2012 International Conference on Indoor Positioning and Indoor Navigation, 13-15th November 2012

A new method to generate and maintain a WiFi fingerprinting database automatically by using RFID

Michael Gunawan, Binghao Li, Thomas Gallagher, Andrew G Dempster
School of Surveying and Spatial Information Systems
University of New South Wales
Sydney, Australia
binghao.li@unsw.edu.au

Günther Retscher
Institute of Geodesy and Geophysics
Vienna University of Technology
Vienna, Austria
gretsch@pop.tuwien.ac.at

Abstract— **Location fingerprinting in WiFi positioning has been widely used in indoor environments. The key issue of the fingerprinting technology is the fingerprint database. The disadvantages of this technology are the database generation and maintenance requirements. The conventional method to create the database is that people carry out the survey manually (that is what the commercial products are doing). When the environment changes significantly (such as after a building renovation, or moving of furniture), the database has to be rebuilt. This paper proposes a new method to build and maintain the database in an efficient manner. This method only requires persons to carry a specific device which consists of a Radio Frequency Identification (RFID) reader and WiFi scanner to log the coordinates and the WiFi signal strengths. The coordinates are provided by some pre-deployed medium range (1-2 meters) RFID tags in the building with a location technique based on 'cell ID'. As the persons conducting the survey are moving around the area of interest for purposes other than the fingerprint survey (such as a security guard who regularly patrols the whole building anyway), so the fingerprint database can be generated "automatically". Also, the database can be refined as the data are being accumulated. When the environment changes, it can be detected by the self-refining database. A preliminary test was carried out for the proposed method. The results show that it works well.**

Keywords- WiFi, Fingerprinting database, RFID

I. INTRODUCTION

Indoor positioning has attracted a huge interest as location based services (LBS) demand a sufficiently accurate positioning system for outdoor as well as indoor environments (in fact the requirement of positioning accuracy for indoor is higher than that for outdoor). Global Navigation Satellite Systems (GNSS) such as Global Positioning System (GPS) provides accurate outdoor positioning solution (several metres accuracy); however, GPS does not have comparable indoor positioning accuracy as the signal from GPS satellites cannot penetrate walls of buildings [1]. Thus, research on indoor positioning is carried out to acknowledge and alleviate the problems regarding conventional method of positioning,

Indoor positioning methods are extensively researched and highly on demand. Several classical range-based methods such as trilateration and triangulation can be used to support positioning [2]. However, there are manifold problems to utilize these methods. There are no perfect suitable infrastructures that are easily available or established to get the accurate range or angle measurements. If the existing infrastructures are used (such as WiFi), only signal strengths (SS) are available.

WiFi is a standard networking technology and WiFi access points (AP) are widely deployed. Modern mobile phones are now equipped with WiFi chips as a standard and WiFi signals are easily available almost in possibly every building which makes using WiFi for positioning a very popular method.

To convert a SS measurement to a range measurement accurately is not an easy task. The nature of an indoor environment leads to non-line-of-sight (NLOS) propagation and multipath of the signal [3]. Instead of converting the SS to a range, fingerprinting is usually employed. The fingerprinting method implements mapping of co-relational information such as location with its characteristics and is considered as a better candidate for ubiquitous indoor positioning as it utilizes the NLOS propagation and multipath [4, 5]. Some commercial systems such as Ekahau and Skyhook Wireless have implemented fingerprinting for positioning.

Positioning using WiFi fingerprint involves two phases: training phase (database generation) and positioning phase (localization). In the former phase, surveys are carried out to obtain the WiFi received signal strength indicator (RSSI) from surrounding WiFi APs as well as the location corresponding to the RSSI. The surveys are usually carried out by a surveyor who has knowledge of the area (required for location input). These measurements are then logged into a database to form a fingerprint database. In the positioning phase, WiFi AP RSSIs are once again measured (by the user's device). This measurement is then queried into the fingerprint database and positioning algorithms are applied to the measurement to find a match of the given RSSI with a location in the database; thus, giving the user's current estimated position [4, 5].

Despite all the benefits of fingerprinting, it also poses two crucial problems that relate directly to the training phase. An entry in the fingerprint database is represented by a location identifier paired with the WiFi RSSI reading of the location.

