RSSI-based Localization through Uncertain Data Mapping for Wireless Sensor Networks

Qinghua Luo, *Member, IEEE,* Yu Peng, *Member, IEEE,* Junbao Li, and Xiyuan Peng

***Abstract*—When localizing the position of an unknown node for wireless sensor networks, the Received Signal Strength Indicator (RSSI) value is usually considered to fit a fixed attenuation model with a corresponding communication distance. However, due to some negative factors, the relationship is not valid in the actual localization environment, which leads to considerable localization error. Therefore, we present a method for improved RSSI-based localization through uncertain data mapping (LUDM). Starting from an advanced RSSI measurement, the distributions of the RSSI data tuples are determined and expressed in terms of interval data. Then, a data tuple pattern matching strategy is applied to the RSSI data vector during the localization procedure. Experimental results in three representative wireless environments show the feasibility and effectiveness of the proposed approach.**

***Index Terms*—Wireless Sensor Networks; Localization; RSSI; Uncertain data mapping**

I. Introduction

THE localization information is of great importance to many systems applying wireless sensor networks (WSNs) [1], for instance, in navigation & emergency response, environmental monitoring, routing and topology control, etc. In general, localization methods can be classified into two groups. One is built on distance estimation, and the other is mapping-based localization. The former consists of two steps: first we perform the distance estimation via a certain technique, i.e., the Received Signal Strength Indicator (RSSI), Time of Arrival (TOA), Time Difference of Arrival (TDOA) or Angle of Arrival (AOA) technique [2]. Then, based on the distance estimation results, we can obtain the location of an unknown node via a specific localization calculation method, which can be the trilateral, maximum likelihood or min-max method [3]. The latter mapping-based method determines the location of an unknown node via a strategy that matches the measured RSSI value to a RSSI database, which is built in advance.

During the distance estimation step, the RSSI-based methods have the advantages of low cost, low power and accessibility, so they are commonly used in a diverse range of systems [2]. However, because of the uncertainty in the measured RSSI values, the accuracy of the distance estimation and localization is low. Although the uncertainty is considered during communication distance estimation [3-4], the localization calculation method may introduce an error, and thus the localization accuracy is not high. Other researchers have also conducted research on mapping-based localization and have proposed methods such as the BP-based [5], SVM-based [6], RADAR [7] and ARIADNE [8] methods. However, the uncertainty of RSSI values is not considered in these works, and the artificial neural network (ANN) and support vector machine (SVM)-based methods [6] are problematic because of their high complexity.

Although the RSSI data have the characteristics of uncertainty in the actual localization environment, through measured RSSI values, we can know that the distributions of the RSSI data for each location point share the same homogenous statistical characteristics. Based on this, and considering the uncertainty of RSSI data, previous work in the literature [2] proposed its distribution in terms of interval data and obtained highly accurate distance estimation through a clustering algorithm, which achieved a significant improvement in the RSSI-based distance (RSSI-D) estimation accuracy. Furthermore, the estimation result was applied to a range-based WSN localization algorithm, resulting in a higher accuracy of the static localization estimation [9].

Because of the problem of low localization accuracy due to uncertain RSSI values and the high complexity of the existing localization methods, it is desirable to find a more effective way to overcome the uncertainty of RSSI values and achieve better WSN localization results. Therefore, in this paper, we present a mapping-based localization method via uncertain data processing. The technical contributions are as follows.

- We consider the uncertainty in the RSSI values and evaluate and express the uncertainty in terms of interval data.

Submitted data: 2015-6-26. This work was supported in part by the National Natural Science Foundation of China (61102038), the Natural Science Foundation of Shandong Province of China (ZR2015FM027), the Space Support Technology Fund Projects (2014-HT-HGD5), the Research Fund of Harbin Institute of Technology (WeiHai) (HIT (WH) 201306, HIT (WH) 201307), the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ14205,YQ15203), the Natural Scientific Research Innovation Foundation in Harbin Institute of Technology (HIT.NSRIF.2015122) and the State Key Laboratory of Geo-information Engineering(SKLGIE2014-M-2-4).

Qinghua Luo is with the School of Information and Electrical Engineering, Harbin Institute of Technology at WeiHai, WeiHai, China (luoqinghua081519@163.com). He is also with GuangXi Key Laboratory of Automatic Detecting Technology and Instruments (GuiLin University of ElectronicTechnology), State Key Laboratory of Geo-information Engineering, and State Key Laboratory of Satellite Navigation Engineering Technology.

Yu Peng is with the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China (pony911@163.com).

Junbao Li is with the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China (junbaolihit@126.com).

Xiyuan Peng is with the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China(pxy@hit.edu.cn)

- To improve the localization accuracy, we utilized a strategy that incorporates uncertain data clustering during the matching.
- To improve the localization efficiency, we determine the location of an unknown node directly using a mapping-based uncertain data matching algorithm.

The remainder of the paper is structured as follows: Section II reviews related work. Section III introduces the uncertain data expression, including the related definitions and the distance computation method in terms of the interval data; Section IV describes the mapping-based WSN localization method using uncertain data matching and its implementation; we evaluate the performance of the localization method in Section V; and Section VI provides concluding remarks.

II. Related Work

Localization methods can be classified into two types: model-based methods and mapping-based localization methods. The former are flexible, but they are sensitive to the uncertainty of the environment and measured parameters, which leads to low localization accuracy, whereas the latter can overcome the uncertainty of the environment and measured parameters to improve the localization accuracy. However, the parameters measurement should be done in advance in the specific localization field.

