Research Article

Outlier-Detection-Based Indoor Localization System for Wireless Sensor Networks

Yu-Chi Chen and Jyh-Ching Juang

Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan

Correspondence should be addressed to Jyh-Ching Juang, juang@mail.ncku.edu.tw

The paper exploits the outlier detection techniques for wireless-sensor-network- (WSN-) based localization problem and proposes an outlier detection scheme to cope with noisy sensor data. The cheap and widely available measurement technique—received signal strength (RSS)—is usually taken into account in the indoor localization system, but the RSS measurements are known to be sensitive to the change of the environment. The paper develops an outlier detection scheme to deal with abnormal RSS data so as to obtain more reliable measurements for localization. The effectiveness of the proposed approach is verified experimentally in an indoor environment.

1. Introduction

Advances in Microelectromechanical Systems (MEMSs), embedded technologies, wireless communication, digital and analog devices, and battery techniques make the wireless sensor networks (WSNs) a prominent and enabling technology in surveillance and exploration applications [1]. One significant attribute of the WSNs is the localization capability. The location information can remarkably enhance the contents of the gathered information in monitoring, tracking, and decision making applications. Indeed, in many applications such as surveillance, target tracking, and intrusion detection, the measurement data are meaningless without the location attributes. To establish a low-cost, easily implementable, and high-reliable indoor positioning capability, WSNs can utilize the received signal strength (RSS) measurements as the baseline for range determination and location estimation. Unfortunately, the propagation behavior of radio signal in indoor environments is complex, which makes the RSS-based indoor localization a challenging issue.

Wireless positioning technology can be roughly divided into two categories: radio positioning, which includes GPS, RFID, WiFi, and ultra-wideband (UWB), and non-radio positioning, which includes video cameras (optical), infrared, ultrasound, and inertial systems [2, 3]. It is worthwhile to note that as each sensor has its limitations, a practical positioning system may often employ sensor fusion techniques to integrate the sensors to yield improved performance. To enhance the overall reliability and accuracy in localization, the basic problem in signal reliability of each sensor needs to be addressed. Although sensors can be well calibrated, each measurement is subject to measurement noise and systematic error. In particular, outliers to sensor data need to be detected and precluded in the signal processing stage; for otherwise the results are prone to significant errors. The paper is dedicated to address the outlier detection problem for the localization in a WSN-based indoor environment. The paper proposes an outlier detection scheme to cope with unreliable measurement data. Together with the fingerprinting method and kernel density estimation, the approach is shown to be effective in achieving robust localization results. The scheme is applied at the database construction phase for the establishment of a reliable database. It is also used at the localization/tracking phase to detect and remove unreliable measurement. The method can thus pave a way for applications to heterogeneous sensor networks.

The organization of the paper is as follows. The related works of localization algorithms and outlier detection techniques are provided in Section 2. The main results are provided in Section 3 in which the localization and outlier
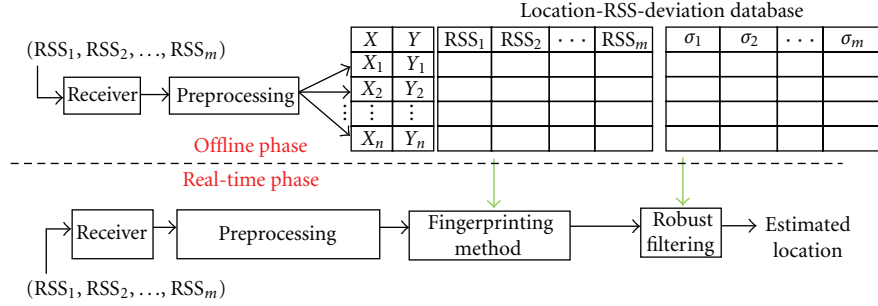
Figure 1: Flowchart of fingerprinting method.

detection techniques used in an RSS-based localization system are described. More precisely, the localization algorithm used in the indoor localization system is investigated and a novel outlier detection technique is proposed to cope with outliers in the localization procedure. Furthermore, quality control with the proposed outlier detection is employed in database management. Section 4 is dedicated to the implementation and verification of the aforementioned methods in a ZigBee-based WSN. Several experiments are conducted and the results are discussed. Finally, Section 5 concludes the paper with some discussions and remarks.

2. Related Works

In a WSN system, sensor nodes or devices are scattered in the field to perform sensing, computation, and communication tasks. Depending on the roles, the nodes are further classified as coordinator, router, and end device. The coordinator is in charge of constructing the whole network and coordinating all devices in the network such as the management of devices for the participation of the network. There can be only one coordinator in a network. For the WSN under consideration, the routers which are also referred to anchors are located at fixed and known locations for the relay of information. The end devices which may be fixed or mobile are responsible for the information collection task. The collection information is transmitted from the end devices to the coordinator directly or through routers. Typically, the locations of the end devices are not known when the WSN is deployed. Hereafter, for clarification, the nodes refer to the end devices whose positions are to be determined. The localization system being a feature of the WSN aims to establish the location information of the nodes in a WSN based on some measurements and a priori information [1, 4]. In an indoor environment, the design of a localization system is extremely challenging. On one hand, the computation, communication, memory, and energy resources of each node in a WSN are limited. On the other hand, the multipath and shadowing effects may degrade the quality of the measurements for position determination. Several indoor localization systems such as "Active Badge" [5], "Cricket" [6], and "RADAR" [7] have been investigated. It is, however, noted that there remain many challenges in the realization of a high-accuracy, high-reliable, and easily implementable indoor localization system. In

particular, the mitigation of measurement outliers in localization has seldom been addressed.

Measurement techniques in radio positioning can be roughly classified in four categories: angle of arrival (AOA), time of arrival (TOA), time difference of arrival (TDOA), and RSS; see [3] for further discussions. The paper adopts the low cost, low power consumption, and widely available RSS measurements for localization in a WSN.

2.1. RSS-Based Localization Algorithms. The RSS-based localization can be implemented using either a range-based approach or a range-free approach. The former locates the nodes by using the distance information between two nodes, while the latter is independent of the range measurements [3, 8, 9].

The range-based approaches typically consist of two steps: conversion of the RSS measurements into equivalent range measurements through a path loss model and estimation of the location through mutilateration [2, 10]. Several different path loss models have been proposed to better represent the propagation phenomena. As for position determination, in addition to multilateration, methods including multidimensional scaling (MDS) localization algorithm [11], semidefinite Programming (SDP) [12], DV-Hop, DV-distance and their refinements CDV-Hop, and CDV-Distance [13] have been investigated.

