International Global Navigation Satellite Systems Society
IGNSS Symposium 2015

Outrigger Gold Coast, Qld Australia
14-16 July, 2015

Vector Distance Measure Comparison in Indoor Location Fingerprinting

Vahideh Moghtadaiee (1)
School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia
v.moghtadaiee@unsw.edu.au
Andrew G. Dempster (2)
School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia
a.dempster@unsw.edu.au

ABSTRACT

Non-satellite-based technologies are important for indoor localization due to the difficulties that satellite-based technologies such as the Global Positioning System (GPS) experience operating in such areas, e.g. low received signal power and low visibility of satellites. The most popular positioning method indoors is fingerprinting, which is a technique that records vectors of received signal strength (RSS) from several transmitters at some reference points (RPs) into a database, and later for localization of the user, the users device records its own vector of signal strength and matches it against the pre-recorded database of vectors by applying pattern matching algorithms. In this work we deal with deterministic fingerprinting algorithms based on the nearest neighbour (NN) matching algorithm. For this purpose, similarity measurements between the recorded RSSs and the new measurements are necessary, which can be calculated using various signal distances. Manhattan and Euclidean distances are the most common. In this paper, however, responding to the lack of research on the performance of utilizing the other signal distances on indoor positioning, nine different types of similarity measures are investigated. The examined distances are Manhattan, Euclidean, Chebyshev, Canberra, Cosine, Sorensen, Helinger, Chi-square, and Jeffrey. For this purpose, we employ an analytical framework for indoor positioning systems that enables us to analyse the accuracy performance of the system based on the probability of error, here defined as the chance of the distance between the RSS of the test point (TP) and the correct location fingerprint being larger than the distance between the RSS at the TP and an incorrect location fingerprint. Such similarity measures are compared in terms of the probabilities of error and the distance errors that they provide so that we can identify which provides least error. The results are presented for both a simulated environment and real experiments using Wi-Fi and FM (Frequency Modulation) signals.

KEYWORDS: Indoor positioning, fingerprinting, vector distance measure.

1. INTRODUCTION

Indoor positioning has become highly important because of the difficulties that satellite-based technologies such as the Global Positioning System (GPS) experience operating in such areas, e.g. low received signal power and low visibility of satellites. Non-satellite-based technologies, therefore, are important for indoor localization. Utilising signals of opportunity is a viable alternative to GPS due to much higher power levels and wider coverage in indoor environments (Raquet and Martin, 2008).

Many studies have effectively employed wireless networks for indoor localization based on the Received Signal Strength (RSS)-based location fingerprinting technique (Li *et al.*, 2007). Fingerprinting is capable of alleviating some of the problems caused by multipath and Non-Line of Sight (NLOS) propagation in an indoor environment (Li, 2006). Fingerprinting requires a survey of Radio Frequency (RF) signal strength vectors to be made ahead of the system's use for localization. Fingerprinting has two stages: "training" and "positioning". It stores the location dependent characteristics of a signal recorded at Reference Points (RPs) in a database in the training stage, and in the positioning stage, pattern matching algorithms are applied to find the best match between the fingerprint of the user and the database, and eventually estimates the position of the user based on good matches.

The matching methods can be implemented in various ways from a mathematical point of view. They are based on deterministic (Bahl and Padmanabhan, 2000) and probabilistic (Roos *et al.*, 2002) algorithms both of which have been used in Wi-Fi (Li *et al.*, 2007), Frequency Modulated (FM) radio (Moghtadaiee *et al.*, 2011a), and mobile phone networks (Li *et al*, 2005). The measurements of the RSS values at one location can vary considerably, but in deterministic location fingerprinting the average value is stored for the post processing and position determination stage. Nearest Neighbour (NN), K-Nearest Neighbour (KNN), and K-Weighted Nearest Neighbour (KWNN) methods (Bahl and Padmanabhan, 2000) are the most popular matching methods based on deterministic analysis of fingerprinting.

In this work we deal with the simple and popular NN deterministic method which has been broadly used in fingerprinting. In this algorithm, the position of the nearest RP is considered as the estimated position of the user. In order to select this nearest RP, the signal distance measures between the recorded RSS fingerprints at the test point (TP) and the RPs (in the database) are necessary, which can be calculated using various signal distances. Most authors dealing with fingerprinting algorithms use Euclidean distance (Kaemarungsi and Krishnamurthy, 2004), (Lin *et al.*, 2009), (Kjærgaard, 2011). However, there is a lack of research on reasons for choosing one distance measure over the others and no analytical criteria have been proposed for choosing the best. Therefore, without having proper information about the effect of using a distance measure, unknown performance errors can result.

This paper develops an analytical framework for indoor positioning systems that enables us to analyse the accuracy performance of fingerprinting. By employing this framework, we investigate different signal distances so that we can identify which provides least distance error. We will show that results from the simulation agree with the real experiments results.

The rest of the paper is organized as follows. Section 2 presents the related work on the analytical model for indoor fingerprinting systems and on utilising different cases of Minkowski distance between fingerprints. An overview of the indoor location fingerprinting system model and the signal distances used are described in section 3. In section 4 the probability of returning the wrong position using the NN method is explained. The analytical model of indoor location fingerprinting is employed in section 5 and the Monte Carlo simulation is also carried out to obtain numerical results. The relationship between the probability of error (POE) in positioning and mean distance error (MDE) is also analysed. Section 6 shows the results of a real experiment in both Wi-Fi and FM positioning systems. Finally, the conclusions of the work are discussed in section 7.

2. RELATED WORK

There are limited studies that partly analyse and estimate the accuracy and precision performance of indoor positioning techniques based on location fingerprinting (Kaemarungsi and Krishnamurthy, 2004), (Wallbaum, 2007). The authors in (Kaemarungsi and Krishnamurthy, 2004) introduce a model for predicting performance of an indoor positioning system based on the probability of correct positioning which is the chance of the distance between the RSS at TP and the correct location fingerprint being less than the distance between the RSS at TP and the wrong location fingerprint. However, their work is based on Euclidean distance only. They also do not discuss the relationship between the probability of correct positioning and distance error.

