International Global Navigation Satellite Systems Society
IGNSS Symposium 2007

The University of New South Wales, Sydney, Australia
4 – 6 December, 2007

Use of Fingerprinting in Wi-Fi Based Outdoor Positioning

Ishrat J. Quader
School of Surveying and Spatial information Systems, UNSW, Australia
Phone 93854208 Fax 93137493 Email: ishrat.quader@student.unsw.edu.au

Binghao Li
School of Surveying and Spatial information Systems, UNSW, Australia
Phone 93854174 Fax 93137493 Email: b.li@unsw.edu.au

Wendi (Patrick) Peng
School of Surveying and Spatial information Systems, UNSW, Australia
Phone 93854201 Fax 93137493 Email: pwdstudye @gmail.com

Andrew G. Dempster
School of Surveying and Spatial information Systems, UNSW, Australia
Phone 93856890 Fax 93137493 Email: a.dempster@unsw.edu.au

ABSTRACT

Though GPS is the most popular positioning system at presents it does not perform well in indoor environments and urban canyons or metropolitan city areas. Wi-Fi has been accepted as one of the good candidates for indoor positioning, its performance in outdoor environment is also of interest as a Wi-Fi based positioning system can overcome the shortcomings of GPS. An experiment has been carried out in the Sydney CBD area, where Wi-Fi access points (AP) are densely deployed. It is a cost effective system as the signal strength of the existing APs are used. Fingerprinting technique is utilized to determine the user's position, 173 reference points were selected to create the fingerprint database and 26 test points were used to analyse the accuracy of positioning.

KEYWORDS: Wi-Fi, positioning, fingerprinting, direction, outdoor

1. INTRODUCTION

GPS is the fully functional positioning system at present. But the accuracy of location provided by GPS suffers under certain environments. It does not perform well indoors or in urban canyons. GPS receivers use trilateration to find their location, hence in environments where the sky is blocked, by any means, positioning becomes difficult; in worst case conditions positioning may not be possible all together. Some alternative systems, like active batch, cricket etc (Hightower and Borriello, 2001); have been developed for positioning in indoors, but they cannot be used widely due to their inherent problems. However, Wi-Fi technology has been proven to be a promising candidate for indoor positioning and thus its performance in metropolitan areas is of interest. The test described in this paper was carried out at the heart of Sydney CBD and accuracy level achieved was 35 meters on an average.

Wi-Fi positioning technology based on signal strength (SS) basically uses two techniques: *trilateration* and *fingerprinting*.

Trilateration is a method of determining relative positions of objects using the known locations of two or more reference points and the measured distance between these points and the point of which location is being measured. Only using trilateration to accurately determine the location of a point on a 2D plane at least 3 reference points are needed. But Trilateration is not useful in urban environments due to the fact that 3 reference points are not always available. As Wi-Fi was not originally designed for positioning distance is not available directly and base stations are not always in the line of sight. Mobile phone positioning is also a significant area of research, but it suffers rigorously from the non line of sight (NLOS) error and is not a good candidate to perform well in urban canyons.

Therefore, fingerprinting technique was used to carry out the experiment. Researchers have used different deterministic and probabilistic algorithms to estimate the unknown location.
The straightforward is the deterministic method (Bahl and Padmanabhan, 2000). Though the probabilistic approach produces more accurate results, it increases the database size and the computation burden. Thus for straightforwardness, only the deterministic approach is discussed in this paper. The paper also describes a new technique to improve the accuracy produced by NN. This technique using a few candidate RPs reduces error to 35 meters on an average.

2. FINGERPRINTING TECHNOLOGY

Two phases of fingerprinting technology are: ‘training’ and ‘positioning’.

A fingerprinting database is created during the training phase. Reference points (RP) are carefully selected. Some important considerations are made while selecting a RP. They must be in an area of high interest (where position is likely to be sought). They are evenly spread out in the area of interest, so that they provide a good reference when determining position. Also they have to be at such a location that their positions can be easily determined (like at an intersection).

The number of RPs chosen is balanced between level of accuracy and labour burden in fabricating a database (Li et al., 2005a). The location of a mobile user (MU) is determined by measuring the SS of all the APs. The vector of SS for that particular RP is established by these measurements and the vector is recorded in the database. This process is repeated until all RPs are recorded. The process is demonstrated in Figure 1.

