Abstract: Recently, indoor localisation techniques that use wireless local area network (WLAN) beacons signals have gained much attention by the research communities. Many localisations methods are used to estimate the user of mobile device in indoor environments. However, the accuracy of these methods is affected by the nature of the test-bed environment. In this paper, we introduce an experimental test-bed in a typical indoor environment. We used a fingerprinting-based localisation algorithm to estimate the user location. The fingerprinting technique consists of two phases: offline phase and online phase. In the offline phase, calibration points are collected at certain places in floor to build a radio map. In the online phase, deterministic and probabilistic approaches are applied in order to get the correct estimated position of a mobile device. In deterministic approach, the position of mobile device estimated by K-nearest neighbour (KNN). In probabilistic approach, the position of mobile device estimated by Bayesian network (BN). Clustering technique is proposed to improve the system’s accuracy and reduce the radio map size in the offline phase. We present experimental results that improved the system accuracy and reduce the size of radio map by using the proposed clustering technique.

Keywords: Bayesian network; deterministic approach; K-nearest neighbour; KNN; probabilistic approach; RF fingerprinting; clustering technique.


Biographical notes: Abdulraqb Alhammadi received his Bachelor degree in Electronics Engineering majoring in Telecommunication from Multimedia University, Melaka, Malaysia in 2011. He is currently working toward the Master degree with the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. His research interests include localisation systems and indoor propagation waves in wireless communication.
Mohamad Yusoff Alias obtained the Bachelor of Science in Engineering (Electrical Engineering) degree from the University of Michigan, Ann Arbor, in May 1998. He then received his PhD degree in December 2004 from the School of ECS, University of Southampton, UK. He is currently a Professor at the Faculty of Engineering, Multimedia University in Malaysia. His research interests cover the field of wireless communications especially in OFDM, multiple antenna system, multiuser detection, genetic algorithms in communications, multimedia applications and visible light communications.

Su-Wei Tan holds BEng (Electrical), MEngSc and PhD degrees from University of Malaya (Malaysia), Multimedia University (Malaysia) and University of Kent (UK) respectively. His research interests include overlay networks, channel assignment and routing and indoor localisation in wireless communications.

Chamal Sapumohotti obtained his MSc from Multimedia University where he researched on indoor localisation systems based on WLAN fingerprints. He has a keen interest on machine learning techniques and telecommunication technology.

This article is a revised and expanded version of a paper entitled ‘Enhanced indoor localisation system for multi-floor environment using clustering techniques’ presented at International Conference of Recent Trends in Information and Communication Technology 2014, UTM, Johor, Malaysia, 12–13 September 2014.

1 Introduction

The global positioning system (GPS) is a great tool for accurate localisation in outdoor environments. However, GPS does not work well in certain places such as indoor environment and urban canyons due to low received signal strength (RSS). Tracking the user of mobile device in indoor environments has given considerable attention in recent works. In past years, many localisation systems are introduced such as GSM-based on indoor localisation (Otsason et al., 2005), infrared-based systems (Want et al., 1992), Bluetooth (Aalto et al., 2004), radio frequency identification (RFID), (Hightower and Borriello, 2000) and ultrasonic-based systems (Priyantha et al., 2000). Most of these systems have different wireless transmission technology. Moreover, some of the localisation systems required expensive equipment such as RFID and GSM. On the other hand, there are simple indoor localisation systems which depend on the availability of RSS in the indoor environment such as RADAR (Bahl and Padmanabhan, 2000), Horus (Youssef and Agrawala, 2008).

Recently, most of indoor localisation techniques used wireless local area networks (WLANs) for indoor environment. WLAN based on IEEE 802.11 standard has been considered a suitable solution because of its widespread availability in indoor environments. Moreover, it is deployed in many indoor environments such as malls, hospital and companies. The advantage of using WLAN for localisation system is very simple and it does not need any additional equipment to be deployed with the mobile device in order to receive the RSS. However, the RSS degrades by different types of
obstacles and people movements as well. The fluctuation of RSS is slightly affected the system’s accuracy.

There are many techniques that can be used for localisation system such as proximity sensing, triangulation and fingerprinting (Farid et al., 2013). In this paper, we use fingerprinting technique because it is very simple and has less error distance, which is divided into two stages namely offline and online stage. There are two existing methods of fingerprinting techniques which are deterministic and probabilistic approaches. In probabilistic method, presented by Mahtab et al. (2007) the location can be estimated by the probability distribution function of RSS. However, deterministic method which is proposed by Bahl and Padmanabhan (2000) estimated the position based on the averaged of RSS. K-nearest neighbour (KNN) is one of the algorithms of deterministic method which estimated the mobile user location by using Euclidean distance algorithm (Moghtadaiee et al., 2011).