Model-based: Methods of this type are composed of two steps. First, based on RSSI, TOA, TDOA and AOA information [2], model-base methods are utilized to obtain distance estimation results. Then, one of the localization calculation algorithms [2] (including trilateral, maximum-likelihood, min-max, weighted min-max and other related algorithms) can be employed to obtain the localization results. The advantage of methods of this type is that they are flexible, i.e., they can be applied in arbitrary unknown environments. However, because the distance estimation model is difficult to implement accurately in practice, the distance estimation accuracy and localization accuracy are very poor. Although the uncertainty of RSSI data and the distance estimation model are considered in many research studies and many have proposed improved distance estimation methods and localization algorithms, the localization accuracy is not adequate. E.g., an angle estimation-based WSN localization method was proposed [10]. In the literature [11], polynomial modeling and trajectories were first applied to estimate the distance and then to conduct the localization calculation. The simulation results indicated higher localization accuracy, but the method was not validated in real localization environments. A bias-reduction localization method that mixes Taylor series and a maximum-likelihood estimate was proposed [12] and validated in simulation and in a SDR platform. Hong Shen et al. presented a TOA-based multiple source localization method, and utilized joint optimization to enhance the efficiency [13]. Joe et al. presented an improved version of the trilateration localization algorithm using a dynamic circle expanding mechanism [14], but the resulting improvement was not significant.

Mapping-based: Due to the uncertainty of the RSSI, TOA, TDOA, AOA and distance estimation models, the distance estimation accuracy and localization accuracy are very low. To obtain satisfying distance estimation and localization results, mapping-based methods have been proposed. E.g., [9] presented a mapping-based distance estimation method and validation of the method in real localization environments, and an improved version for achieving higher efficiency has been presented [2]. In addition to the static distance estimation, a mapping-based dynamic distance is also considered. The DDEUDC method can obtain an accurate estimation efficiently [15]. For obtaining localization results, the RADAR mapping-based method was presented [7]. The ARIADNE algorithm utilizes a search strategy to achieve higher efficiency [8]. However, these works did not consider the uncertainty of RSSI data. Boon et al. presented a hybrid RF-mapping and Kalman filtered spring relaxation based localization [16], in which the RF-mapping provides the initial position information, and the Kalman filter is utilized to refine the position information; however, the localization model should be accurate. An Isomap and partial least squares method to improve localization has also been proposed [17], but it was not evaluated in an actual localization system. In addition, machine learning-based methods, including the BP-based [5] and SVM-based [6] methods, have been proposed for accurate distance estimation and precise localization, but the complexity of the system training is very high.

In this paper, to improve the localization accuracy and efficiency, we consider the uncertainty of RSSI values, discard the two-step localization strategy, and utilize mapping to localize an unknown node directly. We first apply distribution-based mapping to represent the RSSI-Localization relationship. Then, a match strategy is explored to determine the location with high efficiency.

III. Related Equation And Definitions

In this paper, uncertain RSSI values are represented in the form of interval data, which is a powerful tool to express uncertain data. Some definitions related to interval data are provided as follows.

1) **Interval data** [18-19]: For two real numbers A_L, $A_R \in R$ and $A_R \geq A_L$, the set $A = [A_L,\ A_R] = \{u \mid A_L \leq u \leq A_R\}$ is interval data, where A_L is the lower bound of the interval data, and A_R is the higher bound. If $A_R = A_L$, the higher bound is equal to the lower bound, and the interval data becomes exact data.

2) **Mid-point and width:** [18-19]: For interval data $A = [A_L,\ A_R]$, we define $r_A = (A_R - A_L)/2$ and have the following relationship:

$$A_L = m_A - r_A\,,\ A_R = m_A + r_A \tag{1}$$

Where m_A and r_A ($r_A \geq 0$) are the mid-point and width, respectively, of interval data A. Therefore, the interval data A can be represented in the following form: $[\,m_A - r_A\,,\ m_A + r_A\,]$.

3) **Interval data vector (multi-dimensional interval data) [20]:** For an arbitrary dimension, element A_i ($1 \leq i \leq m,\ i \in \mathrm{N}$)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

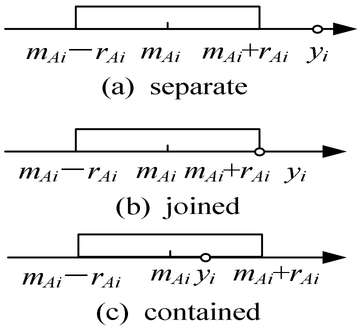

Fig. 1. Relationship between the interval data and the exact data.

of an interval data vector $X=\{A_1,A_2,...,A_i,...,A_m\}$ is represented as $A_i = [m_{A_i}-r_{A_i}, m_{A_i}+r_{A_i}]$.

As we perform localization based on the exact RSSI data tuple during the WSN localization procedure, we need to define the dissimilarity between the exact data vector and the interval data vector as follows.