This fact creates a problem if the RSSI reading of the location has changed since the last survey (it might be caused by furniture movement or new installation of a WiFi AP), thus creating an inconsistency between the database and the information supplied by the user in the positioning phase; this will ultimately result in accuracy loss, or in the worst case, erroneous location output by the server. To resolve this condition, the database has to be updated periodically in order to have the latest RSSI values or detect new WiFi APs in the area [6]; updating the database requires constant new surveys to be taken, which fuels the research of automating the surveys as the surveying process is proven to be a very tedious and resource consuming task. Moreover, it can be argued that fingerprinting has advantage over the range-based methods only in static environment where no moving objects (such as people) are around.

In order to make the system more effective and practical, effort needed to preserve the accuracy and reconstruct the database has to be lowered. Research regarding this topic has been carried out before by Gallagher et al., which proposes a crowd sourcing system, which prompts user to input the correct location upon receiving error position [7]. This system also applies input filtering to eliminate erroneous input, which is a crucial factor to determine whether the accuracy of the database can be preserved. Other similar research includes an organic positioning system introduced by Park et al. [8], which similarly proposes to prompt user to input correct location if an error occurred.

This paper will present the idea and initial experiment results of using a Radio Frequency Identification (RFID) based system to support the training phase by supplying location information automatically. Using the location information, the experiment device will trigger a scan of WiFi RSSI in the area and log the data; this will automate the surveying to a certain extent. In the system, a surveyor is still needed, but the person carrying out the survey for the training phase does not necessarily have to be an expert in surveying. For instance, the surveyor can be a security guard who regularly patrols the whole building anyway. He/she will just walk around not knowing about the survey at all; the device will take care of the surveying.

II. METHODOLOGY

This section will describe the usage of RFID in the system, setup of test bed, and algorithm used for the positioning.

A. RFID Integration

RFID is an established and well known tracking infrastructure that is mostly used to identify or track inventory. A typical RFID system consists of RFID tags and RFID reader; the tags (containing unique ID) are deployed into individual items to give it an identity similar to barcode, those tags are then readable by the RFID reader device. RFID tags are classified according to its frequency and range [9]; in the experiments, we are using Ultra High Frequency (UHF) with range of 862-955 MHz in order to avoid signal interference with WiFi signals (2.4 GHz frequency).

The whole system consists of RFID tags, a RFID reader, a WiFi scanner and a processor. In the test, a modular RFID reader connected to a development board with USB communication to a PC was utilised. Fig. 1 shows the functional diagram of the system. The PC was used as a WiFi scanner and a processor. Upon coming in proximity of an RFID tag, the RFID reader detects the tag and reads the ID of the tag (step 1 in Fig. 1). On each detection the reader triggered a WiFi scan and collected RSSI from surrounding WiFi APs (step 2 in Fig. 1). This scan was then tagged with the ID to form a fingerprint and stored in the memory (step 3 in Fig. 1). A collection of fingerprints are called fingerprint database, which is the database referred to when positioning algorithms are used.

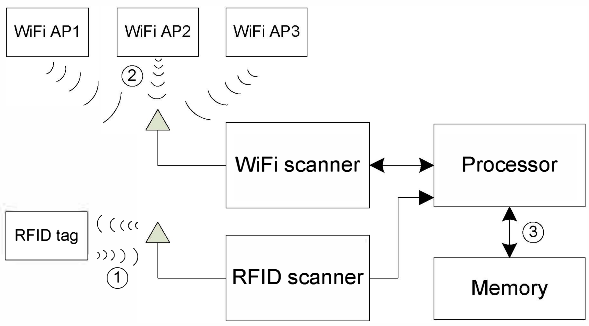

Figure 1. RFID system illustration.

The experimental test bed is located in level 4 of the Electrical Engineering Building of the University of New South Wales. A corridor was chosen to simplify the measurement (surveying) process; the corridor is about 30 metres in length and 2 metres in width. We deployed the tags in interval of 3 metres, starting and ending in both ends of the corridor; this accounts to a total of 11 tags used along the corridor. Each RFID tags were labeled with a unique ID corresponding to its location and each ID can be related to a coordinate location in the location database. Fig. 2 depicts the map of the corridor with the tags position; the pentagram symbol in the figure shows the starting and ending position of the surveys. Later in the experiment, the numbers of tags were reduced in order to simulate a more practical and efficient tag deployment model.