This paper proposes a clustering technique that can be used to reduce the size of radio map in the offline phase. Our clustering technique is called signal space clustering which depends on the signal space of the RSS. The rest of this paper is organised as follows. In Section 2, we provide the related works that have been done in indoor localisation system in the past years. Section 3 describes the localisation approaches which used to estimate the user of mobile device in the indoor environment. In Section 4, we provide overview of clustering techniques and present the details of the proposed clustering algorithm. Section 5, we describe the experimental test-bed and the obtained results from the localisation approaches. Finally, Section 6 concludes the paper and provides recommendations for future work.

2 Related work

RADAR (Bahl and Padmanabhan, 2000) is considered the first indoor positioning system used nearest neighbour in signal space (NNSS) algorithm in order to predict the user’s location for single floor. Later, the NNSS is improved by Tarn et al. (2006) to enhance the system’s accuracy by using probabilistic approach. NNSS works as searching or matching centre in the online phase to compare the current RSS with RSS stored in radio map and chose the matched RSS. The matching is done by using metric technique such as Euclidean distance. The estimated location accuracy is affected by the size of reference points in online phase. The radio map of RADAR contains four parameters which are: \((x, y)\) represents the coordinate of the estimated location, \(t\) represents the timestamp and user’s direction, and \(d\) represents the orientation of the user. RADAR system is adapted for single floor environment since the floor attenuation factor is neglected which is introduced by Seidel and Rappaport (1992). However, the system’s accuracy is affected by the size of the radio map and the samples size. The accuracy of user’s location is about two to three metres.

The authors Li et al. (2006) used KNN algorithm in order to predict the user’s location for the single floor environment. KNN works as searching or matching centre in the online phase to compare the current RSS with RSS stored in radio map and chose the matched RSS. The matching is done by using metric technique such as Euclidean distance. The estimated location accuracy is affected by the size of reference points and k-value. In Lakshmanan and Tomar (2014) and Sun et al. (2015) used particle swarm algorithm (PSO) to improve the localisation accuracy in wireless sensor networks.
Some of the indoor localisation systems based on the probabilistic approach. One of these systems introduced by Al-Ahmadi et al. (2010), the authors designed indoor localisation system by using Bayesian graphical model. This model depends on the set of conditional independence relationships. The first model of Bayesian network (BN) is used by Madigan et al. (2005), which was just a simple model to estimate the user’s location for single floor. Many sampling techniques are used by using collected data. One of the most important sample technique used is a Markov chain Monte Carlo (MCMC). This technique works as a sample generator to generate a huge number of samples from the posterior distribution, so that these numbers of sampling are used to estimate the mean of the posterior. Over-relaxation MCMC approach was introduced by Neal (1998), Adler (1981) and Green and Han (1990) which used to reduce the high autocorrelations and improving the convergence of Gibbs sampler. This can be obtained by generating multiple set of random values and choosing the one that is negatively correlated with the values of the current iteration (Ntzoufras, 2009).

Horus system (Youssef and Agrawala, 2008) is based on the probabilistic method using WLAN for localisation system. The system is developed to improve the main objective of developing or designing a localisation system for indoor environment such as accuracy and computational cost. Horus used Joint clustering technique to reduce the computational cost of the system. The main idea of this clustering technique is to group the location points that have common properties of AP.

3 Localisation algorithms

The localisation algorithms are used in many systems or models to estimate the user’s location at any certain place. The classification of localisation algorithms is divided into subcategories which are: deterministic approach and probabilistic approach (Kolodziej and Hjelm, 2006).

3.1 Deterministic location approach

The location system collects samples of signal strength from each AP and presents that as a single value of samples mean in radio map. One of the deterministic methods uses in localisation system is NNSS which defined as the smallest value of Euclidean distance metric. In localisation system, the Euclidean distance is applied to get the smallest value of matching the current location with the singles stored in radio map. The first system used the deterministic location method is RADAR (Bahl and Padmanabhan, 2000) which based on NNSS.

To implement this system, two phases should be taken, namely the offline and online phase. In the offline phase, the calibration points should be collected for certain positions in the building. These calibration points contain information such as RSS from each access point (AP) and \((x, y)\) coordinates. The collected data in this phase will be stored as a radio map which will be used in the next phase. In the online phase, the current RSS compared with the RSS values that have been stored previously in the offline phase and get the matched value and then estimate the mobile’s user. The mobile’s user location can be estimated by using \(k\) nearest neighbour which is stated by Euclidean distance in equation (1):
An enhanced localisation system for indoor environment

\[
D = \sqrt{\sum_{i=1}^{N} (s_o - s_{ci})^2}
\]  

(1)

where \(s_o\) is RSS which is stored in radio map, \(s_{ci}\) is current RSS which is measured in online phase and \(N\) is the number of APs.