4) **Dissimilarity calculation between the interval data vector and the exact data vector**: For an interval data vector $X=\{A_1,A_2,...,A_i,...,A_m\}$ and an exact data vector $Y=\{y_1,y_2,...,y_i,...,y_m\}$, each element of interval data can be expressed as $A_i=[m_{Ai}-r_{Ai}, m_{Ai}+r_{Ai}]$, where m_{Ai} , r_{Ai} , $y \in$ R . The distance relationship in each dimension between the two data vectors is illustrated in Fig. 1. When A_i is separated from y_i, as shown in (a), the minimum distance is $|m_{Ai}-y_i|-r_{Ai}$, and the maximum distance is $|m_{Ai}-y_i|+r_{Ai}$; when the two data vectors are joined, as illustrated in (b), the minimum distance is 0, and the maximum distance is $|m_{Ai}-y_i|+r_{Ai}=2*r_{Ai}$; when the interval data contains the exact data, as shown in (c), the minimum distance is 0, and the maximum distance is $|m_{Ai}-y_i|-r_{Ai}$. Therefore, the maximum distance $d_{i,\max}$ and the minimum distance $d_{i,\min}$ between A_i and y_i in dimension i can be calculated as

$$d(X,Y)_{i,\min}=\begin{cases}|m_x-y_i|-r_x, |m_x-y_i|-r_x\geq 0\\ 0, |m_x-y_i|-r_x<0\end{cases} \quad (2)$$

$$d(X,Y)_{i,\max}=|m_x-y_i|+r_x \quad (3)$$

To perform the data matching analysis, we introduce a correlation factor λ [21], where 0 ≤ λ ≤ 1, and utilize it to combine these two distance extremes to calculate the distance $d(X,Y)_i$ as

$$d(X,Y)_i=\lambda * d(X,Y)_{i,\min}+(1-\lambda)*d(X,Y)_{i,\max} \quad (4)$$

Thus, the distance $d(X,Y)$ between the interval data vector X and the exact data vector Y can be calculated as

$$d(X,Y)=\sqrt{\sum_{i=1}^{m}(d(X,Y)_i)^2} \quad (5)$$

In this way, the uncertainty and statistics of the measured RSSI values can be captured in terms of interval data. Furthermore, the uncertainty is also considered during the

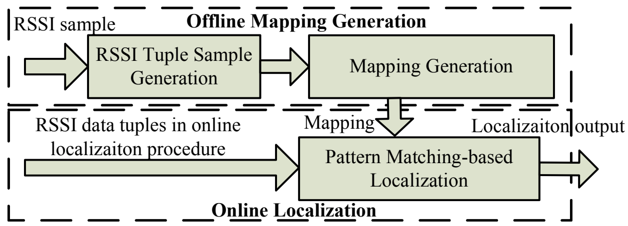

Fig. 2. The architecture of LUDM.

distance metric computation, which could result in it being more comprehensive and reasonable during localization.

IV. LUDM: Localization Using Uncertain Data Mapping

A. Architecture of LUDM

The localization system is composed of a set of unknown nodes and anchor nodes with localization information (X_m, Y_m), respectively, where m =1, 2, 3, ..., M . Suppose that the unknown node is in the communication range of some anchors (at least three). The RSSI value can be measured through the unknown node sending a localization request message to the surrounding anchor nodes. All the surrounding anchor nodes measure the RSSI data and then respond by returning it to the unknown node. The localization in our paper can be achieved by the system and steps detailed in the following.

Fig. 2 illustrates the architecture of LUDM. It comprises two main components: The **Offline Mapping Generation** module and the **Online Localization** module. The **Offline Mapping Generation** module comprises the **RSSI Tuple Sample Generation** sub-module and the **Mapping** Generation sub-module. Differing from the RADAR method [7], we express the RSSI-Location in terms of the distribution information of RSSI data tuples and interval data. Thus, we can improve the localization accuracy.

RSSI Tuple Sample Generation: To obtain the necessary RSSI data tuple sample and corresponding location information dataset, we perform RSSI sampling measurements with specific times at each localization point in the wireless localization application system.

Mapping Generation: Based on the RSSI data tuple sample dataset and the corresponding location information dataset, we conduct a statistical calculation to obtain the distribution information of the RSSI data tuples and represent the RSSI data tuple-Location mapping in terms of the distribution of RSSI data tuples at each localization point.

Pattern Matching-based Localization: For an RSSI data tuple measured by a wireless sensor node (e.g., a CC2530 WSN node), we implement a pattern matching strategy on the RSSI data tuple to determine the nearest distribution and treat the location related to the distribution as the localization estimation result of the current RSSI data tuple.

In the pattern mapping strategy, the RSSI Tuple Sample Generation sub-module and Mapping Generation sub-module achieve Offline Mapping Generation before Localization, which is used to determine the distribution characteristics of the RSSI data tuples of each localization point.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

In this way, the negative impact of the uncertainty in RSSI values can be overcome, so the localization accuracy can be improved with low calculation complexity. A selected set of typical experimental results are chosen to validate the proposed method and to express the potential of the approach.

B. RSSI Sample Generation

In the localization field, we locate the unknown node at the nth location sample point (x_n , y_n), (n =1, 2, 3, ..., N), and collect J ($J \in \mathbf{N}, J > 0$) sets of the RSSI data between the unknown node and the surrounding m ($3 \le m \le \mathrm{M}$) anchor nodes. Then, we can form the RSSI data and location information tuple: ($RSSI^1_{j,n}$, $RSSI^2_{j,n}$, $RSSI^3_{j,n}$,…, $RSSI^i_{j,n}$, …, $RSSI^m_{j,n}$,(x_n , y_n)), where, i denotes the serial number of the anchor nodes, with $3 \le i \le m$, j denotes the serial number of the RSSI sample data tuple, with $j \in \mathrm{N}$, $1 \le j \le J$, and n denotes the serial number of the location sampling point. To obtain the distribution of the RSSI data for each sampling point, we set the value of J to 150.

