

Article

A Hybrid Stochastic Approach for Self-Location of Wireless Sensors in Indoor Environments

Jaime Lloret *, Jesus Tomas, Miguel Garcia and Alejandro Canovas

Departamento de Comunicaciones, Universidad Politécnica de Valencia. Camino Vera s/n, 46022, Valencia, Spain; E-Mails: jtomas@dcom.upv.es; migarpi@posgrado.upv.es; alcasol@epsg.upv.es

* Author to whom correspondence should be addressed; E-Mail: jlloret@dcom.upv.es; Tel.: +34-609-549-043; Fax: +34-962-849-313

Received: 2 April 2009; in revised form: 4 May 2009 / Accepted: 14 May 2009 /

Published: 15 May 2009

Abstract: Indoor location systems, especially those using wireless sensor networks, are used in many application areas. While the need for these systems is widely proven, there is a clear lack of accuracy. Many of the implemented applications have high errors in their location estimation because of the issues arising in the indoor environment. Two different approaches had been proposed using WLAN location systems: on the one hand, the so-called deductive methods take into account the physical properties of signal propagation. These systems require a propagation model, an environment map, and the position of the radio-stations. On the other hand, the so-called inductive methods require a previous training phase where the system learns the received signal strength (RSS) in each location. This phase can be very time consuming. This paper proposes a new stochastic approach which is based on a combination of deductive and inductive methods whereby wireless sensors could determine their positions using WLAN technology inside a floor of a building. Our goal is to reduce the training phase in an indoor environment, but, without an loss of precision. Finally, we compare the measurements taken using our proposed method in a real environment with the measurements taken by other developed systems. Comparisons between the proposed system and other hybrid methods are also provided.

Keywords: indoor location system; positioning system; WLANs

1. Introduction

Currently, sensor networks are the main part of many monitoring and control systems. Many of them tend to be wireless because it allows them to be spatially distributed. Wireless Sensor Networks (WSNs) [1] are formed dynamically because the connectivity between nodes depends on their position and their position variation over the time. These kinds of networks are easy to be deployed and are self-configuring. A sensor node is a transmitter, a receiver, and it offers services of routing between nodes without direct vision, as well as recording data from other sensors.

Since the 1950s, location systems have been incorporated into our lives. In the 1980s the Global Position System (GPS) [2] began as a location method for outdoor environments. This system is based on a triangulation system of variables, where we know the position of a device thanks to the existing satellite network. Now, this system remains the most used because of its good performance and the low price of the required devices. At the end of the year 2000 other location systems appeared, based on cellular networks [3]. These systems are designed for emergency situations. In this case, the base stations are used as a reference point and the location is made through the distance and the angle of the signal. These systems are developed to work in outdoor environments and the devices must also have several communication skills.

Previous technologies are not adequate for indoor environments. This is mainly due to the signal characteristics. Other wireless technologies, such as IEEE 802.11a/b/g [4], Radio Frequency IDentification (RFID) [5], Ultra Wide Band wideband (UWB) [6], Bluetooth [7] or Zigbee [8], must be used for indoor locations. The main problem in these environments is the multipath effect and signal variability. The location in these types of environments can be centralized [4] or distributed [9]. The centralized location uses reference devices. These devices tend to have a greater capacity and they are located in a fixed position, for example a base station (BS) or an Access Point (AP). In contrast, in the distributed location there are no reference devices. The devices interact with their neighbors in order to know their position.

A study about the oscillation of the received signal is shown in reference [10]. The RSS variation is introduced in our approach in order to obtain smaller degree errors in the location service. The three main issues that make variations in the RSS are the following ones:

- Temporal variations: when the receiver remains in a fixed position, the signal level measured varies as time goes by.
- Small-Scale variations: the signal level changes when the device is moving over small distances (less than the wavelength). In IEEE 802.11 b/g technologies the wavelength is 12.5 cm.
- Large-Scale variations: the signal level varies with the distance due to the attenuation that the radio frequency (RF) signal suffers with the distance.

Besides these typical variations of the RF signal together with the receiver mobility, we have also considered the temperature and humidity variations, the effect of opening and closing doors, the changes in the localization of the furniture, and the presence and movement of human beings, which are all characteristics of indoor environments. These variations have already been analyzed in [11].

All these systems can be used in different applications. Localization in sensor networks has attracted a large research effort in the last decade. Some WSNs location systems application areas are:

- **Emergencies:** When we want to locate an individual in the case of an emergency (injury or criminal attacks) or in a life-threatening situation. Both can be located using the positioning capability of the mobile device.
- **Information:** It can be used in public places such as swimming pools, museums, conferences, etc. in order to provide information service about this place to the user depending on his position.
- **Navigation:** When the user needs to meet the situation of addresses, positions, directions in an indoor places such as big supermarkets, commercial centers, etc.
- **Discovery:** When it is necessary to find or locate things or persons in indoor places. It is very useful to locate people with Alzheimer or to locate disabled people with very little motion.
- **Security:** It can be used to avoid theft, to move unwanted items, etc. Wireless sensors would be in specific places, when the sensors transfer a position threshold they send an alarm.
- **Tracking:** When it is required to track a device or person inside a building.

There are two main methods to estimate the position in indoor environments. On the one hand, there are the so-called deductive methods. These take into account the physical properties of signal propagation. They require a propagation model, topological information about the environment, and the exact position of the base stations. On the other hand, there are the so-called inductive methods. These require a previous training phase, where the system learns the signal strength in each location. The main shortcoming of this approach is that the training phase can be very expensive. The complex indoor environment makes the propagation model task very hard. It is difficult to improve deductive methods when there are many walls and obstacles because deductive methods work estimating the position mathematically with the real measures taken directly from environment in the training phase [12].