Another category of localization algorithms is the range-free approach, also known as fingerprinting, proximity-based, database matching, or pattern recognition method, which builds up a reference model or database for localization. The fingerprinting method typically consists of two phases, as shown in Figure 1. The first phase, called the offline phase, is the construction of the database. In this phase, the nodes are placed at some reference points in the environment and the signal strengths with respect to anchors are measured. The reference points are presurveyed points in the environment that are used to facilitate the construction of the database, calibration of devices/algorithms, and assessment of localization performance. The RSS database at all reference points with respect to anchors is thus constructed. The second phase, called online phase or real-time phase, is to perform localization. The signal strengths of a node at an unknown position with respect to anchors are measured and the measurement vector is compared against the database for

the determination of the location. The range-free approach, being a two-phase approach, does not rely on a path loss model and, consequently, the environmental effect can be better accounted for. However, it is pointed out that outliers in the offline phase and online phase may have a significant effect on the resulting database quality and positioning error.

In the paper, the range-free approach is adopted. The database contains two maps which represent two different probability density functions for position determination. One map is the conditional RSS probability $P(\xi \mid a_m, s_n)$ which stands for the condition probability of RSS ξ for the anchor a_m and reference point s_n. Another map in the database is $P(f_m \mid s_n)$ that is used o reflect the reception condition of the environment. Here, f_m stands for the relative frequency for a device at the reference point s_n to receive signal from the anchor a_m. In the following, it is assumed that the number of anchors is M and the number of reference points is N. The left plate in Figure 2 depicts a representative histogram of $P(f_m \mid s_n)$ at s_n with respect to five different anchors. The right plate is a representative RSS distribution $P(\xi \mid a_m, s_n)$ for some anchor and reference point.

2.2. Bayesian Inference. The localization algorithm used in the paper is Bayesian Inference [14, 15]. Unlike most existing range-free algorithms, the Bayesian inference does not take the RSS values for positioning directly; instead, the Bayesian inference views the RSS distributions as probabilities and matches the RSS vector with the database by finding the entry that results in maximal likelihood. The data in the Bayesian inference is represented as statistics of signal strengths, in terms of histogram, not the signal strength itself [3]. This property sets the Bayesian inference method apart from other fingerprinting methods due to the fact that the statistics of signal strength can prevent the location estimations from single or multiple abnormal measurements during the real-time phase; henceforth, the localization error can be mitigated.

Let z be the measurement vector of the node at an unknown location with respect to anchors. A key step in the Bayesian inference is to infer the a posteriori probability. Let $P(s_n)$ be the a priori probability and let $P(z \mid s_n)$ be the conditional probability; then the a posteriori probability can be expressed as

$$P(s_n \mid z) = \frac{P(z \mid s_n)P(s_n)}{\sum_{n=1}^{N} P(z \mid s_n)P(s_n)}. \qquad (1)$$

In localization, the a posteriori probabilities are calculated for different s_n and the location is estimated as

$$\hat{s} = \arg \max_{s_n} P(s_n \mid z). \qquad (2)$$

In the application of Bayesian inference technique for localization, the a priori probability $P(s_n)$ is typically set as $P(s_n) = 1/N$ and the maximum a posteriori estimation is the same as the maximum likelihood estimation. For target tracking application, the probability $P(s_n)$ can be updated through the time propagation model. The condition probability (likelihood) $P(z \mid s_n)$ can be computed as follows.
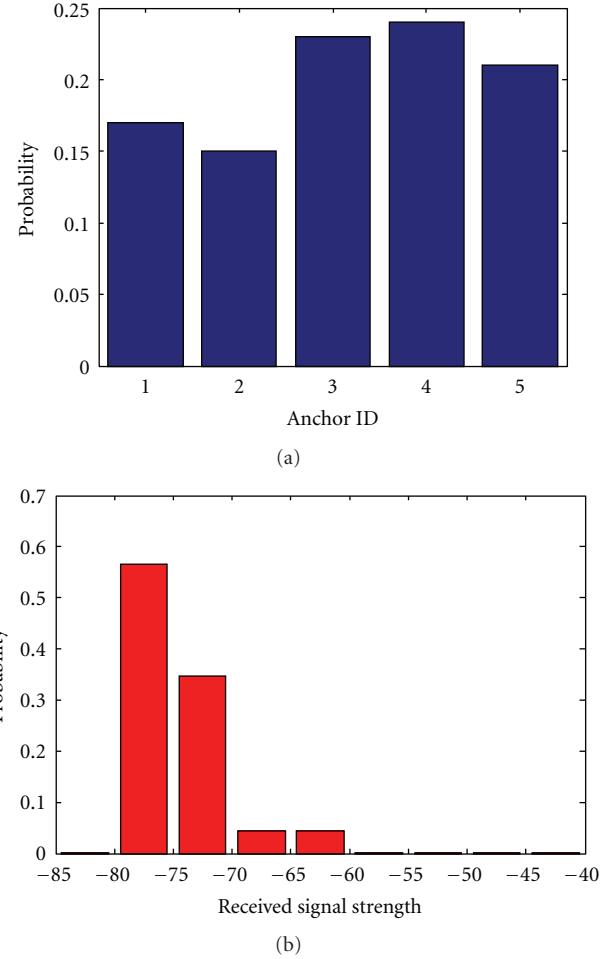

(a)

(b)

FIGURE 2: Histogram of probability of relative frequency (a) and RSS distribution (b).

Suppose that the measurement vector z is a $J \times 1$ vector. Each entry of z indeed contains the RSS measurement ξ_j between the node and the anchor a_i. From the database, the conditional probability $P(z \mid s_n)$ can be computed as [3]

$$P(z \mid s_n) = \prod_{m=1}^{M} P(f_m \mid s_n) \prod_{j=1}^{J} P(\xi_j \mid a_m, s_n). \qquad (3)$$

More precisely, with respect to each entry of z, the corresponding anchor a_m is extracted and the probability $P(f_m \mid s_n)$ is obtained. In addition, the RSS measurement ξ_j is used for the determination of $P(\xi_j \mid a_m, s_n)$. Thus, the condition probability (3) can be computed and, consequently, the maximum a posteriori estimate can be obtained.

3. Localization with Outlier Detection

Due to multipath, shadowing effects, and some hardware constraints, the outliers in RSS measurements are not uncommon. If the measurement is contaminated with outliers, whether at the offline or online phase, the resulting position estimate is likely to be misleading. Several outlier detection

techniques have been investigated in [16–18]. The section briefly reviews some outlier detection techniques, then proposes a new outlier detection scheme, and integrates it into the localization procedure to enhance the robustness of the localization system.

