Fuzzy-based Wi-Fi localisation with high accuracy using fingerprinting

Michael Ibrahim, Tarek Nabil, Hassan H. Halawa, Hany M. ElSayed, Ramez M. Daoud*, Hassanein H. Amer, Mohamed Ismail, Mostafa Wahish and Mohamed Ibrahim

Electronics and Communications Engineering Department, The American University in Cairo, P.O. Box 74, AUC Avenue, New Cairo 11835, Egypt
Email: mmibrahim@aucegypt.edu
Email: tarek.n@aucegypt.edu
Email: hassan@aucegypt.edu
Email: helsayed@ieee.org
Email: rdaoud@ieee.org
Email: hamer@aucegypt.edu
Email: mt.22@aucegypt.edu
Email: mostafa_13@aucegypt.edu
Email: m_baha2@aucegypt.edu
*Corresponding author

Abstract: In this paper, a localisation system based on Wi-Fi fingerprinting and fuzzy data analysis is presented. Three localisation techniques were used, Euclidean distance, K-nearest neighbours (KNN), and weighted K-nearest neighbours (WKNN), to get three independent estimations of a user’s location. Then fuzzy analysis is used to combine the three estimates to achieve highly-accurate localisation. Two experiments were conducted in order to test the proposed new technique and compare it to the traditional fingerprinting techniques present in the literature. The results of the experiments proved that the proposed technique outperforms the traditional techniques.

Keywords: localisation; fingerprinting; fuzzy; Euclidean distance; K-nearest neighbours; KNN; weighted K-nearest neighbours; WKNN.


Biographical notes: Michael Ibrahim received his BSc in Electrical and Communications Engineering from AUC, the American University in Cairo, Egypt, in 2015, with highest honours, in addition to minors in computer science and visual arts from the same university.
Tarek Nabil is a Senior .net and Cross-Platform Software Developer. He graduated Magna Cum Laude from the American University in Cairo in 2015 with BSc in Electronics and Communications Engineering, and minors in computer science and mathematics. His research interests are automatic control, home automation, as well as web design and mobile development.

Hassan H. Halawa received dual BSc degrees (2013) and MSc degree (2015) from the American University in Cairo. His research interests include networked control systems, wireless sensor networks, wireless body area networks, fault-tolerance and reliability modelling.

Hany M. Elsayed received his BSc (1980), MSc (1984) and PhD (1988) degrees from Cairo University in Electronics and Communications Engineering. He worked as a Visiting Researcher in LAAS, France from 1984 to 1987. He held several visiting professor and consultant positions in Egypt and abroad. He is an IEEE member since 1989 and is currently the Chair of the IEEE Control systems chapter in Egypt. His research interests include wireless sensor networks, networked control systems and real-time embedded control.

Ramez M. Daoud received his BSc and MSc in Control Engineering in 2000 and 2004 respectively from Cairo University, Egypt and his DSc in Control and Industrial Informatics with High Honors from Universite de Valenciennes, France in 2008. He received the Best Industrial Application Certificate in 2005 from the IEEE SMC. He is a founding member of the SEAD research group. He is a member of the IEEE IES NBCS, the IEEE IES TCFA-FT. His research interests include digital circuit design, wireless communications, automotive control, vehicular technology, car automation, real-time systems, networked control systems, and industrial informatics.

Hassanein H. Amer is a Professor in the Electronics Engineering Department, American University, Cairo, Egypt. He is the Founding Chair of this department. He received his BSc in Electronics Engineering from Cairo University in 1978 and his MSc and PhD degrees in Electrical Engineering from Stanford University, CA USA in 1983 and 1987 respectively. He founded the SEAD research group in 2003. His research interests include the reliability and testing of digital and mixed-signal circuits, reliability modelling, fault modelling in VLSI, self-checking circuits, networked control systems and wireless sensor networks.