Various distance measures have been introduced (Vorst and Zell, 2010), (Machaj and Brida, 2011). However, most researchers dealing with fingerprinting algorithms limit their work to a single distance measure between TP measurements and RP fingerprints stored in a database, such as Manhattan distance (Li *et al*., 2007), (Ching *et al*., 2010), (Kjærgaard, 2007), Euclidean distance (Kaemarungsi and Krishnamurthy, 2004), (Lin *et al*., 2009), (Kjærgaard, 2011). Furthermore, although the significant part of fingerprinting is the matching technique, no criteria have been proposed for choosing the optimal distance measures. Regarding Manhattan and Euclidean distances, authors in (Machaj and Brida, 2011) compare two scenarios of the same and different indoor environment conditions for training and positioning stages of fingerprinting. Their simulation results show that the lowest error is provided by Manhattan and Euclidean in the same and different environment conditions scenarios respectively, but the difference is not significant. This work, however, highlights a framework that helps utilize the most suitable distance for location fingerprinting. The simulation and experimental results are also discussed.

3. INDOOR LOCATION FINGERPRINTING

In this section we describe the modified preliminary model of indoor location fingerprinting described in (Kaemarungsi and Krishnamurthy, 2004).

3.1 System Model

It is assumed there are P APs and R RPs with known locations in a 2D area. The whole area here is divided into R small regions where the centre of each region is a RP. Fingerprinting is carried out for each RP, and a fingerprint vector which consists of the statistical mean of the RSS from each AP is saved in a database. We denote the mean RSS vector of the rth RP by RP^r, and the fingerprint vector associated with that is $\bar{\mathbf{s}}_{RP}^r = [\bar{s}_1^r, \bar{s}_2^r, \dots, \bar{s}_i^r, \dots, \bar{s}_P^r]$, where $\bar{s}_i^r$ is the mean of RSS random variables measured at RP^r from AP^i. Some small regions can be left out as they do not contain any RP.

A TP can be inside any of the small regions in the network. Based on the position of the TP, its nearest RP is denoted by RP^l and other RPs are denoted by $RP^r, r = 1, 2, , \dots, R$ and $r \neq l$. Figure 1 shows the system model described in this section. To find the position of the TP, we first measure the TP vector which consists of samples of the RSS random variables from each AP. The TP vector is denoted as $\mathbf{s} = [s_1, s_2, \dots, s_i, \dots, s_P]$, where s_i contains the RSS random variables measured at TP from AP^i.

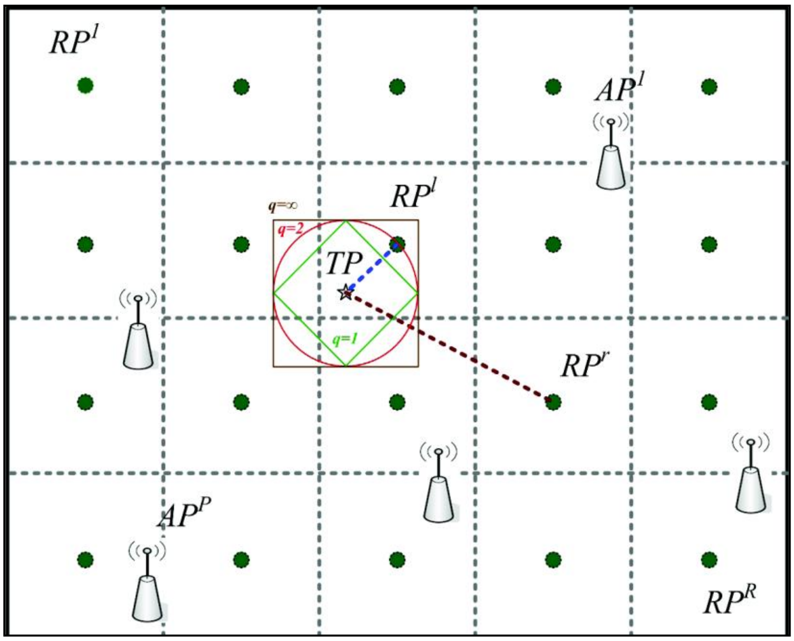

Figure 1. The system model geometry including the RPs and the TP.

3.2 Radio Channel Model

Due to the complexity of signal propagation in indoor environments, and its dependency on various environmental parameters such as multipath, refraction, blockage, and noise, deriving a unique and fixed distribution function for the RSS signals is challenging. Although research is still being conducted to derive an accurate distribution model for RSS random variables in a variety of different scenarios, one of the most common models for the distribution of RSS random variables, that is also empirically validated and is widely used in wireless and positioning literature, is the normal or Gaussian distribution model (Wang *et al.*, 2006), (Chan *et al.*, 2009), (Miliorisl *et al.*, 2011). Therefore, each component in a TP vector, s_i, is assumed to be a normally (Gaussian) distributed random variable i.e, $s_i \sim \mathcal{N}(\bar{s}_i, \sigma^2)$ where $\bar{s}_i$ is the mean RSS vector at TP location (in dBm). It is also assumed that all the components in the TP vector are mutually independent. The standard deviation of all the random variables s_i is assumed to be identical and denoted by σ (in dB). For the radio propagation model, hence, a commonly used large scale model which is the log-distance path loss model with log-normal shadowing is used, similar to the models utilized in (Rappaport, 1996), (Malaney, 2007).

According to this model, the elements of the RSS vector **s** at any point are given by

$$s_i = s_{AP}^i - \left(s_{d_0} + 10\alpha \log\left(\frac{d_i}{d_0}\right)\right) + n_\sigma, \quad i = 1,2,\dots,P \tag{1}$$

where s_{AP}^i is the power transmitting from AP^i, s_{d_0} is the measured path loss at the reference distance, α is path loss exponent, d_i is the geometrical distance from AP^i. The last parameter, $n_\sigma \sim \mathcal{N}(0, \sigma^2)$, is a measurement error vector and assumed to be a zero mean Gaussian random variable with a standard deviation given by σ (in dB). This parameter models the path loss variation at distance d_i due to shadowing caused by obstacles in propagation. All the powers are in dBm and all the distances are in meters. In order for the Gaussian assumption to hold, the impact of fast fading is assumed to be absent in the measurements since it can be removed with averaging filters (Yousef *et al*., 2003), (Goldsmith *et al*., 1994). Although the large-scale log-distance path loss model is not perfect for all cases, it provides means to incorporate fingerprinting into network simulation and analysis which we attempt in section 5.