C. Mapping Generation

For the RSSI data tuples of each location sampling point, we perform statistical calculations on them and obtain the distribution information of the RSSI data tuples of each location point, i.e., the mean $RSSI\_u$ and standard deviation $RSSI\_d$ in (6) and (7). Then, we can express the RSSI distribution-based mapping in terms of interval data, i.e., for the kth location sample point, the mapping is follows.

A_n ={ [$RSSI^1_n\_u - k * RSSI^1_n\_d$, $RSSI^1_n\_u + k * RSSI^1_n\_d$], [$RSSI^2_n\_u - k * RSSI^2_n\_d$, $RSSI^2_n\_u + k * RSSI^2_n\_d$], …, [$RSSI^i_n\_u - k * RSSI^i_n\_d$, $RSSI^i_n\_u + k * RSSI^i_n\_d$], …, [$RSSI^m_n\_u - k * RSSI^m_n\_d$, $RSSI^m_n\_u + k * RSSI^m_n\_d$], (x_n , y_n)}, where the k is the coverage factor [2], ($k \in \mathrm{R}, 0 \le k \le 3$).

$$RSSI^i_n\_u = \frac{1}{J}\sum_{j=1}^{J} RSSI^i_{j,n} \tag{6}$$

$$RSSI^i_n\_d = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(RSSI^i_{j,n} - RSSI^i_n\_u)^2} \tag{7}$$

The distribution-based mapping can be utilized to overcome the uncertainty in RSSI values and improve the accuracy of the RSSI-d estimation.

D. Matching-based Estimation

During the process of online localization, we measure the RSSI data tuple between the unknown node and the anchor nodes: T = (R_1 , R_2 , R_3 , ..., R_i , ..., R_m). We calculate the distance (here, we adopt the Euclid distance) d_i between T and A_n according to formula (2) - (5). Thus, the distance $d_n(A_n, T)$ can be computed as follows.

$$d_n(A_n,T) = \sqrt{\sum_{i=1}^{m}(\lambda d(A_n,T)_{i,\max} + (1-\lambda)d(A_n,T)_{i,\min})^2} \tag{8}$$

$$d(A_n,T)_{i,\max} = \max(0, |R_i - RSSI^i_n\_u| - k * RSSI^i_n\_d) \tag{9}$$

$$d(A_n,T)_{i,\min} = |R_i - RSSI^i_n\_u| + k * RSSI^i_n\_d \tag{10}$$

where λ [5] is the correlation factor, and $0 \le \lambda \le 1$. It is used to combine the two distance extremes to calculate the distance, and thus, we can express all the possible values of the distance $d_n(A_n,T)$. Finally, we select $d_r(A_r,T)$ = min{ $d_n(A_n,T)$ }, $1 \le r \le m$, and treat the location (x_r , y_r) of the sampling point related to $d_r(A_r,T)$ as the localization result ($\hat{x}_T$, $\hat{y}_T$). The pseudocode of the matching-based estimation process is described in the following.

Algorithm 1 LUDM

Input: $T = (R_1, R_2, R_3, ..., R_i, .., R_m)$, { A_n }
($0 \le n \le N$, $0 \le j \le m$)

Output: ($\hat{x}_T$, $\hat{y}_T$)

Loop
  If (input T)
    For j =1 to m
      Calculate $d_n(A_n,T)$
    End for
    $d_r(A_r,T)$=min{ $d_n(A_n,T)$ }
    ($\hat{x}_T$, $\hat{y}_T$)=(x_r , y_r)
    Return ($\hat{x}_T$, $\hat{y}_T$)
  Else
    Continue
  End If
End Loop

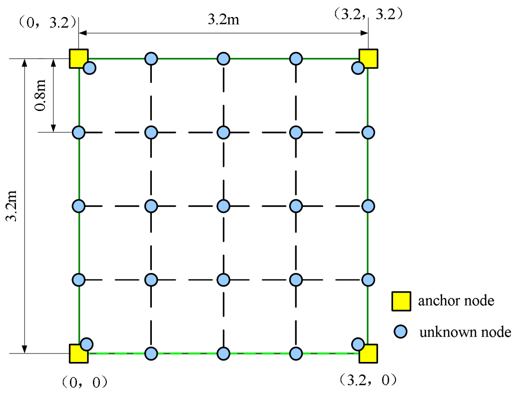

Fig. 3. The test bed for the localization method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

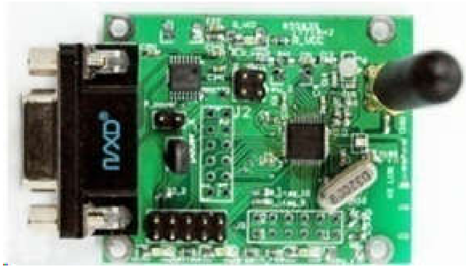

Fig. 4. The sensor node adopted in the test bed

V. Performance Evaluation

The mapping-based wireless sensor localization method LUDM will be validated and evaluated in an actual localization environment. First, we conduct a feasibility evaluation of the localization method. Second, we evaluate the performance of LUDM in terms of the localization accuracy and localization efficiency and make comparisons with other related methods.

A. Evaluation Setting

1) Evaluation Conditions

To comprehensively validate the localization method, we evaluate the performance of the LUDM method in three typical localization environments, i.e., indoors, hall and open air. We set up a test bed for the mapping-based localization method. In a 3.2 m×3.2 m field, four sensor nodes are deployed in the four corners of the field to work as anchor nodes with known location information, and an unknown node is located at different points (25 points), as shown in Fig. 3. We set the value of N to 25, the value of J to 200, and the value of M to 4.