Mohamed Tarek Ismail is a Software Development Engineer at Mentor Graphics Egypt. He received his BSc in Electronics and Communications Engineering (and minors in Computer Science and Economics) with highest honours from the American University in Cairo in 2015. His research interests are machine learning, Internet of Things, and embedded systems.

Mostafa Wahish is an Embedded Software Engineer. He received his BSc in Electronics and Communications Engineering from the American University in Cairo in 2015, with honours. His research interest is in automotive embedded systems and Internet of Things.

Mohamed Bahaa Ibrahim is a Business Intelligence Analyst. He received his BSc in Electronics and Communications Engineering from the American University in Cairo in 2015. His research interests are artificial intelligence and Internet of Things.
1 Introduction

There are currently many applications and location-aware services that depend on knowing users’ locations, in order to allow users to control the surrounding devices automatically. In order to supply such services with accurate user location information, different localisation techniques were developed (Jiang, 2010; Kuntal and Karmakar, 2014; Le Dortz et al., 2012; Sabbour, 2007; Eisa et al., 2013; Sun et al., 2014; Socha et al., 2015; Moreno et al., 2016). Many of these techniques were implemented for outdoor spaces, such as triangulation used in global positioning system (GPS) and GLObal NAvigation satellite system (GLONASS). These systems are being deployed in current applications, such as navigation systems, military applications, monitoring geological activities, and smart advertisement. Although such systems perform sufficiently in outdoor environments, they do not perform well in indoor areas. This is mainly for two reasons: first, signal reception for such systems is often too weak indoors; second, accuracy required in indoor environments is typically harder to achieve than outdoors. For example, a two to three metre range of error is definitely enough outdoors for navigation systems to detect which road the user is on. The same error, however, is not acceptable when it comes to indoor environments, as it might locate the user in a different room than the one he is actually in (Boonsriwai and Apavatjrut, 2013).

In order to reach the level of accuracy required for indoor applications, many systems rely on sensors, or Radio Frequency IDentification (RFID) technologies. The disadvantage of these techniques is the additional hardware that needs to be installed into the system before it can be used. For example, if a sensor is used to detect the users in one room, then at least one sensor per room is needed in order to locate the users in the entire system. A work environment with thousands of offices will result in a huge number of sensors that are expensive, difficult to manage and tiresome to install.

A different approach is to analyse the signal strength received by the users and calculate the location accordingly. There are many techniques that carry out such computations using users’ received signal strength (RSS), among which are the most commonly known methods, triangulation and fingerprinting.

Triangulation requires at least three detected wireless signals whose Access Points’ (APs) locations are known. The idea behind triangulation is that the user receives three signals from three different directions; each signal takes some time \( t \) to propagate with velocity \( v \). If the speed of the wave is defined in the surrounding medium and the time of arrival of the signal to the user is calculated accurately, the distance \( d \) should be easily calculated. The result of each signal is a circular area surrounding the receiving party, whose radius is the distance between the transmitter and the receiver. By knowing the distances between the user and three access points whose locations are known, the location of the user can be estimated by having the intersection of the three circles which yields the final location required.

On the other hand, fingerprinting consists of two phases, an offline training phase where the RSS is measured in different locations across the localisation area and stored in a database, and an online phase where the user is actually located. The RSS sent by the user in the online phase is compared to the saved entries in the database to estimate which location has the closest value to the offline data. The location with the best match is considered to be the location of the user.
The main advantage of the triangulation method over fingerprinting is that it does not require a training phase which could be tedious in large areas. However, it has the disadvantage of being generally less accurate than fingerprinting due to variations in the environment, obstacles, reflections and refractions that create multi-path effects that affect the received signal.

In this paper, a fuzzy-based localisation system is presented based on fingerprinting Wi-Fi signals to locate devices with a high accuracy, suitable for indoor environments. The proposed system does not require the installation of any extra hardware, and it can be used in any environment with deployed Wi-Fi APs. Fuzzy data analysis is used to combine and improve the accuracy of traditional localisation algorithm: Euclidian, K-nearest neighbours (KNN) and weighted K-nearest neighbours (WKNN) (Le Dortz, 2012).

