

Research Article

Nonparametric Bootstrap-Based Multihop Localization Algorithm for Large-Scale Wireless Sensor Networks in Complex Environments

Yongji Ren,^{1,2} Ning Yu,¹ Xiao Wang,¹ Ligong Li,¹ and Jiangwen Wan¹

¹ School of Instrumentation Science and Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics (Beihang University), Beijing 100191, China

² Department of Command, Naval Aeronautical and Astronautical University, Yantai 264001, China

Correspondence should be addressed to Yongji Ren; renyongji1006@sina.com

Received 8 November 2012; Revised 20 March 2013; Accepted 27 March 2013

Academic Editor: Wenzhong Li

Copyright © 2013 Yongji Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a nonparametric bootstrap multihop localization algorithm for large-scale wireless sensor networks (WSNs) in complex environments. Unlike most of the existing schemes, this work is based on the consideration that it is not feasible to obtain a lot of available distance measurements sample for estimation and to get exact noise distributions or enough prior information for conventional statistical methods, which is a situation commonly encountered in complex environments practically. For the first time, we introduce a nonparametric bootstrap method into multihop localization to build confidence intervals for multihop distance estimation, which can eliminate the risk of small sample size and unknown distribution. On this basis, we integrate the interval analysis method with bootstrap approach for ordinary nodes localization. To reduce the computational complexity, boxes approach is utilized to approximate the irregular intersections. Simulation results show that our proposed scheme is less affected by the variation of unknown distributions and indicate that our method can achieve high localization coverage with relatively small average localization error in large-scale WSNs, especially in sparse and complex network with smaller connectivity and anchor percentage.

1. Introduction

In the last few years, there has been a rapidly growing interest in extensive monitoring applications of wireless sensor networks (WSNs). One important reason is that, as opposed to traditional solutions, WSNs can be rapidly and randomly deployed in a large-scale monitoring region by plane or unmanned aerial vehicle (UAV), of which the environments of monitoring region are always complex even inaccessible.

As new technologies, WSNs can provide the means for long-term, all-weather, real-time, accurate, and extensive monitoring unattended, which brought us a convenient way to sense and monitor the complex environments. It is considered as an ideal system for this type of extensive environment monitoring and can fulfill the needs of various practical applications, for example, marine surveillance, ocean scientific exploration, and commercial exploitation.

For most WSNs applications, localization service is an indispensable part and an essential task. The location of the sensor nodes should be determined for meaningful interpretation of the sensed information. In recent years, a certain amount of research work has been conducted in this interesting research area, and a comprehensive survey is provided in [1–3] and the references therein. Some of the solutions are mainly designed for small-scale networks, and although this is also an interesting field, we will not contribute to the region in this paper; instead, we focus on the large-scale localization context.

As we know, out of the consideration of economic rationality, generally, it is too hard to deploy the sensor nodes in a large-scale area with relatively high density and large percentage of anchor nodes. If it is used for monitoring ocean environment, the network will be deployed sparsely and only a minority of anchor nodes can be employed for localization. In

these cases, it will be harder for ordinary nodes to communicate with enough anchor nodes directly so as to measure distances, which are necessary for localization.

To solve this problem, three types of localization schemes are proposed for large-scale localization, namely, mobile anchor assisted localization algorithms [4–6], recursive algorithms [7–9], and multihop algorithms [10–12], respectively. Hereinto, mobile anchor assisted localization methods commonly employ some expensive mobile equipment such as Autonomous Underwater Vehicle (AUV) which can roam across the monitoring region to assist localization. In recursive algorithms, some ordinary nodes that have been localized become secondary anchor nodes and broadcast coordinates to assist other nodes in estimating their locations. The localization process will inevitably suffer from the delay and adverse effects of error propagation and accumulation. Whereas for multihop localization, the ordinary node can infer the distances to its nonneighboring anchor nodes by approximating the length of the shortest path to the Euclidean distance, so as to get enough anchor nodes with known distances for localization. They not only can provide better real-time performance but also have no additional equipment requirement. In comparison, the characteristics of multihop methods are more suitable for large-scale WSNs localization. Therefore, multihop localization method has received more and more attention in recent years.

Certainly there are still many challenges for the multihop localization, especially in the complex environments. Here we define the complex environments as multiple complex terrain conditions (e.g., forest, rocks, marsh, underwater, etc.), random and sparse sensor node distribution, irregular radio propagation pattern, various unknown ambient noises, and multiple anisotropic network situations (e.g., H-type network, C-type network, etc.). There is no doubt that these adverse factors will seriously affect the accuracy of the multihop distance estimation which is the basis of the accurate sensor location estimation. To address this problem, most of the existing schemes have adopted conventional statistical methods and considered distance measurements affected by normal distribution noise and then determined the localization uncertainties through Monte Carlo analysis or nonlinear transformation techniques. For example, in [13], a dynamic localization scheme, the Monte Carlo localization (MCL), has been proposed. In [14], based on the discovery that the multihop distance errors obey normal distribution with various biases, a multihop localization algorithm that incorporates the distance estimation bias has been presented. But for the previous methods, only when the measurements sample size is large and the precondition that the errors obey normal distribution is satisfied, they could perform well in online localization.

However, for the large-scale WSNs in complex environment, it is too hard to get exact distribution characteristics of distance measurements and enough a priori information which is necessary for conventional statistical estimation method, and it is impossible to guarantee repeatable ranging conditions for reproducible measurements so as to obtain a large number of available measurements sample for Monte Carlo analysis or other similar methods. Hence, there is a

need to develop a novel algorithm for producing coordinates through multihop localization approach as only very few measurements are available. And it is better that this methodology only needs less prior information and even no statistical assumptions on disturbances.