3.1. Hampel Filter. Perhaps, the most well-known method for outlier detection is the so-called 3σ edit rule which is based on the concept that if a data sequence is approximately normally distributed, the probability of observing a data point farther than three standard deviations from mean is approximately 0.3%. The rule classifies the data as "normal" or "suspicious" by using estimated mean and standard deviation of the dataset. Unfortunately, due to the outlier-sensitivity of the mean and standard deviation, the masking effect [19] will affect the results of 3σ edit rule seriously.

The Hampel filter [17] is similar to 3σ edit rule in principle but the Hampel filter replaces the outlier-sensitive mean and standard deviation with outlier-resistance median and median absolute deviate from median (MAD), respectively. In this approach, for a set of data $P = \{p_i\}$, let median(P) be the median. The MAD or MAD-scale estimate R is defined as

$$R = 1.4826 \times \text{median}\{|\,p_i - \text{median}\,(P)\,|\}. \qquad (4)$$

Here the factor 1.4826 is chosen so that the expected value of R is equal to the standard deviation for normally distributed data. The MAD-scale estimate is easy to evaluate and the median is obtained by a simple sorting procedure; therefore, it is suitable for the resource-limited WSNs.

A limitation of the Hampel filter occurs when more than half data are the same. Indeed, when more than half data are the same, the MAD-scale estimate is zero and all other data in this dataset are classified as outliers. This situation may happen when the dataset is coarsely quantized. A simple example that illustrates this situation is shown in Figure 3 in which 6 data are observed as -50, 3 data are observed as -51, and 2 data are observed as -49. When the Hampel filter is applied to process the data, data observed as -51 and -49 are regarded as outliers even though they are close to the median value -50.

3.2. Kernel Density Estimator. In order to estimate the distribution of sensor observations as well as overcome the aforementioned shortcoming of the Hampel filter, the paper utilizes kernel density estimator (KDE) to estimate the data distribution of the data sequence. Kernel density estimation is a nonparametric way of estimating the probability density function of a random variable [15]. The estimator $\hat{f}(p)$ is defined as

$$\hat{f}(p) = \frac{1}{|P|} \sum_{h_i \in P} k(p - p_i), \qquad (5)$$

where $|P|$ is the number of samples in the data set P. Each p_i is a sample drawn from some distribution with unknown density f and the estimator attempts to estimate the distribution through a kernel function $k(\cdot)$ [20]. The paper adopts the Epanechnikov kernel for KDE as the kernel is optimal in the minimum variance sense while its efficiency
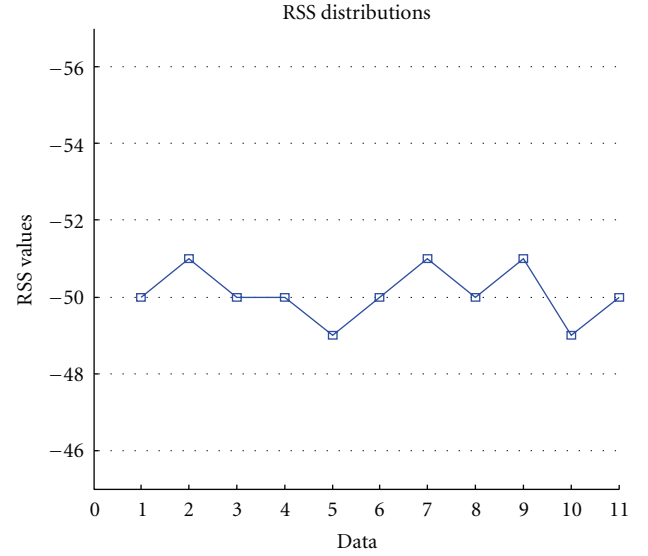

FIGURE 3: A pathology of Hampel filter.

loss is comparatively low among other kernel functions. The Epanechnikov kernel function is

$$k(p) = \begin{cases} \dfrac{3}{4}\dfrac{1}{B}\left(1 - \left(\dfrac{p}{B}\right)^2\right), & \left|\dfrac{p}{B}\right| < 1, \\ 0, & \text{otherwise,} \end{cases} \qquad (6)$$

where B is the bandwidth of the kernel function [21]. In the RSS-based localization problems, the observed RSS data from each sensor can be treated as random data samples; the KDE can then be employed to estimate the distribution of the RSS values.

The KDE may also be subject to erroneous behavior when RSS data are varied significantly in an indoor environment. An example of such a circumstance is illustrated in Figure 4. In the figure, the squares are the observed RSS data. The probability of the RSS value equals to -73 is relatively high in comparison with those with RSS value being equal to -53 or -50. As a result, the RSS data of -73 may be classified as normal data although they appear to be outliers.

The paper exploits the properties of Hampel filter and KDE to develop a new method for finding outliers in large datasets. The estimated density from KDE can prevent the Hampel filter from having a zero MAD. On the other hand, the Hampel filter can identify multiple outliers when the density of outliers is relatively high in the KDE.

3.3. Proposed Outlier Detection Technique. Although outliers are often considered as erroneous data, it may also carry some important information; as a result, the outlier detection technique should cope with the outliers instead of just removing them. Storing the entire history of RSS data is not recommended in WSN applications due to the increasing memory requirements. To this end, the paper presents a general framework for estimating the data distribution in view of adjustable window operation. In the RSS-based
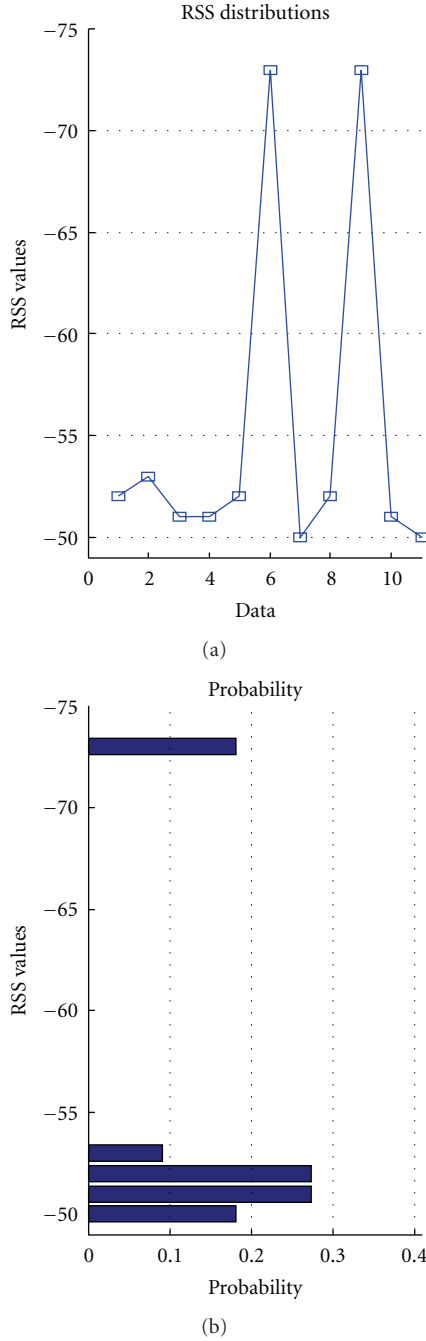
(a)

(b)

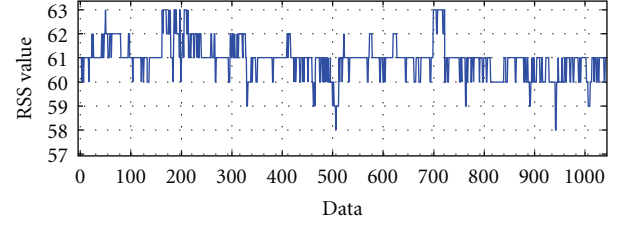Figure 4: A pathology of KDE.