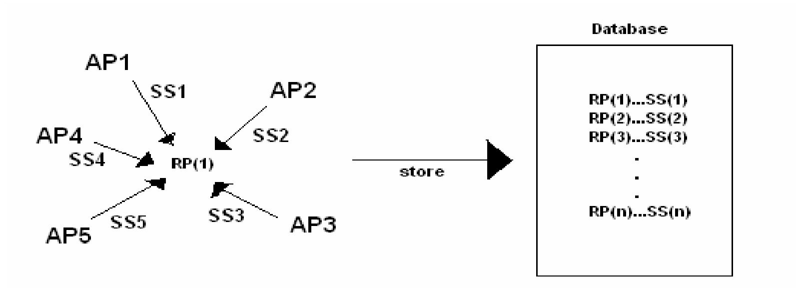

Figure 1 : Training phase of fingerprinting technology

Second is the positioning phase. In this phase, the MU acquires the APs' SSs at a particular place where it requires its location. This vector of the SSs is compared with the database using appropriate algorithm. The end result will be the most likely location of the MU. The process is demonstrated in Figure 2.

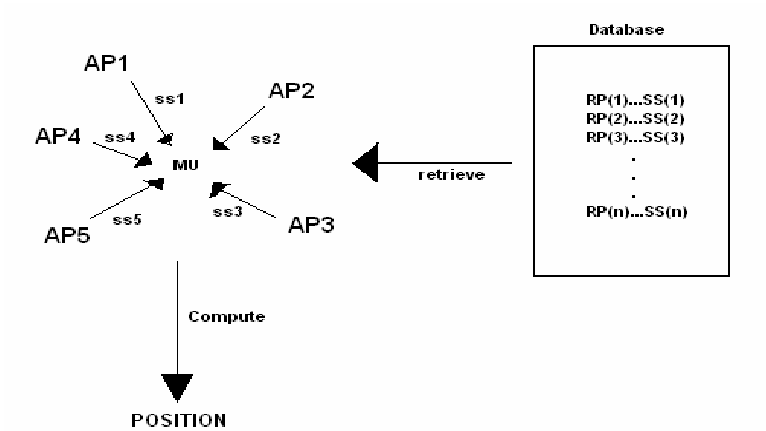

Figure 2 : Positioning phase of fingerprinting technology

However, 'Nearest Neighbour' (NN) algorithm (Bahl and Padmanabhan, 2000) is used in this research, because of its simplicity and reasonable level of accuracy. The method is based on dBm and algorithm is deterministic. The distance (in signal space) between the observed set of SS measurements [ss_1, ss_2, ss_3 …… ss_n] and the SS measurements recorded in the database [SS_1, SS_2, SS_3…… SS_n] are calculated. The distance between these two vectors can be stated as

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

Manhattan and Euclidean distance are the most common distant measurements. Other algorithms like k-NN, k-weighted NN, smallest polygon and other probabilistic approaches (Li *et al*, 2005a; Pandya *et al.*, 2003; Li *et al.*, 2005c), also exist. But they are not used in this paper due to time constrains. Another reason why a k-NN or k-weighted NN may not provide better results is the notion that, as the RPs are along almost a straight line (on a road side) they may not produce better results. Further investigation is needed in this matter.

3. METHODOLOGY USED TO CARRY OUT TEST

3.1 Testing Area

Sydney CBD was chosen to carry out the test. The chosen area has a typical urban setup with tall buildings and towers blocking the sky. Figure 3 shows the CBD area from Hyde Park to Elizabeth St. It contains College St to Hyde Park and Elizabeth St, then down Liverpool St and over the city to Darling Harbour in the background. The streets covered are Liverpool St, College St, Elizabeth St, Bathurst St.

Figure 3 : Sydney CBD area where test was conducted (www.airviewonline.com.au)

Figure 4 : Castlereagh Street

Figure 5 : Martin Place

Figure 4 and 5 shows the normal environment of Sydney CBD area. It is clear from the pictures that positioning using GPS will be difficult in these places.

3.2 Equipment And Software Used For The Test

During the experiment researcher has used a Compaq iPAQ 3970 personal digital assistant or PDA (Figure 6) running Pocket PC 2002 operating system (http://www.compaq.com). A Wireless card from Lucent Technology Wi-Fi Orinoco Wireless Golden Card (Figure 6) (http://www.orinocowireless.com) has also been used.

Figure 6 : Equipment used in this test

The software used to collect and preliminarily process SS data is NetStumbler and a pruned version for the PDA called MiniStumbler (http://www.netstumbler.com). Screenshots of the user interface is shown in Figure 7.