On the basis of analysis on the challenges of complex environments and limitations of existing studies for large-scale localization, a different approach is proposed and investigated, that is, a multihop localization method based on nonparametric bootstrap for large-scale WSNs in complex environments.

The bootstrap is a powerful technique for assessing the accuracy of a parameter estimator in situations where conventional techniques are not valid. The nonparametric bootstrap method was originally introduced by Efron in [15] and used to address CI estimate for statistics based on independent and identically distributed (i.i.d.) random variable from some unknown distribution $F(\mu, \sigma)$ [16]. Compared with conventional techniques, the significant advantages of the nonparametric bootstrap method are that it does not require any modeling or assumptions on the data and it is more suitable for small sample estimation. In this paper, it can be utilized to build confidence intervals effectively for multihop distance estimation online. On this basis, with interval methods, we can compute the bounds of the possible solutions that correspond to measured quantities and determine the guaranteed regions that involve the correct solution.

This paper is organized as follows. Section 2 introduces the related works and challenges for large-scale WSNs localization. Section 3 presents in details the nonparametric bootstrap and interval analysis method. Section 4 evaluates the performance of our algorithm through experiments, and Section 5 concludes this paper.

2. Related Works

2.1. Mobile Anchor Assisted Localization Algorithms. To solve the localization problems of large-scale WSNs, in [4], the authors employ a single mobile beacon to aid in localization. The sensor locations are maintained as probability distributions that are sequentially updated using Monte Carlo sampling as the mobile anchor node moves over the monitoring area. This method relies on the more powerful anchor to perform the calculation and relieves most localization tasks from the less powerful ordinary nodes.

The authors in [5] proposed a range-free localization scheme for WSNs using mobile anchor nodes equipped with four directional antennas. Therein, each mobile anchor node can determine its position via Global Position System (GPS), and then it broadcasts its coordinates as it moves through the region. The ordinary nodes detect these messages and utilize a simple processing scheme to determine their own coordinates based on those of the anchors.

In the AUV-aided localization scheme proposed in [6], AUVs keep roaming across the underwater sensor field to aid in localization. The AUVs can get coordinates from GPS while floating periodically and dive into a fixed depth and navigate through a predefined route using compass and dead-reckoning. They can estimate the distance between the AUV

and ordinary nodes through a request/response message pair exchange which contains its coordinates. Therefore, the ordinary nodes can be localized after the message exchange from three different noncoplanar AUV locations.

Although these protocols exploit the mobility of the mobile equipment to overcome the lack of adequate anchor nodes, the major drawbacks of these mobile anchor assisted localization algorithms are that the mobile machine is too expensive for WSNs, and the slow speed of AUV or other machine always introduces high localization delay. In addition, the movement of the anchor nodes will be severely restricted by the complex environments.

2.2. Recursive Localization Algorithms. Liu and Zhang [7] presented an error control mechanism for recursive localization based on the characteristics of node uncertainty and the active selection strategy of anchor nodes. The error control mechanism only utilizes local knowledge and can mitigate the effect of error propagation for both range and directional sensors to a certain extent.

Yu et al. proposed a two-stage localization approach in [8]. Firstly, localization process starts from the nodes with the largest numbers of neighbor anchors which have the priority. Then, the coordinates of all neighbor nodes are exploited to improve localization accuracy. During the procedure, a number of measures are also taken to ensure the reliability of each location estimate to avoid abnormal errors and reduce error propagation.

Vemula et al. [9] formulated the sensor localization from a probabilistic point of view and incorporated anchor position uncertainty to estimate the distribution of node coordinates, including iterative least squares and Bayesian methods, Monte Carlo importance sampling, and cost-based methods.

These schemes try to inhibit the propagation and accumulation of localization errors, but the high computational complexity and increased communication cost limit their application in practice. Additionally, this kind of method is also easy to cause the localization delay.

2.3. Multihop Localization Algorithms. DV-distance and DV-hop algorithms, as the origination of multihop localization schemes for WSNs, are proposed in [10]. In both algorithms, each anchor node broadcasts a message to its immediate neighbor nodes firstly. Then, the message is propagated in a controlled flood manner so that each ordinary node can estimate the lengths of shortest paths to anchor nodes. When the ordinary node obtains the estimates to enough anchor nodes, its position can be calculated.

Wang and Xiao [11] presented an improved multihop algorithm called i-Multihop which has higher computational complexity. At the beginning, the upper bound constraints are used to filter out the incorrect distance estimations and the estimated position is pinpointed to the intersection constrained by the correct distances. And then, the distance fitting is used to fit correct distance measurements, which makes the final estimated position not to be affected by the layout of anchor nodes.

The authors in [12] proposed three multihop localization schemes based on least squares and multilateration, namely, Taylor-LS, weighted Taylor based least squares, and constrained total least squares, respectively. Additionally, a generalized Cramér-Rao lower bound is developed to analyze the performance of multihop localization approaches.

As we can see, most of the previous schemes considered the measurement uncertainty as normal distribution even zero-mean normal distribution noise and employ conventional error-processing approaches, for example, nonlinear least squares estimator (NLSE), and so forth. However, in complex large-scale WSNs, it is practically too hard to fulfill the requirements which are necessary for these approaches, such as enough a priori information, exact distribution characteristics, and large sample size.