### 3.2 Probabilistic location approach

The idea behind the probabilistic location method is to calculate the conditional probability distribution of current location \(l\) given a vector of \(s\) in the online phase. The posterior probability can be calculated by using Bayes’ rule:

\[
p(l|s) \frac{p(s|l) p(l)}{p(s)}
\]

(2)

where \(p(l)\) is the prior probability of previous information of location \(l\) before knowing any information about RSS. \(p(s|l)\) is the likelihood function of obtaining variable is given the information of variable \(l\). \(p(l|s)\) is the posterior distribution of estimated location \(l\) with known vector of signal \(s\).

The posterior of conditional probability is the multiplication of prior probability and the likelihood function

\[
posterior \propto prior \times likelihood
\]

**Figure 1**  Single BN

Single floor Bayesian system is introduced by Madigan et al. (2005) considered the first indoor localisation system using BN with applying MCMC algorithm. This system based on probabilistic approach using Bayesian hierarchical model. It used Gibbs sampling to sample from the posterior distribution. Figure 1 shows that the proposed system of BN
for the single floor environment which contains of nodes linked by arrows. \(X\) and \(Y\) nodes represent the test-bed’s dimension while the node \(D_i\) represents the distance between the coordinates of user’s location and the location of \(AP_i\). The distance can be obtained by using Euclidean distance equation. \(S_i\) represents the RSS at each particular point in the test-bed. It is normally distributed with mean equal to the regression model of independent variables \(b_{00}\) and \(b_{11}\) variance, \(\tau_i\). The nodes of the system are defined as follows:

\[
X \sim \text{Uniform} (0, L) \\
Y \sim \text{Uniform} (0, W) \\
S_i \sim N \left( b_{00} + b_{11} \log D_i, \tau_i \right) \quad \text{for} \quad i = 1, \ldots, D \\
b_{00} \sim N(0,0.001), \quad \tau_0 \sim \text{Gamma} (0.001, 0.001) \\
b_{11} \sim N(0,0.001), \quad \tau_1 \sim \text{Gamma} (0.001, 0.001)
\]

The single floor Bayesian has achieved accuracy with 4.3 metre by using a large number of training points about 253, but this drawback of large number training points is improved by Madigan et al. (2006) with same system’s accuracy.

4 Clustering techniques

There are many clustering techniques are used for improving or enhancing the system’s output and reduce the consuming time by reducing the complexity of computational overhead. Moreover, it has an advantage of reduction of number of calibration points results in smaller size of radio map. The cluster technique in localisation is defined as a set of calibration points sharing common properties of APs. In Youssef et al. (2003) the clustering technique is called joint clustering technique. The calibration points have divided into small groups which are sharing the same properties of RSS. Each of this group or cluster one point is selected to present its cluster, which has a maximum mean of RSS.

K-means clustering algorithm is proposed by Tarn et al. (2006) which is the most popular technique used in data mining. The algorithm starts to initiate the \(K\) point to be centroid point randomly. Then all the points nearby selected \(k\) point are tested by Euclidean distance equation between each point and \(K\) point (Sidong and Tong, 2013). Furthermore, spatial aware signal space clustering (S3C) is another clustering technique used by Sapumohottti et al. (2013) to minimise the number of calibration points in offline phase. Increasing in calibration points results in large size of the database (radio map). S3C clustering is applied to minimise the size of radio map and improve the location error. S3C algorithm used the cluster margin to control the number of calibration points in radio map, but this clustering reduced the huge number of calibration points which effects on the system’s accuracy. In this paper, we proposed a clustering technique which called space signal clustering algorithm.
4.1 Proposed clustering technique

This technique is used to overcome the big size of radio map in offline phase. The uniform radio map contains a large number of calibration points resulting in increased size of radio map. In order to reduce the number of calibration points combination clustering is used. This clustering technique allows combining two adjacent points together in radio map before performing the online phase. It depends on the signal space distance between two points in radio map. The metric used to calculate the distance is given by Manhattan distance in equation (3).

\[ d_{n,n+1} = \left| P - P_{n+1} \right| \]  

(3)

\[ d_{n,n+1} = \sum_{i=1}^{M} \left| S_i - S_{i,(j+1)} \right| \]  

(4)

where \( M \) is the number of APs and \( S_j \) is the RSS at point \( j^{th} \) measured from \( AP^n \).