Once the database of fingerprints of the RPs is formed, deterministic methods of fingerprinting employ real values of the measurements to calculate the signal distance between the points and find the location of the user. We will refer to this distance as the "vector distance (VD)". Such a VD calculation based on several signal distances is described in the following sub-section.

3.3 Various Vector Distances

3.3.1 Minkowski distance

A Minkowski distance between two points, which are TP and a RP RSS vectors, is

$$VD(\mathbf{s}, \bar{\mathbf{s}}_{RP}^r) = \left(\sum_{i=1}^{P} \left|s_i - \bar{s}_i^r\right|^q\right)^{\frac{1}{q}}, \tag{2}$$

where P is the number of APs and q is the norm parameter. It also refers to L_q which is the norm q between two points.

Manhattan distance is a very common distance, known as city block distance, boxcar distance or absolute value distance (Krause, 1986). It is one of the special cases of Minkowski distance when $q = 1$. It represents distance between points in a city road grid. It examines the absolute differences between elements of two vectors. City Block distance is given by

$$L_1 = \sum_{i=1}^{P} \left|s_i - \bar{s}_i^r\right|. \tag{3}$$

Euclidean Distance is the most common use of distance. It is another special case of Minkowski distance when $q = 2$. Euclidean distance represents the shortest distance between two vectors in Cartesian coordinate system (Machaj and Brida, 2011) and is given by

$$L_2 = \sqrt{\sum_{i=1}^{P} \left|s_i - \bar{s}_i^r\right|^2}. \tag{4}$$

Chebyshev distance, minimax metric, or infinity norm between two vectors is the greatest of their absolute magnitude along the vector dimension. It is obtained by putting $q = \infty$ in Minkowski distance. In a system where the calculation time is very crucial, the Chebyshev distance can be a good alternative instead of other distances. This distance is given by

$$L_\infty = \max_i \left| s_i - \bar{s}_i^r \right|, i = 1, 2, \dots, P. \tag{5}$$

3.3.2 Canberra distance

Canberra distance is the weighted version of Manhattan distance in which the weights are the sum of the absolute variable values. Canberra distance expresses the sum of fraction of differences over sum of two vectors. Since it is using the absolute values, the sum of fractions has a value between 0 and 1. If one of coordinates becomes zero, the term turns to unity no matter how much the other value is, thus the distance will not change. It is given by

$$d_{can} = \sum_{i=1}^{P} \frac{|s_i - \bar{s}_i^r|}{|s_i| + |\bar{s}_i^r|}. \tag{6}$$

Note that if both elements become zeros, then we should define $\frac{0}{0} = 0$. When both elements have very small values, this distance is very sensitive to a small change (Machaj and Brida, 2011).

3.3.3 Cosine distance

Cosine distance in fact measures similarity rather than distance or dissimilarity. A higher value of angular separation represents that the two objects are very similar (Machaj and Brida, 2011). Angular separation is defined as

$$d_{ang} = 1 - \frac{\sum_{i=1}^{P} s_i \cdot \bar{s}_i^r}{\left(\sum_{i=1}^{P} {s_i}^2 \cdot \sum_{i=1}^{P} {\bar{s}_i^r}^2\right)^{\frac{1}{2}}}. \tag{7}$$

Similar to cosine, the value of angular separation is [-1, 1]. It is sometimes referred to as Coefficient of Correlation.

3.3.4 Sorensen distance

Sorensen distance is also referred to as Bray Curtis distance. When the summed differences between the variables are normalised by the summed variables of the objects, Bray-Curtis will become a modified version of Manhattan measurement. In the normalised version, in fact the space is assumed to be a grid similar to the city block distance. If all elements of Sorensen distance are positive, its value is between zero and one. Zero Sorensen distance expresses the exact similar vectors (Machaj and Brida, 2011). The Sorensen distance can be calculated as below

$$d_s = \frac{\sum_{i=1}^{P} |s_i - \bar{s}_i^r|}{\sum_{i=1}^{P} (s_i + \bar{s}_i^r)}. \tag{8}$$

Sorensen distance does not define when both vectors are zero. The normalization is done by dividing absolute difference by summation.

3.3.5 Hellinger distance

The Hellinger distance is a criterion that satisfies the triangle inequality. It is in fact the L_2 norm of the difference of the square root vectors (Vorst and Zell, 2010)

$$d_H = \frac{1}{\sqrt{2}} \left\| \sqrt{s_i} - \sqrt{\bar{s}_i^r} \right\|, \tag{9}$$

where $\sqrt{2}$ guarantees that $d_H \leq 1$ for all probability distributions. It is a pretty similar concept to a quantity known as Fidelity or the Bhattacharya coefficient of two probability distributions $d_{BH} = \sum_{i=1}^{P} \sqrt{s_i . \bar{s}_i^r}$ by the relation

$${d_H}^2 = 1 - d_{BH}. \tag{10}$$

3.3.6 Chi-square distance

The Chi-square distance is defined as (Chardy *et al*., 1976)

$$d_{CHI} = \sum_{i=1}^{P} \frac{(s_i - \rho_i)^2}{\rho_i}. \tag{11}$$

where $\rho_i = \frac{s_i + \bar{s}_i^r}{2}$. It should be noted that this distance measure is similar to Euclidean distance but it is weighted by ρ_i.

3.3.7 Jeffrey distance

The Jeffrey divergence is defined as

$$d_{Jeff} = \sum_{i=1}^{P} \left(s_i \log\left(\frac{s_i}{\rho_i}\right) + \bar{s}_i^r \log\left(\frac{\bar{s}_i^r}{\rho_i}\right) \right). \tag{12}$$

4. PROBABILITY OF ERROR (POE) IN POSITIONING

By utilizing the NN method, if the TP is located in the RP^l region (see Figure 1), the correct decision about the position of the TP is made when the VD between TP and RP^l, i.e., $VD\left(\mathbf{s}, \bar{\mathbf{s}}_{RP}^l\right)$ is less than the distance between the TP and the rest of $RP^r, l \neq r$. Otherwise, the wrong position is returned which causes an error in positioning. The NN method could be mathematically written as

$$VD\left(\mathbf{s}, \bar{\mathbf{s}}_{RP}^l\right) < VD(\mathbf{s}, \bar{\mathbf{s}}_{RP}^r), \tag{13}$$

or equivalently

$$Y_r = VD\left(\mathbf{s}, \bar{\mathbf{s}}_{RP}^l\right) - VD(\mathbf{s}, \bar{\mathbf{s}}_{RP}^r) < 0, \tag{14}$$

for all $r = 1,2, \dots, R, l \neq r$.