Specifically, for the ordinary nodes, there are not so many anchors that can communicate with them to accomplish distance measurement. The ranging process also cannot be performed again and again due to the limited energy of sensor nodes. Even when cost is not a main concern, it is also impossible to guarantee repeatable conditions for reproducible distance measurement. Hence, it is not practically feasible to obtain a lot of available distance measurements sample for estimation. Moreover, in complex environments, for example, underwater environments, the ranging process will suffer from various non-Gaussian noises, for example, the multipath and waveguide effects, surface scattering, and so forth. Obviously, for the previous noises, it is too hard to get exact distribution characteristics and enough prior information which are necessary for statistical methods. It is also unreasonable to consider that the uncertainties always obey normal distribution. Hence, in this paper, the bootstrap localization method is proposed for large-scale WSNs in complex environments.

3. Nonparametric Bootstrap-Based Localization Algorithm for Large-Scale WSNs

To solve the problems mentioned earlier, in this paper, we propose a novel multihop localization scheme which integrates a nonparametric bootstrap distance estimation method with an interval analysis approach. We firstly utilize the nonparametric bootstrap method to provide a confidence interval (CI) as the estimation for distance measurement. On that basis, interval analysis approach is employed to determine the feasible set and search the optimal estimation of coordinates.

3.1. Short Review of CI Estimate. As we know, CI is an interval about the measurement result within which the values that could reasonably be attributed to the measurand may be expected to lie with a given level of confidence. A common approach to address CI estimate based on the normal distribution can be described as follows. Let X_1, X_2, \dots, X_n be n i.i.d. Gaussian random variables with known σ from distance measurements, and suppose that we wish to find a confidence level $1 - \alpha$ interval for the mean θ . As we know, the random variable $(\hat{\theta} - \theta)/(\sigma/\sqrt{N})$ obeys standard normal

distribution, and its probability density function (PDF) $f(x)$ can be computed by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (1)$$

and its cumulative distribution function (CDF) $F(z)$ can be computed by

$$F(z) = \int_{-\infty}^z f(x) dx. \quad (2)$$

For a given confidence level $1 - \alpha$, there exists constant $c > 0$ such that

$$\Pr \left\{ -c < \frac{\hat{\theta} - \theta}{\sigma/\sqrt{N}} < c \right\} = 1 - \alpha = 2F(c) - 1. \quad (3)$$

It follows that constant c satisfies $F(c) = 1 - (\alpha/2)$, and thus, $c = z^\alpha = F^{-1}(1 - (\alpha/2))$, where F^{-1} has been tabulated. Equality (3) can be rewritten as

$$\Pr \left\{ \hat{\theta} - z^\alpha \frac{\sigma}{\sqrt{N}} < \theta < \hat{\theta} + z^\alpha \frac{\sigma}{\sqrt{N}} \right\} = 1 - \alpha. \quad (4)$$

Interval

$$N_\alpha = \left[\hat{\theta} - z^\alpha \frac{\sigma}{\sqrt{N}}, \hat{\theta} + z^\alpha \frac{\sigma}{\sqrt{N}} \right] \quad (5)$$

is the CI corresponding to confidence level $1 - \alpha$.

However, σ must be practically estimated from the distance measurements, and this method performs well only in the case when n is large as per the central limit theorem. In the case where n is small ($n \leq 30$), this method could be invalid.

If sample size $n \leq 30$, the CI based on the Student's t -distribution is often utilized to replace the previous method. Under the circumstances, the CIs can be obtained as the following.

Let the sample variance S^2 be defined by

$$S^2 = \frac{\sum_{k=1}^N (X_k - \hat{\theta})^2}{N - 1}, \quad (6)$$

and the random variable $T = (\hat{\theta} - \theta)/(S/\sqrt{N})$ obeys Student's t -distribution with $(N - 1)$ degrees of freedom. The PDF of T , which is denoted by $f_N(t)$, can be obtained by

$$f_N(t) = C(N) \left(1 + \frac{t^2}{N - 1} \right)^{-N/2}, \quad (7)$$

where $C(N)$ is a normalizing constant. Let $F_N(z)$ be the CDF of Student's t -distribution with $(N - 1)$ degrees of freedom and constant $c > 0$ such that

$$\Pr \left\{ -c < \frac{\hat{\theta} - \theta}{S/\sqrt{N}} < c \right\} = 1 - \alpha = 2F_N(c) - 1. \quad (8)$$

Then, constant c must satisfy $F_N(c) = 1 - (\alpha/2)$, and c can be obtained by $c = t_{N-1}^\alpha = F_N^{-1}(1 - (\alpha/2))$, where F_N^{-1} has been tabulated.

Equality (8) can be rewritten as

$$\Pr \left\{ \hat{\theta} - t_{N-1}^\alpha \frac{S}{\sqrt{N}} < \theta < \hat{\theta} + t_{N-1}^\alpha \frac{S}{\sqrt{N}} \right\} = 1 - \alpha. \quad (9)$$

Interval

$$T_\alpha = \left[\hat{\theta} - t_{N-1}^\alpha \frac{S}{\sqrt{N}}, \hat{\theta} + t_{N-1}^\alpha \frac{S}{\sqrt{N}} \right] \quad (10)$$

is the CI corresponding to confidence level $1 - \alpha$. There into, both $\hat{\theta}$ and S can be estimated from the distance measurements.

