples using the HTC mobile phone on a table.

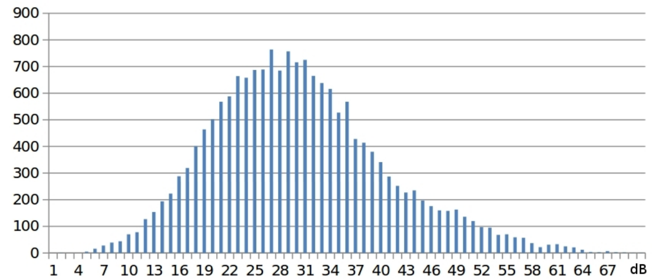

Figure 9. Histogram of the distances in the Signal Domain for all data samples using the HTC mobile phone held by the user.

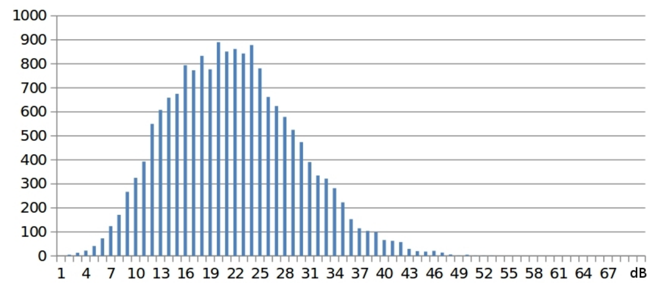

Figure 10. Histogram of the distances in the Signal Domain for all data samples using the Sony Ericsson mobile phone held by the user.

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the histogram for the HCT held by the user (Fig. 9) was shifted to the left (Fig. 11). The new value for the precision was $5.0693m$, which is better than the previous value.

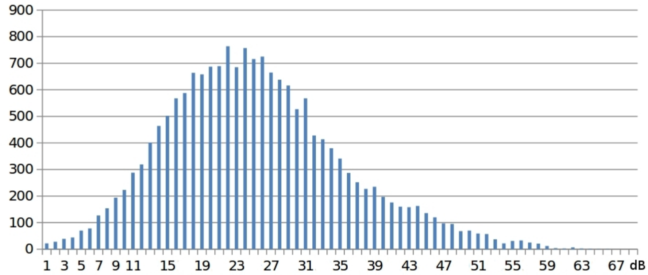

Figure 11. New Histogram for the distances in the Signal Domain for all data samples using the HTC mobile phone help by the user.

V. Conclusion and Future Work

In this paper a comparison between different algorithms for location using IEEE802.11 fingerprinting were presented. Although this work was centred in a particular standard, other wireless network technologies (e.g. IEEE802.15.4) can also be used. In fact in previous works [11],[12] it has been demonstrated that different types of technologies can even be used together to improve the accuracy of the location system.

The Location Estimation Algorithm with the best performance was the method based on Fuzzy Logic, either considering the precision or the standard deviation as metric. With this method a precision of $3.719m$ was achieved when the location is being made with a mobile terminal other than the one used to build the FM. Also the type of mobile terminal has influence on the performance of the location system. However it can be concluded that, for the tested mobile terminals, the used algorithms can be used, and Fuzzy Logic is a feasible solution.

As future work it is planed to study the feasibility of using Fuzzy Logic to do the location using pattern search, based on the information on the FM (RSSI and standard deviation). Also decision algorithms to fine tune the fuzzy rules will be subject to research.

Matlab was used in this work due to its fast prototyping features that allowed a fast, and yet very efficient, development of the algorithms here presented. In the future, the algorithms based on Fuzzy Logic will be converted to Java code and included in the Location Framework.

References

[1] P. Bellavista, A. Kupper, and S. Helal, "Location-based services: Back to the future," *Pervasive Computing, IEEE*, vol. 7, no. 2, pp. 85 –89, 2008.

[2] R. Want, A. Hopper, a. Veronica Falc and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91–102, 1992.

[3] Cricket Project, *Cricket v2 User Manual*, ma 02139 ed., MIT Computer Science and Artificial Intelligence Lab, Cambridge, January 2005, 9-11.

[4] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *Personal Communications, IEEE*, vol. 4, no. 5, pp. 42–47, Oct 1997.

[5] R. J. Orr and G. D. Abowd, "The smart floor: a mechanism for natural user identification and tracking," in *CHI '00: CHI '00 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2000, pp. 275–276.

[6] P. M. Silva, M. Paralta, R. Caldeirinha, J. Rodrigues, and C. Serodio, "Traceme - indoor real-time location system," in *Industrial Electronics, 2009. IECON '09. 35th Annual Conference of IEEE*, 2009, pp. 2721 –2725.

[7] P. Prasithsangaree, P. Krishnamurthy, and P. Chrysanthis, "On indoor position location with wireless LANs," *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, vol. 2, pp. 720–724 vol.2, Sept. 2002.

[8] 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.

[9] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, "Accurate GSM indoor location," in *Mobile Computing (UbiComp 2005)*, Setembro 2005.

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

[11] 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.

[12] P. Mestre, H. Pinto, J. Moura, P. Oliveira, and C. Serodio, "Multiple wireless technologies fusion for indoor location estimation," in *Abstract Volume of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010, pp. 69 –70.

[13] *Wi-Fi Location-Based Services - Design and Deployment Considerations*, Cisco Systems, 2006.

[14] V. Zeimpekis, G. M. Giaglis, and G. Lekakos, "A taxonomy of indoor and outdoor positioning techniques for mobile location services," *SIGecom Exch.*, vol. 3, no. 4, pp. 19–27, 2003.

[15] F. Seco, C. Plagemann, A. Jimé andnez, and W. Burgard, "Improving rfid-based indoor positioning accuracy using gaussian processes," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, 2010, pp. 1 –8.

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

[17] 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.

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)