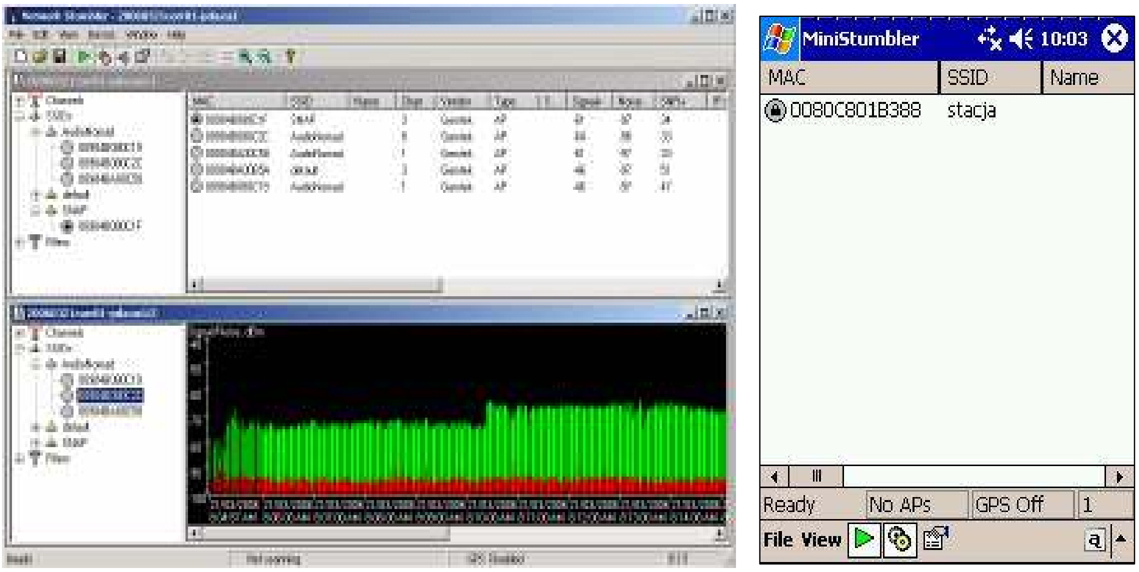

Figure 7 : Software used to collect data, NetStumbler (for PC) and MiniStumbler (for PDA)

The data collected in NetStumbler was processed so that it can be used by MatLab. MatLab has been used to analyse data.

3.3 Data Collection

Figure 8 shows the reference points and test points used to carry out the test. The green pushpins represent reference points (173 RPs) and the red dots represent Test points (26 TPs). As discussed earlier the RPs are evenly spread out in the area along the streets.

Starting from one intersection to the other the researcher has used 50 to 70 paces difference between each of the reference point. The data has been collected from the bottom right corner of the test area till the bottom left corner. In order to cover the whole area (Figure 8), the researcher started collecting data at the first point referring it as RP1 and walked north till the next intersection collecting data at each 50 to 70 paces difference. At the intersection the researcher would keep moving north collecting data to cover the entire test area that has been marked in figure 8. In this way the whole area has been covered walking up and down the streets from north to south. After the whole area has been covered from north to south, the researcher repeated the whole process to cover the area from west to east.

Data for the TPs are also collected in a similar way. These data were processed later. The TPs are marked by red circles in Figure 8.

After collecting all the data Google Earth was used to map the location of the points and local coordinates were calculated.