3.2. CI Based on Nonparametric Bootstrap. In most studies, the CI is commonly and directly obtained through the previous two methods, based on the hypothesis that the distance uncertainty obeys normal distribution and the sample size is sufficiently large. However, as we mentioned before, it is not the case for large-scale WSNs localization in complex environments. When improving accuracy in conventional methods is invalid, bootstrap method can be used to address the CI estimate issues. The CIs of distance measurements can be estimated by employing the nonparametric percentile bootstrap method as follows.

Let $\hat{\theta}_\alpha^*$ represent the 100α th percentile of B bootstrap replications $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$. Percentile limit $\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}$ of intended coverage $1 - 2\alpha$, is directly obtained from the following percentiles:

$$[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}] = [\hat{\theta}_\alpha^*, \hat{\theta}_{1-\alpha}^*]. \quad (11)$$

The nonparametric percentile bootstrap methodology to determine the upper and lower bounds on the CI for distance will be described in detail in the bootstrap distance estimation subsection.

3.3. Nonparametric Bootstrap-Based Distance Estimation Approach. The basic concept of bootstrap method is to produce a large number of independent bootstrap distance estimates by resampling the original distance estimate; that is, $X = (d_1, d_2, \dots, d_n)$, which consists of n measurements at random from unknown probability distribution F .

A bootstrap resample $X^* = (d_1^*, d_2^*, \dots, d_n^*)$ is obtained as a random sample of size n randomly drawn with replacement from the original measurement set X . Sample X has true mean θ ; drawing various other samples X from the distribution F , their means will present distribution $\hat{\theta}$. The distribution of real estimated mean $\hat{\theta}$ is approximated by the distribution of pseudoestimated mean $\hat{\theta}^*$ from bootstrap resample X^* .

Let $X = (d_1, d_2, \dots, d_n)$ represent the set of the corresponding distance measurements. Using the nonparametric percentile bootstrap method on this set, we generate B resamples; $X_j^* = (d_{1j}^*, \dots, d_{nj}^*)$, $j = 1, \dots, B$. And then, we calculate the mean of all measurements in X_j^* to obtain $\hat{\theta}_{X_j^*}^*$ given by

$$\hat{\theta}_{X_j^*}^* = \frac{1}{N} \sum_{k=1}^N d_{k,j}^*, \quad (12)$$

where N is often smaller than 5 (which is a reasonable value and common situation in complex WSNs localization practically), and $j = 1, \dots, B$. The mean of this distribution is given as

$$\hat{\theta}_X^* = \frac{1}{B} \sum_{j=1}^B \hat{\theta}_{X_j}^*. \quad (13)$$

We sort bootstrap estimates $\hat{\theta}_{X_j}^*$ in an ascending order. Let the sorted distance be given by

$$\hat{\theta}_{X_1}^* \leq \hat{\theta}_{X_2}^* \leq \hat{\theta}_{X_3}^* \leq \dots \leq \hat{\theta}_{X_{B-1}}^* \leq \hat{\theta}_{X_B}^*. \quad (14)$$

The desired $100 \cdot (1 - \alpha)\%$ nonparametric CI for the distance is given by $(\hat{\theta}_{X(Q_1)}^*, \hat{\theta}_{X(Q_2)}^*)$, where Q_1 is the integer part of $B\alpha/2$, $Q_2 = B - Q_1 + 1$, and $Q_3 = B/2$ is the integer part of $B/2$.

Furthermore, note that

$$\begin{aligned} \hat{\theta}_{X(Q_1)}^* &= \frac{1}{N} \sum_{k=1}^N d_{k,Q_1}^*, \\ \hat{\theta}_{X(Q_2)}^* &= \frac{1}{N} \sum_{k=1}^N d_{k,Q_2}^*, \\ \hat{\theta}_{X(Q_3)}^* &= \frac{1}{N} \sum_{k=1}^N d_{k,Q_3}^*. \end{aligned} \quad (15)$$

Q_1 , Q_2 , and Q_3 are referred to as the upper, lower, and middle, respectively.

Finally, we can obtain the interval

$$[\hat{\theta}_{\text{lower}}^*, \hat{\theta}_{\text{upper}}^*] = [\hat{\theta}_{X(Q_1)}^*, \hat{\theta}_{X(Q_2)}^*] \quad (16)$$

as the CI of distance corresponding to confidence level $1 - \alpha$. Median point $\hat{\theta}_{X(Q_3)}^* \cong \hat{\theta}_X^*$ is almost identical to the mean. The upper, middle, and lower points are used by the algorithm to estimate the CIs for the distance, respectively. For $\alpha = 0.05$ and $B = 1000$, we can get $Q_1 = 25$, $Q_2 = 976$, and $Q_3 = 500$. And the CI corresponding to confidence level 95% can be described as $[\hat{\theta}_{X(25)}^*, \hat{\theta}_{X(976)}^*]$.

3.4. Determination of Coordinates Based on Interval Analysis Method. When we obtain the CI from a small sample set of distance measurements, the interval analysis method will be utilized to determine the positions of ordinary nodes.

