



Clustering-based location in wireless networks

Luis Mengual, Oscar Marbán *, Santiago Eibe

Facultad de Informática, Universidad Politécnica de Madrid (UPM), Spain

ARTICLE INFO

Keywords:

Indoor location
Wireless lan
802.11 Signal strength
Fingerprint techniques
Clustering

ABSTRACT

In this paper, we propose a three-phase methodology (measurement, calibration and estimation) for locating mobile stations (MS) in an indoor environment using wireless technology. Our solution is a fingerprint-based positioning system that overcomes the problem of the relative effect of doors and walls on signal strength and is independent of network device manufacturers. In the measurement phase, our system collects received signal strength indicator (RSSI) measurements from multiple access points. In the calibration phase, our system utilizes these measurements in a normalization process to create a radio map, a database of RSS patterns. Unlike traditional radio map-based methods, our methodology normalizes RSS measurements collected at different locations (on a floor) and uses artificial neural network models (ANNs) to group them into clusters. In the third phase, we use data mining techniques (clustering) to optimize location results. Experimental results demonstrate the accuracy of the proposed method. From these results it is clear that the system is highly likely to be able to locate a MS in a room or nearby room.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The use of wireless technology has expanded over recent years. Universities, private companies, as well as stations and airports all provide wireless communication facilities. This type of networks can be implemented using what are known as access point devices that provide total connectivity over a surface and connection to Internet. Several such devices are sometimes necessary to cover an office, a university or an airport.

One of the services to be implemented on these networks is user location on the premises. Through location and the visualization of a plan of the coverage area, users can navigate the premises as they would do using GPS navigation in an outdoor environment. Additionally, a network manager would be able to pinpoint users in such environments and interact with them through their positioning. Indoor positioning technology can also complement mobile-telephone-based GPS user positioning to determine the users exact location indoors.

In another context, location tracking is an essential feature for enterprises building business-critical wireless networks. If information technology (IT) staff can identify and track the location of wireless clients and highly mobile assets, they can improve the accuracy of WLAN planning and deployment, optimize ongoing

network performance, enhance wireless security, and improve the usefulness and value of important business applications. Location tracking provides enhanced visibility and control of air space, helping IT staff to deploy wireless networks that are as easy to manage and as effective to deploy as traditional wired networks.

To address lack-of-visibility problems, organizations need a cost-effective, easy-to-deploy solution for tracking and managing thousands of Wi-Fi devices and tags across a variety of business environments.

In heterogeneous environments, e.g. inside a building, though, the received power is a very complex function of distance, wall geometry, building infrastructures and obstacles. Even if you have a detailed model of the building, it takes a lengthy simulation to solve the direct problem of deriving signal strength given the location. This is what has motivated us to consider flexible models based on functions networks (neural networks) to implement a system to locate MSs.

The received signal strength (RSS), a measure of the power received by the client from an access point (AP), is the key parameter for establishing the position of a MS in an indoor environment. Traditionally, WLANs have used different measurement techniques to derive the position of MSs (Kaemarungsi, 2005; Raniwala & Chiueh, 2002). There are three major categories: closest access point-based techniques, location pattern or fingerprint and distance or angle measurement (see Fig. 1).

The *closest AP method* finds devices within the total coverage area of a single AP. It is the simplest but least accurate way to locate a device or user. With the closest AP method, the location tracking system identifies only devices within the total coverage

* Corresponding author. Address: Campus de Montegancedo s/n. DLSIIS, Facultad de Informática, Universidad Politécnica de Madrid (UPM), 28660 Boadilla del Monte, Madrid, Spain.

E-mail addresses: lmengual@fi.upm.es (L. Mengual), omarban@fi.upm.es (O. Marbán), seibe@fi.upm.es (S. Eibe).

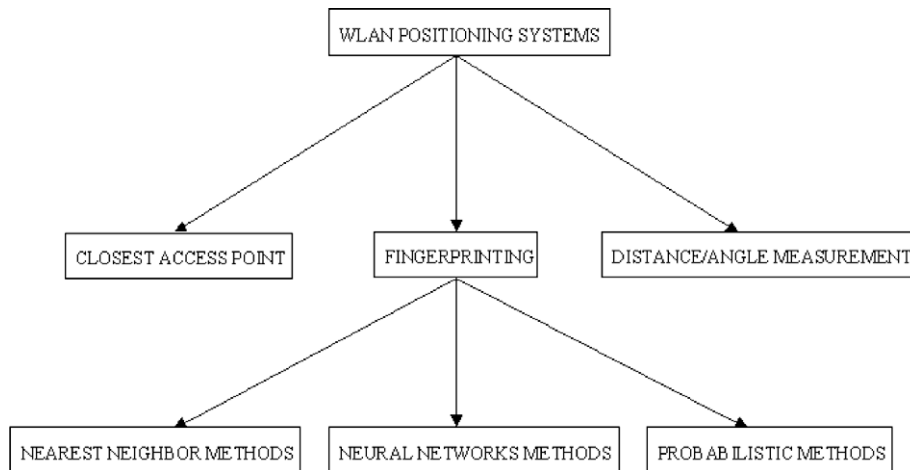


Fig. 1. WLAN positioning techniques.

area of a single AP. This area can be quite large and include multiple rooms. This system considers an AP to which a terminal connects as the user location.

The main distance measurement-based approach is known as *TDOA* (*Time Difference of Arrival*). This technique relies on the timing precision between the signal transmitter and receiver. It uses the propagation delay to calculate the distance between transmitter and receiver (Golden & Bateman, 2007). Therefore, a precise synchronization is very important in such systems. By combining at least three distances from three reference positions, triangulation can be used to estimate the MS's location. Such techniques will require a high accuracy clock in the communication system.

Angle of arrival (*AOA*) is a technique based on angle measurement. This methodology locates the MS by determining the signal angle of incidence. The location can be estimated using simple geometric relationships, like triangulation (Sanchez, Afonso, Macias, & Suarez, 2006).

In indoor environments, however, the distance between transmitter and receiver is usually shorter than the time resolution that the system can measure. Additionally, the MS is surrounded by scattered objects, which results in multiple angles of signal reception. Therefore, the *AOA* and *TDOA* approaches are difficult to implement in indoor environments.

Fingerprinting techniques generally only require measurement of received signal strength or other non-geometric features at several locations to form a database of location fingerprints (Kaemarungsi, 2006; Widyawan, Klepal, & Pesch, 2007). To estimate the mobile location, the system needs to first measure the received signal strength at particular locations and then search for the pattern or fingerprints with the closest match in the database.

Generally, the deployment of fingerprinting-based positioning systems can be divided into two phases. First, the location fingerprints are collected in the off-line or calibration phase by performing a site survey of the received signal strength (RSS) from multiple APs. Enough RSS measurements are taken to set up a database or a table of RSS patterns for predetermined points of the coverage area. The database of RSS patterns with their respective locations is called a *radio map*.

Second, a MS will report a sample measured vector of RSSs from different APs to a central server (or a group of APs will collect the RSS measurements from a MS and send it to the server) in the on-line or estimation phase. The server uses a positioning algorithm to estimate the location of the MS based on the radio map and reports the estimate back to the MS (or the application requesting the position information).

