A New Method for Yielding a Database of Location Fingerprints in WLAN

Binghao Li, Yufei Wang, Hyung Keun Lee, Andrew Dempster, Chris Rizos

Abstract

Location fingerprinting in wireless LAN (WLAN) positioning has received much attention recently. One of the key issues of this technique is generating the database of 'fingerprints'. The conventional method does not utilise the spatial correlation of measurements sampled at adjacent reference points (RPs), and the 'training' process is not an easy task. A new method based on kriging is presented in this paper. An experiment shows that the new method can not only achieve more accurate estimation, but can also greatly reduce the workload and save training time. This can make the fingerprinting technique more flexible and easier to implement.

Binghao Li, Yufei Wang, Andrew Dempster, Chris Rizos are with the School of Surveying and Spatial Information System, University of New South Wales, Sydney, Australia

Hyung Keun Lee is with the School of Avionics and Telecommunication, Hankuk Aviation University, Kyunggi-do, Korea

1.Introduction

Knowledge of users' positions is more and more important in today's society. The possibilities and opportunities for Location-Based Services (LBS) has attracted a lot attention from companies and researchers alike. However, one of the key issues to be addressed is the positioning technology itself. Many systems have been developed to determine a user's location under certain scenarios. The most popular, and globally applicable, is of course the NAVSTAR Global Positioning System (GPS). In outdoor areas, a single receiver can easily provide location information with an error typically less than 10m (2D). However, GPS has its shortcomings; for example, it isn't suitable for indoor environments. Other specially developed systems, such as active badge, active bat, cricket, etc., have found some application [1]. However, since they are very expensive in terms of labor, frequency spectrum and capital costs to establish an *infrastructure* just for positioning, many commentators suggest that it is preferable to use existing wireless signals to determine location. Such systems include use of the cellular phone system [2, 3], television [4], and Wireless LAN (WLAN).

WLAN is becoming increasingly popular today, particularly that based on the IEEE 802.11b standard (also known as "Wi-Fi"). Wi-Fi uses radio frequencies in the 2.4GHz band [5]. Many signal strength (SS) based techniques have been proposed for location estimation in environments in which WLAN is deployed. There are essentially two categories of such techniques. One uses a signal propagation model and the information about the geometry of the building to convert SS to a distance measurement. 'Triangulation' (more correctly *Trilateration*) then can compute the location of the mobile user (MU) [6, 7]. This approach is simple to implement; however it does have difficulties. Indoor radio signal propagation is very complicated, because of signal attenuation due to distance, penetration losses through walls and floors, and the effect of multipath propagation [8, 9]. Interference from other signals is also a serious problem. In the 2.4 GHz frequency band, microwave ovens, Bluetooth devices, etc., can be sources of interference. Furthermore, the orientation of the receiver's antenna, and the location and movement people inside the building, can affect the SS significantly [10]. It is extremely difficult to build a sufficiently good model of signal propagation that is adequate for real world applications.

The other category of WLAN positioning is 'Location Fingerprinting'. This class of technique has received more attention recently, able to address some of the problems related to non-line-of-sight and multipath propagation [11]. The basis of location fingerprinting is first to establish a database that contains the measurements of wireless signals (that is, the SS) at some RPs in the area of WLAN coverage. Then the location of the MU can be identified by comparing its SS measurements with the reference data. The disadvantages of this approach are the database generation and maintenance requirements. For example, when the environment changes significantly (such as after a major building renovation), the database has to be rebuilt.

In general the location fingerprinting technique consists of two phases: 'training' and 'positioning'. During the training phase, a database of location fingerprints is established using measurements of the received SS (RSS) at some known points. During the positioning phase, the measurement of the RSS by the WLAN user at an unknown location is performed. The database is accessed and the MU's 'fingerprint' data is compared to the library of such data from the RPs in order to identify the likeliest MU location. There are in fact two ways to estimate the unknown location. The simplest one is the deterministic approach [6, 12, 13]. The average SS of each WLAN Access Point (AP) measured at each RP is used to create the fingerprint database. Since the variation of the SS measured at each point is large, in order to achieve more accurate results, the probabilistic approach [10, 14] has also been developed. Unfortunately, the distribution of the SS is non-Gaussian. Even worse, it varies at different locations, and at the same location when the orientation of the antenna changes [7, 10]. Hence many measurements are necessary, and this takes more time to generate the RSS distribution at each RP. Furthermore, this increases the database size and the computational burden. However, the establishment of the location fingerprint database is an essential prerequisite. The conventional method of generating the database does not utilise the spatial correlation of measurements sampled at adjacent RPs. To achieve a good estimation of user location, the more RPs, or in other words, the smaller the *granularity*, the better. And since the measured SS is affected by so many factors, the variation of the RSS at each point can be as large as 10dB to 15dB. The more measurements obtained at each point the better. However, more RPs and more measurements mean that the training phase is a significant task in terms of labor and time.

Thus the challenge is to build the location fingerprint database in as efficient a manner as possible. In this paper a new method based on interpolation algorithms is used to generate the database. The result shows kriging can significantly reduce the number of RPs needed, and improve the accuracy of estimation. An experiment to verify the method is described, and the test results are presented.

2. Methodology

2.1 Two phases of fingerprinting

Location fingerprinting has two phases: ‘training’ and ‘positioning’. The objective of the training phase is to build a fingerprint database.

In order to generate the database, RPs must first be carefully selected. Locating a MU at one RP location, the SSs of all the APs are measured. From such measurements the characteristic feature of that RP is determined, and is then recorded in the database. This process is repeated at another RP, and so forth until all RPs are visited. In the positioning phase, the MU measures the RSS at a place where it requires its position. The measurements are compared with the data in the database using an appropriate search/matching algorithm. The outcome is the likeliest location of the MU. The whole process is illustrated in Figure 1.