Fuzzy logic is a kind of logic that takes uncertainty into consideration which makes it more flexible than crisp logic (Ross, 2010). A crisp variable either exists in a set or does not. In fuzzy logic, however, a variable can partially exist in many sets with different degrees of membership. This property makes fuzzy logic well suited for uncertain models.

The rest of the paper is organised as follows: Section 2 provides a background on existing wireless fingerprinting techniques used for localisation, in addition to an overview of fuzzy logic. Section 3 presents the proposed system. Section 4 explains the experiments used to test the proposed system and discusses the results. Finally, Section 5 concludes the paper.

2 Background

The purpose of the proposed system is to locate users with high accuracy using any standard IEEE 802.11 Wi-Fi protocol, without having to add extra hardware. Fingerprinting is used as the localisation technique, in addition to using fuzzy logic to improve the localisation accuracy. In this section, the traditional techniques of fingerprinting are discussed, followed by an overview of fuzzy logic.

2.1 Fingerprinting techniques

Three of the most common fingerprinting-based techniques are used in the proposed system, which are Euclidian, KNN and WKNN. Each of these techniques represents one way of using the RSS collected by the user, in order to calculate an estimate for his location. In this section, the algorithm of each of the techniques is presented, together with some of the results achieved using these basic techniques from the literature.

2.1.1 Euclidean

The Euclidean distance algorithm is a simple choice for RSS fingerprinting that measures the distance between an online RSS value and the offline training database RSS records.

\[
X_j = \sqrt[p]{\sum_{i=1}^{N} (x_{ij} - x_{0j})^p}
\]
In this minimum distance equation (1), $x_{0i}$ is the online RSS value, $x_{ij}$ is the reference fingerprint value in the database for access point $i$ at location $j$, $N$ is the total number of access points, and $j$ is the location number $1 \leq j \leq L$, where $L$ is the total number of locations. For Euclidean distance, $p = 2$ and $X_j$ represents the Euclidean distance between the online RSS value and the offline value at location $j$. The location at which $X_j$ is smallest is the most likely location to be correct according to the Euclidean distance technique (Jiang, 2010). Euclidean distance was used for localisation in (Sabbour, 2007), a mean error of four metres was reached.

### 2.1.2 K-nearest neighbours

The KNN and the WKNN methods are the basic schemes that are generally used for the estimation of users’ location indoors. They are based on Euclidean distance, however the value of $K$ simply determines the number of minimum distances retrieved from (1), for $K$, $1 \leq K \leq N$; the mean of these $K$ minimum distances is computed, resulting in the predicted $(X, Y)$ location. The end result of KNN depends heavily on the number of neighbours $K$ set in the equation (Jiang, 2010).

\[
X = \frac{1}{K} \sum_{j=1}^{K} x_j
\]  

(2)

The KNN technique was used in (Le Dortz et al., 2012), with $K = 3$, and 15 APs were utilised per location. A median error of 2.4 metres was reached.

### 2.1.3 Weighted K-nearest neighbours

For the WKNN algorithm, the weight is calculated by getting the inverse of the corresponding minimum distance $D_{ij}$ between the user location and the actual location from the database.

\[
w_{ij} = \frac{1}{D_{ij}}
\]  

(3)

The same equation (1) is used similar to KNN, but the weighted average is calculated in the case of WKNN, producing the predicted $(X, Y)$ location.

\[
X_j = \frac{1}{\sum_{j=1}^{K} w_{ij}} \sum_{i=1}^{K} w_{ij} \cdot x_i
\]  

(4)