At this point we have several possibilities. The most common algorithm used to estimate the location computes the Euclidean distance between the sample measured RSS vector and each fingerprint in the database. For indoor positioning systems, other advanced algorithms and techniques, such as neural networks (Battiti, Le Nhat, & Villani, 2002; Debono & Buhagiar, 2004; Yiming, Biaz, Pandey, & Agrawal, 2006; Youssef & Agrawala, 2005), probabilistic methods (Chai & Yang, 2007; Kontkanen et al., 2004; Pan, Kwok, Yang, & Chen, 2006; Robinson & Psaromiligkos, 2005; Seigo & Kawaguchi, 2005) or fuzzy logic (Astrain, Villadanos, Garitagoitia, González-Mendivil, & Cholvim, 2006), have been introduced to determine the relationship between RSS samples and the location fingerprint in the radio map.

Other approaches have been presented. For example, some authors have proposed simulating the calibration phase with a ray-tracing model (Nezafat, Kaveh, Tsuji, & Fukagawa, 2004; Nuño & Páez-Borrillo, 2006; Zhong, Bin-Hong, Hao-Xing, Hsing-Yi, & Sarkar, 2001). However, this model requires a highly detailed model of the building. It could be tricky, if not unfeasible, to get such a detailed model. One of the main drawbacks of both of these models is that signal strength prediction is dependent not just on building layout, but also on the position of many other hard-to-model components, such as furniture, equipment and human beings.

So far, fingerprinting techniques with a real calibration phase have attracted more attention because they are a simple and the most effective solution for indoor environments.

Of the solutions that use RSSI to locate a mobile user, some scan the radioelectric spectrum of every square metre of the coverage area. Additionally, they only use one card brand. It is unfeasible to position another mobile user with a different Wi-Fi receiver using this solution.

The solution proposed here is a fingerprint-based positioning system that gets significantly better results than other similar systems because it normalizes measurements and optimizes clusters formed according to the physical features of the ground plan. In this way, the proposed solution is valid for any Wi-Fi receiver, as we account for the relative effect of walls and obstacles as a relative power increase or decrease.

Additionally, our technique does not require an exhaustive scan of the coverage area. Actually we present the results of sampling just one point per location. Our system can adapt to the geometry of each scenario and operate with any Wi-Fi device (irrespective of the manufacturer).

Finally, we present the results of implementing our system in practice and the evaluation tests run in a School of Computing

building (Department of Computer Languages and Systems and Software Engineering). The experimental results confirm the effectiveness of our methodology.

2. Location estimation system issues

Unfortunately, commercial 802.11 hardware does not provide positioning functionality. Even so, positioning can be implemented using RSS measurements provided by an 802.11 wireless card. RSS is a measure of the power received by the client from an AP and provides information as to the clients location.

In heterogeneous environments, e.g. inside a building, though, the received power is a very complex function of distance, geometry, and materials. This is further complicated by the fact that the signal propagation is influenced by environmental factors like, for example, the number of people in the working area, the position of walls and other building infrastructures and the materials they are made of, as well as the multi-path effect (Ali & Nobles, 2007).

Another problem is the RSSI gathered from wireless stations using their device driver. RSSI is an optional parameter that has a value of 0 through to RSSI_Max. This parameter is a measure of the physical sublayer of the energy observed at the antenna used to receive the current frame. RSSI shall be measured from the beginning of the start frame delimiter (SFD) to the end of the frame, i.e. header error check (HEC). RSSI is intended to be used in a relative manner. Absolute RSSI reading accuracy is not specified.

There is nothing in the 802.11 standard that stipulates a relationship between the RSSI value and any particular energy level as could be measured in mW or dBm. Individual vendors have chosen to provide their own levels of accuracy, granularity, and actual power range (measured as mW or dBm) and their RSSI value range (from 0 to RSSI_Max). Some vendors do not provide an RSSI value and instead convert directly from dBm to percent.

In this study we used our own RSS collecting software to test different wireless cards. Each card collected 300 samples over a period of 5 minutes (1 sample/s) on the 3rd Floor of Building 4 at the School of Computing (Universidad Politcnica Madrid). Fig. 2 shows the measurable ranges for the different wireless cards in a room. We found that there was a difference of 20 dB between different cards (3Com/Ovislink). There is even a card (USROBOTICS) that implements RSSI as a percentage and not in dBm units.

From these results we conclude that the mapping between the actual RF energy and the RSSI range in a wireless device varies

from one vendor to another. Since the range and the measurement of RSS depend on the WLAN card, most fingerprint-based positioning systems have used the same wireless card vendor to collect the location fingerprints and determine the location.

The results of our experimental tests match the study run in Kaemarungsi (2006), showing the spread of values for five card manufacturers. In Kaemarungsi (2006), Kaemarungsi even goes as far as to recommend the use of the same card manufacturer for radio map construction-based positioning systems. However, our methodology is capable of working with different wireless cards because we normalize all RSS measurements in the calibration phase.

3. Location estimation methodology

The methodology that we propose can locate any mobile user in any indoor environment equipped with several symmetrically distributed APs. In the following we describe the proposed system.

3.1. Outline of the system

Our system is a three-phase fingerprint-based positioning solution, as shown in Fig. 3.

1. In phase 1, RSS measurements are collected from multiple APs positioned in different rooms on one floor. These will be the location fingerprints used to create the RSS database, called the radio map, of that floor.
2. This information is processed in a calibration phase. The calibration phase is divided into several stages. First, we normalize the collected measurements to identify the relative effect of walls and obstacles. Second, we use neural networks and the computed normalized values to group the measurements in clusters. Finally, we use the physical topology to optimize the clusters formed.
3. In phase 3, called the estimation phase, our system can locate any MS from its current RSS measurements in one or two rooms on the floor. This is possible because of the cluster optimization logic and the applied normalization methodology. In our system we chose to use Kohonen networks as a clustering tool. Kohonen networks represent a type of self-organizing map (SOM), which is actually a special class of neural networks. SOMs are based on competitive learning, where the output nodes compete with each other to be the winning node (or neuron), the only node to be activated by a particular input observation. This

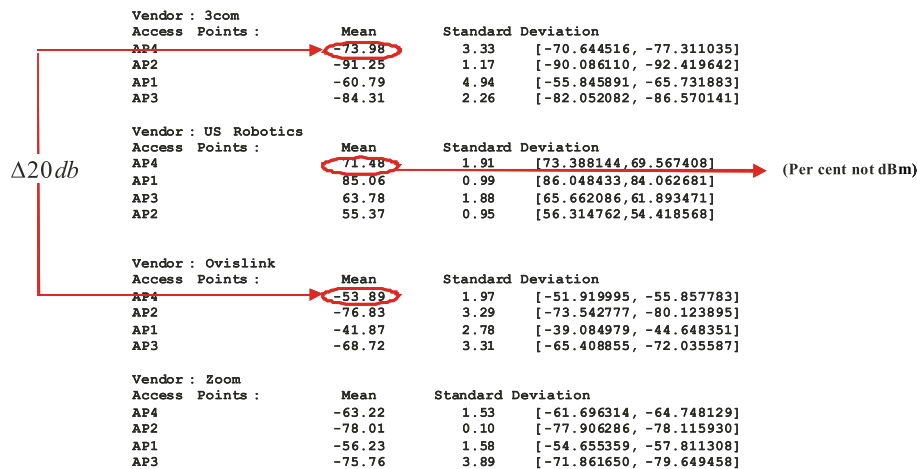


Fig. 2. Measurable RSS range of WLAN cards.