The wireless sensor node adopted in the test bed is the CC2530 shown in Fig. 4. To eliminate the negative effect of different heights of the sensor nodes, we set the height of each sensor nodes to 0.1 m.

2) The Experimental Dataset

In the test bed field, for each of the 25 localization points of the unknown node, the RSSI values of communications between each anchor node and the unknown node are sampled 200 times, of which 150 RSSI sample values are used to generate the mapping, and the other 50 sample values are used to evaluate the RSSI-based localization methods.

A statistical calculation is applied to the first part of the sample RSSI data tuple set to obtain the distribution information of the RSSI data tuples on these 25 localization points. Then, the map between the RSSI data tuple and the location is generated. Finally, the last 50 RSSI sample data tuples of the 25 localization points are utilized to evaluate the RSSI-based localization method.

3) Evaluation Indexes

The performance of the mapping-based localization method is validated in terms of its localization accuracy and localization efficiency. The absolute mean localization error (e_{MAE}) is adopted as an accurate index, which can be calculated according to (11).

$$e_{MAE}=\frac{1}{N}\left(\sum_{n=1}^{N}\sqrt{(x_n-\hat{x}_n)^2+(y_n-\hat{y}_n)^2}\right) \quad (11)$$

where e_{MAE} is the mean absolute localization error, ($\hat{x}_n$, $\hat{y}_n$) is the localization result of the unknown node, (x_n, y_n) is the reference location of unknown node, n is the serial number of the localization point, and N is the total number of localization experiments.

The smaller the e_{MAE} is, the more accurate the localization result becomes.

The estimation efficiency is evaluated according to the processing time and the computational time complexity. The processing time is expressed in terms of the system training time T_t, statistical computing time T_s, localization computation time T_L and total processing time T_T. The smaller these indexes are, the more efficient the method is.

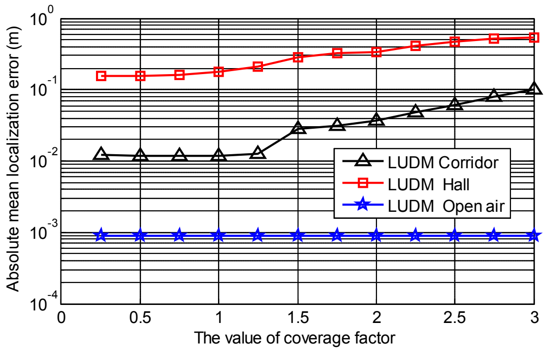

Fig. 5. The affection evaluation of coverage factor.

4) Experiment Design

First, we conduct the validation of the localization method LUDM, and second, we evaluate the efficiency of the LUDM method in terms of the processing time and computational complexity compared with other related methods. Third, the LUDM method is implemented to perform localization in a test field, and the estimation accuracy is assessed and analyzed under different conditions.

B. Validation of the LUDM Localization Method

In this section, we will conduct validation of the LUDM localization method, i.e., we evaluate the effects of the coverage factor k and the correlation factor λ, and determine the appropriate set values of these parameters.

1) Evaluation of the Effect of the Coverage Factor

In the three representative localization environments, we adopt different values of the coverage factor k, apply the LUDM method to conduct localization, and compute the mean absolute localization error e_{MAE} of the 25 localization points, as illustrated in Fig. 5. During experiment, the value of the correlation factor takes the value of 0.5.

Fig. 5 illustrates that the absolute mean localization error e_{MAE} increases with the coverage factor in complex localization environments, whereas in the open air localization, the absolute mean localization error e_{MAE} is very small with little deviation. This results mainly due to the many uncertain factors, which lead to a high localization error. Therefore, the appropriate set

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

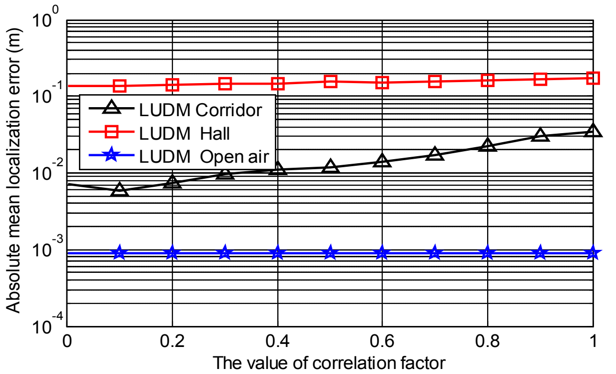

Fig. 6. Affection evaluation of the correlation factor.

value of the coverage factor k should be determined through experiment.

In this paper, from Fig. 5 and the analysis above, we set the value of the coverage factor to 1, 0.25 and 0.25 in the corridor, hall and open air, respectively.

2) Evaluation of the Effect of the Correlation Factor

Based on the determination of the set values of the coverage factor, we conduct an impact analysis of the correlation factor λ, i.e., we apply the LUDM method to perform localization with different values of the correlation factor λ and compute the mean absolute localization error of the 25 localization points. We show the localization error in the three typical environments in Fig. 6.

TABLE I
COMPUTATIONAL TIME COMPLEXITY ANALYSIS OF LUDM

Processing	Computational Time Complexity
RSSI Tuple Sample Generation	$O(NJ)$
Statistical Computational	$O(NJ)$
Mapping Expression	$O(N)$
Matching-based Localization	$O(MN)$
Total	$O(NJ)+O(NJ)+O(N)+O(MN)$

From Fig. 6, we can see that correlation factor plays an important role in the localization performance in the corridor, while it has no effective impact on the localization performance in the other two types of environments. We can set the value of the correlation factor through experimental evaluation. From Fig. 6 and the analysis above, we set the value of the correlation factor to 0.1, 0 and 0 in the corridor, hall and open air, respectively.