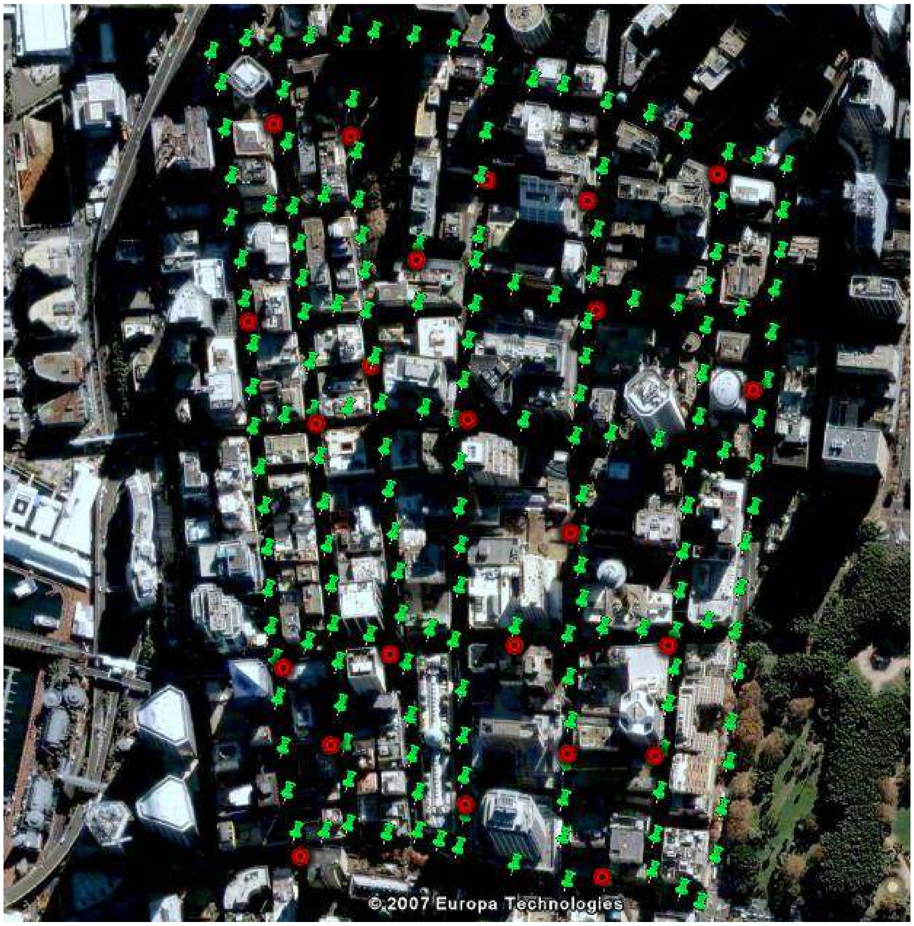

Figure 8 : Reference points and test points in Sydney CBD area

3.4 Creating Database

In the first stage a fingerprinting database should be created. The researchers have collected data from all four directions (east, north, west and south). However the directional data was ignored at this stage and the average SS was used to represent the feature at the specific RP. The database has all the mean SSs from all the APs for one particular RP. If the SS could not be received, a default value, -100dBm, was set. All the 173 RPs were processed in this way and logged in the database.

3.5 Calculating Position

The mean value of the SSs collected at each TP were computed and matched with the database to find the likeliest position. This was repeated for all the 26 TPs.

3.6 Using GPS For Positioning In Sydney CBD Area

The test area has a typical urban setting and the number of visible satellites are limited by tall buildings, towers etc. GPS needs at least 3 satellites to calculate a position (2D) and this requirement is not meet in most of the Reference Points (RPs) or test points (TPs). Figure 9 shows the eTrex GPS receiver used in this test (www.garmin.com/products/etrex/).

Figure 9 : eTrex GPS receiver

Figure 10 shows the number of visible satellites from all the TPs. Thus this indicates the situation if GPS was used to calculate position. It is clearly noticeable that 4 satellites are not visible from any of the TPs. 3 satellites are visible from 8 TPs. And even less satellites are visible from the rest of the TPs. It is clear that at most of the TPs the number of visible satellites is not sufficient and thus a position could not be calculated. Furthermore, in the TPs where position can be calculated with 3 visible satellites, the geometric distribution of the satellites are bad – that is the dilutions of positioning (DOP) values are quite large (Understanding GPS : Principles and Applications, Elliott D. Kaplan) The consequence is that the error is quite high (average 30.25 meters). Another disadvantage is that the time of first position is quite long. Therefore, some other techniques of positioning are required to achieve satisfactory level of accuracy.