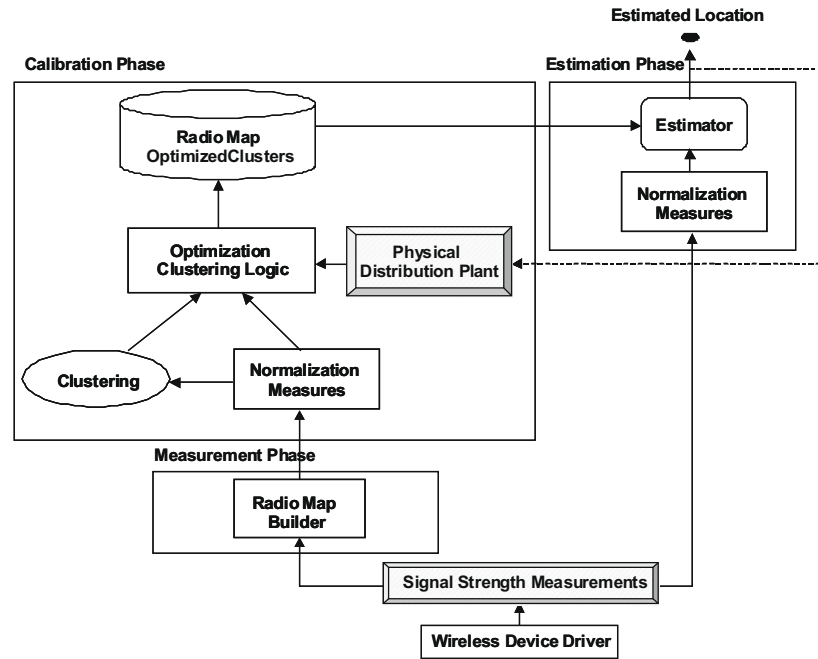


Fig. 3. Location estimation methodology structure.

clustering technique has the advantage of converting a complex high-dimensional input signal into a two-dimensional discrete map.

In the next section we detail the location estimation system.

3.2. Location estimation algorithm

Our location estimation system consists of the following phases (see Fig. 3).

3.2.1. Measurement phase

As already mentioned, the RSSI measurements database is built in this phase. To do this, we systematically measure the signal values relative to the available APs. The calculations, which we will describe later, call for at least three APs arranged to form a triangle. If there are more than three APs, they should be positioned symmetrically to optimize coverage. As many measurements will be taken as necessary to assure that the results are significant.

At the end of the process, a vector S is output for each measurement. This vector represents the signal power values received from k surrounding APs (in this case $k = 4$):

$$S = (s_1, s_2, \dots, s_k) \quad (1)$$

In this paper we will demonstrate that, according to our methodology, the make of the card used to build the radio map is irrelevant. Generally, as explained in Section 2, the values output would be different. However, later data processing makes them valid for estimating the position of any card.

3.2.2. Calibration phase

This phase can be divided into the following stages illustrated in Fig. 4.

Normalization. The normalization of measurements is a key aspect in our methodology. Thanks to normalization, we can take into account the relative effect of walls and obstacles on the signal loss from the surrounding APs. Additionally, the normalization of

measurements allows us to use any wireless device to estimate the position of a mobile terminal.

To understand this key point, let's look at the measurements in Fig. 2. The table data are evidence of the spread of the received RSSI values. This rules out the use of a clustering technique. However, you can solve this problem by referencing all the power values to fixed coordinate sources on the map.

The normalization of measurements involves first setting the normalization reference points. We will set both the reference points and APs existing in the region under analysis will be set. Actually, the coordinates of the measurement point nearest to each AP will be taken as the normalization reference. Consequently, we considered four reference point coordinates x_1, x_4, x_6 and x_7 , respectively, in this study. They correspond to the symmetric positions in which the APs were located.

Second, we calculate the mean value of the measures taken using the different cards in each of the rooms with the set reference points. This way, we get the following data.

Let k be number of APs = number of reference points, and let's take n measurements as follows:

$$m_i = (S_1, S_2, \dots, S_n) \quad \text{where } i = 1, \dots, n \quad (2)$$

measured at reference point:

$$m_j^R = (S_1^R, S_2^R, \dots, S_k^R) \quad (3)$$

e.g. at $R = P1$:

$$m_j^{P1} = (S_1^{P1}, S_2^{P1}, \dots, S_k^{P1}) \quad (4)$$

Mean of the measurements at the reference points:

$$M^R = \bar{m}^R = \left(\frac{\sum_{j=1}^n S_{1j}^R}{n}, \frac{\sum_{j=1}^n S_{2j}^R}{n}, \dots, \frac{\sum_{j=1}^n S_{kj}^R}{n} \right) \quad (5)$$

e.g. at $R = P1$:

$$M^{P1} = \bar{m}^{P1} = \left(\frac{\sum_{j=1}^n S_{1j}^{P1}}{n}, \frac{\sum_{j=1}^n S_{2j}^{P1}}{n}, \dots, \frac{\sum_{j=1}^n S_{kj}^{P1}}{n} \right) \quad (6)$$

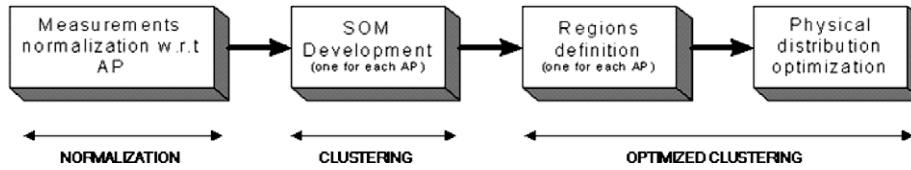


Fig. 4. Detailed calibration phase.

Then we proceed to expand the measures by making:

$$\forall m_i = (S_1, S_2, \dots, S_k), \quad \forall R = 1, \dots, k$$

$$e_i^R = (\bar{m}^R - m_i) = (M^R - m_i) \quad (7)$$

This would give, for one reference point, the vector e_i^R of size $1 * k$ and, for k reference points, we would have the vector E_i of size $1 * (k * k)$

$$E_i = (e_i^{R_1}, e_i^{R_2}, \dots, e_i^{R_k}) \quad (8)$$

Example. If $R = P1, P4, P6, P7$:

$$e_i^{P1} = (M^{P1} - m_i); \quad e_i^{P4} = (M^{P4} - m_i)$$

$$e_i^{P6} = (M^{P6} - m_i); \quad e_i^{P7} = (M^{P7} - m_i) \quad (9)$$

the expansion vector of the measurements would be

$$E_i = (e_i^{P1}, e_i^{P4}, e_i^{P6}, e_i^{P7}) \quad (10)$$

This expansion process is motivated by the analysis of the experimental results, as we will see in the evaluation section.

Clustering and optimized clustering. Having built the radio map of normalized measures our system applies artificial intelligence techniques to estimate the location. We will actually use clustering techniques to form different SOMs from the values of the normalized measurements with respect to each reference point. This way,

we will get as many SOMs as reference points were set in the system.

A classification method validation process can be included in SOM construction, although, as we will see in the examples, this is not always necessary. The process involves dividing the input data into two sets: a training set to form the SOMs and a test data set to test the SOMs and check they are capable of locating a device whose measurements have not been used to form the SOM. With this test set, we can evaluate the performance of the SOMs in terms of percentage of correct locations. The original data from the experiments run are partitioned at random as follows: 90% for the training set and 10% for the test set.