The advantage of the WKNN algorithm over the KNN is that it increases the weight of the values that are closer to the desired result and excludes the values that would significantly damage the accuracy of the system. In (Kuntal and Karmakar, 2014), the WKNN algorithm was used in simulations on MATLAB. Seven APs were simulated, and a shadowing factor was added to account for environmental changes. The algorithm reached an error less than two metres (90% percentile) without the shadowing factor, and 6 metres (90% percentile) with the shadowing factor.
2.2 Fuzzy logic

Fuzzy logic is a form of many valued logic in which each value has a certain degree of validity. Instead of rigid fixed truths, it presents rather flexible ones. This is achieved since fuzzy logic variables have a truth value that ranges from zero to one. Zero means completely false or invalid while one means completely true or valid. For example if a fuzzy logic variable is used to represent temperature, then the temperature does not have to be binary (either hot or cold). It can be hot to a certain degree and cold to another degree at the same time. This property of fuzzy logic makes it perfect for uncertain models (Lee, 2005; Gautam et al., 2014).

It follows from above that a fuzzy set is different from a crisp set. A variable belongs to a fuzzy set with a certain degree ranging from zero to one called the grade of membership. The most frequently used way of processing fuzzy knowledge is defining a set of ‘if-then’ fuzzy rules. These are logical statements that are used to reach a conclusion based on the fuzzy information collected. For example, if a variable $x$ belongs to a fuzzy set $A$ then the fuzzy rule produces an output $Y$. If variable $x$ belongs to fuzzy set $A$ with membership 0.4 then the rule applies with degree 0.4. The inverse process which aims for reaching a crisp conclusion from a set of fuzzy rules is called defuzzification, which is normally the last step of fuzzy systems (Ross, 2010).

3 Proposed technique

As mentioned earlier, accuracy is crucial when it comes to indoor localisation. The proposed technique aims at improving the accuracy of existing localisation techniques, using the existing Wi-Fi protocols, without using extra hardware. The technique has two phases: offline and online. The offline phase is a training phase that needs to be executed once before the system could be used to locate any users. Afterwards comes the online phase when the users are actually located using their received Wi-Fi signals by comparing with the data collected in the offline phase. The proposed algorithm for locating the users employs the traditional techniques (discussed in Section 2) to get preliminary estimates for the user’s location. Then, fuzzy logic is used to combine these estimates into one final location that is more accurate than the preliminary estimates. This section presents the proposed technique in detail.

3.1 Overview

The proposed system is comprised of two phases:

1. **Offline training phase** where RSS readings are collected from various points in a certain area. The collected data is stored in a database. In our experiments, this process was repeated 20 times, and the average was calculated at the end, in order to get a better estimate of each location’s RSS fingerprint (and to minimise variations in the environment and measurements).

2. **Online determination phase** which is the actual running system that reads the user’s RSS and compares it to the database by applying minimum distance algorithms and computes the predicted user location. During the online phase the computation process goes through three stages, namely:
a Collecting data: the user device’s Wi-Fi readings are sent to a server application.

b Applying localisation algorithms: the server takes the online RSS and runs the three basic localisation algorithms in parallel; Euclidean distance, KNN and WKNN. Each technique compares the online RSS with the mean (average) of the previously collected 20 offline maps.

c Applying fuzzy rules: takes the output of stage two and inserts it in a fuzzy logic Unit that combines the three techniques and produces a final output corresponding to the user’s location.

The flow of the proposed implementation is shown in Figure 1, displaying the three stages and the output at each stage.

Figure 1  Proposed system’s block diagram (see online version for colours)

Since the techniques used in the determination phase are discussed in Section 2, this section will explain the proposed fuzzy logic unit and how it is used to combine the outputs of the three basic localisation techniques to get a more accurate final output.

3.2 Fuzzy techniques

Fuzzy logic is a form of logical reasoning that has loose decision-making criteria. It removes any strictness when seeking a deterministic ‘crisp’ output. Therefore fuzzy logic was used in the localisation algorithm, in order to provide each predicted output the flexibility to have a certain degree of validity in different locations simultaneously.