The probability of correct positioning of the TP between all regions can be written as

$$\begin{aligned}\Pr^{\text{crt}} &= \Pr[Y_r < 0, r = 1,2, \dots, R \text{ and } r \neq l] \\ &= \Pr[Y_1 < 0, Y_2 < 0, \dots, Y_{l-1} < 0, Y_{l+1} < 0, \dots, Y_R < 0,]. \end{aligned} \quad (15)$$

By using the fact that the random variables $Y_r, r = 1,2, \dots, R$ are independent, the probability of correct positioning of TP is written as (Kaemarungsi and Krishnamurthy, 2004)

$$\Pr^{\text{crt}} = \prod_{\substack{r=1 \\ r \neq l}}^{R} \Pr[Y_r < 0]. \quad (16)$$

Thus given the $\Pr^{\text{crt}}$, the POE in positioning of the TP can be written as

$$\begin{aligned}\Pr^{\text{err}} &= 1 - \Pr^{\text{crt}} \\ &= 1 - \prod_{\substack{r=1 \\ r \neq l}}^{R} \int_{-\infty}^{0} f_{Y_r}(y_r) dy_r. \end{aligned} \quad (17)$$

5. SIMULATION AND NUMERICAL RESULTS

The environment layout is illustrated in Figure 2, which is a 6 m x 6 m indoor area including 36 RPs and 20 uniformly distributed TPs. The RP separation is 1 meter. There are four APs at the corners of the environment. Note that the distributions of RPs and APs do not have to be uniform and can be randomly grided. For simplicity we assume the RPs are evenly distributed in the area. The nearest RP for one TP, RP^l, is the RP located at the same grid as that TP, and other RPs are considered as RP^r.

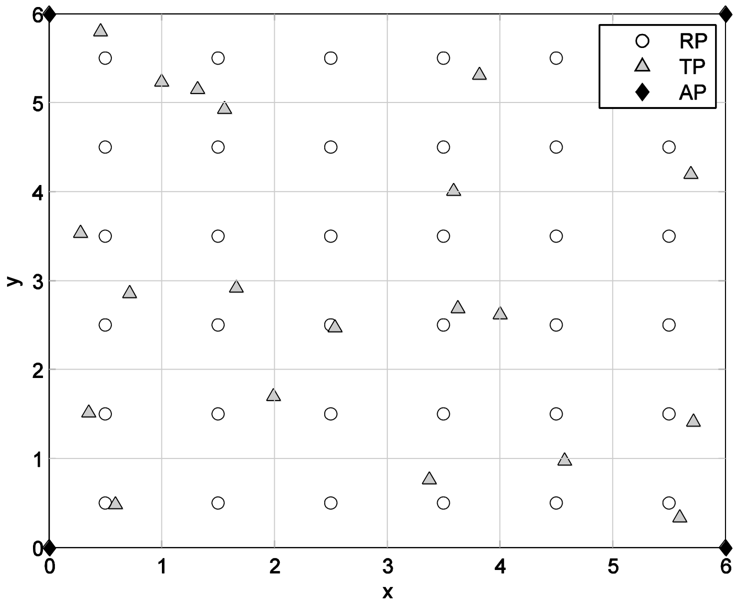

Figure 2. The system model layout.

In order to generate the RSS values at all points to build the database of fingerprinting, (1) is used with this assumption that all APs have the same transmission power. By knowing the values of s_{AP}, s_{d_0}, α and σ, we generated 5000 random Gaussian RSS values for all RPs and TPs. The values used for these four parameters are derived from the measured data in our real experiment presented in section 6, which are 18 dBm, 44.90 dBm, 3.14, and 3.54 dB for s_{AP}, s_{d_0}, α and σ, respectively. Note that the real measurements can also be utilised for RSS values with no changing in the calculations.

Now that the indoor environment is set up and the location fingerprints are logged to the database, the mathematical expression from (17) can be applied to find the POE when different vector distances are employed. As mentioned before, we have 5000 RSS vectors of $\mathbf{s} = [s_1, s_2, s_3, s_4]$ for each TP. Therefore, 5000 VDs between a TP and each RP are calculated based on different vector distances. We have 5000 Y_r for each of which (17) can be applied and the POE of the TP is then achieved. Taking the average over the POE of all TPs gives the average POE for the whole area.

For the same environment, the fingerprinting positioning procedure is carried out to find the MDE so that we could understand the relationship between the POE and distance error. For this purpose, each TP sample is considered as a user measurement vector and the NN algorithm is run to calculate the distance error. Since there are 5000 samples at TP, 5000 distance errors using NN method can be obtained. By taking the average of them, we can have a good estimate of MDE.

5.1 Vector Distance Comparison

This section investigates the impact of various distance measures defined in subsection 3.3 on localisation error when the NN deterministic method is used for the core fingerprint retrieval and assessment task. For the test bed of Figure 2, the NN fingerprinting is carried out to compute the POE and the MDE based on the mentioned VDs as explained in the previous part. Figure 3 illustrates the calculated POEs and the MDEs. The variation of both POEs and MDEs are similar and when the probability of returning the wrong position is minimum, the MDE of fingerprinting is also in its minimum value. This figure indicates that POE and MDE are almost proportional. As can be seen, employing Euclidean distance shows the lowest POE and MDE and Cosine, Helinger, Chi-square, and Jeffrey distances yielded roughly the same accuracy which is close to Euclidean.