The next step is to optimize the above clusters. To get a preliminary evaluation of the cluster obtained from the positions with respect to their arrangement in the physical space, the cluster is plotted on a chart. Fig. 5 shows this chart.

As the chart shows, all the clusters contain measures taken at different positions. As our ultimate aim is to define regions in which to locate measurements, we decompose the above chart into planes. Each plane contains the measurements of just one position, as shown in Fig. 6. The regions containing points from each position are easier to identify in Fig. 6. The result of the classification will now be associated with a region instead of a position, as had been the case up to this point. Obviously, the drawback of this solution is that it outputs an area rather than pinpointing an exact location. Therefore, the location estimation problem has need of further processing.

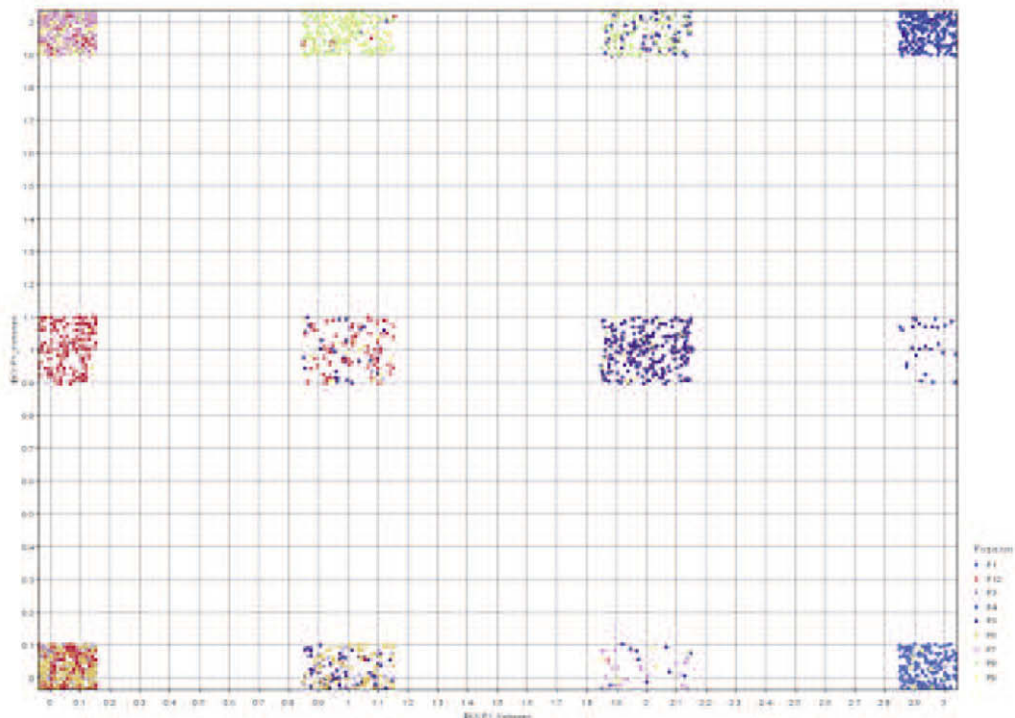


Fig. 5. Cluster visualization.

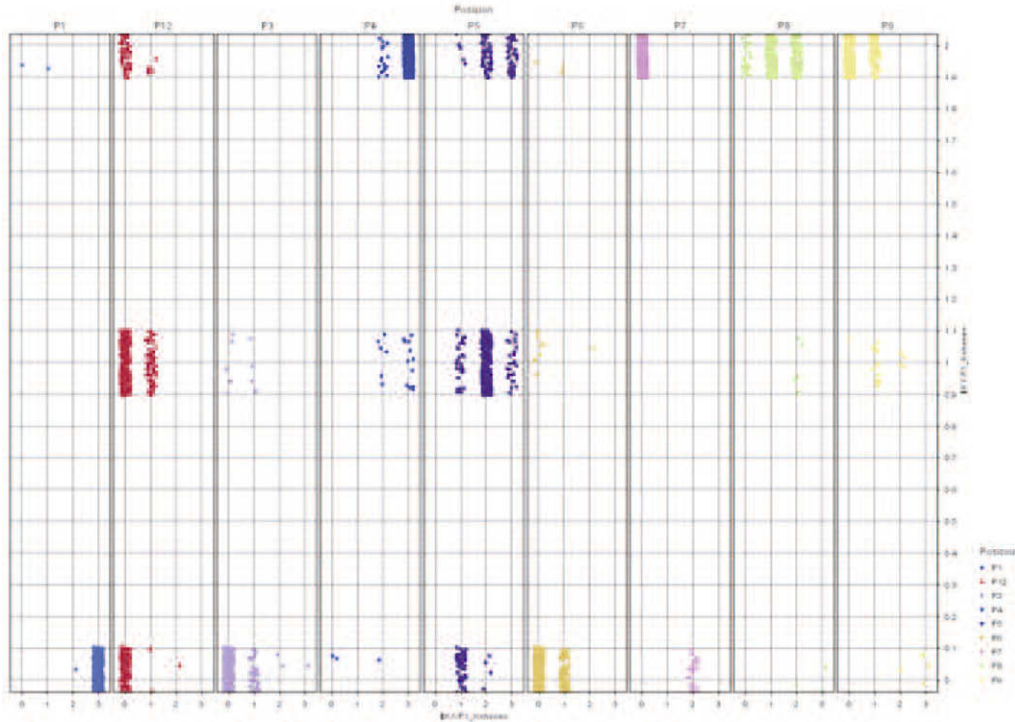


Fig. 6. Decomposition by cluster positions.

3.2.3. Estimation step

The MS location estimation phase gathers information about the power actually received from the nearby APs. With this information and the information about the optimized regions, we position the MS in one of these regions. There are three possible ways of doing this:

1. Exact location estimation: if the n SOMs output one position, this is the MS location. If the chart plots a point located within a region, we do not know the exact location of the MS.
2. Region estimation: the n SOMs output a region, and this region is possibly different for each SOM. This conflict is settled by a voting algorithm, where the position/region that is the most common across all the SOMs is the final result.
3. Optimized region estimation according to the physical distribution of the space: as above, but only the regions that have points that are close together in the physical space are taken into account before the result is shown. For this purpose, rules are built into the system to reduce the number of possible regions for each result. The experimental results (see Section 4) show that the likelihood of locating a station is high. Therefore, it is possible to build optimized regions with a single wireless card brand and use this infrastructure to locate any other brand of wireless device.

4. Results of evaluation experiments

We applied our location estimation methodology shown in Section 4 to implement an estimation location system and conducted evaluation experiments on the 3rd floor of building 4 at the School of Computing. Fig. 7 gives an overview of the experimental environment. The figure shows the floor of part of our university building. The total surface area ranges from 300 to 500 m². Faculty members work in separate rooms. The walls are very thin (plaster). This is an obstacle especially for locating MS in adjoining offices because of the reception of similar powers from the APs.

Red circles show the location of each AP that we used in our tests. Green squares show the locations (12) at which we took measurements of the four AP signal strengths. The normalization points are $P1, P4, P6$ and $P7$. In the survey step, we observed signal strength for 4 min and took 240 measurements at each location.

The hardware that we used in our experiments included seven different IEEE 802.11 wireless adapters; three wireless PC cards (3Com, Enterasys and US Robotics); three wireless USB 2.0 adapters (SMC, Zoom and Ovislink) and one integrated internal Intel Wi-Fi. APs are manufactured by US Robotics and 3Com.