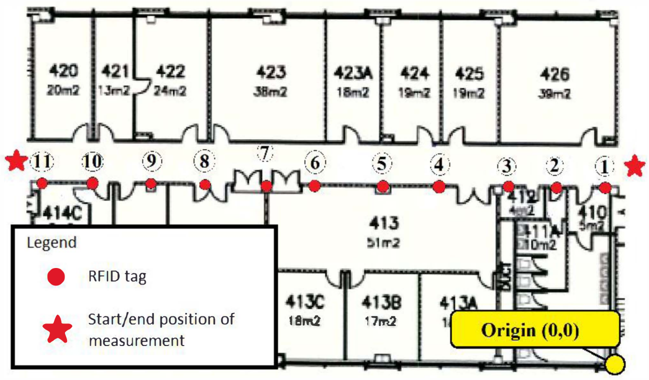

Figure 2. RFID tag placement in the test bed.

B. Database Format

The system used an additional database compared to the traditional system, which only requires a fingerprint database. In our case, we needed to keep the RFID tag database in addition to the fingerprint database; RFID tag entry was in the following form:

$$t_x = \{tid_x, l_x\}$$

where t_x represents an entry of a certain tag in the database, the values consists of tid_x – the unique ID of the tag, and l_x – the location information referenced by the tag.

On the other hand, fingerprint database used the same format found in the system by Gallagher et al. [6], in the form:

$$f_{li} = \{l_i, (MAC_0, SS_0, MAC_1, SS_1, ...)\}$$

where l_i represents location where the fingerprints are taken, and (MAC_0, SS_0, MAC_1, SS_1,...) shows the result of measured WiFi RSSI ordered and associated by the MAC address of the WiFi AP.

The location information in both databases are relational, as each WiFi scan was triggered by a tag detection, which means that the location of the WiFi scan (l_i) is obtained by getting the location of the detected tag (l_x). Obviously, the more tags, the more chances the WiFi scan can be triggered. However, to deploy too many tags is not a trivial work. Ideally, the less tags required, the simpler the system. But too less tag would mean that the accuracy of the system is going to be very low. Later in the section, we will discuss how interpolation theory can be used to ameliorate this problem by creating in-between reference points and generating the RSSI values.

C. Positioning Algorithms

In order to process the WiFi MAC address and RSSI readings given by the users and convert them into a location value, some algorithms have to be executed. There are various algorithms to compute the location of user; one of the basic ones used in this experiment is the 'nearest neighbor' (NN) algorithm. This algorithm requires the distance between the RSSI vector (SS_1, SS_2, SS_3, ...) given by the user and the RSSI vector in the database to be computed. The Manhattan or Euclidean distance can be calculated using the following equations:

$$L_q = \left(\sum_{i=1}^{n} | SSU_i - SSD_i |^q\right)^{\frac{1}{q}} \quad (1)$$

where *SSU* represents the user RSSI vector and *SSD* represents the database RSSI vector. The Manhattan and Euclidean distance were calculated with q=1 and q=2 respectively; the nearest neighbor is the point with a shortest signal distance [4, 5].

One variation from NN algorithm is KNN, where K is the number of neighbors considered to produce the user location. To calculate KNN, the average coordinate of K points are used; this logically will yield better results than using one neighbor only and eliminating others, but in our experiments, the results from KNN is worse than NN. The way the reference points were chosen (one dimensional) and the nature of the corridor (narrow and long) are the reasons. The K weighted nearest neighbor (KWNN) algorithm uses weighted average rather than normal average of coordinates. The inverse of signal distance defines the weight as such:

$$w = \frac{1}{L_{qi} + \partial} \quad (2)$$

where L_{qi} is the signal distance found in equation 1 and ∂ is a small real constant to avoid division by zero [5].

Nearest neighbor based on signal distance and its variant is classified as deterministic method, meaning that it only uses the average of the signal strength as the main variables for calculation. One other method to consider is the probabilistic method, which uses the variation of the signal strength received from APs. The probabilistic method is more complex, but can provide a better estimate of the user's position. The probabilistic approach used here is based on the Bayes Rule [10]:

$$\arg\max_{L_r}[P(L_r \mid M)] = \arg\max_{L_r}[P(M \mid L_r)] \quad (3)$$

where L denotes location and M denotes a measurement.

To estimate the probability of such occurrence, commonly the mean and standard deviation of RSSI at each location is computed; however, the behavior of RSSI is not simple to be expected as propagation of the signals are influenced by several factors (which doesn't make it necessarily Gaussian). Two approaches are generally considerable, i.e., the kernel method and histogram method.