Table 1 depicts the complexity and required memory comparison between the VDs. According to this Table, all the VDs have the same complexity and required memory when they are used for positioning except the Chebyshev distance, which has lower complexity than the others since the operations in each measure are different and the ones with square roots will take longer. We denote P as the number of APs, R as Number of RPs, T as number of points we want to localise, B as number of required units (e.g. bytes) to save RSS from each AP, and S as the number of samples for observable APs at each TP. For the all mentioned distances except Chebyshev, the RSS corresponding to all P AP should be first gathered. Thus the total number of RSS registry is $R \times P$. Moreover, for the TP, a total number of S samples are required. For Chebyshev distance, the RSS calculation is independent from the number of access points. In terms of the total amount of the required memory, all the registered information should be gathered for the whole number of TPs.

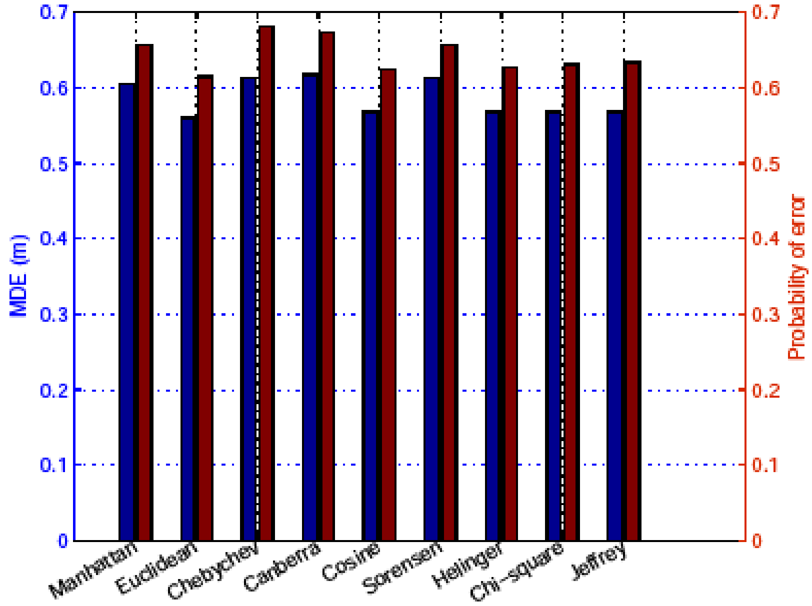

Figure 3. The POE and MDE when different VDs are used in NN method.

Vector Distance	Complexity	Required Memory (Bits)
Chebyshev	$O(R+S)$	$B \times T \times (R \times P + S)$
Others	$O(R \times P + S)$	$B \times T \times (R \times P + S)$

Table 1. Complexity and required memory comparison between the VDs

When designing an indoor fingerprinting system these issues are important. Such comparisons allow the designers to decide based on the computational capacity and power of the devices they use. They should also consider the trade-off between accuracy and complexity of the positioning method. Low complexity is very important especially for the fingerprinting systems localising a high number of users over a vast area.

6. REAL EXPERIMENTS AND RESULTS

Here the experimental fingerprinting measurements from our previous work for Wi-Fi positioning (Li *et al.*, 2007) are used to verify the simulation results. The experimental test bed is the 4th floor of Electrical Engineering building at the University of New South Wales, Sydney, Australia. Its dimensions are 11 m by 23 m, and consists of 7 rooms, which are a typical indoor office environment, and the corridor. In this experiment, there are 119 RPs and 28 TPs. The RPs are distributed as evenly as possible. There are 5 APs available when the Wi-Fi data are collected. Figure 4 illustrates the layout of this experiment and distribution of RPs, TPs, and APs.

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, and the RSS values were logged. To properly take into account the stochastic fluctuations of radio signals, a total of 12 measurements were made at each point and then the average of all the RSS of each AP was calculated, and was inserted into the database as the fingerprint of that location. The data were collected at different time periods, but most of it was logged during the daytime over a few weekends, so that not many people were present. However the environment of the test would be different from that on weekdays, so the results should be more accurate than could be achieved during weekdays. The same method and equipment of RSS collection was used for the TPs. More information on the accuracy of the positioning based on Wi-Fi signals can be also found in (Moghtadaiee *et al*., 2011b).

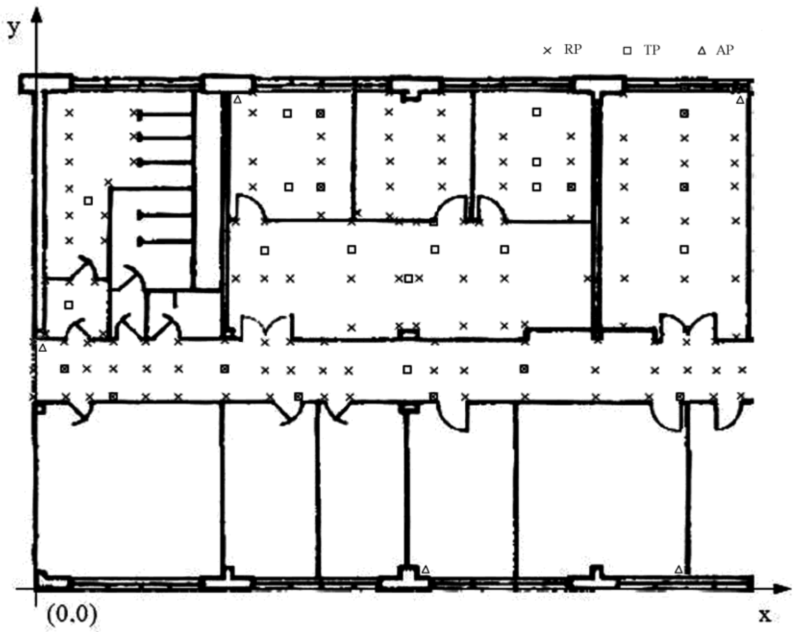

Figure 4. The real experiment layout with APs, RPs, TPs positions.

The NN MDEs obtained when using different VDs in a Wi-Fi positioning system are illustrated in Figure 5. The same result as the simulation results can be seen in the real experiment as well. Among all VDs, Euclidean distance shows the least MDE when NN is carried out, which has been mathematically proven in our previous work (Moghtadaiee *et al*., 2014a). The MDEs obtained from the Helinger, Chi-square and Jeffrey distances are just slightly higher than for Euclidean distance.