There are many algorithms for computing the location of the MU during 'positioning' phase.

The basic one is the 'nearest neighbor' (NN) algorithm [6]. First, the signal distance between the measured SS vector $[s_1\ s_2\ \ldots\ s_n]$ and the SS vector in the database $[S_1\ S_2\ \ldots\ S_n]$ is computed. The generalized distance between two vectors is:

$$L_q = \left(\sum_{i=1}^{n} \left| s_i - S_i \right|^q \right)^{\frac{1}{q}} \tag{1}$$

Manhattan distance and Euclidean distance are L_1 and L_2 respectively [15]. Experiments show increasing q does not necessarily improve the accuracy of location estimation. The nearest neighbor is the point with the shortest signal distance.

If K (K≥2) nearest neighbors (those with the shortest Euclidean distance) (KNN) are chosen, the average of the coordinates of K points can be used as the estimate of the MU location. Intuitively, this method should be better than NN since there is no reason to just pick the nearest one and abandon others nearby.

The third algorithm used in this paper is referred to as KWNN (K weighted nearest neighbor, K≥2). It is similar to KNN, but when the location of MU is computed the weighted average is calculated rather than the average. The inverse of the signal distance defines the weight. It can be expressed as:

$$p = \sum_{i=1}^{K} \frac{1}{L_{qi} + \varepsilon} \cdot p_i \tag{2}$$

where p_i is the position of the K nearest neighbor, _ is a small real constant used to avoid division by zero.

There are also other weighting schemes, such as the standard deviation of the SS samples. However, due to the complexity of the signal propagation environments it is no surprise that this algorithm does not always improve the result.

Other algorithms such as the smallest polygon [16] and neural networks [17] are either too complicated and/or do not necessarily improve the accuracy of the location estimate. In this paper the three simplest algorithms will be used to evaluate the proposed methodology.

2.2 Conventional database generation methodology

The conventional method for building the fingerprint database is comparatively simple. The 'receiver' makes RSS measurements at each RP, and after some processing (generally averaging), the data is logged in the database. In general the more RPs that are chosen in the training phase, the better the accuracy that can be achieved in the positioning phase. The procedure is illustrated in Figure 2.

Many RPs should be selected, and normally these points are gridded. In fact, to give an even sampling of SS, the RPs should be uniformly distributed in the area of interest (coverage of

WLAN signals) [12, 13, 18]. But in reality, the rooms in a building will have different geometry and size. Furthermore, to make the training phase task easier, the RPs have to be selected depending on the structure of the building.

When the RPs are close to a uniform distribution, the spacing between two adjacent RPs, or the granularity, can indicate the quality of position estimation that can be achieved. When the granularity decreases, or in other words the number of RPs increases, it is more likely (though not necessarily) that a better result can be obtained. When the granularity is smaller than a threshold, the SS values at two neighboring points will be almost the same. The variation of the RSS at a point dominates the change of RSS due to the different place and distance from the AP [12]. To get some idea about the level of granularity, assuming the RPs are uniformly distributed, the average granularity can be calculated approximately as:

$$g = \sqrt{S / N} \tag{3}$$

where S is the total area the RPs occupy, and N is the number of RPs.

2.3 Database generation utilising spatial correlation

When measurements at a small number of RPs are made, they not only provide information at these points, but also imply information of the surrounding area. If a more dense database can be generated efficiently by interpolation based on a small number of RPs, labor effort and time can be saved during the training phase. Two methods, weighted distance inverse (WDI) and kriging, are chosen here to generate the database. The methodology is illustrated in Figure 3.

WDI is a simple interpolation method. The estimator can be computed as:

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^{n} \frac{1}{d_i} \times Z(x_i)}{\sum_{i=1}^{n} \frac{1}{d_i}} \qquad (4)$$

where x_0 is the location of the value unknown point, x_i is the location of the point where the value is known; d_i is the distance between x_0 and x_i .

In order to yield a good estimate the algorithm can be enhanced. For example, when d_i is larger than a specific value d_t, the measurement on that reference location can be abandoned. This is reasonable because the correlation between locations a distance larger than d_t apart is very small (effectively zero).

As an estimation procedure, kriging was first used by the mining industry [19]. The basic tool, the variogram, is used to quantify spatial correlations between observations. As there are many advantages of kriging, its application can be found in very different disciplines ranging from the classical fields of mining and geology to soil science, hydrology, meteorology, environmental sciences, agriculture, and so on [20, 21]. Theoretically, kriging can also be used wherever a continuous measure is made on a sample at a particular location in space or time.

A classical assumption is second order stationarity, but in practice a slightly weaker assumption is more widely used, i.e. the intrinsic hypothesis. This consists of two conditions:

$$E[Z(x)] = \mu \qquad (5)$$

where $Z(x)$ is the random function in domain D, and for all $x \in D$

$$\frac{1}{2} Var[Z(x+h) - Z(x)] = \frac{1}{2} E\left[(Z(x+h) - Z(x))^2\right] = \gamma(h) \qquad (6)$$

where $\gamma(h)$, called the variogram, depends only on the vector h and not on the locations x and

$x+h$ [22, 23].