We have divided the results into two parts for presentation, depending on the measurements used to create the clusters. The first part presents the results of using the measurements collected from all the wireless cards. The second part uses measurements from just one SMC wireless card. In this case, the data do not have to be divided into a training and a test set, as the SMC card data form the training set and the data of the other cards (not used to define the SOMs) form the test set. In both cases, we have created four different SOMs using the normalized measurements with respect to the normalization points $P1, P4, P6$ and $P7$ in Fig. 7.

As the first step in the explanation of our method, we are going to discuss an example using the experimental values (absolute or pre-normalized values) obtained by different cards in the same place (any place) on the floor in question:

$$\begin{aligned}
 m_{i,3Com} &= (-92, -74, -67, -87) \\
 m_{i,Enterasys} &= (-86, -70, -52, -83) \\
 m_{i,Ovislink} &= (-75, -54, -49, -64) \\
 m_{i,SMC} &= (-82, -71, -53, -79) \\
 m_{i,USRobotics} &= (+56, +70, +86 + 63)
 \end{aligned} \tag{11}$$

Note that, as mentioned in Section 2, there is a wide spread of values. This makes it impossible to directly apply the clustering technique.

However, we can transform these absolute values into relative values as explained in Section 3.2.2. This way, by normalizing the

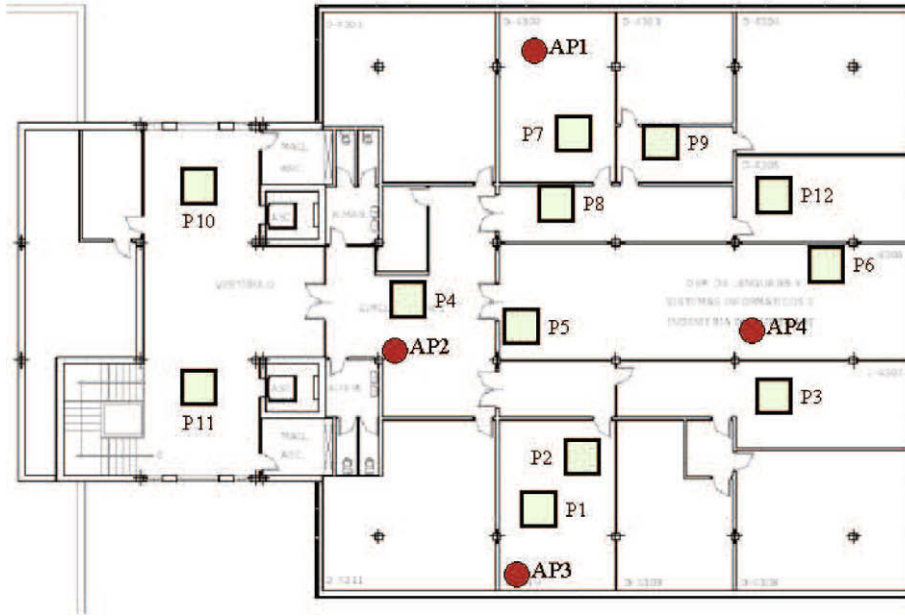


Fig. 7. Floor plan.

values with respect to the reference value $P1$ in each case, we get the expansion vectors:

$$\begin{aligned} e_{i,3Com}^{P1} &= (-0.75, -0.02, -6.21, -2.69) \\ e_{i,Enterasys}^{P1} &= (-2.19, -0.94, -2.40, +1.83) \\ e_{i,Ovislink}^{P1} &= (+1.83, -0.11, -7.13, +4.72) \\ e_{i,SMC}^{P1} &= (-0.29, +0.67, +2.27, +1.99) \\ e_{i,USRobotics}^{P1} &= (+0.63, -1.48, +0.94, -0.78) \end{aligned} \quad (12)$$

By normalizing the values with respect to the reference point $P4$, we get:

$$\begin{aligned} e_{i,3com}^{P4} &= (-5.52, -21.87, +19.63, +0.12) \\ e_{i,Enterasys}^{P4} &= (-6.61, -39.14, +21.03, +1.18) \\ e_{i,Ovislink}^{P4} &= (-6.46, -24.61, +19.49, +3.04) \\ e_{i,SMC}^{P4} &= (-7.07, -31.10, +22.51, -0.01) \\ e_{i,USRobotics}^{P4} &= (-8.18, -34.09, +22.76, +1.58) \end{aligned} \quad (13)$$

By normalizing the values with respect to the reference point $P6$, we get:

$$\begin{aligned} e_{i,3com}^{P6} &= (-7.34, +0.83, +22.18, -28, 67) \\ e_{i,Enterasys}^{P6} &= (-6.9, -7, 59, +25, 34, -22.46) \\ e_{i,Ovislink}^{P6} &= (-13.61, +1.34, +23.74, -19.89) \\ e_{i,SMC}^{P6} &= (-12.67, -5.43, +26.58, -25.12) \\ e_{i,USRobotics}^{P6} &= (-11.83, -7.51, +29.24, +22.24) \end{aligned} \quad (14)$$

and, finally, by normalizing the values with respect to the reference value $P7$, we get:

$$\begin{aligned} e_{i,3com}^{P7} &= (-23.92, +6.12, +35.59, +2.01) \\ e_{i,Enterasys}^{P7} &= (-30.28, +3, 21, +40.38, +1.02) \\ e_{i,Ovislink}^{P7} &= (-35.82, +9.74, +34.58, +3.03) \\ e_{i,SMC}^{P7} &= (-31.08, +4.74, +37.37, +1.78) \\ e_{i,USRobotics}^{P7} &= (-33.73, +2.09, +34.87, +2.24) \end{aligned} \quad (15)$$

After analysing the experimental results, we find that we have managed to get very similar relative, card-independent values using this method. This is because we measure the relative difference between two points across all the cards. This way, we can enter these new relative data into our clustering system. The deviations in the measurements, as shown in Fig. 8, are insignificant compared with the absolute values of the analysed cards.

Additionally, we could undertake the clustering process from k perspectives (see Section 3.2.2). This will make the mobile user location more precise.

4.1. Experiment 1: clustering data for all wireless cards

This is the first experiment we did by calibrating seven different wireless devices and collecting received power data from the specified APs at the locations of interest.

After building the four SOMs, we named the clearly separate regions in each one. For naming purposes, the SOM is plotted on a chart. The measurement position indicated in the data is superposed, and, as you can see, the positions are clearly grouped (see Fig. 9).

Having named the regions, each SOM will classify the measurements in a particular location region or area where the wireless device could be located. Therefore, as we have created four SOMs, we will have four regions or locations for the device depending on the AP on which the measurement is normalized. This way, we have located the wireless device. If the result of the four SOMs is the same position, this is the device's location. Otherwise we have to analyse the different SOM responses to be able to locate the device.

Fig. 10 shows the confusion matrix for the SOMs using the test set data. The table rows list the positions predicted by the system, whereas the columns list the real positions. Clearly, the system correctly predicts position $P1$, for example, in 98.15% of the cases, and these points are incorrectly classified in 1.85% of the cases.

These results are quite acceptable for some positions, like $P1$, $P12$ or $P4$, for example, but not for others, like $P3$ or $P7$, where most of the test cases are classified incorrectly. Many of these incorrect classifications correspond to the region defined as *multiple*. This means that the four SOMs indicate that the wireless device is equally likely to be located in several positions.