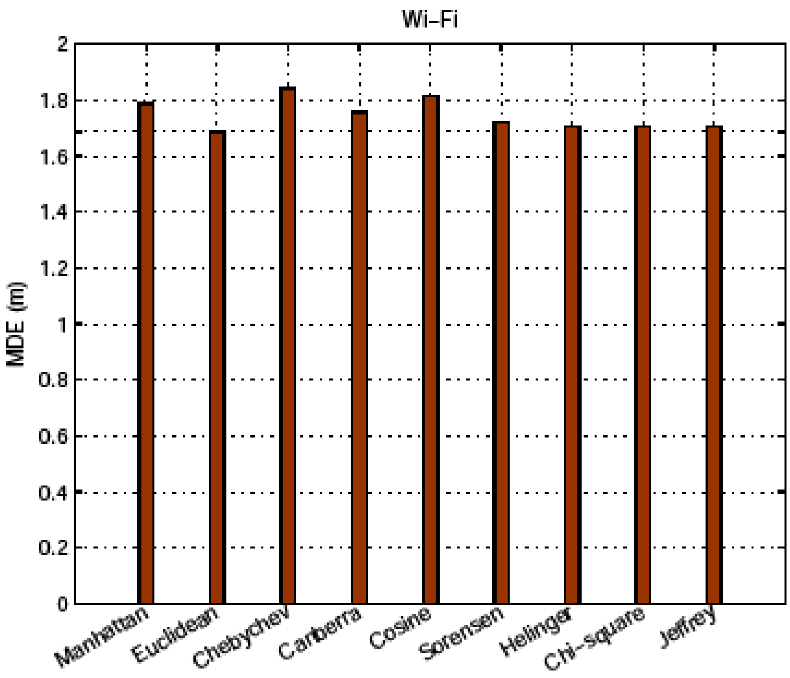

Figure 5. NN MDE in Wi-Fi positioning when different VDs are used.

All the analysis and discussions up to this point were based on the NN method which has low complexity and can be easily used. However, the effect of using various VDs on KNN and KWNN algorithms in a real environment is also of interest. We first examine the Wi-Fi fingerprinting system to find out how MDE is affected by root coefficient in broadly-used Minkowski distance. Figure 6 shows the variation of Minkowski distance for all NN, KNN, and KWNN. The NN MDE reaches the minimum at $q = 2$. Based on this figure, however, the least MDE for Wi-Fi positioning system when the KNN and KWNN methods are used is achieved when $q = 0.5$. It should be noted that when the special cases of Minkowski distance are just considered, Manhattan distance is the best distance measure for the KNN and KWNN in Wi-Fi positioning system.

Figure 7 compares the performance of all three deterministic algorithms under the choices of different VDs. In contrast to the NN fingerprinting system in which utilising Euclidean distance produces the least MDE, positioning results for KNN and KWNN ($K = 4$ based on (Li *et al.*, 2007)) indicates that Manhattan distance achieved the highest accuracy over other distance measures. The Canberra and Sorensen distances in combination with the KNN and KWNN algorithms give almost the same MDE as Manhattan distance. Such findings are needful for a designer of the indoor fingerprinting systems to optimise the systems and get the least positioning error out of the existing fingerprinting data.

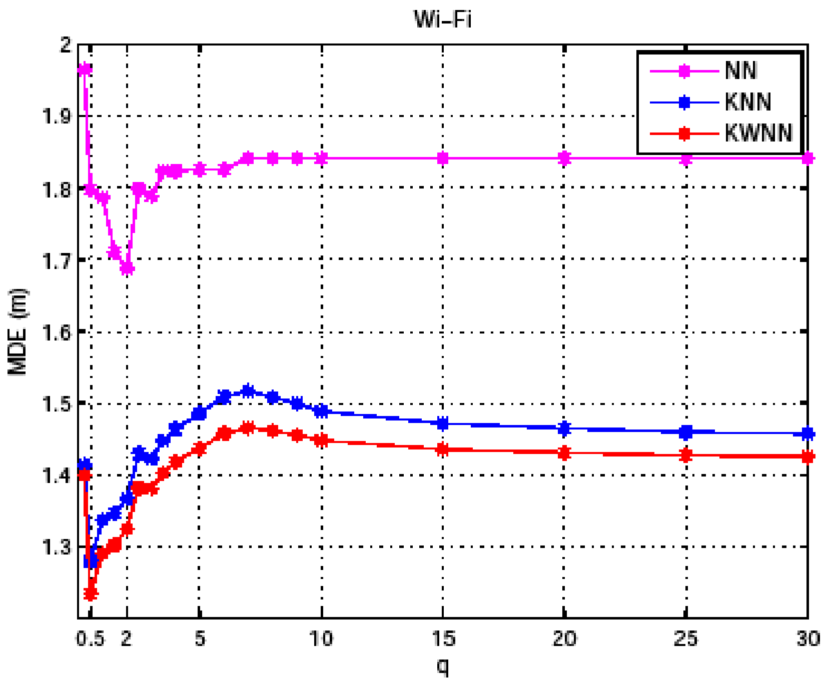

Figure 6. MDEs in the Wi-Fi positioning system using different values of q in Minkowski distance for the choices of NN, KNN, KWNN method.