Kriging provides a solution to the problem of estimation based on knowledge of the variogram and the above assumption. Here is the simple case that the mean is constant across the entire area. Unfortunately, in reality it is common that the mean is not constant. For example, the received SS is weaker when the distance between the measured point and the AP is greater. The SS relation to the distance from the AP has a significant 'trend' rather than being a constant. Assuming the mean is a function of the site coordinates:

$$Z(x) = f_0(x)\beta_0 + f_1(x)\beta_1 + \ldots + f_p(x)\beta_p + \delta(x)$$

where $\beta_0,\ldots,\beta_p$ are unknown parameters; $\delta(x)$ is intrinsic and $E[\delta(x)]=0$.

In matrix notation, the above expression can be written as:

$$Z = X\beta + \delta \tag{7}$$

In order to deal with the drift, UK (universal kriging) is proposed. (Ordinary kriging can be treated as a subset of UK when $f_0(x)=1$, $\beta_1= \ldots = \beta_p=0$.) To predict $Z(x_0)$ the UK predictor is a linear combination of values of the sample $Z(x_i)$:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{8}$$

where $\_{i}$ is the weighting factor.

For the purpose of making this predictor unbiased for all possible vectors β, the following conditions need to be satisfied:

$$E\left[\sum_{i=1}^{n} \lambda_i Z(x_i) - Z(x_0)\right] = 0 \tag{9}$$

As the estimation variance is:

$$\sigma_K{}^2(x_0) = Var\left[Z(x_0) - \hat{Z}(x_0)\right] = -\sum_{j=1}^{n}\sum_{i=1}^{n}\lambda_j\lambda_i\gamma\left(x_i - x_j\right) + 2\sum_{i=1}^{n}\lambda_i\gamma\left(x_i - x_0\right) \tag{10}$$

the best unbiased linear estimator is the one which minimizes $\sigma^2_k(x_0)$ under the constraint of the sum of the coefficients in (9). Introducing the Lagrange multipliers leads to a straightforward linear equation:

$$\begin{bmatrix} \Gamma & X \\ X' & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ m \end{bmatrix} = \begin{bmatrix} \gamma \\ f \end{bmatrix} \tag{11}$$

where

$$\Gamma = \begin{bmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) \\ \vdots & \gamma(x_i - x_j) & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) \end{bmatrix} \qquad X = \begin{bmatrix} 1 & f_1(x_1) & \cdots & f_p(x_1) \\ \vdots & \vdots & f_i(x_j) & \vdots \\ 1 & f_1(x_n) & \cdots & f_p(x_n) \end{bmatrix}$$

$$\gamma = \left[\gamma(x_0 - x_1) \quad \cdots \quad \gamma(x_0 - x_n)\right]' \qquad f = \left[1 \quad f_1(x_0) \quad \cdots \quad f_n(x_0)\right]'$$

So, the result is

$$\lambda' = \left[\gamma + X\left(X'\Gamma^{-1}X\right)^{-1}\left(f - X'\Gamma^{-1}\gamma\right)\right]'\Gamma^{-1}$$

$$m' = -\left(f - X'\Gamma^{-1}\gamma\right)'\left(X'\Gamma^{-1}X\right)^{-1}$$

and

$$\sigma_k{}^2(x_0) = \gamma'\Gamma^{-1}\gamma - \left(f - X'\Gamma^{-1}\gamma\right)'\left(X'\Gamma^{-1}\gamma\right)^{-1}\left(f - X'\Gamma^{-1}\gamma\right) \tag{12}$$

Kriging is the best linear unbiased estimation (BLUE) that has the following features: (a) this estimator is a linear function of the data with weights calculated according to the specifications of unbiasedness and minimum variance, and (b) the weights are determined by solving a system of linear equations with coefficients that depend only on the variogram that describes the structure of a family of functions. A major advantage of kriging is that it is more flexible than other interpolation methods. The weights are dependent on how the function varies in space, rather than on the basis of some arbitrary rule that may be applicable in some cases but not in others. Another advantage of kriging is that it provides the means to evaluate the magnitude of the estimation error.

But there is a classical problem of UK: _ is unknown, and in order to estimate it efficiently, knowledge of _ is needed. However, _ is unknown, bringing the question right back to where it started [22]. Assuming the isotropy of the phenomenon, if a direction in which there is no shift can be found, the variogram in this direction can be used for all directions. However, it is not always so. Like the RSS of AP, the shift exists in each direction. One approach to solving this problem was suggested by Neuman & Jacobson [24]. Starting with the standard least squares estimator of β in (7), this approach computes a variogram estimator from the residuals, fits a variogram model, then obtains the general least squares estimator of β based on the fitted model, and so forth. Normally the process will converge after a few iterations. In this paper, this iterative approach is used.

There are several ways to estimate the variogram [22]. The classical formula is:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{x_i - x_j = h} \left(Z(x_i) - Z(x_j)\right)^2 \tag{13}$$

Most of the time the points are irregularly spaced. In order to have more pairs, the summation x_i-x_j=h has to be weakened:

$$\left\| x_i - x_j \right| - h \right| \leq \varepsilon \quad Angle(x_i - x_j, h) \leq \delta$$

3. Experiment

The experimental test bed was located on the 4th floor of Electrical Engineering Building at the University of New South Wales, Sydney, Australia. The layout of the test bed is indicated in Figure 4. The test bed has dimensions of 17.5m by 23m. The crosses are RPs while the squares are the test points. The figure shows the test area is the upper part of the test bed with dimensions 11m by 23m, and consists of 7 rooms and a part of the corridor. There are in total 132 RPs, and 30 test points.