In the kernel method, a probability mass is assigned to a kernel around the data observed; the probability is then calculated using a kernel function. The histogram method (used in our experiments) on the other hand, uses bins or value categorization to cover all measurement range; according to these bins, we can then calculate the probability (existence of an AP in a certain location), thus each AP will appear with different probability and we can estimate the location according to the probabilities [11].

D. Database Interpolation using Kriging

Kriging is an interpolation method that utilises spatial correlation, which means that every location in the area of interest contains information about the surrounding locations as well. This method will enable interpolation to generate a more entries to the database; kriging originally comes from mining industry, using a variogram to quantify spatial correlation between each measurement [12]. The feature of kriging includes:

- It is a linear function of the data with weights calculated according to specifications of non-biased and minimum variance.
- The weights are determined by solving a system of linear equation with coefficient that depends only on the variogram.

Kriging is generally more flexible than other interpolation methods as the weight are dependent on the variation of the function in space. Another advantage of kriging is that the estimation error can be measured [5, 12].

The classical way to estimate the variogram is by using the following formula:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{x_i - x_j = h} (Z(x_i) - Z(x_j))^2 \qquad (4)$$

As most of the points (location) are irregularly spaced, to have more pairs, $x_i - x_j = h$ has to be weakened:

$$\left||x_i - x_j| - h\right| \leq \varepsilon \qquad Angle(x_i - x_j, h) \leq \delta \quad (5)$$

In our experiments, interpolating the database using kriging yielded very accurate results and the method was applied for extremely scarce tag deployment model to help with the accuracy of positioning.

III. TEST AND RESULTS

This section explains the method of taking the measurements, results of positioning using automatically generated database, and comparison between manual database and automatic database positioning accuracy.

A. *Measurement Method*

While we were performing the measurement, we tried to simulate how the model surveyor would move along the corridor, which was just walking straight from end to end of the corridor without needing to deal with any system; Fig. 3 depicts the location of the test points. Some details about the measurements are:

- There were 10 WiFi APs used in the experiment. More AP are actually available and detected but they are very unstable (might be because of range factor), thus we opted to leave them out.

- As the RFID reader antenna in our test device reads only in a particular direction, we panned the antenna to face the side of wall where all the tags were installed.

- Two training phase measurement models were performed: non-stop slow movement, and non-stop fast movement. To reduce outlier, each measurement models were taken more than ten times and the values averaged. In reality, there could be more measurements taken as the 'surveyor' walks around all the time.

- The influence of the human body moving in the environment was considered in the test. We did tests during work hours (many traffic) and after hours (no traffic) and the results were compared.

- In the positioning phase, NN, 3NN, 3WNN, probabilistic method, and kriging + NN method where utilised where applicable.

- We were mostly only considering one dimensional result from the positioning phase as the test bed is a long and narrow space (corridor).

On the later stage of analyzing, we reduced the number of tags to simulate a more effective implementation of the system. To do this, several tags were removed and the fingerprint entries associated with these tags in the database were removed. In total, three tests were conducted: 11 tags, 5 tags and 3 tags. Fig. 1 depicts the test with 11 tags (one tag every 3 metres); the 5 tag model takes off tag number 2, 3, 5, 7, 9, and 10 (one tag for each 6 metres); the 3 tag model takes off tag number 2, 3, 4, 5, 7, 8, 9, and 10 (one tag for each 10 metres). Please note this is not an optimised deployment of the tags since tag 1 only covers a small area.

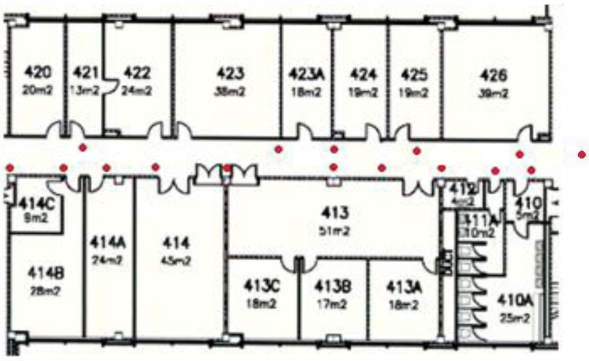

Figure 3. Positioning test reference location.