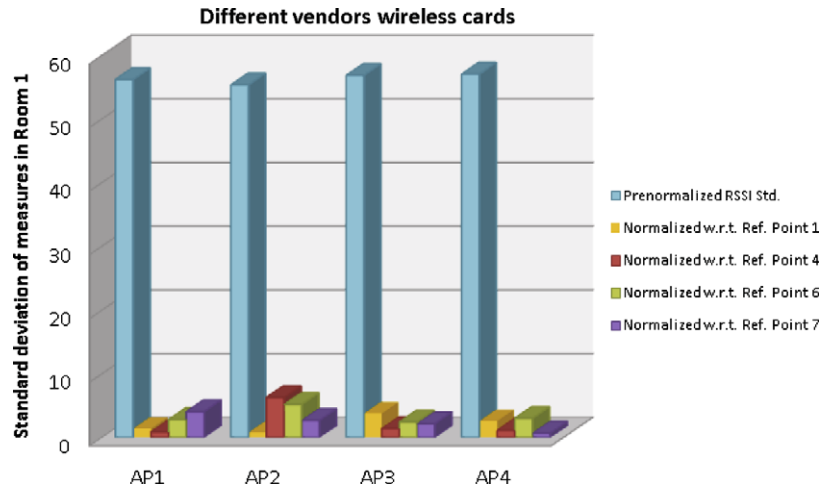


Fig. 8. Comparison of absolute/normalized measurement STD.

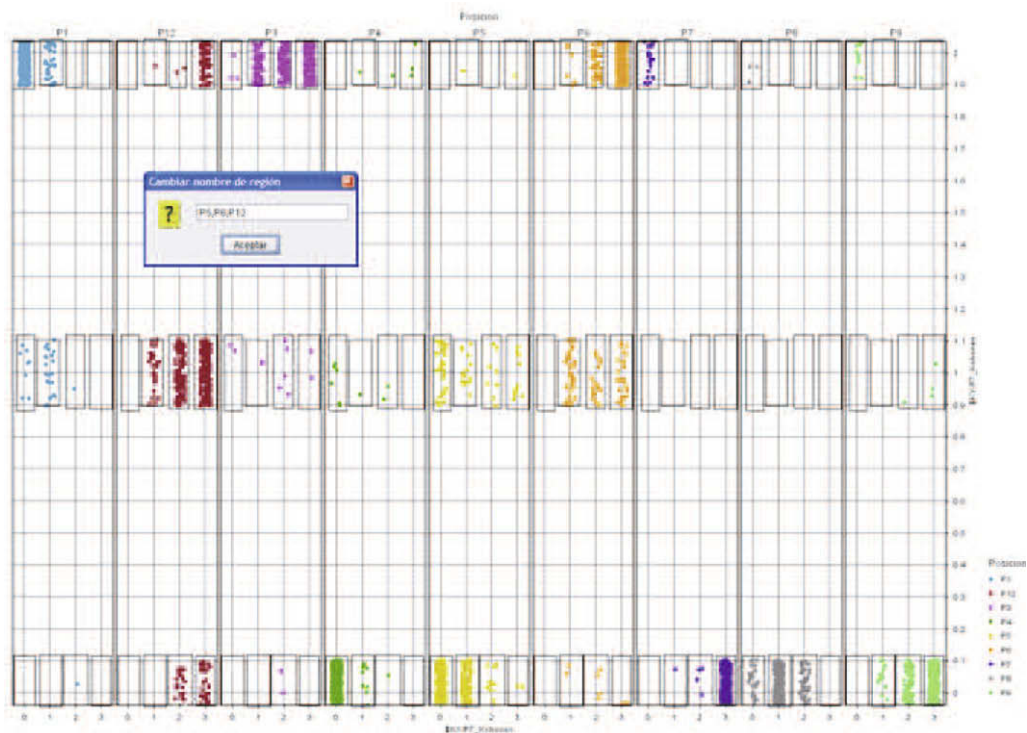


Fig. 9. Region naming.

\$Position Count	Position								
\$Position	P1	P12	P3	P4	P5	P6	P7	P8	P9
Multiple	1,85%	29,63%	88,89%	0,00%	38,89%	72,22%	84,44%	33,33%	77,78%
P1	98,15%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	3,70%
P12	0,00%	70,37%	0,00%	0,00%	1,85%	5,56%	0,00%	0,00%	1,85%
P3	0,00%	0,00%	1,85%	0,00%	0,00%	16,67%	0,00%	0,00%	0,00%
P4	0,00%	0,00%	0,00%	98,15%	16,67%	0,00%	0,00%	4,44%	0,00%
P5	0,00%	0,00%	1,85%	0,00%	33,33%	0,00%	0,00%	0,00%	0,00%
P6	0,00%	0,00%	7,41%	0,00%	0,00%	5,56%	0,00%	0,00%	0,00%
P9	0,00%	0,00%	0,00%	1,85%	9,26%	0,00%	15,56%	62,22%	16,67%
Corrected	98,15%	70,37%	1,85%	98,15%	33,33%	5,56%	0,00%	0,00%	16,67%

Fig. 10. Confusion matrix for SOM predictions.

A voting algorithm has been used to find out the position or positions (also termed regions) of the wireless device that the SOMs together locate in multiple positions. In this algorithm, each

SOM indicates the region that it assigns to the device. The final region is the most voted by all the SOMs. The percentage of correctly assigned regions is calculated using the following formula:

$$\sum_{Red}^{Reg \Rightarrow Pos} \%Corrected_{Reg} \tag{16}$$

Fig. 11 shows the result of using the location regions instead of the exact measurement position. Comparing the total results (corrected line) in Figs. 10 and 11, we find that the predictions have improved in all cases as a result of using the regions. Taking position P3, for example, we find that the system gives a correct response in 1.85% of the cases using positions, whereas the success rate is 90.74% if we use regions.

Fig. 12 is a graph comparing the success rate for positions and regions. It shows that the system's success rate increases significantly using regions.

The system's success rate increases using the regions, but the approach has the drawback of there being very extensive regions covering a large part of the physical space in which the wireless device can be located, e.g. [P4, P5, P8, P9] or [P5, P9, P12]. To reduce the positioning area, at this point we enter information about the layout of the physical space of the premises where the system is going to be used, e.g. information related to the proximity of the