Five WX-1590 SparkLAN 11Mbps WLAN Wireless Multi-Mode APs were installed at the locations (pentagram symbols) indicated in Figure 4. The APs are essentially base stations transmitting signals for positioning. The MU is a Compaq iPAQ 3970 installed with the Pocket PC 2002 operating system. The network card used in this test is the Lucent Technology Wi-Fi Orinoco Wireless Golden Card, which can exchange information with the APs. The SS information (in units of dBm) of the received signal can be extract and logged. Figure 5 displays the AP, wireless card and PDA used for these tests.

RF propagation in indoor environments is very complicated, and in most of the cases it is a non-line-of-sight (NLOS) scenario. For example, in the corridor there is an AP on the left side (see Figure 4). The signal propagation of this AP is a line-of-sight (LOS) case. But, for all the other APs, it is a NLOS scenario. NLOS error is the dominant error for positioning techniques based on time-of-arrival (TOA), time-difference-of-arrival (TDOA), angle-of-arrival (AOA), etc. However, for fingerprinting, NLOS propagation can help to extract the profile of the SS at a specific location.

Before the training phase was carried out a local coordinate frame was established, and the RPs selected carefully based on the structure of the building/rooms. The RPs were close to or aligned with the corners, doors and windows, which could be easily identified on the map and in the real test bed. The map in Figure 4 is accurate enough so that the coordinates of RPs or test points can be determined by the pixel locations on the screen of a PC. Then the pixel values were converted to meters. The RPs were gridded as well as possible (although in UK it is not necessary). At each RP the user faced east first, and recorded the RSS of each AP. Then the orientation was changed to north, west and south consequently, the RSS values were logged. A total of 12 measurements were made at each point.

The average of all the RSS of each AP was calculated, and was inserted into the database as the reference SS at that location. The data were collected at difference time periods, but most of it was logged during the daytime over a few weekends, so that not many people were present. But the environment of the test would be different from that on weekdays. This should lead to a more accurate result than could be achieved during weekdays. However, the results cannot be used to evaluate how good a WLAN-based location system really is in an absolute sense.

After the data collection for the training phase was finished, the parameters for the variogram and WDI model can be obtained. For kriging, the iteration method mentioned in section 2.2 was used to determine the variogram model for UK. The different parameters were chosen visually. In order to get a conservative estimation, the function given by (14) would approximately over bound the different empirical variograms for each AP:

$$\begin{cases} \gamma(0) = 0 \\ \gamma(0 < h < a) = C0 + C\left(\frac{3}{2}\frac{h}{a} + \frac{1}{2}\left(\frac{h}{a}\right)^3\right) \\ \gamma(h \geq a) = C0 + C \end{cases} \tag{14}$$

The parameters retained were: $C0$=9 dBm^2, C=22 dBm^2, a=6m. The variogram reaches it (C0+ C) at the range (a=6). This means that when the distance between two points is larger than 6m, the measurements at these two points are uncorrelated. γ(0)=0 shows a discontinuity at the origin, which is called the 'nugget effect'. This is caused by the unknown microscale variation. Figure 6 is a plot of the experimental residual variograms for each AP in dash-dot, and the chosen spherical model in solid.

For WDI, d_t=6m was chosen.

In WLAN, RSS is reported in units of dBm. It can be converted to P (in units of mW) easily ($P=10^{\,dBm/10}$). In free space, the received signal power is proportional to the inverse of the square of distance from the transmitter. However, in a real application, especially in an indoor environment, the appropriate model is difficult to define. Intuitively, the further the MU is from the AP, the weaker the received signal should be. If only the signal from one AP is taken into account, the difference of RSS between the test point and RP is not necessarily inversely proportional to the square of the distance between them. However, when signals from many APs are considered simultaneously, the sum of the difference of RSS is a monotonically increasing function of the distance between them. In [6, 12, 13, 16], dBm is used to compute the signal distance between a test point and an RP. We compared instead the received power P (in units of mW), $1/P$, $1/P^2$ and $1/P^4$ to calculate the signal distance. The results show that the dBm measure achieves the best estimation, and using $1/P^2$ or $1/P^4$ can ensure reasonable accuracy of positioning. Investigation of this question will be carried out in future research. In this paper, dBm is used as the distance measure.

Different algorithms were applied to compute the locations of the 30 test points based on a different size of the database (varied from 132 RPs to 16 RPs). In KNN, K equals 2, 3, 4, 5 or 6. In KWNN, K equals 2, 3, 4 or 5. Figure 7 depicts the relationship between mean distance error and different signal distance (see (1)). No matter which database is utilized, when q is close to 1, the smallest distance error can be achieved. The average curve shows that q equals 1 (Manhattan distance) is the best choice. But the difference using other signal distance measures (i.e. q with other values) is not significant. The Manhattan distance is used in this paper to evaluate the new method.

Table 1 lists all the mean distance errors computed using the different algorithms for the different cases. In general, the KNN and KWNN can achieve better accuracy than the simple NN algorithm. Nevertheless, when the granularity of the RPs is large, the NN even performs better than some of the more complicated algorithms. When KNN is used, in general K equals 3 or 4 will yield the best result. This indicates that only using the two nearest neighbors is not enough (some of the useful information has been ignored), but too many nearest neighbors could decrease the accuracy of the estimator since some of the nearest neighbors are too far from the estimated points. KWNN slightly improves the accuracy of estimation. But none of these algorithms can always provide the best result. To evaluate the proposed method, five algorithms have been selected: the NN (Manhattan distance), KNN (K=3,4) and KWNN (K=3,4).