Consider a network consisting of m anchors with known positions and n ordinary nodes which need to be localized. The location of node N_i can be described by $\mathbf{X}_i = [x_i, y_i, z_i]^T$. The CI of measured distance from the ordinary node N_a to one of its anchor nodes N_i

$$[\hat{d}_{ai_{\text{lower}}}, \hat{d}_{ai_{\text{upper}}}] \quad (17)$$

can be obtained through nonparametric bootstrap distance estimation approach, and the Euclidean distance can be defined as

$$D_{ai} = \|\mathbf{X}_a - \mathbf{X}_i\|_2. \quad (18)$$

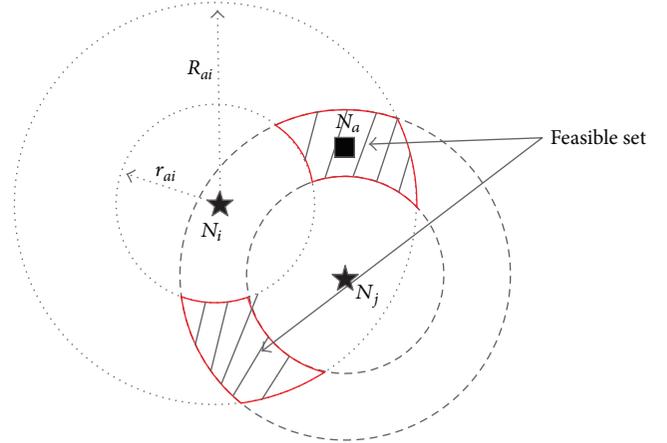


FIGURE 1: Feasible set of ordinary node N_a (top view).

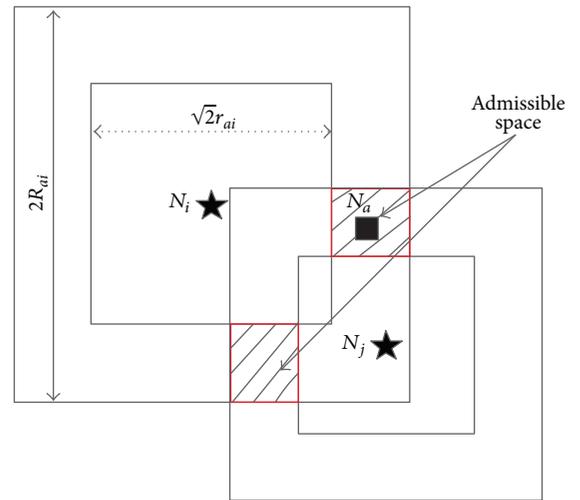


FIGURE 2: Approximate intersections (top view).

Obviously, the desired position estimate is guaranteed to be bounded by an admissible space which can be denoted as

$$S(x) = \bigcap_{i=1}^k \left\{ \hat{d}_{ai_{\text{lower}}} \leq \|\mathbf{X}_a - \mathbf{X}_i\|_2 \leq \hat{d}_{ai_{\text{upper}}} \right\}. \quad (19)$$

$S(x)$ is the intersection of all the spherical caps of center X_i and of internal and external radii $r_{ai} = \hat{d}_{ai_{\text{lower}}}$ and $R_{ai} = \hat{d}_{ai_{\text{upper}}}$, respectively. Clearly, $S(x)$ has a complicated geometrical shape (see Figure 1).

To reduce the computational burden, boxes approach will be utilized to approximate the irregular intersection of all feasible solutions. For each anchor ordinary pair, their admissible space represents a region between two boxes of length $\sqrt{2}r_{ai}$ and $2R_{ai}$, respectively (see Figure 2).

As we know, interval analysis is usually used to model quantities that vary around a central value within certain bounds. With simple operations, the interval analysis allows to consistently deal with problems involving interval data. Based on the interval analysis theory, the admissible space

of N_a to N_i can be denoted by $\vartheta^I = [\vartheta^{I-}, \vartheta^{I+}]$, where ϑ^I is a closed and connected subset of \mathbb{R}^3 , and ϑ^{I-} and ϑ^{I+} are the minimal and maximal bounds of ϑ^I , respectively. Set theory operations, such as intersection, can be applied to intervals. Consider two intervals ϑ_{a1}^I and ϑ_{a2}^I , and their intersection is always an interval which can be denoted as $\vartheta_{a1}^I \cap \vartheta_{a2}^I = [\{\vartheta \mid \vartheta \in \vartheta_{a1}^I, \vartheta \in \vartheta_{a2}^I\}]$. The intersection can be computed by

$$\vartheta_{a1}^I \cap \vartheta_{a2}^I = \left[\max \left\{ \vartheta_{a1}^{I-}, \vartheta_{a2}^{I-} \right\}, \min \left\{ \vartheta_{a1}^{I+}, \vartheta_{a2}^{I+} \right\} \right]. \quad (20)$$

Specifically, the area ϑ_{ai} encircled by the circumscribed and inscribed squares as shown in Figure 2, that is, N_a 's admissible space to N_i , can be computed by:

$$\begin{aligned} \vartheta_{ai} &= [x_a - R_{ai}, x_a + R_{ai}] \times [y_a - R_{ai}, y_a + R_{ai}] \\ &\times [z_a - R_{ai}, z_a + R_{ai}] - \left[x_i - \frac{1}{2}\sqrt{2}r_{ai}, x_i + \frac{1}{2}\sqrt{2}r_{ai} \right] \\ &\times \left[y_i - \frac{1}{2}\sqrt{2}r_{ai}, y_i + \frac{1}{2}\sqrt{2}r_{ai} \right] \\ &\times \left[z_i - \frac{1}{2}\sqrt{2}r_{ai}, z_i + \frac{1}{2}\sqrt{2}r_{ai} \right]. \end{aligned} \quad (21)$$

Using the same previous method, we can get N_a 's another admissible space. Finally, the feasible set W , an interval vector set contains all the possible coordinates, can be obtained by simple enumeration of the intersection points among the two boxes.