\$Region Count	Position								
\$Region	P1	P12	P3	P4	P5	P6	P7	P8	P9
P1	98.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.70%
P1,P5	1.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P12	0.00%	70.37%	0.00%	0.00%	1.85%	5.56%	0.00%	0.00%	1.85%
P3	0.00%	0.00%	1.85%	0.00%	0.00%	16.67%	0.00%	0.00%	0.00%
P3,P6	0.00%	0.00%	53.70%	0.00%	0.00%	7.41%	0.00%	0.00%	0.00%
P3,P6,P12	0.00%	16.67%	35.19%	0.00%	0.00%	51.85%	0.00%	0.00%	0.00%
P4	0.00%	0.00%	0.00%	98.15%	16.67%	0.00%	0.00%	4.44%	0.00%
P4,P5	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%
P4,P5,P8	0.00%	0.00%	0.00%	0.00%	9.26%	0.00%	0.00%	2.22%	0.00%
P4,P5,P8,P9	0.00%	0.00%	0.00%	0.00%	1.85%	0.00%	0.00%	0.00%	0.00%
P4,P9	0.00%	0.00%	0.00%	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%
P5	0.00%	0.00%	1.85%	0.00%	33.33%	0.00%	0.00%	0.00%	0.00%
P5,P6	0.00%	0.00%	0.00%	0.00%	0.00%	1.85%	0.00%	0.00%	0.00%
P5,P8,P9	0.00%	0.00%	0.00%	0.00%	1.85%	0.00%	0.00%	13.33%	0.00%
P5,P9	0.00%	0.00%	0.00%	0.00%	9.26%	0.00%	0.00%	2.22%	0.00%
P5,P9,P12	0.00%	0.00%	0.00%	0.00%	1.85%	0.00%	0.00%	0.00%	0.00%
P6	0.00%	0.00%	7.41%	0.00%	0.00%	5.56%	0.00%	0.00%	0.00%
P6,P12	0.00%	7.41%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%
P7,P9,P12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	82.22%	0.00%	51.85%
P8,P9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.33%	3.70%
P9	0.00%	0.00%	0.00%	1.85%	9.26%	0.00%	15.56%	62.22%	16.67%
P9,P12	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	2.22%	2.22%	22.22%
Corrected	100.00%	100.00%	90.74%	98.15%	68.52%	77.78%	82.22%	28.89%	94.44%

Fig. 11. Regions confusion matrix.

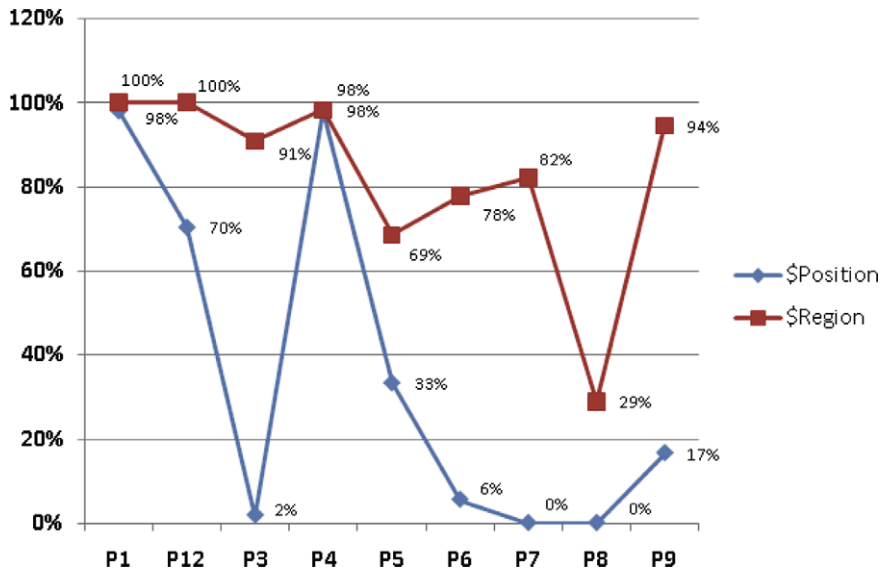


Fig. 12. Position vs. region prediction comparison.

\$Region Count	Position								
\$Region	P1	P12	P3	P4	P5	P6	P7	P8	P9
P1	98.15%								
P4,P5				98.15%	61.11%				
P7,P9,P12		5.56%					97.78%		90.74%
P3,P6,P12		94.44%	90.74%			92.59%			
P6,P9,P12		83.33%				16.67%			40.74%
P8,P9								75.56%	
Corrected	98.15%	94.44%	90.74%	98.15%	61.11%	92.59%	97.78%	75.56%	90.74%

Fig. 13. Optimized region results.

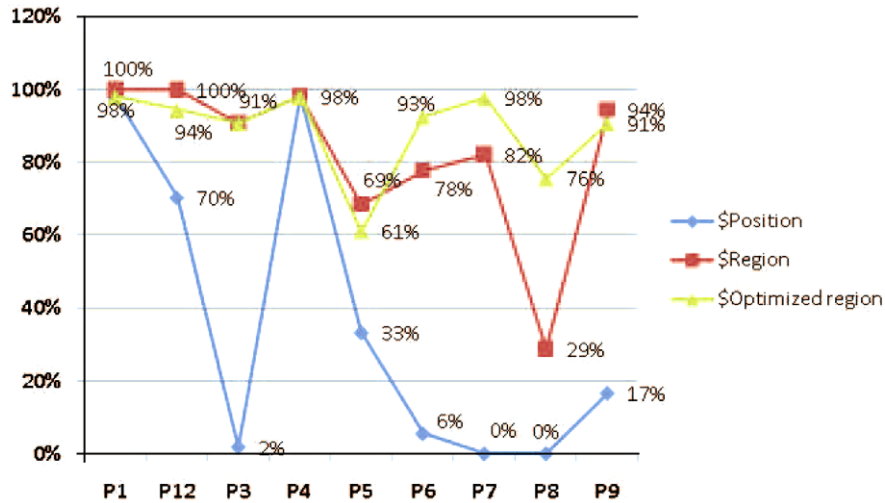


Fig. 14. Comparison between estimated position, region and optimized region.

regions. The results with the optimized regions are shown in Fig. 13. Comparing the results with the non-optimized regions (see Fig. 14), we find that the success rates drops in some cases and increases in others. The advantage is, though, that the regions are quite a lot smaller, and the positioning is much more accurate in most cases.

Fig. 14 is a comparison of the three wireless device location estimation methods. In view of these results, we conclude that the best and most accurate way to do things is to use regions optimized according to the physical arrangement of the premises.

4.2. Experiment 2: clustering data for one wireless card

This experiment addresses a real operating environment where there is at first just one type of wireless device on which to take the

measurements. As wireless devices of other brands are acquired, they are located without modifying the SOMs or the regions defined for the first device.

Therefore, unlike the last experiment, where all the measurements on the available wireless devices were used, we have not divided the data into a training set and a test set. In this case, the SMC card data form the training set and the data on the other cards (not used to define the SOMs) make up the test set. Fig. 15 shows the results of locating different devices from the data used to create the SOMs. As you can see, success rate is over 90% in most cases.

Fig. 16 shows the comparison between using the position, the region and the optimized region if a single card is used to create the SOMs.

Finally, we can compare the results of using data from a single wireless device and using the data from all the devices to create

\$Region Count	Position	P12	P3	P4	P5	P6	P7	P8	P9
\$Region	P1	97,78%							
P1			81,11%						
P7,P8,P9						95,33%			
P3,P6			94,89%						
P4,P5				96,89%	74,00%				
P7,P9			64,00%				88,06%		91,11%
P8,P9			60,00%					91,39%	
Corrected		97,78%	81,11%	94,89%	96,89%	74,00%	95,33%	88,06%	91,39%

Fig. 15. Optimized region results.