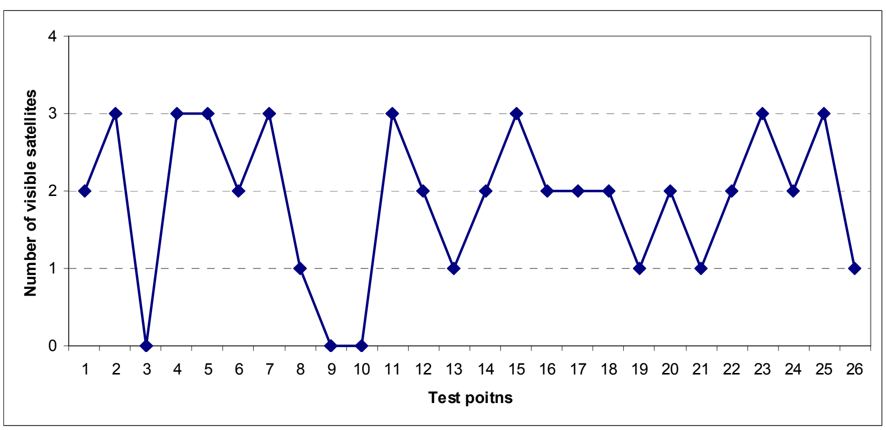

Figure 10 : Number of visible satellites from test points

4. TEST RESULTS AND ANALYSIS

At each of the 26 TPs, data were collected in four directions (east, north, west and south). Thus if direction is ignored, 104 independent TPs should be considered. However, in this test only 56 TPs are calculated. This is because some points only have one directional data as it was not possible to collect any data from the other directions. Again some points have two or three directional data. For example, TP 11 has only one directional data, TP14 had data from two directions and data from TP12/13/15 have no specific direction. Further research is being carried out on the impact of directional information over the directionless data.

Google Earth only provides the latitude and longitude of a location which is not convenient for processing. Thus a local coordinate system was created and the local coordinates of all the RPs and TPs were computed. Both Manhattan and Euclidean distances can be used in the algorithms, but Euclidean distance was used in this test.

In the beginning, a normal NN algorithm for indoor environment was used to process the data. Fingerprinting is an eminent approach when it comes to indoor positioning. But its performance outdoors is still to be investigated. So a normal NN was used and distance was calculated. The algorithm calculates the signal distance based on SS, but error in locations was quite high (average 119 meters). This is due to the fact that when all the RPs are considered, there are many elements in the calculation. It means that the distance (in signal space) between the observed set of SS measurements of the TP and the SS measurements recorded in the database for the RPs is calculated. When the number of APs increases the nearer RP may end up having a larger value of distance vector than a far away RP. Hence it produces a large error in positioning. The worst position is calculated in the TP47 (TP 17 in the where total number of TP is 26, marked as the yellow square in Figure 11) and this point is further investigated to detect the problem. While using normal NN algorithm, nearest neighbour is found to be RP 74 (marked by red balloon) and error in positioning was 711.25 meters. This is because the nearer RPs are calculated to have larger value of distance than the farther ones and hence a RP far away from the TP is detected to be the nearest neighbour.

Figure 11 : Algorithm finds erroneous nearest neighbour

Hence, a technique was used to improve positioning accuracy. First a few candidates, who received signals from the similar APs were chosen. Similar APs mean that a certain portion of the APs which can be 'heard' at the RPs can also be found at the TP and if this happens it is considered to be a 'match' of APs. The steps used to implement this are :

1) find total number of APs, from which singal can be received, for a TP
2) repeat step 1 for RP
3) find out how many of the APs match (that is same APs can be 'heard' at the RPs and also be found at the TP)
4) if a certain portion (may be half, one third or one fourth) of the total number of APs for one TP matches that of the ith RP, select this RP to be a candidate, else reject this RP
5) repeat steps 2,3 and 4 until all the RPs are checked
6) apply NN using only the candidate RPs and find the nearest neighbour
7) calculate position

When this is done the erroneous RPs are rejected and for TP 47 the nearest neighbour is found to be RP 57 (marked by yellow balloon).

The proportion of the similarity of the APs is varied to see the impact it has on the result. Results improve while the similarity of the APs increases (Figure 12).