Therefore, the feasible set $S(x)$ of ordinary node N_a can be rewritten as

$$\Omega_a = \bigcap_{i=1}^k \left\{ \vartheta \in \vartheta^I : \vartheta_{ai}^I = [\vartheta_{ai}^{I-}, \vartheta_{ai}^{I+}] \right\}. \quad (22)$$

Clearly, the intersection of all subfeasible set ϑ_{ai}^I is a set of boxes. Regarding the coordinates of all subboxes' centers as samples of \mathbf{X}_a , we can get a sample set

$$\Omega = \{\Theta_1, \Theta_2, \dots, \Theta_n\}, \quad (23)$$

and the centre of Θ_n can be found by $\vartheta_n^* = (\vartheta^- + \vartheta^+)/2$. The optimum point estimate, that is, the desired coordinates, can be obtained by

$$\widehat{X}_a = \arg \min_{X_a} \sum_i^k (\|\vartheta_n^* - X_i\|_2 - d_{ai})^2 \quad (24)$$

$$\text{subject to } X_a \in S(x).$$

Finally, the coordinates can be given as

$$\widehat{X}_a = [\widehat{\vartheta}_n^*(1), \widehat{\vartheta}_n^*(2), \widehat{\vartheta}_n^*(3)]^T. \quad (25)$$

4. Performance Evaluation

In this section, we conduct extensive simulations to evaluate the performance of our proposed localization scheme, called Nonparametric Bootstrap Based Multihop Localization Algorithm (NBMLA). All simulations are run in Matlab R2012a. To reduce the influence of outliers, we take the average of 100 simulation runs as the final data points.

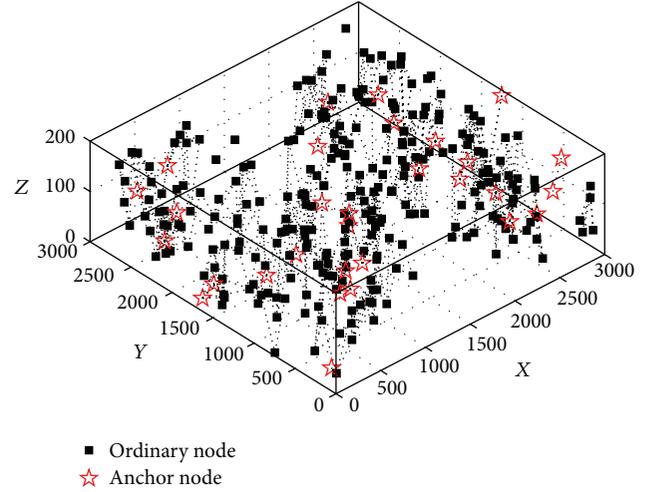


FIGURE 3: Topology of large-scale WSNs.

4.1. Simulation Settings. In our simulation experiments, 400 nodes with adjustable transmission range R are randomly distributed in a large-scale three dimensional region with a size of $3000 \text{ m} \times 3000 \text{ m} \times 200 \text{ m}$ (see Figure 3). We control the density and connectivity of the network by changing the transmission range while keeping the area of deployment the same. Different anchor percentages are also considered in our simulation. Besides our scheme, we also simulate an improved localization algorithm named Taylor-LS for comparison which is almost the same as [12].

We mainly consider three performance metrics: localization coverage, localization error, and time complexity. Localization coverage is defined as the ratio of the localized nodes to the total sensor nodes. Average localization error is the average distance between the estimated positions and the real positions of all ordinary nodes. Distribution boxplots of node localization errors show the average errors, median errors, and maximum outliers. The analysis of time complexity mainly focuses on the communication cost and computation complexity.

4.2. Localization Coverage. In this subsection, we analyze the performance of localization coverage with different network connectivity and anchor percentage, respectively. Figure 4 shows the performance of localization coverage with changing network connectivity while anchor percentage is 15%. We can observe that the localization coverage of our scheme increases monotonically with the connectivity ranging from 4 to 13. But when the connectivity reaches a high value, the coverage rate becomes relatively large and will not change much after that. For example, when the network connectivity is 10, the localization coverage becomes 97% and then changes slower. The reason is that when the network connectivity reaches a certain point, most nodes can get enough anchor nodes by multihop approach and localize themselves.

As shown in Figure 5, the localization coverage of our proposed scheme increases with the anchor number when we change the anchor nodes from 20 (5% anchors) to 80 (20% anchors) with step 10, while keeping the network connectivity

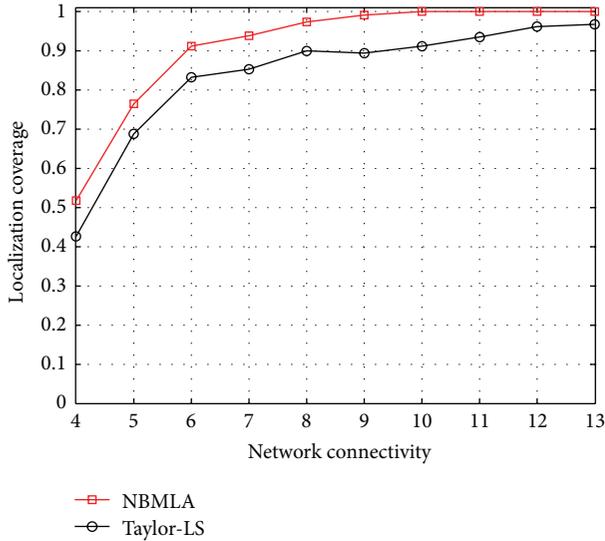


FIGURE 4: Localization coverage versus network connectivity.