In this work we present a hybrid location system using a new stochastic approach which is based on a combination of deductive and inductive methods. This system has been developed for wireless sensor networks using the IEEE 802.11b/g standard in order to use a deployed wireless access network that is also used for internet access and data transfer. On the other hand, the aforementioned technology allows us to cover a hard indoor environment without many base stations. The goal of this work is to reduce the training phase without losing precision.

The remainder of this paper is organized as follows. Section 2 presents the best known related work on location methods in sensor networks. Our hybrid location system is described in Section 3. Section 4 shows the efficiency of our system. It shows real measurements and compares our proposal with other systems. A comparison between our proposal and other hybrid systems proposed in the literature is shown in Section 5. Section 6 concludes the paper and discloses our planned future work.

2. Related Works

The main location systems in WSNs are based either on the GPS [2], on localization algorithms based on different measurement techniques [13] or systems based on known sensor positions [14]. The system proposed in reference [14] uses the physical layer and the Medium Access Control (MAC) to transmit location information between pairs of sensors using IEEE 802.15.4.

The well-known analytical techniques used in localization algorithms are the angle of arrival (AOA), the time of arrival (TOA) and the time difference of arrival (TDOA) [15-17], relative distance [18] and the RSS [13]. This section discusses some works related with deductive and inductive location systems and other hybrid location systems.

2.1. Deductive Methods

As shown in reference [4], the effectiveness of the measurement techniques based on location algorithms in indoor environments is limited by multiple reflections. That paper describes the RADAR system which is based on RF. The system uses the received signal strength information for a trilateration system and signal propagation models to locate the device. While the empirical model has higher precision, the signal propagation method is easier to use.

In [14], measurement based statistical models of TOA, AOA and RSS are presented and used to generate localization performance boundaries. Such boundaries are useful as design tools to help choose among measurement methods, select neighborhood size, set minimum reference node densities, and compare localization algorithms.

R. Schmidt proposed, in reference [15], a system that processes the signals received from the emitter on an array of sensors. The system considers sensors with arbitrary locations and arbitrary directional characteristics (gain/phase/polarization) in a noise interference environment of arbitrary covariance matrix. The author proposes the multiple signal classification (MUSIC) algorithm to determine the parameters of multiple wave fronts arriving and then the method calculates the position.

A paper where the authors explain another trilateration based location system is [19]. Niculescu and Nath present an indoor positioning architecture that does not require a signal strength map, simply requiring the placement of special VOR (VHF Omnidirectional Ranging) Base Stations (VORBA). The accuracy of the positions obtained by their system, of 2.1 meter median error, is comparable to the original RADAR. Its basic idea is to find the strongest maximums in the signal strength, and use them as the most likely directions in which the device can be placed.

Reference [20] describes two approaches developed by the same authors of this paper, where wireless sensors could find their position using WLAN technology. The scenario is an indoor environment that contains walls, several sources of interferences, multipath effects, humidity and temperature variations, etc. Both approaches are based on the RSS. The first approach uses a training session and the position is calculated based on a heuristic model using the data obtained from the training session. The second approach uses a triangulation model with some fixed access points and a signal propagation model based on wall losses to calculate the position.

The techniques based on RSS are easier to implement. This is due to the fact that standard wireless devices possess features for measuring this value. As indicated in [21], we must take into account that the location error in triangulation systems is very low when the number of base stations is higher than seven and the triangulation analysis is in three dimensions. As we have said previously, when the number of base stations is high, the location system provides more precision, but it makes the sensor estimate more distances (for every base station), so the sensor could need more processing time, diminishing the overall system performance [20].

2.2. Inductive Methods

Inductive methods use location techniques based on RSS profiles. This technique consists of building a map according to the signal strength behavior with respect to the coverage area. A sensor location can be estimated with the gathered information from several base stations. These access points or base stations work with the signal strength vector. The vector is obtained from the RSS model and probabilistic techniques or various methods based on neighbors. With this information the system can estimate the possible area where the sensor could be located.

This method uses two parameters: a) the likelihood that an object is in this area and b) the precision of the signal strength. This second parameter depends on, for example, the size and the type of the location area. With these types of systems, the final user does not require any additional hardware for the localization process. These algorithms give very low localization errors using the IEEE 802.11 technology [21]. According to reference [21], we can consider three different techniques used in algorithms based on patterns or areas: a) single point adaptation, b) likelihood based on areas and c) Bayesian networks.

There are statistical models based on the signal strength, where the distance between different sensors is obtained by the calculation of a Cramér-Rao bound (CRB) on the location estimation precision possible for a given set of measurements (see reference [14] for more details). This is a useful tool to help system designers and researchers select measurement technologies and evaluate localization algorithms.

Reference [22] shows a localization system based on the RSS, which considerably reduces the number of broadcasting stations. This system is called Location Estimation Assisted by Stationary Emitters (LEASE) for indoor RF wireless networks and uses a few stationary emitters and sniffers in a novel way to solve the location estimation problem. The estimation engine uses non-parametric modeling techniques that automatically capture the anisotropy of the RSS encountered in indoor environments.

Siddiqi *et al.* present in [23] another work where we can see the use of Bayesian networks. In this paper, the authors use a robot to retain samples. These samples are used to know the location by means of a probability density function. Each time the robot moves or senses the signal strength of an AP, a Bayes filter is used to recursively update the belief function (Monte Carlo localization (MCL) algorithm). Their results show that accurate localization (~ 2 m) is achieved in most test cases and the average localization error decreases with time.

Another important inductive location system is LANDMARC (see reference [5]). It is a localization prototype that uses RFID technology to locate objects inside the buildings. Although RFID is not designed for indoor location, the authors demonstrate that active RFID is a viable and cost-effective candidate for indoor location. References [4,13,20,21] can be consulted to examine some inductive methods in depth.

2.3. Hybrid Methods

There are other papers where the authors propose hybrid systems. In [24], the authors propose a radiolocation scheme based on the AOA and the TOA in multipath environments with a single base

station. This scheme is used in macrocellular networks (such as the Code Division Multiple Access cellular network) and Global System for Mobile communications (GSM) networks. In [25], the authors combine RSS measurements and the TDOA measurements. This model is sturdy to variations of measurement noise and quantization. The error is lower than ones based on individual measures.