Each location is assigned only one ‘rule’, and this rule is applied on its corresponding location with the corresponding degree of membership. If the online reading indicates place A, the user may still be in place B with some degree as a result of imprecise measurements, which means that if a certain vector is received during the online phase, there is a possibility that it would map to a different location, with a similar RSS.

The degree of membership is calculated by predicting a location using three minimum distance techniques, Euclidean, KNN and WKNN, and then defuzzified. Defuzzification is the process of producing a quantifiable result in fuzzy logic, given the fuzzy sets and
the corresponding membership degrees. Two defuzzification algorithms were implemented: weighted fuzzy average and fuzzy sum for location prediction in the system.

In both approaches, it is important to note that we do not compare values between locations one-by-one; however, we construct a dataset to best match the input RSS value with the location that is most likely correct. We take the RSS reading at any random location with known coordinates, and then we compare it to all the locations on the 20 offline maps. For each iteration, the output is collected and used to build a histogram. The histogram represents the number of times, the predicted output correctly matched with the input RSS signal. Then we repeat this experiment for all $N$ locations available, and each one of them will have its own histogram; in other words, each location will have its unique fuzzy dataset. Therefore, any online input in this system becomes in turn a fuzzy member of this set with some degree, which means that the RSS signal does not compare to one single location, but to the histogram created during the offline training phase.

In weighted fuzzy average, the predicted locations from the three techniques are weighted by a membership grade of an associated fuzzy set stored in a database. This database contains a number of fuzzy sets (for each location) derived from a histogram of percentages of accuracy at each location during the offline phase. By multiplying the predicted coordinates of each location by its degree of validity, a weighted average algorithm is implemented that provides a crisp output and hence deciding upon which predicted location has a higher degree of being correct.

Fuzzy sum falls under the category of ‘composition’, where all of the fuzzy subsets assigned to each output variable are combined together to form a single fuzzy set (Ross, 2010). Here, ‘sum composition’ is implemented by having $N$ grade vectors for each of the $N$ locations on the offline map, resulting in an $N \times N$ grid. This means that when the user device sends its RSS value to the system, the undefined location could map to any vector in the training set, however each location has its own $N$ degrees of validity relative to the user’s RSS. Hence, if the point-wise sum at each location is taken over all of the fuzzy subsets assigned from the histogram database, the output will be a vector $1 \times N$ representing the combined validity of each of the $N$ locations. Finally, the index of the location having the maximum weight over the sum of subsets represents the predicted location that is correct to a higher degree. That is why sometimes this technique is referred to as ‘fuzzy max’ or ‘fuzzy sum’. The defuzzification is then applied in order to obtain the final crisp output, representing the predicted location. This is done by determining the index of the maximum location and then mapping this location to its $(X, Y)$ coordinates on the offline grid.

4 Experiments and results

The aim of the proposed technique is to improve the accuracy of existing localisation techniques present in the literature without requiring additional hardware. That is why it was important to test the proposed localisation technique against traditional ones in order to draw a conclusion about the proposed system’s accuracy. In order to test the localisation techniques that were developed, two environments were chosen, a lab room and a corridor. To facilitate reference to any of the two experiments, they will be referred to as experiment one and two, respectively.
4.1 Equipment and simulation tools