C. Evaluation of the LUDM Localization Efficiency

We evaluate the efficiency of the LUDM method in terms of computational time complexity analysis and the processing time and compare the results with those of other related localization methods.

1) Computational Time Complexity Analysis

The mapping-based localization method LUDM is composed of the mapping generation part and the on-line localization part. In the RSSI-Location mapping generation, suppose the number of localization points is N, the number of samples for each localization point is J, and the number of anchor nodes is M. Then the computational time complexity of the localization method LUDM is shown in Table I.

TABLE II
COMPUTATIONAL TIME COMPLEXITY OF DIFFERENT LOCALIZATION METHODS

Localization methods	Computational Time Complexity
DEDF-based	$O(NJ)$
DEUDC-based	$O(NJ)$
BP-based	$O(m^3k^3)$
LS-SVC-based	$O(m^3k^3)$
RADAR	$O(NJ)$
LUDM	$O(NJ)$

From Table I, we observe that the computational time complexity of the **RSSI Tuple Sample Generation** is $O(NJ)$, that of the **statistical computation** including the means and standard deviation is $O(NJ)$, that of the **mapping expression** is $O(N)$, and that of the **matching-based localization** is $O(MN)$. Therefore, the total computational time complexity is $O(NJ)+O(NJ)+O(N)+O(MN)$, and when $J \gg M$, $N \gg M$, the total computational time complexity is simplified as $O(NJ)$.

To evaluate the efficiency in terms of the computational time complexity, as shown in Table II, we analyze and compare the computational time complexity of these localization algorithms, including the DEDF-based [22], DEUDC-based [2], BP-based [5], LS-SVC-based [6], and RADAR [7] algorithms.

In Table II, N is the number of localization points, and J is the number of samples for each sample distance point. Table II, shows that the computational time complexity of the BP and LS-SVC methods are largest among these algorithms, which is explained by the fact that the two algorithms require more computations when training the model. The table also indicates that the computational time complexity of the LUDM method proposed in this paper is the lowest of all the methods.

2) Processing Time Evaluation

The efficiency of the LUDM method is analyzed according to the system training time T_t, statistical computing time T_s, estimating time T_e and total processing time T_T, and the results is compared with those of the other related localization methods, including the DEDF-based [22], DEUDC-based [2], BP-based [5], LS-SVC-based [6], and RADAR [7] methods, as shown in Table III (based on Matlab 2011b).

In the DEDF-based [22] and DEUDC-based [2] localization methods, the communication distance estimation between the anchor nodes and the unknown node is conducted first through an uncertain or certain data clustering algorithm, respectively. Then, based on the distance estimation results, the localization computation algorithm is utilized to obtain the localization result of the unknown node. The RADAR method obtains the localization result directly using a RSSI-Localization map, which is built in advance. As opposed to the RADAR method, the LUDM method considers the uncertainty in the RSSI tuples

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

TABLE III
PROCESSING TIME OF DIFFERENT LOCALIZATION METHODS

Localization method	T_t(s)	T_s(s)	T_e(s)	T_T(s)
DEDF-based	0	2.050E-3	1.501E-3	3.551E-3
DEUDC-based	0	1.300E-2	1.506E-3	1.450E-2
BP-based	2.737E1	0	8.842E-2	2.745E1
LS-SVC-based	1.005E2	0	7.826E-1	1.012E2
RADAR	0	6.66E-04	2.400E-3	3.066E-3
LUDM	0	7.800E-4	3.200E-3	3.980E-3

and takes measures to process it.

Analysis of the System Training Time

Table III indicates that the machine-learning-based localization methods (including BP-based and LS-SVC-based) require a specified processing time during systematic training. On the contrary, there is no system training time for the DEDF-based, DEUDC-based, RADAR and LUDM mapping-based methods, so the training time is 0 s.

Analysis of the Statistical Calculation Time

From Table III, we can see that the statistical calculation time of the RADAR and LUDM methods is the same and is less than that of the DEDF-based and DEUDC-based localization methods, whereas that of the BP-based and LS-SVC-based methods is zero because these machine learning methods do not need statistical computation.

TABLE IV
THE ABSOLUTE MEAN ERROR OF LOCALIZATION METHODS IN THREE TYPICAL ENVIRONMENTS

Localization methods	Corridor	Hall	Out door
DEDF-based	0.6255 m	0.3951 m	0.1698 m
DEUDC-based	0.524 m	0.4069 m	0.1785 m
BP-based	0.1324 m	0.1743 m	0.0445 m
LS-SVC-based	0.0953 m	0.1547 m	0.0481 m
RADAR	0.0124 m	0.1532 m	0.0009 m
LUDM	0.0058 m	0.1373 m	0.0009 m

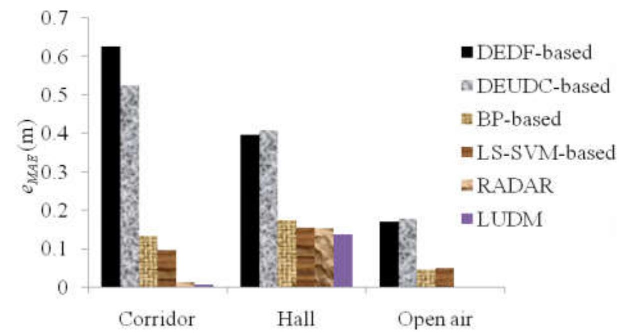

Fig. 7. The absolute mean localization error of methods in test fields.