There are few works related with the hybrid location techniques in sensors networks. In reference [26] Cramer-Rao Bound (CRB) is used for location estimation using of two different hybrid schemes: TOA/RSS and TDOA/RSS. These techniques provide improved location accuracy with respect to TOA and TDOA schemes for networks with devices having communications ranges of 30 meters or less.

In [27] Sahinoglu and Catovic developed a hybrid location estimation scheme for heterogeneous WSNs with unsynchronized short range simple relays and mobile sensor nodes, and synchronized stations. In this work, the authors use RSS measurements as well as TOA and TDOA measurements. These measurements are used to filter out the clock offset that appears due to the lack of synchronization. They quantify the estimation accuracy of the scheme by deriving the Cramer-Rao bound (CRB), and discuss the performance trade-off between the number of synchronized and non-synchronized devices involved. This work takes into account the heterogeneity of sensor networks, in terms of communication range, time synchronization and routing capabilities of network devices.

If we analyze the hybrid location systems in WLAN, regardless of whether they are sensor networks or any network, there are several proposed systems. In [28] the authors propose a hybrid location system based on three stages. Firstly, it establishes a database that contains a distance-signal strength map. Next, the system uses the database to obtain the distances between mobile terminal and base stations. Thirdly, this proposal applies trilateration to calculate the mobile terminal position. In this case, we see that the hybrid modeling has better accuracy than propagation modeling.

Finally, another hybrid location method is proposed in [29]. This hybrid method has two stages. In the first stage, it uses the fingerprinting method with a fast training phase to obtain an estimate of the mobile user (MU) position. In the second stage, trilateration is used to compute the MU location more accurately. The result shows their proposed method is better than the simple trilateration method based on general propagation mode, but worse than the fingerprinting method with a medium training phase.

The Euclidean models are optimum when there are multiple access points. Although some works show that the statistical properties of the RSS signal is stationary under certain circumstances, the distribution of the RSS is not usually Gaussian, it is often left-skewed and the standard deviation varies according to the signal level. Signals from multiple APs are mostly independent and the interference from other APs using the same frequency does not have a significant impact on the RSS pattern. Consequently, the coverage areas can be grouped together as a group of clusters. More than one cluster may represent one location because of the multimodal distribution of the RSS. In such a case, using a simple Euclidean distance to determine the location may easily classify some patterns into a wrong location. Our proposal combines the advantages of the deductive and inductive methods in order to provide more accurate measurements in hard environments (few base stations and/or few trained points).

3. Hybrid Stochastic Approach to Location Estimation

In this section we explain the mathematical assumptions used in our proposal. We analyse the inductive and deductive methods from a statistical point of view. In this way, we can describe our hybrid model. Table 1 shows the variables used in the analysis.

Table 1. Variables.

Parameter	Description	Parameter	Description
l	found location	l^j	location of base-station j
o	current observation	d_0	reference distance for signal strength measured
b	number of base station	o_0^j	mean signal strength measured to d_0 distance from base-station j
o^j	signal strength from base-station j of o ; $j \in \{1, \dots, b\}$	d^j	Euclidian distance between l and l^j
T	set of training data	n	attenuation variation index
t	number of training samples	L_w^j	attenuation caused by the obstacles from base-station j
T_i	training sample i ; $i \in \{1, \dots, t\}$; $T_i = (l_i, o_i)$	w	number of wall crossed
l_i	location of training sample i	L_0	wall average attenuation
o_i	observation of training sample i	X_σ	zero-mean normal distributed random variable with standard deviation σ .
o_i^j	signal strength of training sample i from base-station j	d_i^j	distance from l_i to l^j
B^j	base-station j ; $j \in \{1, \dots, b\}$; $B^j = (l^j, o_0^j)$	L_w^j	wall attenuation obtained from l_i to l^j

3.1. Stochastic Approach for Location Estimation

The location estimation problem can be statistically stated as follows. For simplicity, the true distribution $\Pr(X = x)$ and $\Pr(X = x \mid Y = y)$ are denoted as $\Pr(x)$ and $\Pr(y)$. The model parameters are denoted by $p(\cdot)$.

Let b be the number of base-stations. We denote o as an observation. The observation variable is a b -dimensional vector; one for each signal strength from each base-station. We denote as o^j the signal strength from base-station j for $j \in \{1, \dots, b\}$. We have a location l associated to each observation. In this work we use bi-dimensional locations for simplicity, but it can be used in three dimensions easily.

The methodology used is based on the definition of a function $\Pr(l|o)$ that returns the probability of the location l , given the observation o . This nomenclature has been used in other proposals [22, 28, 29]. Once this function is estimated, the problem can be formulated to find the location l that maximizes the probability $\Pr(l|o)$ for a given observation o . Using Bayes' theorem, we can write:

$$\Pr(l | o) = \frac{\Pr(l)\Pr(o | l)}{\Pr(o)} \quad (1)$$

The denominator in equation (1) does not depend on the location variable l . And, therefore, the location estimation problem can be presented as:

$$l' = \arg \max_l \Pr(l)\Pr(o | l) \quad (2)$$

where $\Pr(l)$ is the *prior probability* of the location l , knowing the observation. This probability can be used to incorporate information such a more training locations [22] or tracking [29] to our statistical model. The tracking information will not be taken into account in this work, so, for the prior probability, we use the uniform distribution.

In equation (1), $\Pr(o|l)$ is the so called *likelihood function*. It estimates the probability of one observation given a location. In the literature, we find two main approaches to estimate this function: inductive approach and deductive approach. The next two subsections explain these methods analytically in order to propose our approach.