Table1. Mean distance error using different algorithm for different cases (Unit: m)

	NN	2NN	3NN	4NN	5NN	6NN	2WNN	3WNN	4WNN	5WNN
Test1 (132 RPs)	1.75	1.47	1.29	1.23	1.38	1.31	1.49	1.29	1.19	1.31

Test2 (99 RPs)	1.63	1.52	1.38	1.31	1.36	1.39	1.53	1.37	1.27	1.30
Test3 (66 RPs)	1.74	1.47	1.51	1.60	1.52	1.60	1.48	1.44	1.49	1.43
Test4 (33 RPs)	1.78	1.93	1.94	1.72	1.99	2.12	1.79	1.79	1.64	1.75
Test5 (16 RPs)	2.55	2.34	2.65	2.98	3.41	3.99	2.11	2.28	2.45	2.69

Three groups of databases were generated. They are the original group (using the conventional method), the WDI group (database based on WDI), and the kriging group (database based on kriging). In each group, there are five different versions of the databases. In the case of the original group, the five different databases contain the average RSS of 132, 99, 66, 33, 16 RPs respectively. In the case of the WRI group and the kriging group, all the databases have the same size (288 RPs), but they are generated from five different original databases. The granularity roughly equals 1m.

Five algorithms are used here to compute the location of the MU. The mean of the five distance errors are computed to compare the different methods. Figure 8 shows the positioning error using the different databases: the original databases, the databases generated by the WDI and the ones generated by kriging.

Basically, when the average granularity reduces (or the number of RPs increases), the accuracy of the MU's location estimate increases. But when the density of the RPs is high, the rate of increase of the accuracy decreases. For example, using the original database, the average distance error reduces from 2.58m to 1.56m, when the number of RPs increases from 16 to 33. The ratio of accuracy increase is 1.46, and the number of RPs doubles. But when the number of RPs changes from 66 to 132, the ratio is 1.16 (small change in accuracy). In the case of the 95 percentile distance error, the situation is similar (when the database has 99 RPs is an exception). Increasing the number of RP does not in itself ensure high accuracy positioning when the granularity is already adequate.

When information on the spatial correlation of the reference measurements is utilized, the accuracy of the estimation can be improved. Using WDI, when the number of RPs is less than a threshold value, the performance of the estimation is better than using the original database. However, based on the WDI database, the performance is not as good as based on original database if the number of RPs is larger than 66. This means that using WDI cannot successfully estimate the RSS from the known information.

An impressive result is obtained using kriging, where better estimation can always be achieved. When the number of RPs is 66, 99 and 132, the mean distance error is 1.21m, 1.22m and 1.22m respectively. All are better than using the original database. Even when only using 16 RPs, the new method can achieve distance error of 1.49m, only 10% larger than the best estimation that can be achieved based on the original database (1.35m). But the number of RPs used is only 12.5% of the original database. Comparing the 95 percentile distance errors also indicates the advantage of kriging.

From Figure 8 two conclusions can be drawn. Firstly, kriging can efficiently estimate the RSS using the information of some of the RPs. This means that kriging can yield a database of location fingerprints with good quality. Secondly, when the density of RPs reaches a particular value (in this case around 66 points, implying an average granularity of around 1.96m), kriging cannot provide better estimation. On the other hand, it is unnecessary to measure so many RPs (66 here) to achieve the best estimation. This implies the training phase can be shortened significantly. In the worst case, when only very few RPs can be measured, kriging can also obtain reasonable location estimates. In this experiment, when only 16 RPs were used, the estimation error is less than using the original database of 66 RPs, and only slightly worse than 99 or 132 RPs. One thing must be emphasised is that 16 RPs means in small room (of around $20m^2$) there is only one RP, and in a large room (of around 50 m^2) only 2 or 3 RPs are required. This makes the training phase very flexible, and if the environment is changing, fast training can be carried out and a new database of location fingerprints can be generated quickly.

4. Concluding Remarks

In this paper a new method based on kriging for obtaining a database of location fingerprints for WLAN-based positioning has been presented. It differs from the conventional method in that it utilizes the spatial correlation of measurements to generate the database during the training phase. An experiment was carried out and the results indicate that the proposed method does work efficiently. Not only is more accurate estimation possible, but in addition the proposed method can greatly reduce the workload and save on training time. Since in real world situations the signal propagation environment may vary significantly, it is very hard to estimate the degree of granularity of RPs required to ensure a specified accuracy. However, on the basis of these results, when the granularity is around 2-3m, the new algorithm can still achieve the best quality results. Only 1/4 even 1/8 of the number of RPs are needed using the proposed method compared to other methods that do not take into account spatial correlation.

5. References

1 Hightower, J., and Borriello, G.: ‘Location systems for ubiquitous computing’, IEEE Computer, August 2001, 34, (8), pp. 57-66

2 Caffery, J.J., and Stuber, G.L.: ‘Overview of radiolocation in CDMA cellular systems’, IEEE Communications Magazine, April 1998, 36, (4), pp. 38-45

3 Drane, C., Macnaughtan, M., and Scott, C.: ‘Positioning GSM telephones’, IEEE Communications, April 1998, 36, (4), pp. 46-59

4 Eggert, R.J., and Raquet, J.F.: ‘Evaluating the navigation potential of the NTSC analog television broadcast signal’, ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, California, September 2004, pp. 2436-2446

5 ‘White paper IEEE 802.11g’ http://itresearch.forbes.com/detail/RES/1051538795_191.html

6 Bahl, P., and Padmanabhan, V.N.: ‘RADAR: An in-building RF-based user location and tracking system’, IEEE Infocom 2000, Tel Aviv, Israel, March 2000, 2, pp. 775-784

7 Wang, Y., Jia, X., Lee, H.K., and Li, G.Y.: ‘An indoor wireless positioning system based on WLAN infrastructure’, 6th Int. Symp. on Satellite Navigation Technology Including Mobile Positioning & Location Services, Melbourne, Australia, 22-25 July 2003, CD-ROM proc., paper 54

8 http://people.deas.harvard.edu/~jones/es151/prop_models/propagation.html

9 Hashemi, H., ‘The indoor radio propagation channel’, Proc. IEEE, July 1993, 81, pp. 943-968

10 Ladd, A.M., Bekris, K.E., Rudys, A., Marceau, G., Kavraki L.E., and Dan, S.: ‘Robotics-based location sensing using wireless ethernet’, Eighth ACM Int. Conf. on Mobile Computing & Networking (MOBICOM), Atlanta, Georgia, September 2002, CD-ROM proc.