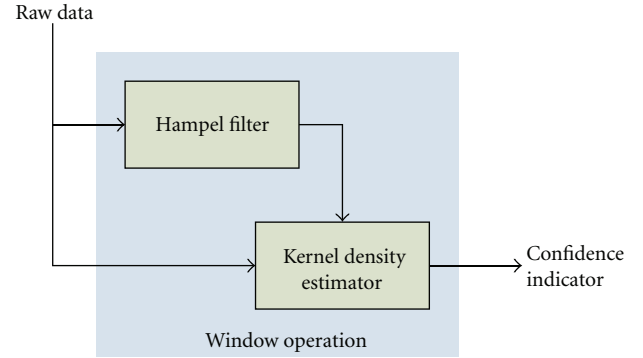

Figure 5: RSS Measurements at a fixed node.

Figure 6: Proposed outlier detection scheme.

environments. Hence, the proposed outlier detection scheme assigns each data a confidence indicator which indicates the degree of the reliability of the corresponding data instead of just identifying or removing it. This is depicted in Figure 6.

To combine the Hampel filter with KDE, the MAD-scale score is introduced. The MAD-scale score m_i for each data sample p_i is defined as

$$m_i = \frac{|p_i - \text{median}(P)|}{R}, \quad (7)$$

where R is the MAD-scale estimate as computed in (4). The MAD-scale score addresses how far the data sample is deviated from the median of the data set in terms of MAD scale. Then, combining the Hampel filter and probability density estimation from KDE, one can obtain the confidence indicator c_i of the data sample p_i as

$$c_i = \frac{\text{Prob}(p_i)}{m_i}, \quad (8)$$

where $\text{Prob}(p_i)$ is the probability of the data computed from kernel density estimator. From (8), it is clear that when the probability is high and the MAD-scale score is low, the confidence indicator is high, implying that the data sample is trustworthy. On the other hand, when the probability is low and MAD-scale score is high, the confidence indicator is low and the data must be used judiciously.

An example of the relationship between inputs and outputs of the proposed outlier detection technique is illustrated in Figure 7 in which the square-dash line is the raw RSS measurement data which are the inputs of the outlier detection scheme, and the asterisk-solid line is the confidence

localization problems, the values of RSS are integers and the values of RSS from a fixed node are usually not a constant as time varies, as shown in Figure 5. Thus, the dataset of RSS values is coarsely quantized and the distributed density of outliers in a small size window may be relatively high in comparison with normally distributed data.

In the indoor environments, the characteristics of radio signal and the obstacles often cause some data to deviate to become outliers. These outliers, however, may provide the information about the walls or obstacles in the indoor
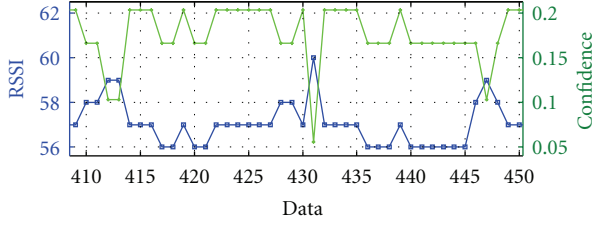
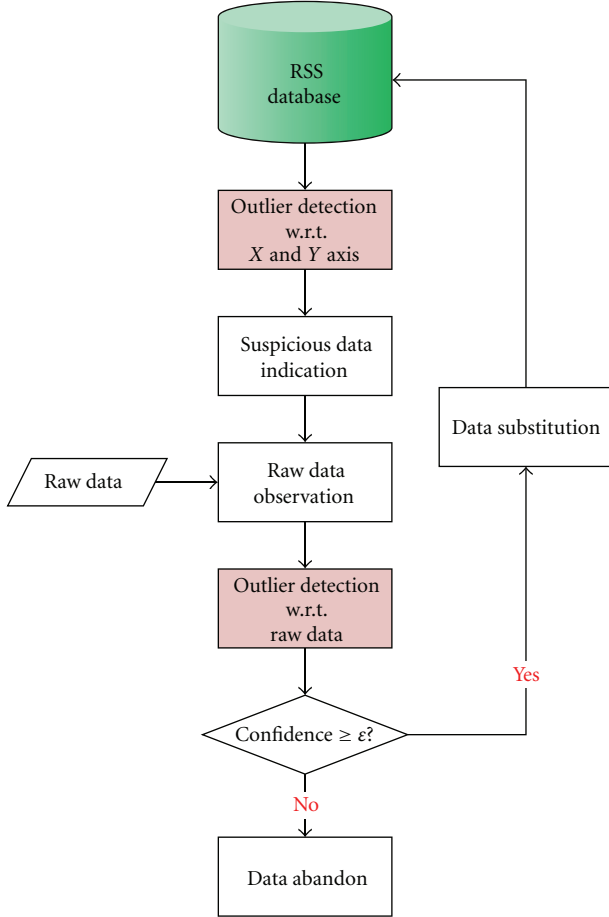FIGURE 7: Inputs and outputs of the proposed scheme.

FIGURE 8: Flow chart of database quality control.

indicators of the data. In the figure, the confidence indicator of data 431 is significantly lower due to the fact that the RSS measurement is notably higher than other data in the dataset.

In the localization procedure, the proposed outlier detection scheme can be applied in two ways: censoring sensor reading sequences and RSS database, respectively. The proposed outlier detection scheme can censor raw RSS sequences and give each reading a confidence indicator; then the fingerprinting methods can use the confidence indicators as weightings in the position determination process. The RSS database can also be censored by the proposed outlier detection scheme and the overall mechanism is described in the following section.