3.2. Inductive Approach for Location Estimation

On the one hand, the *inductive approach* estimates the *likelihood function* measuring directly the signal strength in each place. That is, several measurements are taken for each training place; then, the function $p(o|l)$ is estimated. The main drawback of this approach is the time consuming nature of the training phase. We denote as T the set of training data, formed by t observations with their respective locations. Each training data, T_i , it is represented as (l_i, o_i) , where i can be from 1 to t . Several alternatives has been proposed in the literature to estimate $p(o|l)$ from T : the histogram method [30, 32], the Bayesian method [33] or the kernel method [30, 34]. Another drawback is that this model only returns one of the locations from the training set. In order to solve this problem, several proposals can be found in reference [35].

3.3. Deductive Approach for Location Estimation

On the other hand, the *deductive approach* estimates the *likelihood function* by using empirical formulas about the signal propagation in an indoor environment. In this approach, we need to know the location of each base-station, the described map of the environment (walls, obstacles, etc.) and a propagation model. Several propagation models can be seen in references [35, 36]

If we assume that each observation o^j , from the vector o , is mutually independent, we can write:

$$\Pr(o | l) = \prod_{j=1}^b \Pr(o^j | l, B^j) \quad (3)$$

where b is the number of base-stations; B^j is the base-station j ; and, o^j is the observation signal strength from base-station j . In our study the base-station B^j is characterized by two variables (l^j, o_0^j) . The variable l^j denotes the location of base-station j , and o_0^j denotes the mean signal strength measured to d_0 distance from base-station j .

In this work, we assume that $(o^j|l, B^j)$ follows a Gaussian distribution with standard derivation σ . The following empirical propagation model, which supposes that signal strength is measured in dB, is used [36, 37]:

$$p(o^j | l, B^j) \approx N(o_0^j + 10n \log(\frac{d^j}{d_0}) + L_w^j, \sigma^2) \quad (4)$$

where, d^j is the Euclidian distance between the observation location (l) and the base-station j (l^j), n is the attenuation variation index (n value depends on the specific propagation environment) and L_w^j is the attenuation caused by the obstacles. In our study the value of L_w^j depends on the number of walls that the line of sight crosses from the base-station j to the location. We will use $L_w^j = wL_0$. Where w is the number of wall crossed and L_0 is the wall average attenuation. Finally, let $N(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2 .

3.4. Stochastic Hybrid Approach for Location Estimation

In the inductive approach we assume that the signal distribution for each training sample location is known in advance. Taking a sample for all possible locations is not a realistic assumption. However, for a given location we can have several training samples near to our location. In the hybrid approach we are interested in combining the information of both previous approaches to improve the system. That is, we know the signal distribution for several training samples near our location and we know how the signal is attenuated from the location of these samples to our actual location. Without loss of generality, we can write:

$$\Pr(o | l) = \sum_{i=1}^t \Pr(T_i | l) \Pr(o | l, T_i) \quad (5)$$

We assume that $\Pr(T_i|l)$ is uniformly distributed. Then, we are only interested in defining the second term. Using the same assumption that in equation (3), we can write:

$$\Pr(o | l, T_i) = \prod_{j=1}^b \Pr(o^j | l, B^j, T_i) \quad (6)$$

Now, we define the random variable $(o^j|l, B^j, T_i)$ in the same manner as $(o^j|l, B^j)$ has been defined in equation (4). But, instead of o_0^j (the signal strength measured in the reference distance d_0), we use o_i^j (the signal strength measured in the training sample location i).

$$(o^j | l, B^j, T_i^j) = o_i^j + 10n \log(\frac{d^j}{d_i^j}) + \frac{L_w^j}{L_{wi}^j} \quad (7)$$

where o_i^j is a random variable that represent the signal strength of training sample i from base-station j . d_i^j is the distance from training sample i to the base station j . And, L_{wi}^j is the wall attenuation from training sample i to the base station j . Note that in this equation, X_σ has been eliminated because the variability is included in the random variable o_i^j .

3.5. Implementation Details

From equation (7) the random variable o_i^j can be expressed as follows:

$$o_i^j = o^j - 10n \log\left(\frac{d_i^j}{d_w^j}\right) + \frac{L_w^j}{L_{wi}^j} \quad (8)$$

In the training phase, we have estimated $p(o_i^j | l_i, B^j)$. In this stage several methods such as histogram or kernel can be used. Then, using equation (8), we can write:

$$p(o^j | l, B^j, T_i^j) = p(o^j - 10n \log\left(\frac{d_i^j}{d_w^j}\right) + \frac{L_w^j}{L_{wi}^j} | l, B^j) \quad (9)$$

where $n = 2$ in free space. In order to obtain the optimal location for equation (2), the proposed algorithm is written in pseudocode in Figure 1. Its explanation is as follows: given the input signal strength the location probability for each point is evaluate using 0.5 meter greed. For each point the k nearest samples are taken. The probability of this location is calculated using equation (5), but, using only these k samples instead the all training set. Farther samples will distort the results. For each sample (of k nearest samples), first we use equation (9) to combine the deductive approach, to take into account the shift from the actual location to the sample location, and, second the inductive approach, to obtain the signal probability in a well known place.

Figure 1. Proposed algorithm.