B. *After Hour Measurements*

Fig. 4 and 5 show the average error (in metre) of the positioning result. It can be seen that NN are actually yielding better result than 3NN or 3WNN. This is caused by the one dimensional deployment of the reference points. In the slow measurements, probabilistic (Bayes histogram) method has the most accurate result among the rest while in fast measurements interpolated database yields the best result. In general, the accuracy is sufficient for most of the indoor applications (1.8 metre is the largest error that can be seen). The fact that slow measurement yields almost double the data compared to fast measurement does not make slow measurement better; as we can see in Figs. 4 and 5 that slow measurement outputs has worse accuracy. Thus, more data collection at the same time does not necessarily improve the accuracy of the system.

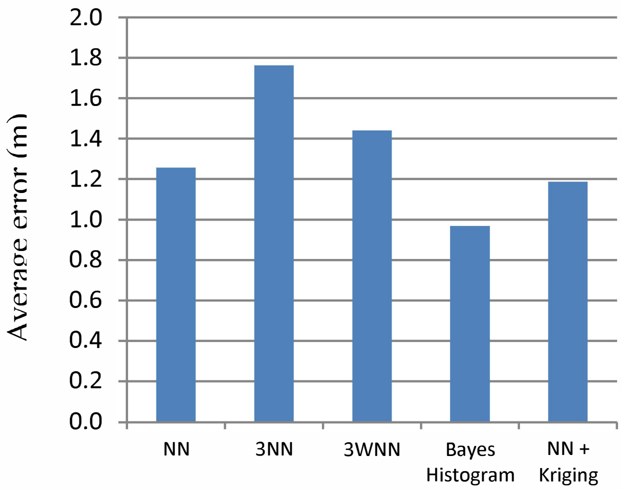

Figure 4. Positioning errors (in metre) of slow measurement on after hour setting.

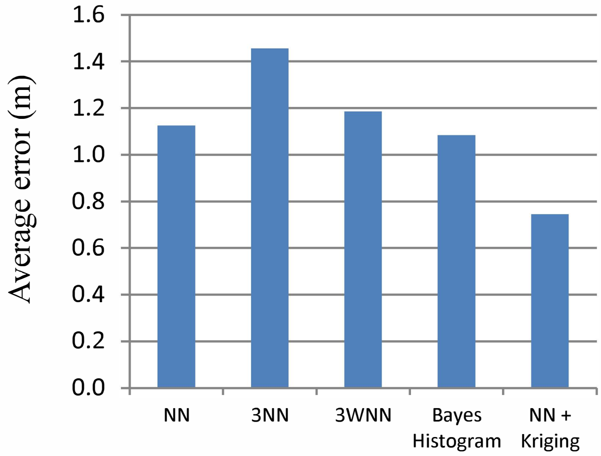

Figure 5. Positioning errors (in metre) of fast measurement on after hour setting.

C. Working Hour Setting

Fig. 6 shows the average error (in metre) of the fast measurements during working hours. Once again, 3NN and 3WNN has lower accuracy level compared to the other methods. Probabilistic method shows the best accuracy performance out of all, although kriging with NN has a very low (0.8 metre) level of error.

D. Efficient Tag Placement Model

Fig. 7 and 8 show the average positioning error (in metre) of the 5 tag and 3 tag tests. The scarce tag placement practically reduces the accuracy of positioning for all methods. This is where database interpolation using kriging is very crucial; we can see from the two graphs that on interpolated fingerprint database, the error are still very low (1.2 metre for 5 tag tests and 1.8 metre for 3 tag test).

E. Comparison with Manual Measurements/Surveys

The results from this experiment were compared with previous work by Li et al., which was performed in the similar location (level 4 of Electrical Engineering in UNSW) [5, 13]. The test area of the previous work was about five times bigger than that of automatic survey test (including the main part of the corridor and several office rooms). In the comparison, we used kriging and nearest neighbor method as it was the most stable and had a viable accuracy even with low number of RFID tags. For the setting, the 3 tag model was used because it was the setting with least tag number required for an acceptable accuracy level. Fig. 9 shows the comparison of average error between the database constructed with automatic survey and manual survey. The results from the manual survey were taken from the setting that has the smallest number of reference points (16 RPs) in the test.