11 Haldat, E.: 'Fingerprinting based-technique for positioning', 2002, http://www.telecomlab.oulu.fi/home/Radiotekniikka/ materiaali/fingerprinting-report.pdf

12 Prasithsangaree, P., Krishnamurthy, P., and Chrysanthis, P.K.: 'On indoor position location with wireless LANs', 13th IEEE Int. Symp. on Personal, Indoor, & Mobile Radio Communications, Lisbon, Portugal, 15-18 September 2002, vol.2, pp. 720-724

13 Saha, S., Chaudhuri, K., Sanghi, D., and Bhagwat, P.: 'Location determination of a mobile device using IEEE 802.11b access point signals', IEEE Wireless Communications & Networking Conference (WCNC), New Orleans, Louisiana, 16-20 March 2003, vol.3, pp.1987-1992

14 Youssef, M., Agrawala A., and Shankar, A.U.: 'WLAN location determination via clustering and probability distributions', IEEE Int. Conf. on Pervasive Computing & Communications (PerCom) 2003, Fort Worth, Texas, 23-26 March 2003, pp.143-150

15 Cormen, T.H., Leiserson C.E., and Rivest, R.L.: 'Introduction to Algorithms', MIT Press, 1990

16 Pandya, D., Jain, R., and Lupu, E.: 'Indoor location estimation using multiple wireless technologies', 14th IEEE Int. Symp. on Personal, Indoor, & Mobile Radio Communications (PIMRC), Beijing, China, 7-10 September 2003, vol. 3, pp. 2208-2212

17 Battiti, R., Nhat, T.L., and Villani, A.: 'Location-aware computing: a neural network model for determining location in wireless LANs', Technical Report DIT-02-0083, February 2002. See also http://eprints.biblio.unitn.it/archive/00000233/01/83.pdf

18 Laitinent, H., Lahteennmaki, J., and Nordstrom, T.: 'Database correlation method for GSM location', Proceedings of IEEE 53rd Vehicular Technology Conference, Rhodes, Greece, May 2001, Vol. 4, pp. 2504-2508

19 Krige, D.G.: ‘A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52, 119-139

20 Clark, I.: ‘Practical geostatistics’, Applied Science Publishers, London, 1979

21 Armstrong, M.: ‘Basic linear geostatistics’, Springer, Berlin, 1998

22 Cressie, N.: ‘Statistics for spatial data’, John Wiley & Sons, INC., New York, 1991

23 Kitanidis, P.K.: ‘Introduction to geostatistics: applications to hydrogeology’, Cambridge University Press, Cambridge, 1997

24 Neuman S.P., and Jacobson, E.A.: ‘Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels’, Journal of the International Association for Mathematical Geology, 1984, 16, pp. 499-521

List of Figures

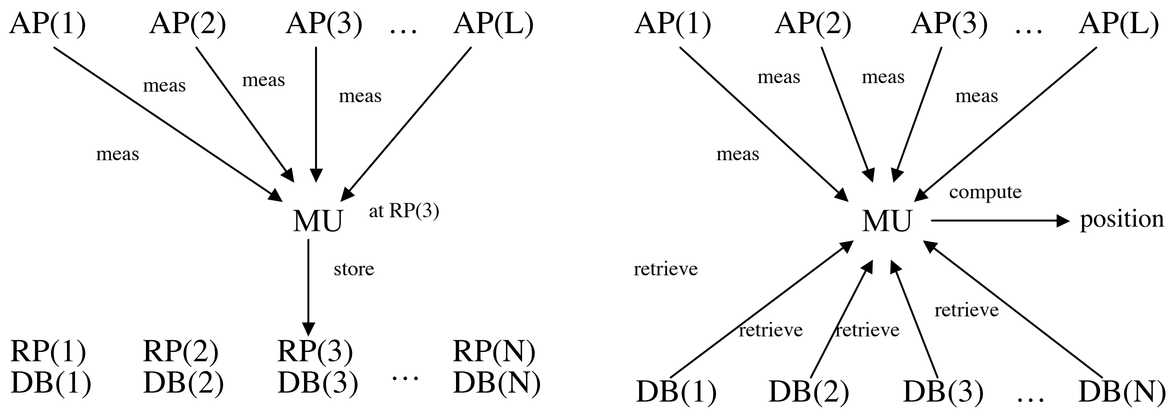

Fig. 1 Training phase (left) and positioning phase (right)