3.4. Quality Control.

The RSS database plays a critical role in the fingerprinting methods. However, in the indoor environments, the change of environment such as addition/removal of furniture or the variation of hardware such as low battery may affect the quality of RSS database seriously. In order to overcome this problem and maintain the localization quality, a quality control scheme should be employed for the warning of suspicious RSS distribution in the database.

The flow chart of the quality control system based on outlier detection is shown in Figure 8. During the quality control, the outlier detection scheme is applied on the RSS distribution map with respect to the two axes; and then the confidence of RSS data below certain threshold ε will be viewed as suspicious data. In the establishment of the RSS distribution map, additional measurements can then be conducted at those reference points upon which the RSS data are suspicious. The newly obtained RSS data are examined by the outlier detection scheme again and the RSS distribution map is updated once the RSS data with a high confidence are obtained.

Combining the database quality control system and the localization system, this paper constructs a localization system which can localize the nodes and also update and maintain the RSS database. The localization system views the static localization data points as a kind of training data; after localizing the static unknown nodes, the system adopts the RSS information into training data and passes to the quality control procedures.

## 4. System Implementations and Experiments

In this section, a WSN is set up to evaluate the proposed localization technique in an office environment. Both static localization and dynamic tracking are considered.

4.1. Localization Platform. A WSN is composed of a set of nodes that are capable of performing sensing, computation, and communication. The experiment adopts the Texas Instrument (TI) CC2431 ZigBee Development Kit (ZDK) [22] which includes the ZigBee Evaluation Module (EM), Battery Board (BB), and Evaluation Board (EB), as shown in Figures 9 and 10, respectively, for the construction of the WSN. The CC2431 SoC chip is on the EM board, which can be programmed and compiled by using IAR EW8051 C compiler and be connected to BB or EB in this development kit to perform different functionalities and message formats in a network that is based on the ZigBee protocol [23].

Each device in the WSN can be programmed as a coordinator, router, or end device. In this paper, the routers are programmed as anchors that are fixed and located at known positions. In contrast, the end devices which may be carried by users are nodes of which the positions are to be determined. The network architecture of the WSN localization system is shown in Figure 11. The coordinator which is in charge of coordinating all devices in the network serves as the communication interface between the server and the WSN. The end devices also referred to as unknown nodes or mobile nodes are responsible for searching the anchors in the whole network and broadcasting the requests of RSS

(a)
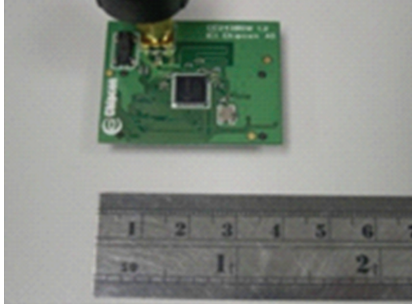
(b)

Figure 9: ZigBee evaluation module (EM) of CC2431ZDK.

(a)

(b)

Figure 10: (a) Battery board (BB) and (b) evaluation board (EB).

measurements when they receive the location request messages from the coordinator. After receiving the responded messages of RSS measurements from anchors, the unknown nodes transmit the whole messages to respond to the request of the coordinator. The anchors take charge of measuring the RSS values when they receive the requests from the unknown nodes. After measuring the RSS values, the anchors send the corresponding RSS values to the coordinator for further localization processing at the server.

The experiments are conducted at the 8th floor, Department of Electrical Engineering, National Cheng Kung University, Taiwan. The size of the sensing area is 11 meters by 12 meters. The environment layout is depicted in Figure 12 which contains three regions (bottom left: *Room 1*; top left: *Room 2*; right: *corridor*). The WSN localization system platform consists of one coordinator, 12 anchors, and a number of mobile nodes. The red marks in Figure 12 are the locations of anchors. In the experiment, in order to mitigate the shadowing effects in indoor environments, the anchors are fixed at the ceiling which is 2.5 meters high from the floor.

The RSS database is created first. In the offline phase, the RSS information is collected by placing the end devices at some predefined reference points. Each reference point is about 90 centimeters from the ground since 90 centimeters is approximately the height of the wrist of a human from the ground. The distance between two nearby reference points in the database is about 60 centimeters. In this offline phase, as anchors and end devices are placed at known locations, a set of training data is obtained.

In establishing the RSS database, the RSS measurements at each reference point are censored by the outlier detection
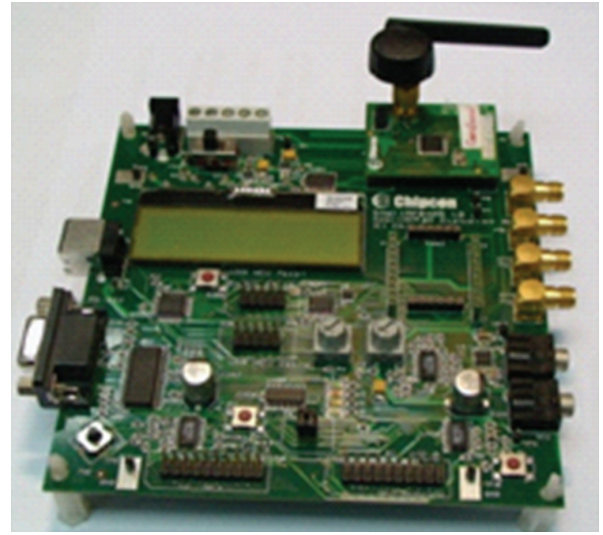
scheme and tagged by a confidence indicator. Afterwards, by taking the confidence indicators as weighting factor, the weighted mean of RSS at each reference point is computed and saved as RSS distribution. Further, by adopting the Kriging method, the database is enhanced to cover the whole area [9]. Such a database is termed as the refined database hereafter. In contrast, the database that is established without using the outlier detection scheme is termed as the original database.