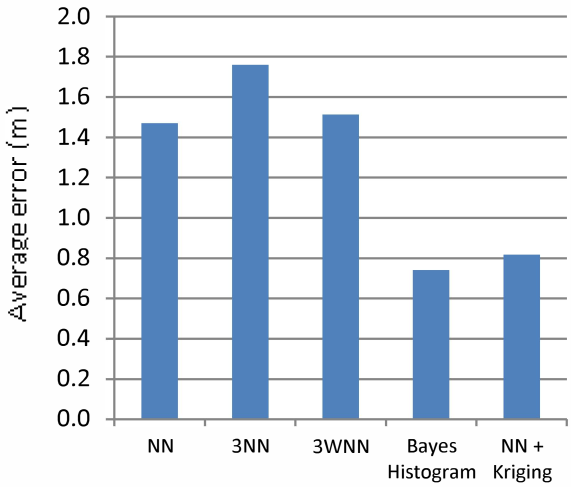

Figure 6. Positioning errors (in metre) of fast measurement on working hour setting.

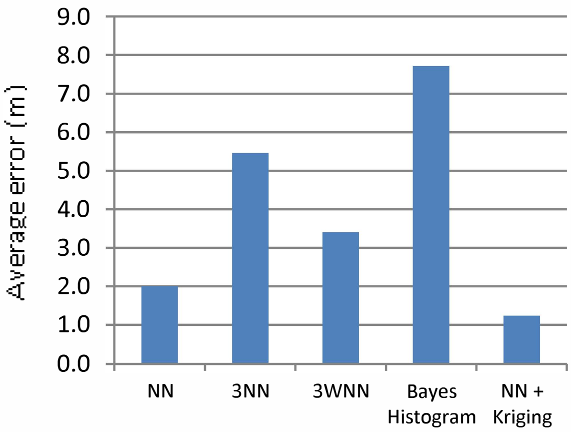

Figure 7. Positioning errors (in metre) of 5 tag model.

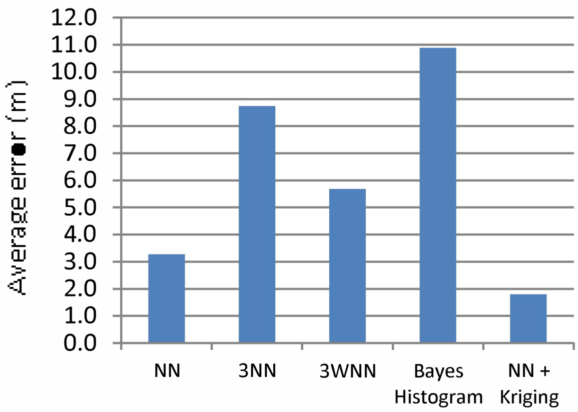

Figure 8. Positioning errors (in metre) of 3 tag model.

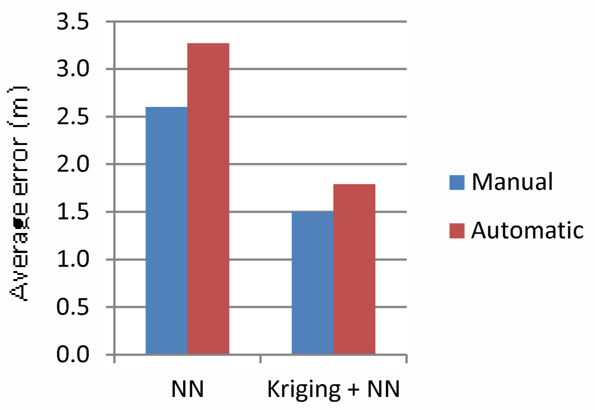

Figure 9. Comparison of positioning errors (in metre) between using database generated automatically and manually.

It can be seen from Fig. 9 that by applying kriging to the automatic database, we can achieve similar results compared to manual database generation. The ratio of number of test points to the test area is lower (about 85%) in the automatic test than that in the manual test. Hence by installing the RFID tags and automatically collect WiFi data, we can reduce the effort needed to survey a location, thus overcoming one of the main hurdles in WiFi fingerprint positioning.

IV. Concluding Remarks

Training phase (surveying) of WiFi fingerprinting requires significant amount of labour and time, and any environmental changes requires the database to be updated. This change includes moving furnitures, renovation, and WiFi AP installation/removal. This poses a crucial problem where the area of interest needs to be surveyed repetitively, and will reduce the system's overall effectiveness and viability. This condition is the main reason that fuels the research in this paper, which targets to automate the surveying. RFID has a history of being an established tracking technology that is very cost efficient and customizable. RFID systems has also been researched in localization field [14] and is very capable of assisting WiFi automation with the ability of supplying identification of location to surveyor without having the surveyor to recognise the surrounding and inputting the location into the system.