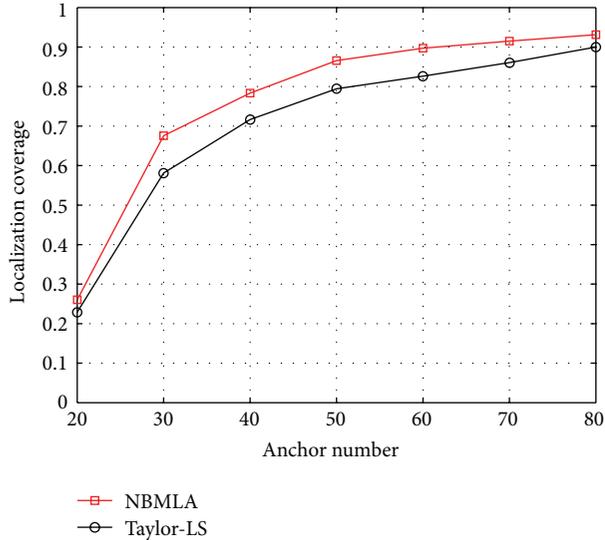


FIGURE 5: Localization coverage versus anchor number.

as 9. We can see that when the anchor number reaches a certain value, for example, 50 anchors, the coverage rate will not change much after that. We can also see that when the anchor rate is 15% and the connectivity is 9, our scheme can localize more than 95% nodes in the large-scale WSNs. Compared with the results of Taylor-LS, NBMLA can achieve better performance in terms of localization coverage, especially in the low anchor percentage situation.

4.3. Localization Error. As mentioned before, in complex environment the ranging process usually encountered some non-Gaussian noises which stem from multiple factors, for example, multipath effect, multihop distance estimation, and anisotropic network condition. This makes it harder for ordinary nodes to get exact distribution characteristics of the distance measurement noises. For comparison with classical

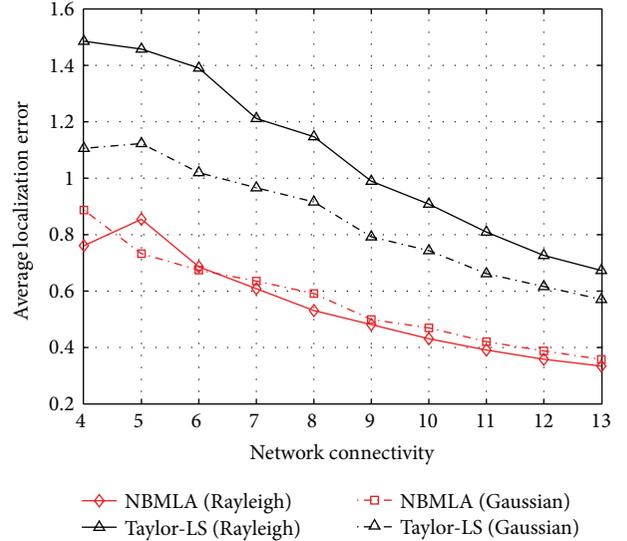


FIGURE 6: Average localization error versus connectivity.

approaches and showing the performance of our algorithm in non-Gaussian noises context, the measurement noise has been assumed to follow Rayleigh distribution, with three percent of real distances as the standard deviations. In other words, the knowledge of the error bound roughly equals nine percent of real distances. This is a reasonable assumption and can be satisfied by the existing distance measurement technologies when suffering multiple adverse factors.

Figure 6 plots the relationship between the average localization error and network connectivity when the anchor percentage is 15%. Besides the given Rayleigh distribution, we also simulate a normal distribution noise situation with mean value and standard deviation (10 m, 10 m) for comparison with the classical approaches relying on normal distributions hypotheses. We can observe that the average localization error of our scheme decreases significantly with the increase of network connectivity. It should be noted that our scheme can achieve relatively high localization accuracy even with low network connectivity. This indicates the good localization performance of our proposed scheme in sparse region. And it can be seen that our scheme outperforms the Taylor-LS method in both Gaussian and non-Gaussian cases. It indicates that our scheme is less affected by the variation of distributions.

In Figure 7, we vary the anchor percent, which is ranging from 5% to 20%, and get the accuracy comparisons with different anchors. For our scheme, with the number of anchor nodes varying from 20 to 80, the localization error decreases by 40%. However, the localization error of Taylor-LS decreases only by 20%, when the network connectivity is 9. This suggests that in sparse networks, our scheme can achieve higher localization accuracy just by increasing a little number of anchor nodes.