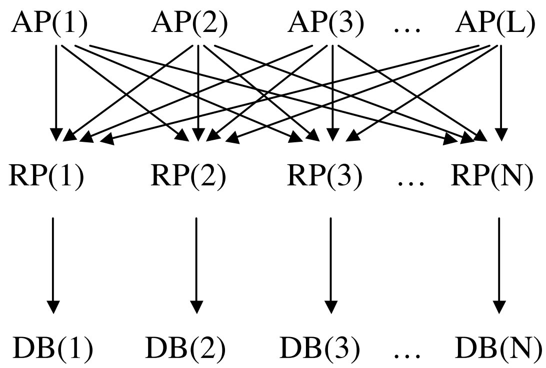

Fig. 2 Conventional DB generation methodology

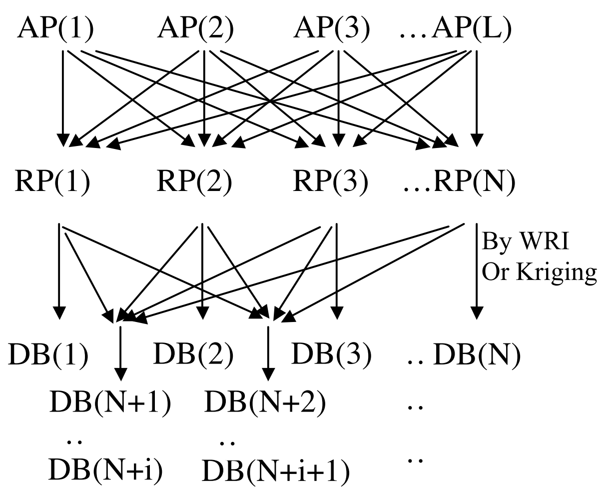

Fig. 3 Proposed DB generation methodology

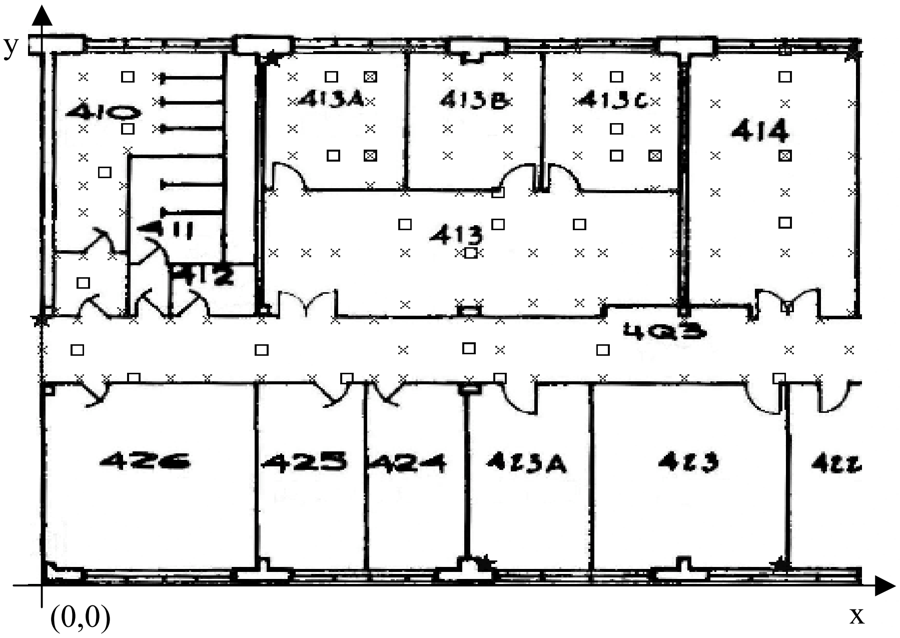

Fig. 4 Experimental test bed, the orientation is North (up) and East (right)

Fig. 5 Access Point device, Wireless card and PDA (from left to right)

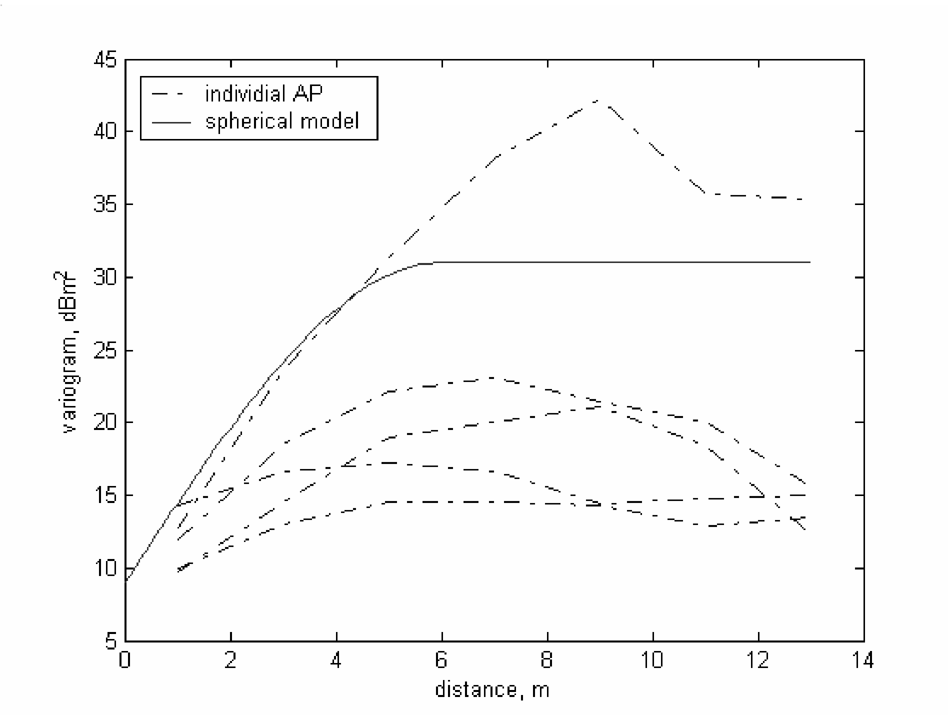

Fig. 6 Variograms of the residual (different dash-dot line means different base station)