4.2. Static Localization Experiments. In the static localization experiment, mobile nodes are placed in the area to collect RSS measurements with respect to anchors. Once the data are obtained, the outlier detection and kernel density estimation schemes are applied to tag each measurement with a confidence indicator. The data are then compared against
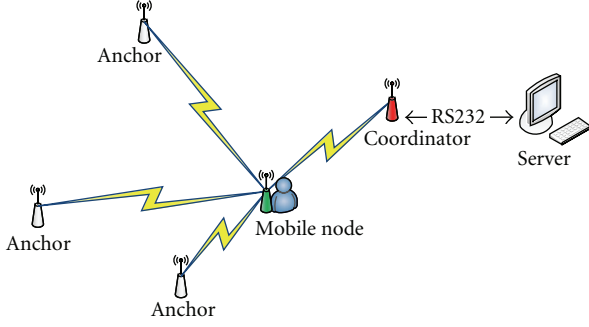
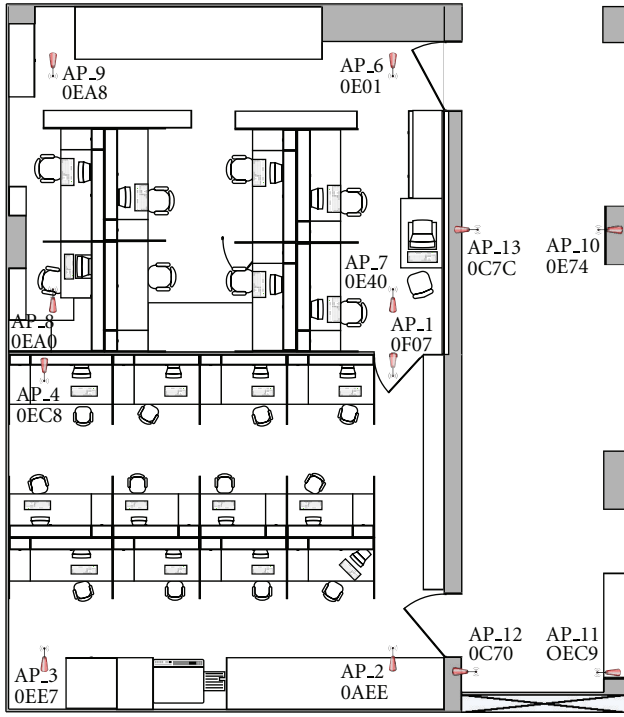FIGURE 11: Network architecture.

FIGURE 12: Environment layout.

FIGURE 13: Localization results without outlier detection.
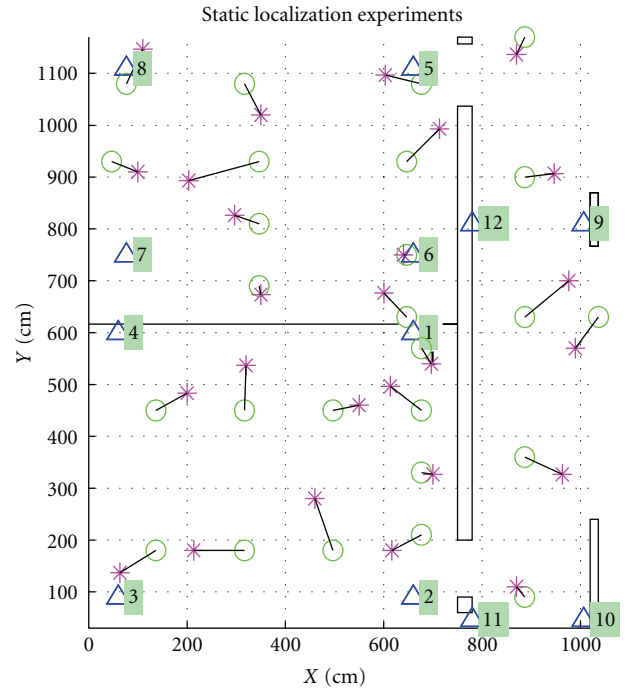
FIGURE 14: Localization results with outlier detection.

the database by using Bayesian inference to obtain the maximum a posteriori estimate of the position.

To assess the performance of the quality control scheme, the static localization experimental results are obtained based on the original database and the refined database, respectively. The static localization result based on original database is depicted in Figure 13. In the figure, the circles represent the true positions of unknown nodes, the asterisks represent the estimated locations, the lines between circles and asterisks are the error distances between true positions and estimated positions, and the triangles are the positions of anchors. In contrast, when the outlier detection scheme is applied on the received RSS data, the localization results based on the refined database are shown in Figure 14.

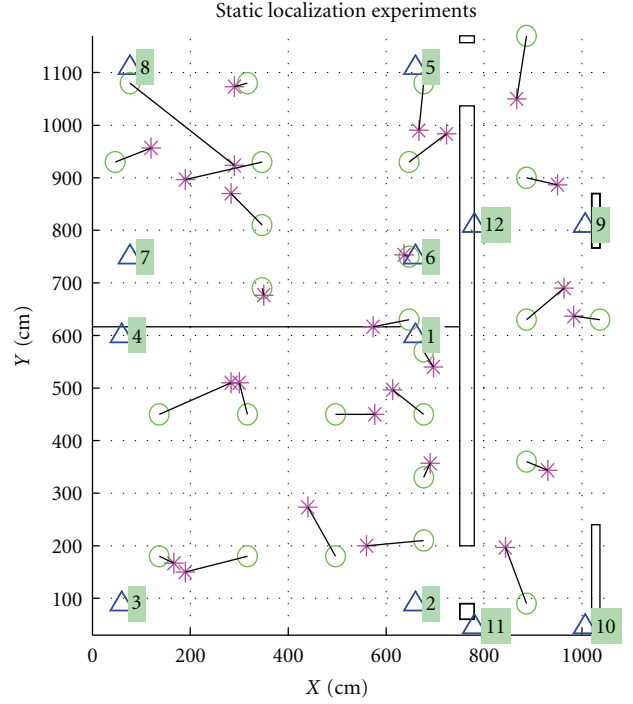By comparing the two experimental results, it is clear that with the outlier detection scheme in enhancing the refined database and improving the data quality, the results with the outlier detection are better than those without the outlier detection. The average localization errors and standard deviations are summarized in Tables 1 and 2, respectively. The improvement in static localization by using the outlier detection scheme ranges from 13.95% to 31.11%.

TABLE 1: Comparisons of average localization error.

	Localization area			
	Room 1	Room 2	Corridor	Overall
Without outlier detection scheme (unit: cm)	82.86	89.90	83.47	85.41
With outlier detection scheme (unit: cm)	71.30	61.93	66.23	66.49
Improvement	13.95%	31.11%	20.58%	22.15%

TABLE 2: Comparisons of standard deviation.