4.4. Discussions. Finally, we analyze the time complexity of the schemes which mainly focuses on the communication cost and computation complexity. For a network with M

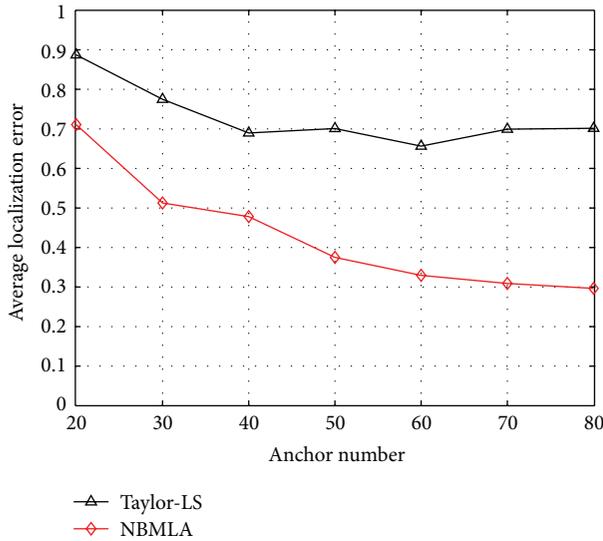


FIGURE 7: Average localization error versus anchor number.

anchor nodes and N ordinary nodes, the overall communication complexity of our scheme is $O(MN)$. Firstly, NBMLA needs the M times flooding initiated by M anchors to make the ordinary nodes know their hop counts, whose overhead is $O(MN)$. Besides the flooding, our scheme requires each anchor to propagate its hop counts to three or four hops' neighboring nodes by a confined flooding. This also has the overhead of $O(MN)$. The communication complexity of Taylor-LS equals to that of NBMLA. But when the algorithms need more distance samples for parametric statistics or location refinement, our scheme can achieve much lower communication cost compared with Taylor-LS. This is due to the fact that NBMLA obtained lots of distance samples by bootstrap resampling approach rather than repeating communication with anchors which introduces a large communication cost for Taylor-LS. This characteristic is particularly important when the network is sparse. Of course, in this case, the computation complexity of our scheme will be larger than Taylor-LS, because the resampling process is necessary when the original distance sample size is very small. For example, if there are only 5 effective distance measurement samples, the NBMLA needs to resample at least B times (>200) for estimating parameter and constructing the CI. Thus, the computation complexity of NBMLA is $O(BN)$. Fortunately, in our scheme, the location estimate can be obtained by simple enumeration instead of complex matrix arithmetic. And compared with the communication cost, the computation cost of our method is moderate and acceptable.

5. Conclusions

In this paper, we present a novel multihop localization algorithm which integrates the nonparametric bootstrap method with interval analysis approach for large-scale WSNs in complex environments. In the first phase, the nonparametric bootstrap method is utilized to build confidence interval

of distance from a small sample set of available measurements with unknown noise distributions. On that basis, we adopt interval analysis approach to estimate the positions of ordinary nodes. The simulation results demonstrate that our algorithm can get high localization coverage and accuracy while resisting against the variation of unknown noise distributions. Compared with the traditional method, the proposed method can greatly improve the average localization accuracy. Further studies are being conducted to extend the method to mobile WSNs in complex environments.

Acknowledgments

The authors are grateful to the anonymous reviewers for their industrious work and insightful comments. This work is supported by the National Natural Science Foundation of China under Grant nos. 61001138 and 61201317.

References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.
- [2] C. Buratti, A. Conti, D. Dardari, and R. Verdone, "An overview on wireless sensor networks technology and evolution," *Sensors*, vol. 9, no. 9, pp. 6869–6896, 2009.
- [3] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," *IEEE Communications Surveys & Tutorials Home*, vol. 13, pp. 487–502, 2011.
- [4] R. Huang and G. V. Záruba, "Monte Carlo localization of wireless sensor networks with a single mobile beacon," *Wireless Networks*, vol. 15, no. 8, pp. 978–990, 2009.
- [5] O. Chia-Ho, "A localization scheme for wireless sensor networks using mobile anchors with directional antennas," *IEEE Sensors Journal*, vol. 11, no. 7, pp. 1607–1616, 2011.
- [6] M. Erol, L. F. M. Vieira, and M. Gerla, "AUV-aided localization for underwater sensor networks," in *Proceedings of the 2nd Annual International Conference on Wireless Algorithms, Systems, and Applications (WASA '07)*, pp. 44–54, August 2007.
- [7] J. Liu and Y. Zhang, "Error control in distributed node self-localization," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 162587, 2008.
- [8] K. Yu, Y. J. Guo, and M. Hedley, "TOA-based distributed localisation with unknown internal delays and clock frequency offsets in wireless sensor networks," *IET Signal Processing*, vol. 3, no. 2, pp. 106–118, 2009.
- [9] M. Vemula, M. F. Bugallo, and P. M. Djurić, "Sensor self-localization with beacon position uncertainty," *Signal Processing*, vol. 89, no. 6, pp. 1144–1154, 2009.
- [10] D. Niculescu and B. Nath, "DV based positioning in ad hoc networks," *Telecommunication Systems*, vol. 22, no. 1–4, pp. 267–280, 2003.
- [11] C. Wang and L. Xiao, "Sensor localization in concave environments," *ACM Transactions on Sensor Networks*, vol. 4, no. 3, article 1, 2008.
- [12] J. W. Wan, N. Yu, R. J. Feng, Y. F. Wu, and C. M. Su, "Localization refinement for wireless sensor networks," *Computer Communications*, vol. 32, pp. 1515–1524, 2009.

- [13] L. Hu and D. Evans, "Localization for mobile sensor networks," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (MobiCom '04)*, Philadelphia, Pa, USA, 2004.
- [14] R. J. Feng, "Multihop localisation with distance estimation bias for 3D wireless sensor networks," *Electronics Letters*, vol. 48, no. 14, pp. 884–886, 2012.
- [15] B. Efron, "Bootstrap methods: another look at the jackknife," *Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.
- [16] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing: an attractive tool for assessing the accuracy of estimators and testing hypothesis for parameters in small data-sample situations," *IEEE Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, 1998.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