Analysis of the Localization Time

Table III illustrates that the DEDF-based, DEUDC-based, RADAR and LUDM localization methods require less estimation time, whereas the BP-based and LS-SVC-based localization methods need more time when performing localization. This is because the DEDF-based, DEUDC-based, RADAR and LUDM localization methods just need a simple calculation with the parameters determined in step (2) to determine the location of the unknown node. However, more complex calculations are required in the BP-based and LS-SVRM-based methods when conducting localization, so these two methods need more localization time.

Analysis of the Total Processing Time

From Table III, we can see that the total processing of the LUDM localization method is less than that of the BP-based and LS-SVC-based localization methods. The total processing time of the LUDM method is 1/6897 and 1/25427 of that of the BP-based and LS-SVC-based localization methods, respectively, and it has almost the same processing time as the DEDF-based, DEUDC-based and RADAR localization methods. Therefore, the LUDM method has less processing time and can enhance the efficiency to a large degree relative to the other methods, and this result is in accord with the results of the analysis in part C of section V.

D. Evaluation of the LUDM Localization Accuracy

For evaluation of the localization accuracy, some typical related methods are utilized to perform localization in different test fields. These methods include the DEDF-based [22], DEUDC-based [2], BP-based [5], LS-SVC-based [6], RADAR [7] and LUDM methods. The absolute mean localization error e_{MAE} according to statistical methods is shown in Table IV and Fig. 7.

Accuracy analysis of the localization methods

Table IV and Fig. 7 illustrate that the LUDM method has more accurate localization results than the other methods on average. Relative to the DEDF-based and DEUDC-based localization methods, the LUDM method reduced the localization error by 296 and 289 times, respectively, on average under different test bed conditions because these two localization methods require a localization calculation, which may introduce more error. Relative to the BP-based, LS-SVM-based and RADAR localization methods, the LUDM method reduced the localization error by 71, 68 and 1 times, respectively on average in the three typical environments. This is because the BP-based, LS-SVM-based and RADAR localization methods do not consider the uncertainties in the RSSI values, which leads to greater errors.

Analysis of environmental impact on the accuracy

From Table IV and Fig. 7, we see that localization accuracy of the LUDM method is lowest in the hall and is highest in open air. This agrees with the results in Fig. 5 and Fig. 6. The reason for this trend is that there are less uncertain factors in open air. The results of some of the other mapping-based localization methods including the BP-based, LS-SVC-based, and RADAR methods also follow this trend.

In brief, in terms of the overall performance, the LUDM method obtains higher localization accuracy with lower calculation complexity, so it is appropriate for precise RSSI-based localization applications.

E. Discussion of the Generality of LUDM in WSN Application Systems

The LUDM localization method can be a more accurate localization method with low computational complexity. It should be noted that the Mapping Generation module is necessary for accurate localization. To improve the generality

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2524532, IEEE Sensors Journal

in unknown localization environments, we can resort to the RSSI attenuation model, but the localization accuracy may be very low.

VI. Conclusion

For the problem of accurate RSSI-based localization in wireless sensor networks, we present a mapping-based localization method named LUDM. In this method, distribution-based mapping is utilized to overcome the uncertainty in RSSI values. Then, a RSSI tuple matching algorithm is implemented to determine the localization result. We also perform further validation in a real WSN localization test bed under different conditions. Through experimental results, it is demonstrated that the method yields more accurate RSSI-based localization with higher efficiency, especially in environments with a high level of uncertainty (e.g., the corridor and hall). This method is competitive in WSN localization systems for obtaining better localization accuracy.

Acknowledgment

The research presented in this paper is supported by the National Natural Science Foundation of China (61102038), the Natural Science Foundation of Shandong Province of China (ZR2015FM027), Space Support Technology Fund Projects (2014-HT-HGD5), the Research Fund of the Harbin Institute of Technology (WeiHai) (HIT (WH) 201306, HIT (WH) 201307), the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ14205,YQ15203), the Natural Scientific Research Innovation Foundation of the Harbin Institute of Technology (HIT.NSRIF.2015122) and the State Key Laboratory of Geo-information Engineering (SKLGIE2014-M-2-4).

References

[1] A. Vempaty, O. Ozdemir, K. Agrawak, et al, "Localization in Wireless Sensor Networks: Byzantines and Mitigation Techniques," *IEEE Trans. on Signal Process.*, vol. 61, no. 6, pp. 1495-1508, Jun., 2013.

[2] Q. H. Luo, Y. Peng, X. Y. Peng, Abdulmotaleb EI Saddik, "Uncertain Data Clustering-based Distance Estimation in Wireless Sensor Networks," *Sensors*, vol. 14, no. 4, pp. 6584-6605, April., 2014.

[3] M. F. Snoussi, H. Richard, "Interval-Based Localization using RSSI Comparison in MANETs," *IEEE Trans. Aerospace and Electronic Systems.*, vol. 47, no. 4, pp. 2897-2910, Apr., 2011.

[4] T. Stoyanova, F. Kerasiotis, A. Prayati, G. Papadopoulos, "Evaluation of impact factors on RSS accuracy for localization and tracking applications," in *MobiWac'07*, October, 22, 2007, Chania, pp. 9-17.