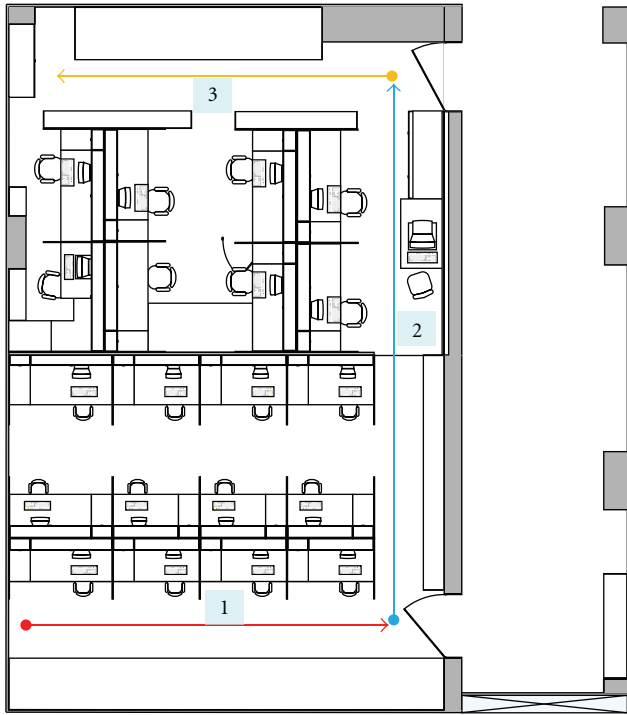|	Localization area			
	Room 1	Room 2	Corridor	Overall
Without outlier detection scheme (unit: cm)	52.60	87.43	27.11	62.78
With outlier detection scheme (unit: cm)	25.03	46.02	36.89	35.52

FIGURE 15: Paths in the tracking experiments.

TABLE 3: Performance comparisons of different outlier detection methods (unit: cm).

Method	Proposed method	KDE	Hampel filter	3σ rule
Average error	166.26	265.43	389.14	405.22

nodes will suffer from severe data loss. On the other hand, by incorporating a simple human motion model, the a priori position estimate can be used in the Bayesian inference for position determination.

To better quantify the performance, the tracking experiments are conducted by considering the case when the human moves along a straight line. The three paths are illustrated in Figure 15. The rounded points in each path are the starting points of the paths and the arrows are the final positions.

Figure 16 depicts the starting point (star) and terminal point (square) of path 1. The estimated trajectories without and with the outlier detection scheme are provided in the top and bottom plates of the figure. The average tracking error without the outlier detection scheme is 103.53 centimeters, while the average tracking error with the outlier detection scheme is reduced to 39.21 centimeters, which implies a 62.1% improvement.

The results of path 2 tracking experiment are shown in Figure 17. In the figure, the top plate is the result without outlier detection and the bottom plate is the result with outlier detection. The former leads to an average tracking error of 118.01 centimeters, while the latter results in an error of 65.91 centimeters. The improvement is about 46.5%. Similar results are also observed in path 3, which are not included due to space limitation.

It is also noted that the use of the outlier detection scheme can reduce the worst case positioning error. For example, the positioning result near Anchor 8 is misleading when the outlier detection scheme is not used. The result is improved after the application of the outlier detection scheme.

4.3. Human Tracking Experiments. The experimental environment of the human tracking experiments is the same as the environment of static localization experiments. The RSS database is also the same. The only difference is that the unknown nodes are carried by user and are movable. A concern in the tracking experiment is that the human body forms the major blocking effect. As a result, the unknown

4.4. Discussions. In the resource-limited WSN, the computational cost is a critical factor. As the computational cost of the fingerprinting method depends on the size of the RSS map or, equivalently, the number of reference points, the localization performance as a function of the number of reference points is discussed.
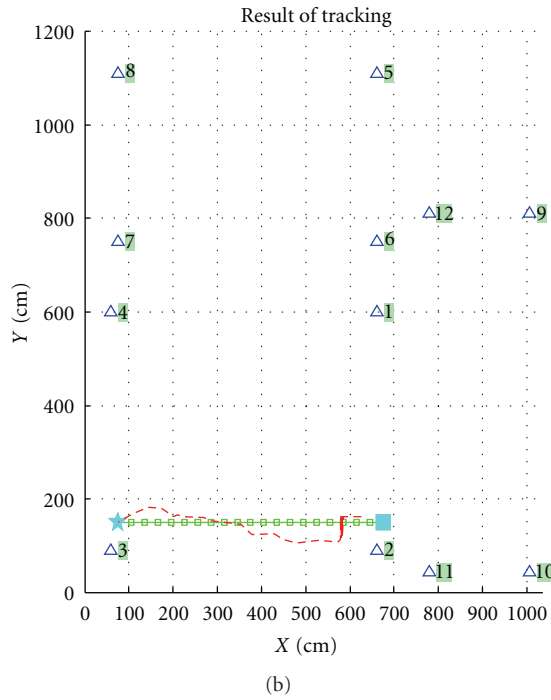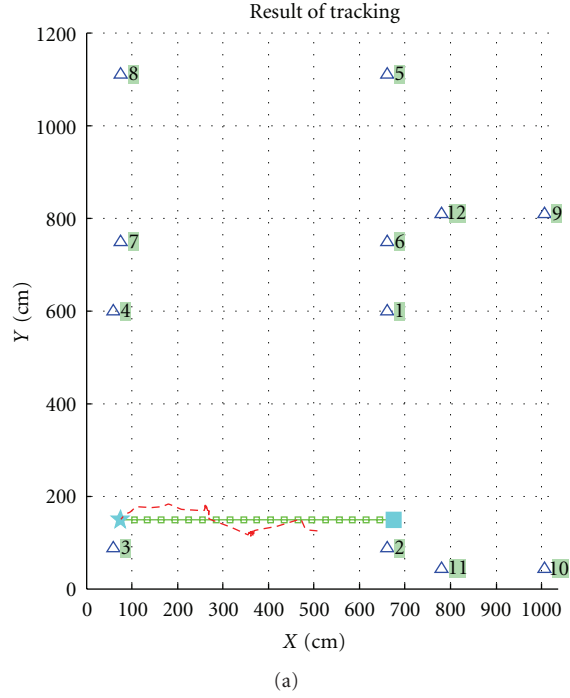
(a)

(b)

FIGURE 16: Estimated trajectory comparison of path 1.

(a)

(b)

FIGURE 17: The estimated trajectory of path 2.