**Figure 2** Signal space clustering algorithm

Figure 2 shows that there are two distances are calculated between three points \( P_n, P_{n+1} \) and \( P_{n+2} \). First and second distant are calculated between \( (P_n, P_{n+1}) \) and \( (P_{n+1}, P_{n+2}) \) respectively by using equation (5). Then both distances are compared together to observe the smallest distance if \( d_{n,n+1} \) is less than \( d_{n+1,n+2} \) then \( P_n, P_{n+1} \) are combined together by obtaining their RSS mean, otherwise the next point needs to be checked as so on.

\[ S_j = \frac{1}{2} \sum_{i=1}^{m} \left| S_i + S_{i+1} \right| \]  

(5)

where \( S_j \) is the new signal space of point \( P_j \) located at \( x_q, y_q \).

The new point will have a new coordination as well. The new coordination is calculated in equation (6)

\[ p(x_q, y_q) = \left( \frac{x_i + x_{i+1}}{2}, \frac{y_i + y_{i+1}}{2} \right) \]  

(6)

where \((x_i, y_i)\) and \((x_{i+1}, y_{i+1})\) the point coordination at \( P_i \) and \( P_{i+1} \), respectively.

Figure 3 shows the flowchart of signal space clustering algorithm that used to minimise the calibration points in each cluster in radio map. \( N \) indicates the number of points in radio map, three points are taken at each computation of cluster in radio map and last two points are not considered since the distance calculation required at least three points.
5 Experimental test-bed

We have conducted measurement experiments on the ground floor (wing B) of Faculty of Engineering, Multimedia University. The area has a dimension of $52 \times 22$ m$^2$. Eight Linksys-Cisco-WRT54G2-802.11b/g APs with operating frequency of 2.4GHz were used with four APs placed on ground floor. Figure 4 shows the locations of the APs that have chosen according to pervious work (Baala et al., 2009) which the APs placed at many different locations in the building. Then, the best location of APs that the signal coverage can cover most of the areas on the floor is chosen to evaluate the user’s location. The
WirelessMon (WiFi scanner.wireless Mon, http://www.passmark.com/) software is used to collect offline data. WirelessMon is a WiFi signal scanner software and it is able to provide information such as RSS, MAC address, time, service set identifier (SSID) and channel used by the AP. At each calibration point, we collect 30 samples in 360 degree rotation with one second time interval. Once all needed information is obtained, a radio map is created to be used in the online phase. The radio map is created by collecting RSS for each coordinate of calibration point on the ground floor as shown in Table 1. For our experiment, 50 calibration points are collected in order to build up the radio map.

**Table 1** Sample of radio map

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>AP1</th>
<th>AP2</th>
<th>AP3</th>
<th>AP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>7</td>
<td>-89</td>
<td>-78.4</td>
<td>-56.25</td>
<td>-85.75</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>-84.35</td>
<td>-58.15</td>
<td>-80</td>
<td>-90</td>
</tr>
</tbody>
</table>

**Figure 4** The floor plan of ground floor (wing b) at faculty of engineering, multimedia university (see online version for colours)

**Figure 5** RSS mean for four APs along the corridor on the ground floor (see online version for colours)
RSS were recorded of four APs along the corridor on the ground floor. Figure 5 shows the mean of RSS of 50 locations over distance. It is observed that the RSS mean is higher when the location is close to APs while the weaker RSS mean when the location is far away from APs. The fingerprint at each location has a unique RSS mean in which make the estimation of the user’s location in indoor environments by using fingerprint technique is the best choice.

5.1 Deterministic approach

In this approach, we used 30 points as uniform radio map in the offline phase. In the online phase, the user location was estimated by using \textit{KNN} algorithm when \((K = 1)\). Then, the clustering technique was used to reduce the radio map size without affecting the system’s accuracy. The uniform radio map was reduced the calibration points from 30 to 15 points by using the proposed clustering technique. On the other hand, the clustering technique reduced the uniform radio map size by 50%. Figure 6 shows the cumulative distribution function of uniform and clustered radio map in \textit{KNN}. The number of calibration points in uniform and clustered radio map is 30 and 15 points respectively. The mean of distance error was calculated by using Euclidean distance. Furthermore, it can be observed that the mean of distance error of uniform radio map is 7.3 metre while the clustered radio map is 6.9 metre. Despite the system accuracy was not significantly improved but the size of radio map was reduced by 50% without affecting the system’s accuracy.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cumulative_distribution.png}
\caption{Cumulative distribution function for the uniform clustered radio map (see online version for colours)}
\end{figure}

It can also be observed that the optimum distance error when the clustering is used has a small value of a distance error (6.9 m). The small distance error results in a better accuracy of location estimation. On the other hand, the high distance error was obtained by using the uniform radio map while the best accurate location is obtained by the clustering technique.
5.2 Probabilistic approach

In this approach, the same number of the calibration points was implemented by using Madigan model that has been discussed in Section 2.3. The location of mobile user was estimated by using OpenBUGS (Bayesian inference using Gibbs sampling) program which is based on BN using Gibbs sampling. This program runs on the computer with windows operating system. The user’s location \((X_i, Y_i)\) that needs to be estimated is indicated by \(NA\) in the radio map as shown in Figure 7. Table 2 shows the specifications of the system’s parameter that used for this work.