[5] X. F. Zhang, Q. M. Cui, Y. L. Shi, X. F. Tao, "Robust localisation algorithm for solving neighbour position ambiguity," *Electronics Letter*, vol. 49, no. 17, pp. 1106-1107, Aug., 2013.

[6] W. Kim, J. Park, J. Yoo, H. J. Kim, C. G. Park, "Target Localization Using Ensemble Support Vector Regression in Wireless Sensor Networks," *IEEE Trans. Cybernetics.*, vol. 43, no. 4, pp. 1189 - 1198, Apr., 2013.

[7] P. Bahl, V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," In *IEEE INFOCOM 2000*, Tel-Aviv, Israel, March, 2000, pp. 775 - 784.

[8] Y. Ji, S. Biaz, S. Pandey, P. Agrawal, "ARIADNE: A dynamic indoor signal map construction and localization system," In *ACM MobiSys'06*, Uppsala, Sweden, June, 2006, pp. 151-164.

[9] Y. Peng, Q. H. Luo, X. Y. Peng, "WSN Localization Method Using Interval Data Clustering,"*Acta Automatica Sinica*, vol. 38, no. 7, pp. 1190-1199, July, 2012.

[10] W. L. Zhang, Q. Y. Yin, H. Y. Chen, et al, "Distributed Angle Estimation for localization in wireless sensor networks," *IEEE Transactions on Wireless communications*, vol. 12, no. 2, pp. 527-5367, Feb., 2013.

[11] P. K. Sahu, E. H. K. Wu, J. Sahoo, "DuRT: Dual RSSI Trend based localization for wireless sensor networks," *IEEE Sensors Journal*, vol. 14, no. 3, pp. 3115-3123, Mar. 2013.

[12] Y. M. Ji, C. B. Yu, J. M. Wei, et al, "Localization bias reduction in wireless sensor networks," *IEEE Transactions on Inductrial Electronics*, vol.62, no. 5, pp. 3004-3016, May, 2015.

[13] H. Shen, Z. Ding, S. DasGupta, et al, " Multiple Source Localization in Wireless Sensor Networks Based on Time of Arrival Measurement," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 1938-1949, August, 2014.

[14] J. A. Jiang, X. Y. Zheng, Y. F. Chen, et al, "A Distributed RSS-Based Localization Using a Dynamic Circle Expanding Mechanism," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3754-3766, October, 2013.

[15] Q. H. Luo, X. Z. Yan, J. B. Li, Y. Peng, "DDEUDSC：Dynamic distance estimation using uncertain data stream clustering in mobile wireless sensor networks," *Measurement*, vol. 55, no.40, pp. 423-433, Sept., 2014.

[16] B. C. Seet, Q. Zhang, C. H. Foh, et al, "Hybrid RF Mapping and Kalman filtered spring relaxation for sensor network localization," *IEEE Sensors Journal*, vol. 12, no. 5, pp1427-1445. 2012.

[17] B. Li, Y. G. He, F. M. Guo, L. Zuo, "A novel localization algorithm based on Isomap and partial least squares for wireless sensor networks," *IEEE Transactions on Instrumentation and Measuremen*t, vol. 62, no.2, pp. 304-314, 2013.

[18] S. J. Ren, "Interval Data-Based Uncertain Data Mining and Its Applications," Ph.D. Thesis, Institute of Intelligent systems and decision making, Zhejiang University, Hang Zhou, Zhejiang, China, 2006.

[19] Z. P. Qiu, "The Analysis Method of Static Responses for Uncertain Parameter Structural and the Problem of Eigen Value," Ph.D. Thesis, Calculation mechanics center, Jilin University of Technology, Siping, Jilin, China, 1994.

[20] Q. H. Luo, Y. Peng, X. Y. Peng, "Multiple Dimensional Uncertain data stream clustering algorithm," *Chin. J. Sci. Instrum.* vol. 34, no. 6, pp. 1330-1338, 2013.

[21] Habich, D.,Volk, P.B., Diittmann, R., Utzny, C., "Error-Aware Density-Based Clustering of Imprecise Measurement Values," In *the 23nd IEEE International Conference on Data Mining (ICDM 2007)*, Omaha, NE, USA, October, 2007, pp. 471–476.

[22] D. Wang, "Research on the Localization Method Based on RSSI for Wireless Sensor Network," MS. Thesis, the school of electrical engineering and automation, Harbin Institute of Technology, Harbin, Heilongjiang, China, 2011.

Qinghua Luo **(corresponding author)** received B.S. and M.Sc. degrees in communication and information systems from Harbin Engineering University (HEU), Harbin, China, in 2005 and 2008, respectively. He received a Ph.D. degree in instrument science and technology from Harbin Institute of Technology (HIT), Harbin, China, in 2013.

From 2011 to 2012, he was a Research Assistant with HongKong Polytechnic University. Since 2013, He has been an Assistant Professor with the school of Information and Electrical Engineering, Harbin Institute of Technology at WeiHai. He is the author of twenty nine articles and more than 20 inventions. His research interests include wireless sensor networks, uncertain data processing and fault diagnosis.

Yu Peng (M'10) received a B.S. degree in measurement technology and instrumentation, and M.Sc. and Ph.D. degrees in instrumentation science and technology from the Harbin Institute of Technology (HIT), Harbin, China, in 1996, 1998, and 2004, respectively. He is currently a Full Professor with the Department of Automatic Test and Control, School of Electrical Engineering and Automation, HIT, where he is also the Vice Dean of the School of Electrical Engineering and Automation. His current research interests include automatic test technologies, virtual instruments, system health management, and reconfigurable computing.