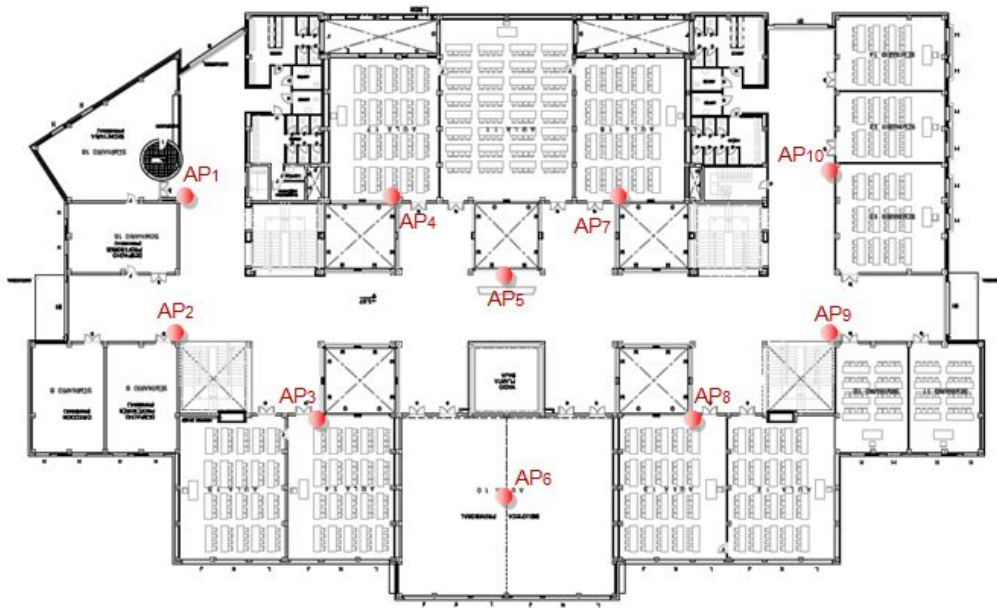
1. **get** the vector of signal strength, o
2. **for** each possible location l using 0.5 meter greed
3. **search** the k nearest neighbor samples from l using Euclidian distance
4. **estimate** $\Pr(o|l)$ using equation (5), (6) and (9)
5. **output** the most probable location

4. Experimental Results

This section shows the results obtained from a real environment to test our proposal. First, we will test the errors based on the number of samples and based on the number of base stations. Then, we compare it with other commercial and implemented location systems.

4.1. Test Bench

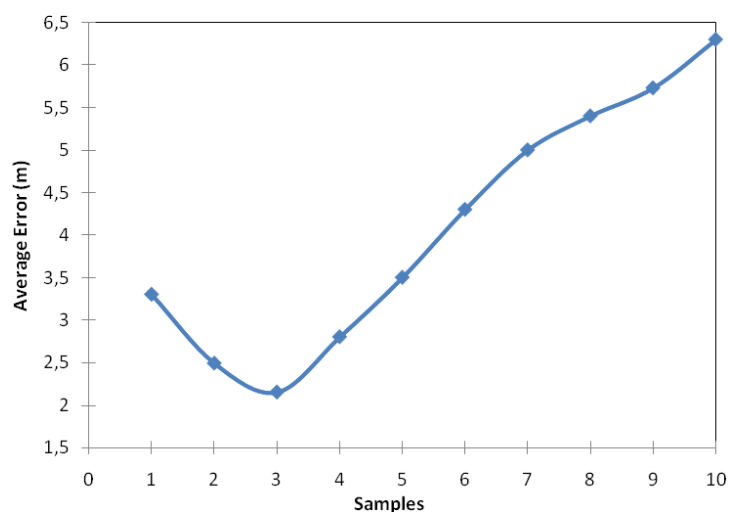
To assess our proposal, we have deployed the approach in an indoor wireless environment. This place is located on the first floor of the “A building” in the “Campus Gandia” of the Polytechnic University of Valencia. The distribution of this floor is shown in Figure 2. There are 10 access points acting as base stations. These base stations have a fixed position and their transmission power is known in advance.

Figure 2. Test bench place used in the experiments.

4.2. Error Measurement Based on the Number of Samples

Our proposal takes into account k nearest neighbour samples from a position using Euclidian distance (see Figure 1). This experiment gives us the optimum number of samples in order to obtain the lowest error. In order to test our approach we took 56 samples spread equitably throughout the floor of the building. The floor was split in a grid where sampling points are placed every 2 meters.

Figure 3 shows that the error is not reduced linearly as regards the number of samples. It has an inflection point in which the error changes its trend. In other words, the error decreases until it has the three closest samples, then, the error begins to increase.

Figure 3. Average location estimation error as a function of the number of samples.

This happens because the method obtains relative distances from the samples to the sensor and when the method begins to use measures that are not close to the sensor the error increases. Obviously,

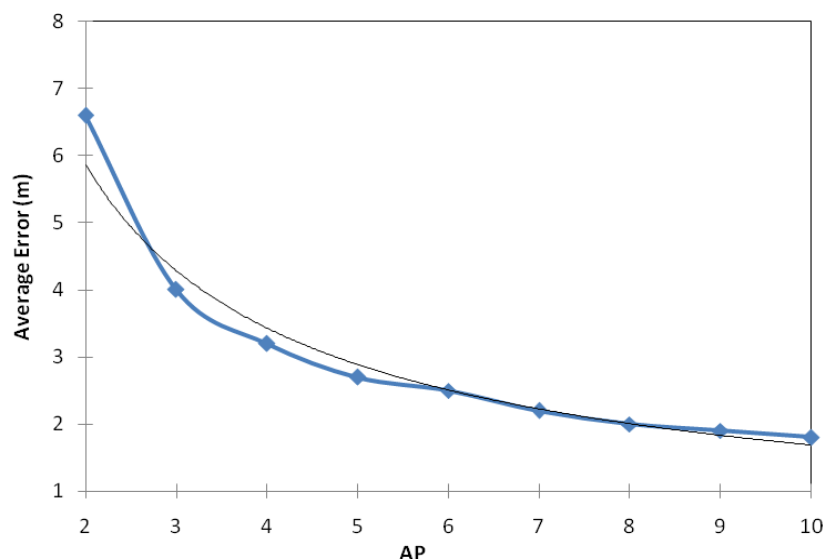
the smaller the area is, where the sensor can be found, the lower the error in its location will be. More samples will give higher relative distances and therefore the error of location will be greater.

Our first conclusion, based on the previous graph, is that given a fixed number of samples, there will be a value of number of samples where the location error will be the optimal. Then, if the number of samples used to train the system is greater, the estimated position will be more accurate because there will be closer samples.