Figure 18 depicts the localization error as a function of reference or candidate nodes. In this analysis, three outlier detection techniques, namely, the Hampel filter (dotted line), KDE (dashed line), and the proposed Hampel+KDE filter (solid line) are compared. It is shown that when the proposed scheme is employed, the positioning error can be significantly reduced when there are only a few reference points. This is due to the fact that high-confidence
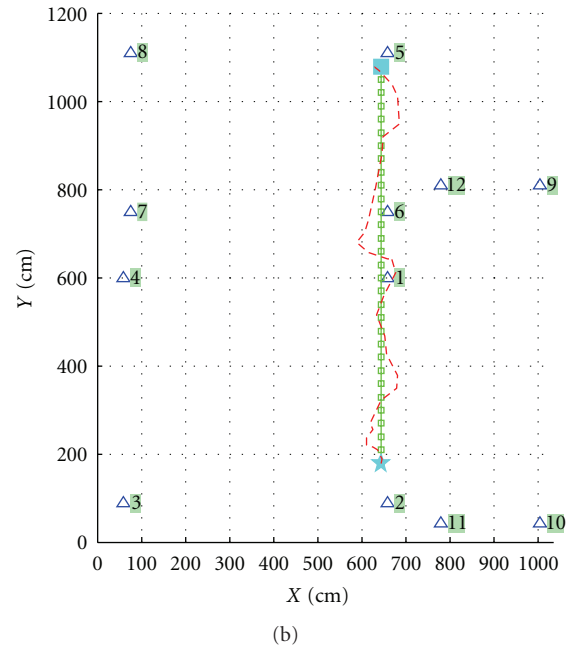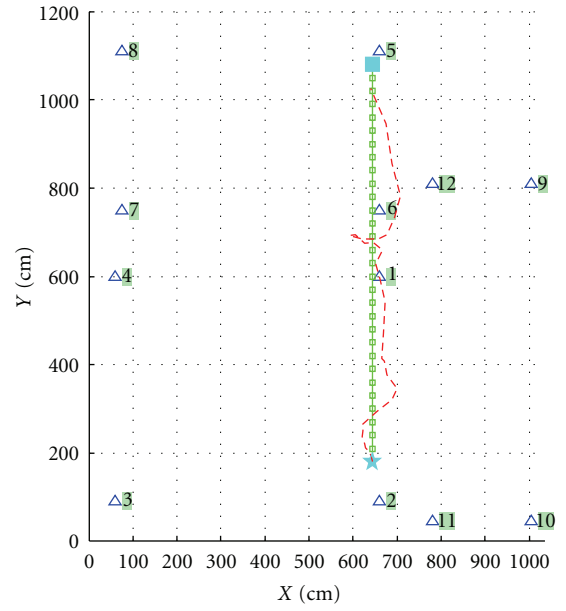
trustworthy data are processed with a heavy weighting and the localization result is not misled. This implies that the proposed outlier detection scheme meets the requirement of the resource-limited environment of WSN. The average localization errors among all numbers of candidate points of different methods are computed and depicted in Table 3. For comparative purpose, the average error of the 3σ edit rule is also provided. The proposed outlier detection scheme has less tracking error than other methods when the number of candidate points is varied.
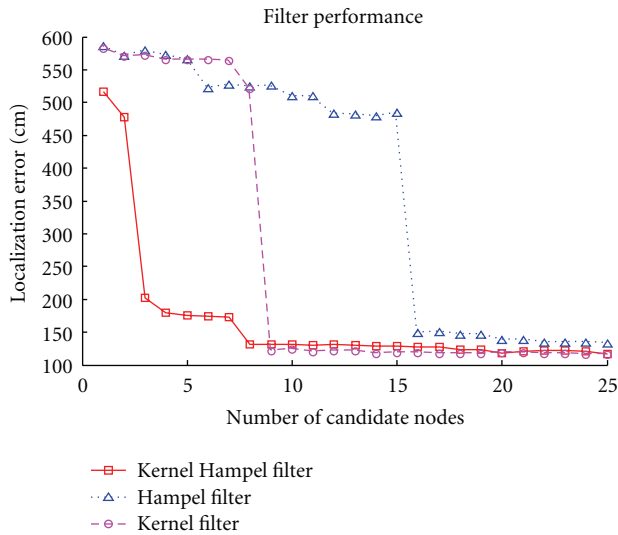
FIGURE 18: Comparison of the filter performances.

5. Conclusions

To account for the complicated RF propagation effects with limited resources in WSN-based indoor localization, the paper proposes an outlier detection scheme to perform quality control of the RSS database and data filtering in real-time localization. The approach is shown to be robust and effective in dealing with data that are subject to anomalies. The experimental results indicate that the incorporation of the outlier detection scheme can improve the localization accuracy by 13~30%. The outlier detection scheme and the localization system can thus pave a way for diverse WSN applications in automated surveillance, exploration, and context awareness.

Acknowledgment

References

[1] S. Phoha, T. LaPorta, and C. Griffin, Eds., *Sensor Network Operations*, John Wiley & Sons, 2006.

[2] R. Mannings, *Ubiquitous Positioning*, Artech House, 2008.

[3] A. Bensky, *Wireless Positioning: Technologies and Applications*, Artech House, 2008.

[4] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529–2553, 2007.

[5] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "Active badge location system," *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[6] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "Cricket location-support system," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM'00)*, pp. 32–43, August 2000.

[7] P. Bahl, A. Balachandran, and V. Padmanabhan, "RADAR: an in-building user location and tracking system," in *Proceedings of the IEEE Computer and Communications*, vol. 2, pp. 775–784, 2000.

[8] J. C. Juang, T. K. Weng, P. V. Cuong, C. Y. Hsu, Y. H. Lee, and S. H. Su, "Experimental assessment of wireless sensor network localization techniques," in *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision*, Hanoi, Vietnam, 2008.

[9] J. C. Juang, Y. C. Chen, W. C. Sun, C. L. Chu, and J. C. Wang, "Development of an RSS-based sensor network localization method: a kriging approach," in *Proceedings of the 5th Workshop on Ad Hoc and Sensor Networks*, Hsinchu, Taiwan, 2009.

[10] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 11, pp. 961–974, 2004.

[11] I. Borg and P. Groenen, *Modern Multidimensional Scaling, Theory and Applications*, Springer, New York, NY, USA, 1997.

[12] S. J. Benson, Y. Ye, and X. Zhang, "Solving large-scale sparse semidefinite programs for combinatorial optimization," *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 443–461, 2000.

[13] D. Niculescu and B. Nath, "DV Based positioning in Ad Hoc networks," *Telecommunication Systems*, vol. 22, no. 1–4, pp. 267–280, 2003.

[14] S. Ito and N. Kawaguchi, "Bayesian based location estimation system using wireless LAN," in *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom'05)*, pp. 273–278, March 2005.

[15] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

[16] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[17] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, 2002.

[18] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: a survey," *IEEE Communication Surveys & Tutorials*, vol. 12, no. 2, 2010.

[19] L. Davies and U. Gather, "The identification of multiple outliers," *Journal of the American Statistical Association*, vol. 88, pp. 782–801, 1993.

[20] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Online outlier detection in sensor data using non-parametric models," in *Proceedings of the 32nd International Conference on Very Large Databases*, 2006.

[21] D. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley & Sons, 1992.

[22] "Texas Instrument, CC2431:System-on-Chip (SoC) Solution for ZigBee/IEEE802.15.4 Wireless Sensor Network Datasheet," http://focus.ti.com/docs/prod/folders/print/cc2431.html.

[23] "ZigBee Alliance, ZigBee Specifications," Tech. Rep. 053474r06, version 1.0, 2005, http://www.zigbee.org/.