Both experiments were conducted in normal everyday conditions, and their offline readings were collected over a long period of time (around two weeks). Three APs were set up and fixed in three different locations in the lab, such that they would provide complete coverage to both areas. The reason for fixing the APs is to resemble a typical Wi-Fi AP deployment. The areas of experiments one and two are 90 tiles each. However, the shape of the area used is different. The area of experiment one is a rectangle of ten tiles of length and nine tiles of width. Area two, which is the corridor, is 30 tiles long and three tiles wide. Twenty offline RSS maps were collected for both experiments one and two as the offline database. It was observed that the means of the RSS readings per location were converging, and that having more offline maps was not going to significantly affect the obtained readings.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Access point 1</th>
<th>Access point 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>TP-LINK</td>
<td>Linksys</td>
</tr>
<tr>
<td>Company</td>
<td>TP-LINK Technologies Co.</td>
<td>Cisco Systems Inc.</td>
</tr>
<tr>
<td>Model</td>
<td>TL-WA701ND</td>
<td>WAP54G</td>
</tr>
<tr>
<td>Version</td>
<td>1.2</td>
<td>3.1</td>
</tr>
<tr>
<td>No. of antennas</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
| Transmit power       | <20 dBm                 | 802.11g: typ. 13.5 ± 2 dBm
@Normal temp range
802.11b: typ. 16.5 ± 2 dBm
@Normal temp range |
| Wi-Fi standards      | IEEE 802.11n            | IEEE 802.11n            |
|                      | IEEE 802.11b            | IEEE 802.11b            |
|                      | IEEE 802.11g            | IEEE 802.11g            |
| No. of APs           | 1                       | 2                       |

Figure 2 presents the tile/location numbers in the lab environment as well as the Wi-Fi heat map representing an overlay of RSS readings seen from each AP. Figure 3 presents the same information for the corridor environment.

In our experiments, three APs were used. Table 1 summarises the specifications of the utilised APs (Cisco Systems Inc., 2008; TP-LINK TECHNOLOGIES CO, 2009).

A laptop was used, with a dedicated WLAN analyser card to collect offline readings to be stored in the database. The card used is an RF protect mobile card produced by network chemistry. It supports IEEE 802.11 a/b/g standards. The provided software enables the user to scan the surrounding area and saves the RSSs of all detected APs. A measurement was collected at every tile of each map until fingerprints of all tiles of the map were collected. Then, the operation was repeated for the number of offline maps needed.
4.2 Procedure

RSS readings were taken at every tile in the two environments. The procedure was then repeated for a certain number of times to minimise variations in the readings as previously specified. An average of 20 rounds per area was collected where RSS readings were captured from all available APs. After completing one round, an extensible markup language (XML) file is exported using the Wi-Fi analyser software. The output file contains the X, Y coordinates for the 90 locations (tiles) as well as the RSS at each tile. During a data collection round, RSS readings are only read once per tile and once all readings are collected for all tiles, the data collection round is repeated for the required number of times. As such, the time between successive RSS readings for the same tile is large enough to capture variations in the wireless environment in addition to minimising the effect of intermittent disturbances.

The orientation of the laptop used to collect the RSS information was fixed every time an RSS reading was taken in order to minimise the errors as much as possible. Also, similar to typical Wi-Fi deployments, the APs were fixed throughout each experiment.
After constructing the offline database, a set of basic localisation techniques were implemented on MATLAB. These are Euclidean distance which is the minimum distance algorithm, KNN, WKNN in addition to a weighted average technique that uses custom weights. These basic techniques are compared to the two proposed fuzzy techniques. These are the weighted fuzzy mean and fuzzy sum techniques mentioned earlier.

Since 20 offline maps were collected, one of the maps was used as the online map and the remaining 19 were used as offline maps in the simulation, looping over the 20 maps. This way, one set of results was obtained for each map being used as online and the other 19 as offline.

4.3 Results

4.3.1 Performance measurement units

In this section, the implemented techniques mentioned earlier are compared through six criteria. The results of the comparison based upon each criterion are presented separately. The criteria are accuracy, average error, maximum error, minimum error, probability density function (PDF) and cumulative distribution function (CDF). Among these units of measurements, the CDF and the average errors are the most commonly used in the literature. For example, the CDF criterion was used in (Le Dortz et al., 2012), (Kuntal and Karmakar, 2014), and (Seifeldin et al., 2013), while the average error was used in (Kuntal and Karmakar, 2014), (Seifeldin et al., 2013), and (Eisa et al., 2013). The accuracy criterion is based on counting the number of times and error less than one metre from the correct location is received.