4.3. Error Measurement Based on the number of APs (Base Stations)

In order to test the influence of the number of APs in our proposal, we measured the error of the approach adding access point one by one in each location (in the same place of the 56 samples previously taken). In Figure 4 we can observe that the localization error tends to decrease exponentially (blue line with squares). Therefore, with higher number of APs we obtain lower error values in the sensor location estimation. This tendency is given because one of the methods used in our hybrid system is based on the triangulation method. This method uses the distance from the sensor to various access points based on RSS. Once the sensor obtains the value of at least three distances, between the sensor and the APs, the sensor estimates its position. Therefore, the more distances the sensor to different APs has, the higher the accuracy of the localization sensor will be, in other words, the error of location will be lower.

Figure 4. Average location estimation error as a function of the number of APs.



Bearing in mind this tendency, we have estimated which function follows the error when the number of APs of the indoor environment varies. Equation (10) gives our estimation:

$$y = 2,8383e^{-0,255x} \quad (10)$$

where x is the number of APs in the indoor environment. Equation (10) is shown in figure 4 with the black-thin line. We can see that it fits the error tend quite well.

It should be noted, that when there are more than five APs the improvement appreciation in terms localization error is minimal when a new AP is added to the indoor environment.

4.4. Comparative Measurement with others Existing Location Systems

In order to compare our proposal with others, we have evaluated five wireless sensor location systems:

- a) Inductive 1. This is an inductive location system which has enough a number of samples for an adequate training.
- b) Inductive 2. It is also an inductive location system, but in this case, the number of samples is very low.
- c) Deductive. This system uses the method based on the equation of spread that we have seen in subsection 4.3.
- d) The Hybrid method. This is our proposed method.
- e) Ekahau, which is the basis of many currently used location systems [38].

For the inductive methods we used a system described in our previous work [11]. As has been previously mentioned, inductive methods need a training phase. For the Inductive 1 and Ekahau methods, we collected 312 samples spread equitably throughout the floor of the building. The floor was split in a grid where sampling points are placed every 2 meters. Thirty observations were taken from each training point; 15 of them were taken one day and the other 15 were taken one week later. In contrast, for the Inductive 2 and hybrid methods we used a subset of 56 samples. For the hybrid method we estimated $p(o_i|l)$ using the histogram method shown in references [30,32].

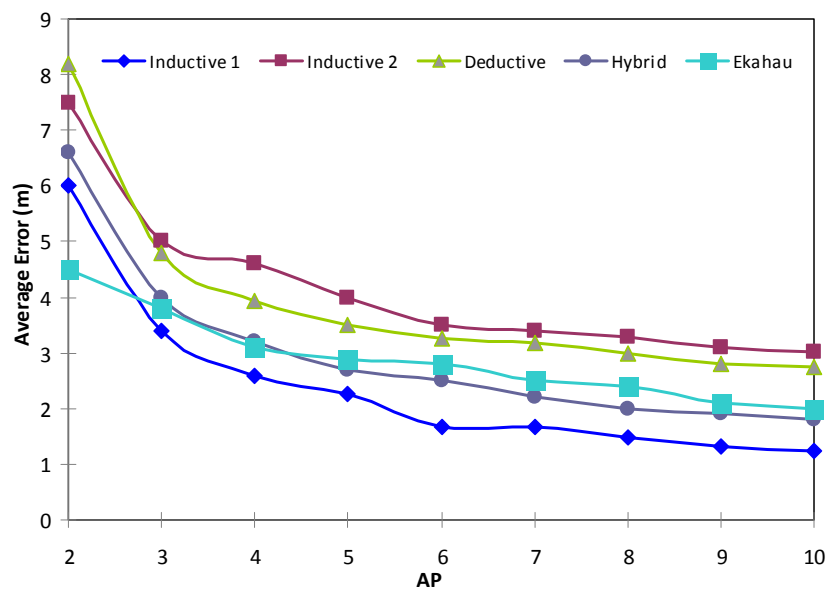
In the test phase, all these systems were tested for 40 locations (all these locations were different that the training ones, they were randomly placed and they were not inside the training grid). For each location we gathered a mean of 15 RSS consecutive values. This let us take into account the signal variability in the measurements. Each one of the test samples has been applied to the different location methods. Then, we estimate the error measuring the Euclidean distance between the output of the method and the real location of the sample. Figure 5 shows the results obtained for all the location systems as a function of the number of APs. Their graph follows an exponential tends approximately.

We note that the Inductive 2 method has a higher localization error than the others. This method had 56 training samples. They were few compared with the Inductive 1 method (312 samples). This difference gives considerably more accuracy in the Inductive 1 than the Inductive 2 method.

With regards to the deductive model, we note that it did not give good results because the floor where the measurements were taken had many walls, so there was very little accuracy when we estimated their loss.

The hybrid model proposed in this paper has a stable and optimal graph compared to the rest of systems (with few training measures low errors were obtained). As noted in Figure 5, for a certain number of AP (five APs) its average error remains the second best.

Finally, the Ekahau system together with the Inductive 2 system are the methods with the worst results.

Figure 5. Comparative of average location error.

In Table 2 we can see the average error and the standard deviation of the approached compared in our experiment. We see that the method with less error is the Inductive 1 (1.23 m), this is because the number of samples is adequate. The model with the worst behaviour is the Inductive 2 (3.02 m). The proposed hybrid method has an average error of 1.80 m, with the advantage that training is minimal. A statistical significance test has been calculated using a paired t-test (the hybrid approach is used as reference). A result labelled with a “▲” means statistical confidence of 99%. “△” means statistical confidence of 90%.

Table 2. Average errors and standard deviation to the surveyed approaches.