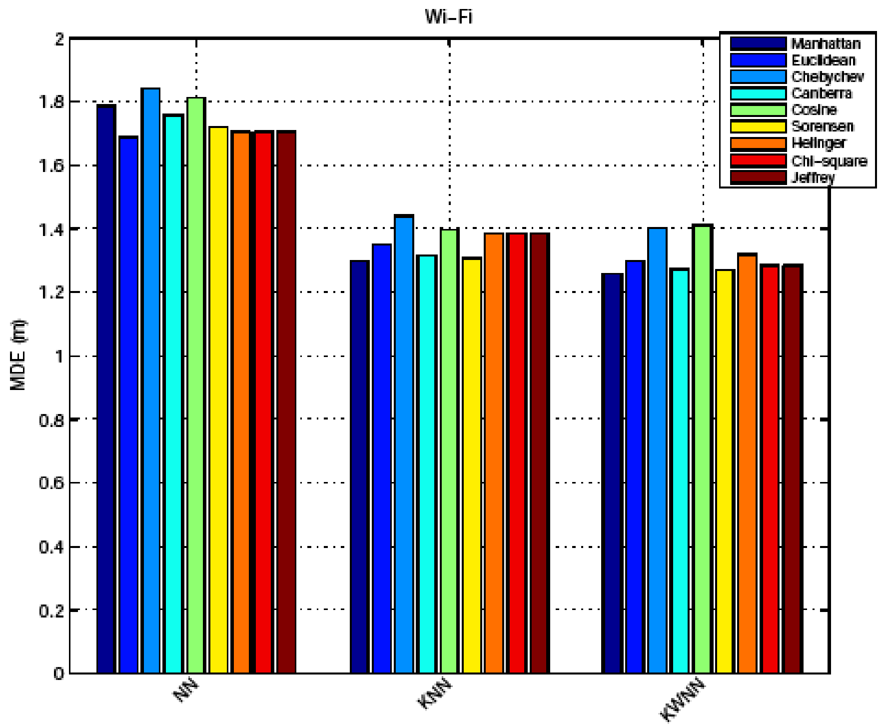

Figure 7. MDEs of the Wi-Fi positioning system for different distance measures, and choices of deterministic fingerprinting algorithms ($K = 4$).

6.1 VDs Influence on FM Positioning

The simulation analysis of this work so far has been for an indoor fingerprinting system where transmitters are close such as Wi-Fi positioning, wheras FM transmitters are much further away. This section examines the behaviour of the proposed FM fingerprinting system which is fully investigated in our previous work (Moghtadaiee *et al.*, 2014b) when all these VDs in combination with NN, KNN, and KWNN algorithms are employed. The FM positioning system includes 150 RPs, 28 TPs, and 17 FM channels.

We first examine the system to find out how Minkowski distance is affected by root coefficient in FM positioning. Figure 8 shows the variation of Minkowski distance for all NN, KNN, and KWNN. It is evident from the figure that, unlike the Wi-Fi case for KNN and KWNN, the least MDE for the FM positioning system is achieved by Euclidean distance for all the NN, KNN, and KWNN methods.

The influence of different distance measures and different deterministic fingerprinting algorithms NN, KNN, KWNN on the FM positioning system are illustrated in Figure 9. The results are calculated based on various sets of RP numbers. These results show that among the special cases of Minkowski distance, Euclidean distance has the least MDE in combination with NN, KNN, and KWNN even for various sets of RPs. The interesting finding of this figure is that almost in all cases of RP sets, the best accuracy is achieved by the Canberra distance. When the RP number is 150, the best MDE was 2.96 m in (Moghtadaiee *et al.*, 2014b), which was given by KWNN ($K = 6$) using Euclidean distance, while the Canberra distance in combination with NN provides 2.44 m localisation error improving the accuracy by 18%. Thus, unlike the Wi-Fi positioning system, FM localisation achieves the lowest MDE by using Canberra distance.

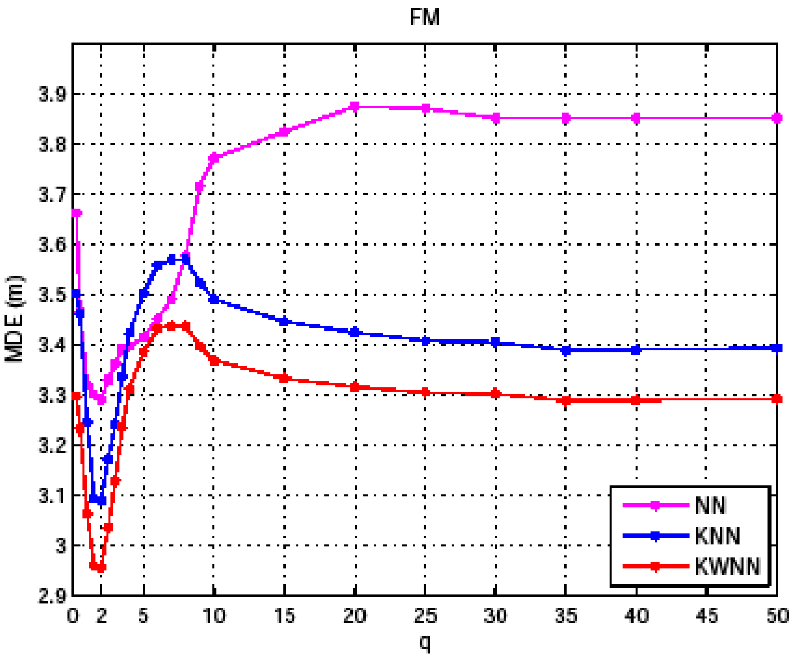

Figure 8. MDEs in the FM positioning system using different values of q in Minkowski distance for the choices of NN, KNN, KWNN method.

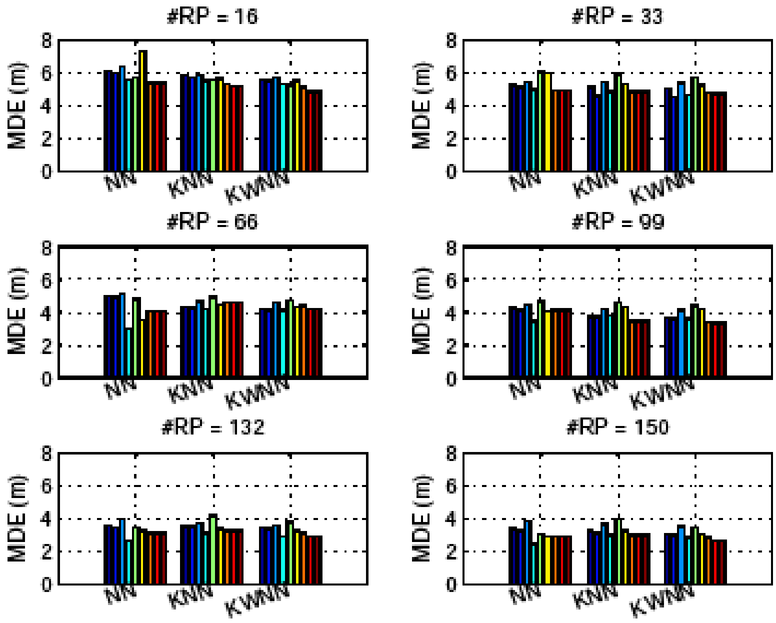

Figure 9. MDEs in the FM positioning system for the different distance measures, choices of deterministic fingerprinting algorithms (NN, KNN, KWNN) for various sets of RP numbers $K = 6$. (definition of the colours are the same as Figure 7)