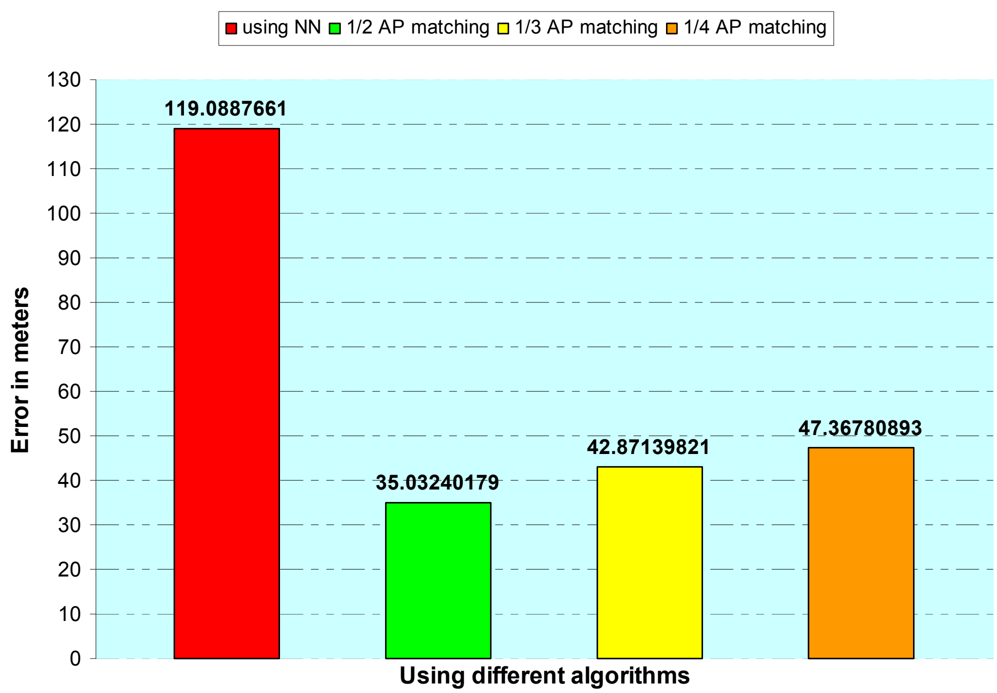

Figure 12 : Different Algorithms vs. Error

Data were processed matching half of the APs, then one third and one fourth of APs. The best result is found matching half of all the APs. After choosing the candidates, the signal distance of those candidates from the TP were computed, and the user's position was estimated. This improves the positioning dramatically. Using this technique the algorithm finds RP 57 (marked by yellow balloon with star) as the nearest neighbour for the same TP (TP 47 marked by red balloon) and error reduces to 15.06 meters. The technique reduces the error to 35 meters or less on an average (Figure 12). Using candidate RPs narrows down the number of calculations and searches that needs to be carried out, hence saves calculation burden and time as well.

5. CONCLUSION

Wi-Fi has become a widely used and ever spreading technology. It is a good choice in calculating position as it is a wide spread well structured network. Positioning technology always tries to depend on existing network rather than installing a new network. Hence Wi-Fi is used to calculate location in environments where GPS cannot perform well. Fingerprinting performs well indoors and thus it can calculate positions at indoor like environments too.

This test proves that fingerprinting method can be used outdoors as well as indoors. In urban canyons the sky is usually blocked by tall buildings and GPS cannot perform well or in worst cases cannot determine position at all, because sufficient satellites are not visible. As a result determining position can be impossible. However, positioning in urban settings is very important as it has commercial as well as social importance. Fingerprinting can be used in these situations. Though the test described in this paper establishes that initially the level of accuracy is poor, approximately 119 meters. However, using the new technique, where a few candidates are used to narrow down the search space significantly improves accuracy to approximately 35 meters. It is also assumed that use of directional data will improve the accuracy even more. Further investigation is being carried out in this matter.

6. REFERENCE

Li, B, Wang, Y, Lee, HK, Dempster, AG, Rizos, C (2005a) Method for Yielding a Database of Location Fingerprints in WLAN, *Communications, IEE Proceedings*, Vol. 152, Issue 5, October 2005, pp. 580-586.

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

Dobkin, D (2002) Indoor propagation issues for wireless LANs, *RF Design*, Sep 1, pp. 40-46.

Hightower, J, Borriello, G (2001) Location systems for ubiquitous computing, *IEEE Computer*, 2001, 34, (8), pp. 57–66

Li, B, Dempster, AG, Barnes, J, Rizos, C, Li, D (2005b) Probabilistic algorithm to support the fingerprinting method for CDMA location, *Int. Symp. on GPS/GNSS*, Hong Kong, 8-10 December, paper 9C-05, CD-ROM procs.

Li, B, Dempster, AG, Rizos, C, Barnes, J (2005c) Hybrid method for localization using WLAN, *Spatial Sciences Conference*, Melbourne, Australia, 12-16 September, pp. 341-350, CD-ROM procs.

Li, B, Salter, J, Dempster AG, Rizos, C (2006) Indoor Positioning Techniques Based on Wireless LAN, *First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, Sydney, Australia, 13-16 March, paper 113.

Xiang, Z, Song, S, Chen, J, Wang, H, Huang, J, Gao, X (2004) A wireless LAN-based indoor positioning technology, *IBM J. Res & Dev*, Vol 48 no. 5/6, September/November 2004

Wang, Y, Jia, X, Lee, HK and Li, GY (2003) 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, CD-ROM proc., paper 54.

Li, B, Jeffrey, K, Johnathan, L, Dempster, AG, (2007) Use of Directional Information in Wireless LAN based indoor positioning