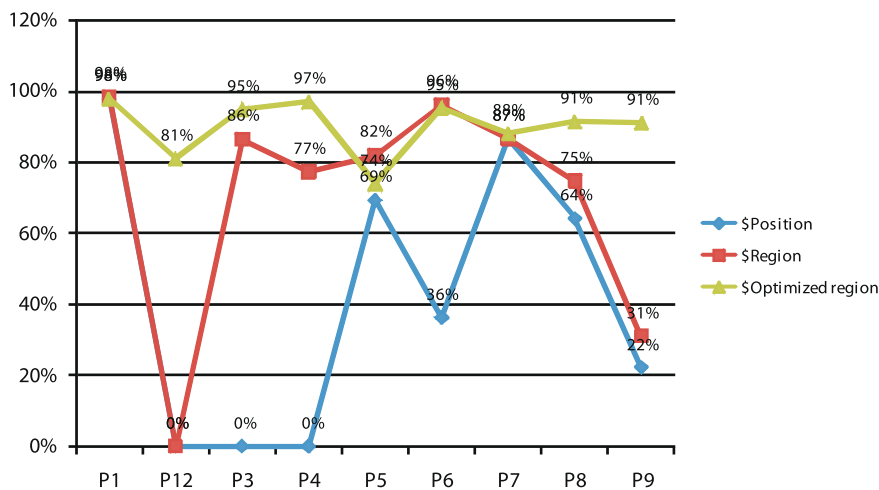


Fig. 16. Comparison between estimated position, region and optimized region.

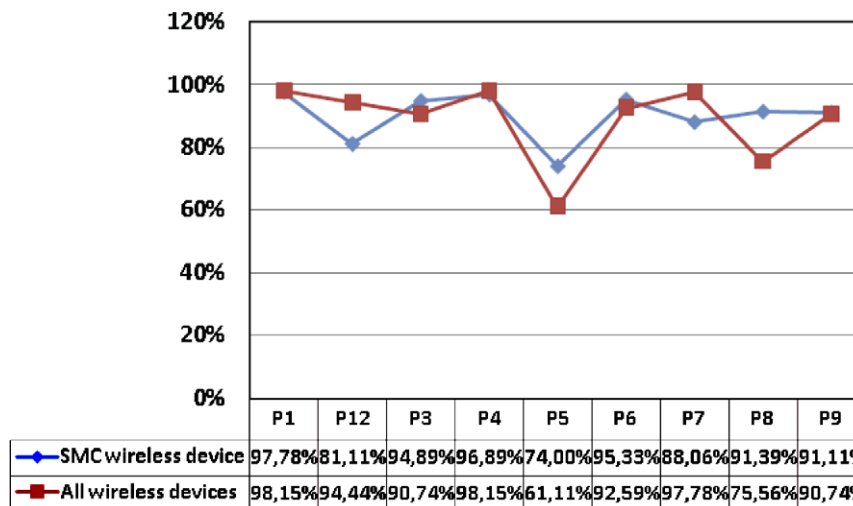


Fig. 17. One wireless device vs. all wireless devices used to build SOMs.

the SOMs. Fig. 17 shows that the results are quite similar. The outcome is sometimes better using all the cards, and the use of just one card gets better results in other cases.

From these results we can conclude that the system can be used in any environment with a wireless device and at least three APs.

5. Conclusions

We conceived and developed a system that is highly likely to correctly estimate the location of a mobile terminal indoors in a room or adjacent rooms.

Unlike many other implemented systems, the resulting system is independent of the card brand and hardware technology (USB, PCMCIA) manufacturer. Our system evaluates the position by relative differences in the measurements taken at places close to the APs. The system is also adaptable to any environment. All it takes is system training with any card at locations on the floor, the normalization of the measurements and the creation of clusters. As of this point, the system is ready to locate any user with any wireless card.

Role of the funding source

This work was conducted as part of the CYCIT-funded Project No. TIN2008-05924.

References

- Ali, S., & Nobles, P. (2007). A novel indoor location sensing mechanism for IEEE 802.11 b/g wireless lan. In *4th Workshop on positioning, navigation and communication (WPNC '07)* (pp. 9–15).
- Astrain, J. J., Villadangos, J., Garitagoitia, J. R., González-Mendivil, J. R., & Cholvim, V. (2006). Fuzzy location and tracking on wireless networks. In *International workshop on mobility management and wireless access* (pp. 84–91).
- Battiti, R., Le Nhat, T., & Villani, A. (2002). Location-aware computing: A neural network model for determining location in wireless lans. Technical Report DIT-5. Università di Trento, Dipartimento di Informatica e Telecomunicazioni.
- Chai, X., & Yang, Q. (2007). Reducing the calibration effort for probabilistic indoor location estimation. *IEEE Transactions on Mobile Computing*, 6(6), 649–662. June.

- Debono, C. J., & Buhagiar, J. K. (2004). Neural location detection in wireless network. In *7th European conference on wireless technology*.
- Golden, S. A., & Bateman, S. S. (2007). Sensor measurements for Wi-Fi location with emphasis on time-of-arrival ranging. *IEEE Transactions on Mobile Computing*, 6(10), 1185–1198.
- Kaemarungsi, K. (2005). Design of indoor positioning systems based on location fingerprinting technique. PhD thesis. University Pittsburg.
- Kaemarungsi, K. (2006). Distribution of WLAN received signal strength indication for indoor location determination. In *1st International symposium on wireless pervasive computing*.
- Kontkanen, P., Myllymäki, P., Roos, T., Tirri, H., Valtonen, K., & Wettig, H. (2004). Topics in probabilistic location estimation in wireless networks. In *15th IEEE international symposium on personal, indoor and mobile radio communications, Barcelona, Spain*. IEEE Press.
- Nezafat, M., Kaveh, M., Tsuji, H., & Fukagawa, T. (2004). Localization of wireless terminals using subspace matching with ray-tracing-based simulations. In *Sensor array and multichannel signal processing workshop proceedings* (pp. 623–627).
- Nuño, G., & Páez-Borrillo, J. M. (2006). New location estimation system for wireless networks based on linear discriminant functions and hidden Markov models. *EURASIP Journal on Applied Signal Processing*.
- Pan, J. J., Kwok, J. T., Yang, Q., & Chen, Y. (2006). Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. *IEEE Transactions on Knowledge and Data Engineering*, 18(9), 1181–1193.
- Raniwala, Ashish, & Chiueh, Tzi-Cker (2002). Deployment issues in enterprise wireless lans. *Wireless Communications*.
- Robinson, M., & Psaromiligkos, I. (2005). Received signal strength based location estimation of a wireless lan client. *IEEE Wireless Communications and Networking Conference*, 4, 2350–2354.
- Sanchez, D., Afonso, S., Macias, E. M., & Suarez, A. (2006). Devices location in 802.11 infrastructure networks using triangulation. In *The 2006 IAENG international workshop on wireless networks, Hong Kong*.
- Seigo, I., & Kawaguchi, N. (2005). Bayesian based location estimation system using wireless lan. In *Proceedings of the third IEEE international conference on pervasive computing and communications workshops* (pp. 273–278).
- Widyawan, K., Klepal, M., & Pesch, D. (2007). Influence of predicted and measured fingerprint on the accuracy of RSSI-based indoor location systems. In *4th Workshop on Positioning, Navigation and Communication (WPNC '07)* (pp. 145–151).
- Yiming, J., Biaz, S., Pandey, S., & Agrawal, P. (2006). Ariadne: A dynamic indoor signal map construction and localization system. In *International conference on mobile systems, applications and services* (pp. 151–164).
- Youssef, M., & Agrawala, A. K. (2005). The Horus WLAN location determination system. In *Third international conference on mobile systems, applications, and services (MobiSys 2005)*, Seattle, WA, USA.
- Zhong, J., Bin-Hong, L., Hao-Xing, W., Hsing-Yi, C., & Sarkar, T. K. (2001). Efficient ray-tracing methods for propagation prediction for indoor wireless communications. *IEEE Antennas and Propagation Magazine*, 43(2).