4.3.2 Results

First, the results in the lab experiment are presented, followed by those in the corridor experiment, and finally a comparison between the results in the two areas is presented together with a summary of the results.

4.3.2.1 The lab experiment

The first performance measurement unit is the percentage of accuracy up to 1 m. From Figure 4, it can be seen that Euclidean performs the worst. Accuracies of the other techniques are close to each other, ranging from 40% to 60%, where in some maps, KNN and WKNN perform better while in other ones, and fuzzy average and fuzzy sum are more accurate.

Average error is based on calculating the error range that is dominant in the proposed system. It helps define how the system behaves in each area. As can be seen in Figure 5, all the fuzzy and KNN techniques yield better results when it comes to average error due to the averaging done in each implementation. For example, if there are two possible locations, their average will be bounded in the test area. Here K = 5 is used, so this biases the results such that the averaged locations are nearer to the centre of the test area in case of error. It can be seen that Euclidean distance has the highest average error, whereas, again the fuzzy max and average techniques reached the least errors around 1 m, especially fuzzy average which give the best average error almost in every map.
**Figure 4** Accuracy (up to 1m) in the lab environment (see online version for colours)

**Figure 5** Average error in the lab environment (see online version for colours)

*Minimum error* is the best case error; it is the smallest error that could be achieved after performing a number of trials. It is important to note that this criterion does not count the number of times this best case has occurred; it just displays the thresholds and best-case capability of the algorithm. As shown in Figure 6, Euclidean is the only technique that could achieve 0 m minimum error. That is true because Euclidean is a deterministic algorithm which compares to the database and retrieves the closest match without averaging. The remaining techniques vary due to averaging; however, the scale is within 0 m to 0.2 m, so the variations are barely significant.
Fuzzy-based Wi-Fi localisation with high accuracy using fingerprinting

Figure 6 Minimum error in the lab environment (see online version for colours)

Figure 7 Maximum error in the lab environment (see online version for colours)

Maximum error determines the maximum error observed in a number of samples, which sets a maximum threshold of error for the system. It is critical to take into consideration the worst case error because it reflects worst-case performance. Since the environment is dynamically changing, it was important to guarantee that the proposed algorithm does not result in huge error offsets from the correct locations. According to the results shown in Figure 7, the proposed fuzzy average technique in the lab environment has the best results (lowest observed error).

Figure 8 presents the PDFs for the two proposed fuzzy techniques constructed from the errors corresponding to every online input. The mean is located at the peak of the red curve, the maximum and minimum on the sides. From the figure, the average errors of both techniques in the lab are at 1 m.
The CDF for the techniques, shown in Figure 9, describes each algorithm’s accuracy in terms of error and distance. The horizontal axis is the localisation error in metres. For example, in Figure 9, 30% of the time, the error is less than at 0.5 m for fuzzy max and 20% for fuzzy avg. At one metre, almost all the techniques are 50% then at the top of the graph, the best technique that has actually reached more than 95% is fuzzy average. This means that any result coming from fuzzy average will fall in the region within zero to two metres. Hence in comparison with the remaining techniques, fuzzy average is the best technique; the fastest technique to reach 100%.

### 4.3.2.2 The corridor experiment

As for the percentage of errors up to 1 m in the corridor environment, the proposed fuzzy sum technique outperforms the proposed fuzzy average technique, as can be seen in Figure 10. The fuzzy average is significantly impacted by the environment. The reason why fuzzy average performs worse in the corridor than in the lab is due to the averaging nature of the technique, which tends to locate the user towards the centre. The shape of the corridor area makes most of its tiles at the edges. That is why average based techniques in general behave better in square shaped areas rather than long thin rectangular ones.
Figure 10  Accuracy (up to 1 m) in the corridor environment (see online version for colours)

Figure 11  Average error in the corridor environment (see online version for colours)

Figure 11 shows the average errors in the corridor environment. It can be seen that Euclidean methods performs the worst, while fuzzy sum produces the least average error. Note that the average errors of all techniques in the corridor environment are generally higher than those in the lab environment. This results from the open nature of the corridor environment which makes it more susceptible to noise and interference.