	Inductive 1	Inductive 2	Deductive	Hybrid	Ekahau
Error (m)	1.23▲	3.02▲	2.75▲	1.80	2.04△
Standard deviation	0.62	2.12	1.73	0.74	1.61

5. Hybrids Methods Comparison

This section compares our proposal with the hybrid methods found in the literature. In Table 3 we can see the performance analysis. First, we analyzed the analytical techniques used. All cases use multiple parameters, except in [28,29] and our proposal. The next feature compared was their working environment (indoor or outdoor environments). Our proposal and the one in reference [28] support both environments. The systems used in WSNs are [26,27] and our application; the others are used for other purposes. Our location system has a better accuracy (1.8 m) than other works, although the systems in [27, 28 and 29] have good features too.

The next analyzed feature is the number of stages to ascertain the final position. In this case, our system has two stages. The best solution is one stage because of simplicity. As we can see in Table 3, only the systems in [26] and [27] have one stage, but these systems need extra messages to estimate the location. But, on the other hand there are several works that demonstrate that sending messages

wastes more energy than computing [39]. Finally we have analyzed the centralization and decentralization of the location systems and if they are recursively updated. Only our proposal and the one in reference [28] are recursively updated. In conclusion, we can see that the proposed location system has more positive features and it improves all analyzed hybrid location systems.

Table 3. Hybrid location systems comparison.

	Gu <i>et al.</i> [24]	McGuire <i>et al.</i> [25]	Catovic <i>et al.</i> [26]	Sahinoglu <i>et al.</i> [27]	Wang <i>et al.</i> [28]	Li <i>et al.</i> [29]	Our system
<i>Analytical techniques</i>	AOA/TOA	RSS/TDOA	TOA/RSS and TDOA/RSS	TDOA/RSS	RSS	RSS	RSS
<i>Outdoor or indoor</i>	Outdoor	Outdoor	Outdoor	Outdoor	Both	Indoor	Both
<i>Is it used in WSNs?</i>	No	No	Yes	Yes	No	No	Yes
<i>Mean accuracy (m)</i>	122	60	2-3.3	2.1	1.86	1.83	1.8
<i>Stages</i>	1	1	1	1	3	2	2
<i>Extra messages</i>	No	No	Yes	Yes	No	No	No
<i>Centralized or distributed</i>	Centralized	Centralized	Both	Distributed	Centralized	Centralized	Centralized
<i>Recursively update</i>	No	No	No	No	Yes	No	Yes

6. Conclusions

In this paper two different approaches had been discussed. The so-called deductive methods, which require a model of propagation, a topological information of the environment, and the exact position of the radio stations, and the so-called inductive methods, which require a previous training phase where the system learns the signal strength in each location. The main shortcoming of this approach is that in some scenarios, the training phase cannot be done or could be very expensive. On the other hand, the Euclidean models are optimum when there are multiple access points and few walls.

In this work we present a hybrid location system using a new stochastic approach which is based on a combination of deductive and inductive methods. This system has been developed for wireless sensor networks using IEEE 802.11b/g standard in order to use a deployed wireless access network that is also used for internet access and data transfer.

We have tested the errors based on the number of samples and based on the number of base stations. We have estimated an experimental equation based on the graph trends of the errors of the measures obtained. We have compared our proposal with several methods in order to check our approach.

Our system uses a small set of training samples (inductive information). Given the actual signal strength, we use the closest training samples as a starting point. Then, the deductive propagation model is used to obtain the shift from the training samples. A stochastic approach is used whereby the optimal location can be estimated as the point that maximizes the product of probabilities from each of the closest training samples

Our proposal combines the advantages of the deductive and inductive methods in order to provide accurate measurements in hard environments (few base stations and/or few trained points). The goal of this work has been to reduce the training phase without losing precision. Now, we are trying to find the proposed model by adding other methods in order to obtain more accurate results.

The proposed method is useful in cases where a good training phase is not practical (very few samples can be taken in advance), and the precise location of some access points is not known. These environments could be military, such as troop deployments inside buildings or discovery squads for hard environments, environments where the radio coverage is not known in advance (unknown deployments), or even environments where there the APs can be on or off at any time (dynamic environments). We are currently working on enhancing the precision of the proposed model. In future work we will evaluate the performance our proposal in hard environments.

References

1. Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. Wireless sensor networks: a survey, *Computer Networks*. **2002**, *38*, 393-422.
2. Hoffmann-Wellenhof, B.; Lichtenegger, H.; Collins, J. *Global Positioning System: Theory and Practice*, 4th Edition; Springer-Verlag: New York, 1997.
3. Tekinay, S. Wireless Geolocation Systems and Services. *IEEE Commun. Mag.* **1998**, *36*, 28-28.
4. Bahl, P.; Padmanabhan, V.N. Radar: An in-building user location and tracking system. In *Proceedings of the IEEE Infocom*, 2000; vol. 2, pp. 775-784.
5. Ni, L.M.; Liu, Y.; Lau, Y.C.; Patil, A.P. Landmarc: Indoor location sensing using active RFID. *Wireless Networks* **2004**, *10*, 701-710.
6. Shreve, G.; Kell, D. A precision location network using Ultra Wide Band WLAN radios. In *Proc. of Third IEEE Workshop on Wireless LANs*, 2001.
7. Gonzalez-Castaño, F.J.; Garcia-Reinoso, J.J. Bluetooth Location Networks. In *IEEE Globecom Int. Conference*, 2002; Vol.1, pp. 233-237.
8. Park, W.; Yoon, M. The Implementation of Indoor Location System to Control ZigBee Home Network. In *SICEICASE. Int. Joint Conference*, 2006; pp. 2158-2161.
9. Harter, A.; Hopper, A. A distributed location system for the active office. *IEEE Network* **1994**, *8*, 62-70.
10. Youssef, M.; Agrawala, A. Small-Scale Compensation for WLAN Location Determination Systems. In *Proc. of the 2003 ACM Workshop on Wireless Security*, 2003; pp. 11-20.
11. Garcia, M.; Tomas, J.; Boronat, F.; Lloret, J. The Development of Two Systems for Indoor Wireless Sensors Self-location. *Ad Hoc Sens. Wireless Netw.* in press.