**Figure 7** Radio map in OpenBUGS with unknown location (see online version for colours)

![Radio map in OpenBUGS with unknown location](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of chain</td>
<td>1</td>
</tr>
<tr>
<td>Burn-in samples</td>
<td>10,000</td>
</tr>
<tr>
<td>Updates(no. of iterations)</td>
<td>100,000</td>
</tr>
<tr>
<td>Refresh</td>
<td>100</td>
</tr>
<tr>
<td>Thin</td>
<td>1</td>
</tr>
<tr>
<td>Over-relax</td>
<td>Yes</td>
</tr>
<tr>
<td>Estimated nodes</td>
<td>X, Y</td>
</tr>
</tbody>
</table>

**Figure 8** Summary of the estimated location (see online version for colours)

![Summary of the estimated location](image)
The average of system accuracy was calculated after 30 testing points were estimated. Figure 8 shows the mean and standard deviation of the error distance for $X$ and $Y$ coordinates with the 25th and the 75th percentile values. It can be observed that the estimated location is $(X = 18.93$ and $Y = 11.33)$ where the actual location is $(X = 21$ and $Y = 10)$. Thus, the distance error of the estimated location is 2.5 metre.

**Figure 9** Kernel density of estimated variables $(X, Y)$ (see online version for colours)

Figure 9 shows the kernel density of estimated variables by using the posterior probability function of two nodes $X[1]$, and $Y[1]$. The bell-shaped by posterior distribution showed that the Markov chain (MC) is reaching the convergence level. According to Al-Ahmadi et al. (2010) the best estimated value when the MC run up to 100,000 iterations with 10,000 burn-in samples.

**Figure 10** Cumulative distribution function for the uniform clustered radio map (see online version for colours)

Figure 10 shows the cumulative distribution function for the uniform and clustered radio map. The clustered radio map reduced the size of radio map from 30 to 15 points. Furthermore, the accuracy of the system was improved by minimising the mean distance error.
Figure 11  Cumulative distribution function for the uniform and clustered (KNN and BN) radio map (see online version for colours)

The uniform radio map was compared with clustered radio map of deterministic approach (KNN) and probabilistic approach (BN). Figure 11 shows the cumulative distribution function for the uniform and clustered (KNN and BN) radio map in localisation system. The both clustered radio map (KNN and BN) are significantly reduced and improved the performance of localisation system compare to the uniform radio map. Moreover, the clustered radio map using BN achieves a good accuracy better than clustered radio map using KNN.

Figure 12  Comparison between uniform and clustered radio map in deterministic and probabilistic approaches
Table 3 and Figure 12 show the effect of using uniform and clustered radio map in deterministic (KNN) and probabilistic (BN) approach. In deterministic approach, the uniform radio map has achieved system accuracy of 7.3 metre with the use of 30 calibration points. A better result has been achieved when the clustered radio map was used giving an average accuracy of 6.9 metre with only 15 calibration points used. Although the improvement of the average accuracy using clustered radio map is slightly small, the size of the radio map was reduced by 50%. In the other hand, the results in probabilistic approach have shown that the uniform radio map has achieved system accuracy of 4.2 metre with the use of 30 calibration points.

6 Conclusions

Localisation systems have been widely used in indoor environments. However, it is still facing localisation accuracy issues. In this paper, we have conducted experiments in indoor environment for estimating the mobile user location by using deterministic and probabilistic approach. These approaches were carried out based on KNN algorithm and BN. The mean of distance error has been computed and evaluated for each approach in the indoor environment. The results have shown the proposed clustering technique has significantly reduced the size of radio map from 30 to 15 calibration points for both approaches. In the other hand, the accuracy in deterministic approach was slightly improved from 7.3 to 6.9 metre while the probabilistic approach has achieved a better average accuracy of 2.9 metre. In future work, this work can be developed to include multi-floor environments. In addition, the floor attenuation factor needs to be investigated and computed in typical multi-floor environments.

References


An enhanced localisation system for indoor environment