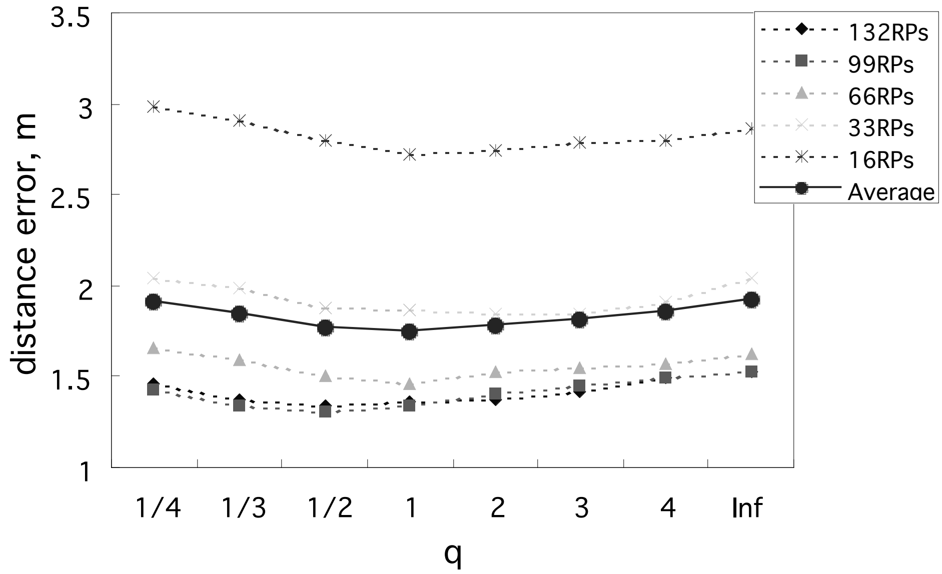

Fig. 7 Mean distance error using different signal distance

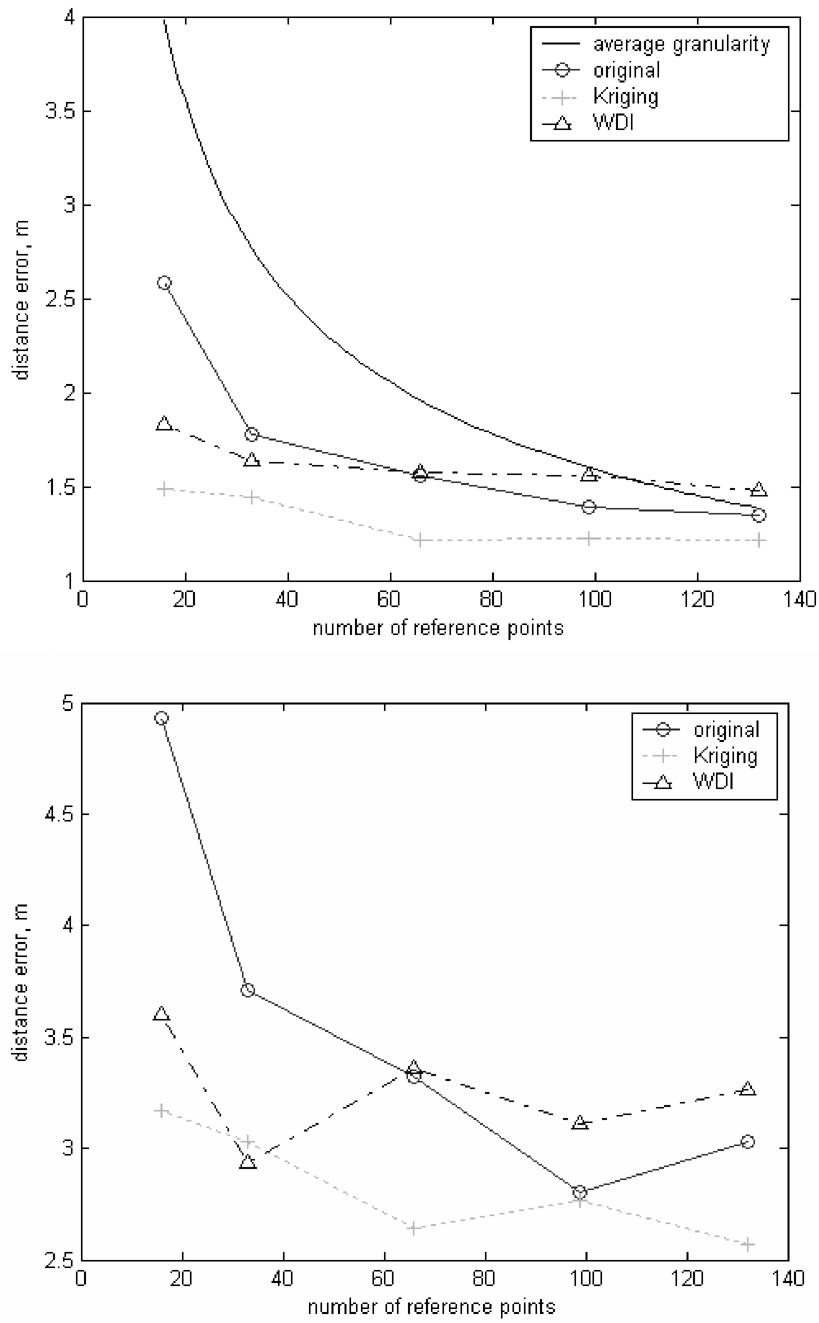

Fig. 8 Mean of average distance errors (upper) and mean of 95 percentile distance errors (lower) of five algorithms using different databases