12. Li, B.; Salter, J.; Dempster, A.G.; Rizos, C. Indoor Positioning Techniques Based on Wireless LAN. In *First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, paper 113. Sydney, Australia, 13-16, March 2006.
13. Mao, G.Q.; Fidan, B.; Anderson, B.D.O. Wireless sensor network localization techniques. *Comput. Networks* **2007**, *51*, 2529-2553.
14. Patwari, N.; Ash, J.N.; Kyperountas, S.; Hero, A.O.; Moses, R.L.; Correal, N.S. Locating the nodes: cooperative localization in wireless sensor networks. *IEEE Signal Process Mag.* **2005**, *22*, 54-69.
15. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276-280.
16. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984-995.
17. Paulraj, A.; Roy, R.; Kailath, T. A subspace rotation approach to signal parameter estimation. In *Proceedings of the IEEE* **74**, July 1986; pp. 1044-1046.
18. Romer, K. The lighthouse location system for smart dust. In *Proceedings of MobiSys 2003 (ACM/USENIX Conference on Mobile Systems, Applications, and Services)*, 2003; pp. 15-30.
19. Niculescu, D.; Nath, B. VOR base stations for indoor 802.11 positioning. In *Proc. of the 10th Annual international Conference on Mobile Computing and Networking. MobiCom '04. ACM*, New York, NY, 58-69. Philadelphia, PA, USA, September 26-October 01, 2004.
20. Garcia, M.; Martinez, C.; Tomas, J.; Lloret, J. Wireless Sensors Self-Location in an Indoor WLAN Environment. In *International Conference on Sensor Technologies and Applications, SensorComm 2007*, 14-20 Oct. 2007; pp.146-151.
21. Elnahrawy, E.; Li, X.Y.; Martin, R.P. The limits of localization using signal strength: a comparative study. In *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004*. 4-7 Oct. 2004; pp. 406-414.
22. Krishnan, P.; Krishnakumar, A.; Ju, W.-H.; Mallows, C.; Gamt, S. A system for LEASE: location estimation assisted by stationary emitters for indoor RF wireless networks, *IEEE Infocom* **2004**, *2*, 1001-1011.
23. Siddiqi, S.; Sukhatme, G.S.; Howard, A. Experiments in Monte-Carlo Localization using WiFi Signal Strength. In *Proceedings of the International Conference on Advanced Robotics*, Coimbra, Portugal, Jul 2003.
24. Gu, Z.; Gunawan, E. Radiolocation in CDMA cellular system based on joint angle and delay estimation. *Wireless Personal Commun.* **2002**, *23*, 297-309.
25. McGuire, M.; Plataniotis, K.N.; Venetsanopoulos, A.N. Data Fusion of Power and Time Measurements for Mobile Terminal Location. *IEEE Trans. Mob. Comput.* **2005**, *4*, 142-153.
26. Catovic, A.; Sahinoglu, Z.; The Cramer-Rao bounds of hybrid TOA/RSS and TDOA/RSS location estimation schemes. *IEEE Commun. Lett.* **2004**, *8*, 626-628.
27. Sahinoglu, Z.; Catovic, A. A hybrid location estimation scheme (H-LES) for partially synchronized wireless sensor networks. In *IEEE International Conference on Communications*, 20-24 June 2004; vol.7, pp. 3797-3801.

28. Wang, H.; Jia, F. A hybrid modeling for WLAN positioning system. In *International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007*, 2007; pp. 2152-2155.
29. Li, B.; Dempster, A.; Rizos, C.; Barnes, J. Hybrid Method for Localization Using WLAN. In *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis*, 2005.
30. Roos, T.; Myllymäki, P.; Tirri, H.; Misikangas, P.; Sievänen, J. A Probabilistic Approach to WLAN User Location Estimation. *Int. J. Wireless Inf. Networks* **2002**, *9*, 155-164.
31. Xiang, Z.; Song, S.; Chen, J.; Wang, H.; Huang, J.; Gao, X. A Wireless LAN-Based Indoor Positioning Technology. *IBM J. Res. Dev.* Sept.-Nov. **2004**, *48*, 617-626.
32. Xie, H.; Tanin, E.; Kulik, L. Distributed Histograms for Processing Aggregate Data from Moving Objects. In *International Conference on Mobile Data Management*, 1-1 May 2007; pp.152-157.
33. Krishnamachari, B.; Iyengar, S. Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Trans. Comput.* **2004**, *53*, 241-250.
34. Nguyen, X.; Jordan, M.; Sinopoli, B. A kernel-based learning approach to ad hoc sensor network localization. *ACM Trans. Sensor Networks* **2005**, *1*, 134-152.
35. Alk, R.; Tummala, D.; li, X. Indoor Propagation Modeling at 2.4 Ghz for IEEE 802.11 Networks. In *The Sixth Lasted International Multi-Conference on Wireless and Optical Communication, Wireless Networks and Emerging Technologies*, Banff, AB, Canada, July 3-5, 2006.
36. Lloret, J.; Lopez, J.J.; Turro, C.; Flores, S. A fast design model for indoor radio coverage in the 2.4 GHz wireless LAN. In *1st International Symposium on Wireless Communication Systems*, 20-22 Sept. 2004; pp. 408-412,
37. Hii, P.; Zaslavsky, A. Improving location accuracy by combining WLAN positioning and sensor technology. In *Proceedings of the Workshop on Real-World Wireless Sensor Networks*, Uppsala, Sweden, 2005.
38. Ekahau. "Ekahau Positioning Engine 2.0: 802.11-based Wireless LAN positioning system", *An Ekahau Technology Document*, Nov 2002, Available online: <http://www.ekahau.com>.
39. Heinzelman, W.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless sensor networks. In *Proceeding of the Hawaii International Conference on System Sciences*, Hawaii, January 2000.