Overall, the method of automatic database generation yields an accuracy level on par with manual surveys in some cases. However, to reduce the number of tags needed to run the system, accuracy has to be sacrificed. By using 3 RFID tags in an area of 30 metres by two metres, we can reach error level of 1.8 metre (with the help of database interpolation method) which is comparable to manual survey. One fact that is very important to note is that this method of automatic database generation uses far less effort, which in turn increases the feasibility of WiFi fingerprinting method itself.

The prototype of a compact, battery-operated device used for the automatic survey is under development. This is preferred rather than a bulky device with a laptop PC as the person performing the survey during the training phase does not need to be an expert. For instance, a security guard patrolling the building can be used to perform the measurements. As the nature of the automatic survey enables the system to get new measurement data constantly, it is possible to create a distinct database for each human traffic situation (heavy traffic or light traffic). WiFi AP installation/removal can also be detected, and thus giving the latest information about the environment. All these information has to be processed and filtered before being inputted into the database, thus the method of database update are yet to be investigated. Furthermore with the help of accelerometer and other sensors to accurately locate the surveying device, it is possible to switch on/off a periodic WiFi scan between to tags and generate dense reference points. Further investigation will be carried out in the near future.

References

[1] B. Li, J. Zhang, A. G. Dempster, C. Rizos, “Open Source GNSS Reference Server for Assisted-Global Navigation Satellite Systems”, The Journal of Navigation, 64(01): 127-139, 2011.

[2] E. Mok and G. Retscher, “Location determination using WiFi fingerprinting versus WiFi trilateration,” Journal of Location Based Services, 1(2):145-159, 2007.

[3] H. Hashemi, “The indoor radio propagation channel”, Proc. IEEE., 1993, 81, pp. 943–968.

[4] P. Bahl and V. N. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system,” in Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies Tel Aviv. Israel, 2000. vol. 2, pp. 775–784.

[5] B. Li, Y. Wang, H.K. Lee, A.G. Dempster, C. Rizos, “A new method for yielding a database of location fingerprints in WLAN,” IEE Proc. Communications, 152(5): 580-586, 2005.

[6] T. Gallagher, B. Li, A.G. Dempster, C. Rizos, “Database updating through user feedback in fingerprinting-based Wi-Fi location systems,” in Proceedings of Int. Conf. Ubiquitous Positioning Indoor Navigation & Location Based Service, Kirkkonummi, Finland, 14-15 October, 2010, paper 1, session3.

[7] T. Gallagher, B. Li, A.G. Dempster, C. Rizos, “A sector-based campus-wide indoor positioning system,” in Proceedings of International Conference on Indoor Positioning and Indoor Navigation, September 15-17, 2010, Zurich, Switzerland.

[8] J. Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. Teller, and J. Ledlie, “Growing an organic indoor location system,” in Proceedings of 8th International Conference on Mobile Systems, Applications, and Services, June 15-18, 2010, San Fransisco, California, USA, pp. 271-284.

[9] K. Finkenzeller, “RFID Handbook: Radio-Frequency Identification Fundamentals and Applications, 2nd ed: Wiley, 2004.

[10] B. Li, A.G. Dempster, J. Barnes, C. Rizos, & D. Li, “Probabilistic algorithm to support the fingerprinting method for CDMA location.” in Proceedings of Int. Symp. on GPS/GNSS, Hong Kong, 8-10 December, paper 9C-05, 2005.

[11] A.M. Ladd, K.E. Bekris, A. Rudys, G. Marceau, L.E. Kavraki, and S. Dan, “Robotics-based location sensing using wireless ethernet”, in Proceedings of 8th ACM Int. Conf. on Mobile Computing and Networking (MOBICOM), Atlanta, GA, September 2002.

[12] M. Armstrong, Basic linear geostatistics, Springer, Berlin, 1998.

[13] B. Li, J. Salter, A.G. Dempster, C. Rizos, “Indoor positioning techniques based on Wireless LAN,” in Proceedings of 1st IEEE Int. Conf. on Wireless Broadband & Ultra Wideband Communications, Sydney, Australia, 13-16 March, , 2006.

[14] H.D. Chon, S. Jun, H. Jung, S.W. An, “Using RFID for accurate positioning,” Journal of Global Positioning System, 3(1-2): 32-39, 2005.