Similar to the minimum errors in the lab environment, Euclidean is the only technique that can reach 0 m error in all maps. Nevertheless, the minimum errors of all the techniques do not exceed 0.25 m which is barely significant.

The maximum errors of fuzzy average are far less than any of the other techniques, as can be seen in Figure 13. In other words, the error given by fuzzy average does not exceed 7 m, which is very good compared to the other techniques which reach 12 m in some maps.

Figure 14 shows the PDFs of the two fuzzy techniques in the corridor environment. The same observation as noted earlier can be observed here, i.e., the average errors are higher than those given by the same techniques in the lab environment. This is due to the increased noise in the corridor than in the lab. The standard deviation from the mean is also higher than in the lab.
Finally, in the corridor, same as in the lab, the fuzzy average technique still gives best percentile error results; the percentile for error below 5 m is 90% for the proposed fuzzy
average technique whereas the other techniques are only able to achieve a percentile of 80%.

**Figure 15** CDF in the corridor environment (see online version for colours)

4.3.2.3 Summary

To summarise the results presented in this section, all techniques are generally more accurate in the lab environment than in the corridor environment, especially non-average-based techniques such as Euclidean and fuzzy sum. The two proposed fuzzy techniques generally outperform the other techniques in both environments where, fuzzy average outperforms all other techniques in the lab environment, and fuzzy sum gives the best results in the corridor environment. The proposed fuzzy localisation techniques were demonstrated to be more accurate than the traditional localisation techniques across both studied environments.

5 Conclusions

Locating users is a crucial part of many applications like navigation applications, automation systems, and smart advertisement. Global positioning systems that rely on satellites, like GPS, are typically used for localisation. Although such systems are sufficient for outdoor applications, where errors in the range of metres are acceptable, they do not perform as well in indoor environments, where higher accuracy is much more important.

Consequently, different techniques were developed for indoor applications; some make use of additional hardware including sensors or RFID for automation. Despite having the advantage of high localisation accuracy, using sensors or RFID requires added hardware which adds to the cost and complexity of the localisation system. Other techniques were introduced which depend on analysing a received wireless signal by the user, and calculating the location according to its strength without required additional hardware specifically for localisation. Examples of such techniques are triangulation and fingerprinting. These have the advantage of being able to use an already existing wireless infrastructure, like Wi-Fi or Bluetooth. However, they have the disadvantage of being less accurate than the sensor or RFID based localisation techniques.
In this paper, a localisation system is proposed that uses unmodified Wi-Fi and fingerprinting to locate users with high accuracy. The system combines three traditional fingerprinting algorithms, Euclidean, KNN, and WKNN, to produce three location estimates for a user’s location. Then two fuzzy systems, fuzzy average and fuzzy sum, are proposed to combine these location estimates to achieve higher localisation accuracy. Since the system uses the existing Wi-Fi infrastructure, it does not require installing any extra hardware. It, however, requires a training phase to be executed once before using the system.

In order to test the proposed system, two experiments were conducted in two different areas. Each of the areas is 10 m² covered by only three access points. The results of the proposed techniques were compared with the traditional ones using six different criteria. From the results, it was shown that the fuzzy average offers the best localisation accuracy among all the tested techniques. The observed mean error is within one to two metres, the maximum error ranges from three to four metres and more than 95% accuracy is achieved before two metres.

References
Jiang, L. (2010) A WLAN Fingerprinting Based Indoor Localization Technique, University of Nebraska, Lincoln, Nebraska.