7. CONCLUSIONS

In this paper the impact of nine different distance measures on probability of error (POE) in positioning and mean distance error (MDE) for NN deterministic method of location fingerprinting is investigated. The relationship between the POE in positioning and distance error is analysed and found to be similar, which has an important role for calculating the accuracy of the positioning systems. The simulation results demonstrate that employing Euclidean distance shows the lowest POE and MDE for the indoor fingerprinting environment. The real test experiments are examined for both Wi-Fi and FM systems based on NN, KNN, and KWNN methods. The results indicate that for a Wi-Fi fingerprinting system, Euclidean distance for NN and Manhattan distance for KNN or KWNN give the least MDE. In FM fingerprinting, however, Canberra distance provides up to 18% higher accuracy than Euclidean distance.

It should be noted that the results prove that the difference between using different VDs are not very significant. However, both accuracy and complexity should be considered by network designers. Depending on the application, a higher accuracy or lower complexity is needful. In some applications, accuracy is critical and even a few centimetres matter, so utilizing the optimum distance measure can play an important role, while to localize a high number of users over a vast area low complexity is essential.

REFERENCES

Bahl P and Padmanabhan VN (2000) RADAR: an in-building RF-based user location and tracking system, *Proceedings of IEEE INFOCOM*, 2: 775-784

Chan E, Baciu G, and Mak SC (2009) Using Wi-Fi signal strength to localize in wireless sensor networks, *Proc. International Conference on Communications and Mobile Computing*, 538–542

Chardy P, Glemarec M and Laurec A (1976) Application of inertia methods to benthic marine ecology: practical implications of the basic options, Estuarine and Coastal Marine Science, 4: 179–205

Ching W, Teh RJ, Li B, and Rizos C (2010) Uniwide Wi-Fi based positioning system, *Proc. IEEE International Symposium on Technology and Society (ISTAS)*, 180–189

Goldsmith A, Greenstein L and Foschini G (1994) Error statistics of real-time power measurements in cellular channels with multipath and shadowing, *IEEE Transactions on Vehicular Technology*, 43(3): 439–446

Kaemarungsi K and Krishnamurthy P (2004) Modeling of indoor positioning systems based on location fingerprinting, *Proc. IEEE Twenty-third Annual Joint Conference of Computer and Communications Societies (INFOCOM)*, 1012–1022.

Kjærgaard MB (2007) A Taxonomy for Radio Location Fingerprinting, *Proceeding of Location-and Context-awareness* 4718: 139-156

Kjærgaard MB (2011) Indoor location fingerprinting with heterogeneous Clients, *Pervasive and Mobile Computing*, 7(1): 31–43

Krause E (1986) *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, New York: Dover Publications Inc.

Li B, Wang Y, Lee HK, Dempster A, Rizos C (2005) A New Method for Yielding a Database of Location Fingerprints in WLAN, *Proceeding of IEE Communications* 152(5): 580-586.

Li B (2006) Terrestrial mobile user positioning using TDOA and fingerprinting techniques, *Ph.D. Thesis*, School of Surveying & Spatial Information Systems, University of New South Wales, Sydney, Australia

Li B, Dempster AG, Barnes J, Rizos C, and Li D (2005) Probabilistic algorithm to support the fingerprinting method for CDMA location, *Proc. Int. Symp. On GPS/GNSS*, Hong Kong, 9C–05

Lin H, Zhang Y, Griss M, and Landa I (2009) WASP: An enhanced indoor locationing algorithm for a congested Wi-Fi environment, *Proc.Mobile Entity Localization and Tracking in GPS-less Environments*, 183–196

Machaj J and Brida P (2011) Performance comparison of similarity measurements for database correlation localization method, *Proc. Intelligent Information and Database Systems*, 452–461

Malaney R (2007) Nuisance parameters and location accuracy in log-normal fading models, *IEEE Transactions on Wireless Communications*, 6(3): 937–947

Miliorisl D, Tzagkarakis G, Jacquet P and Tsakalides P (2011) Indoor positioning in wireless LANs using compressive sensing signal-strength fingerprints, *Proc. 19th EUSIPCO*, Barcelona, Spain, 1776–1780

Moghtadaiee V, Dempster AG and Lim S (2011a) Indoor Localization Using FM Radio Signals: A Fingerprinting Approach, *Proceeding of Indoor Positioning Indoor Navigation IPIN*, Guimarães, Portugal, 1–7

Moghtadaiee V, Dempster AG and Lim S (2011b) Indoor positioning based on FM signals and Wi-Fi signals," *Proc. of IGNSS*, Sydney, Australia

Moghtadaiee V and Dempster AG (2014a) Determining the Best Vector Distance Measure for Use in Location Fingerprinting, Pervasive and Mobile Computing (in press)

Moghtadaiee V and Dempster AG (2014b) Indoor location fingerprinting using FM radio signals, *IEEE Trans. on Broadcasting*, 60(2): 336–346

Rappaport TS (1996) *Wireless Communications: Principles and Practice*, Upper Saddle River, New Jersey: Prentice Hall PTR

Raquet J, Martin RK (2008) Non-GNSS radio frequency navigation, *Proceeding of IEEE ICASSP*, 5308-5311.

Roos T, Myllymaki P, Tirri H, Misikangas P, and Sievanen J (2002) A probabilistic approach to WLAN user location estimation, *Int. Journal Wireless Info. Networks* 9(3): 155–164

Vorst P and Zell A (2010) A comparison of similarity measures for localization with passive RFID fingerprints, *Proc. ISR/ROBOTIK*, Munich, Germany, 1–8

Wallbaum M (2007) A priori error estimates for wireless local area network positioning systems," *Pervasive and Mobile Computing*, 3(5): 560–580

Wang J, Zha H, and Cipolla R (2006) Coarse-to-fine vision-based localization by indexing scale-invariant features," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B: Cybernetics, 36(2): 413–422

Yousef N, Sayed A and Jalloul L (2003) Robust wireless location over fading channels, *IEEE Transactions on Vehicular Technology*, 52(1): 117–126