

Research Article

Incremental Localization Algorithm Based on Multivariate Analysis

Xiaoyong Yan,^{1,2} Huanyan Qian,² and Jiguang Chen²

¹ College of Information Technology, Jingling Institute of Technology, Nanjing 211169, China

² School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Correspondence should be addressed to Huanyan Qian; hyqian@mail.njust.edu.cn

Received 13 May 2013; Accepted 1 July 2013

Academic Editor: Qin Xin

Copyright © 2013 Xiaoyong Yan 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.

In view of the traditional increment localization method, only the heteroscedasticity caused by the error accumulation is considered unilaterally a kind of incremental localization algorithm based on multivariate analysis is proposed. The algorithm combines the feasible weighted least squares (FWLS) in the multianalysis with the canonical correlation regression (CCR) and utilizes the FWLS to solve the heteroscedasticity caused by the error accumulation in the process of incremental localization; in the process of estimation, the CCR is used to solve the topology problems between the original beacon nodes and new beacon nodes. The simulation results show that the method can not only effectively restrain the influence caused by the accumulative errors but also can adapt to the different node topological shapes, so as to improve the positioning accuracy of the nodes.

1. Introduction

In many application problems related with sensor network, location information of nodes is of great importance to the monitoring activity of the whole network, which plays a critical role in many applications [1]. Monitoring data without nodes' location information is often of no use. 80% of information provided by sensor nodes to users related with the monitored area is connected with location [2]. In the application of wireless sensor network, nodes' location information can be acquired by adding GPS to nodes. However, this is only applicable with outdoor and open-sided circumstances. Besides, GPS is large in volume and high in cost and energy consumption. Moreover, GPS also needs stable base installations. These facts have made it difficult to realize the requirements of sensor network, which are "low price, low cost and low energy consumption" [3]. As for this, in the deploying area, only some of the nodes can be installed with GPS. For the rest nodes, their location information can only be calculated via a certain algorithm. However, in monitoring area, some unknown nodes not within communication radius of beacon nodes cannot be estimated out though ordinary

localization algorithm concurrently at once for the reasons that communication radiuses of sensor nodes are limited by energy, randomness of nodes distribution, barrier between nodes, and so on, which will cause deficient monitoring area coverage; as a result, the quality of sensor network service decreases rapidly and is not able to effectively monitor the deployment area. The most frequently used solution is to increase the coverage of nodes through mobile beacon nodes [4], but the path, difficult reachability, and relatively larger power consumption of mobile nodes and others limit the application of this method. Incremental estimation algorithm [5, 6] is another method to enhance nodes coverage that has some advantages, such as, it does not need to consider path problems, and is not limited by actions of mobile nodes, and moreover consumes much less power compared with mobile nodes. Incremental algorithm will estimate unknown nodes near beacon nodes at first; these unknown nodes will act as new beacon nodes once their locations are determined, then locations of the rest unknown nodes will be estimated by newly added beacon nodes together with original beacon nodes and so on in a similar way, locations of all nodes were estimated out.

Incremental localization is a kind of low energy consumption positioning method which can effectively solve the coverage rate of the monitoring area; through outward extension in turn, each node is localized successively. Due to the successive localization, the previous estimation error will be bound to affect the following estimated accuracy. Such error accumulation inevitably leads to the inconsistency of the variance between the previous error term and the following localized error term; such phenomenon is also known as heteroscedasticity [7, 8]. In the process of location estimation, the heteroscedasticity appears, then the traditional ordinary least squares (OLS) is used to estimate the location of unknown nodes; the estimated value of the obtained node coordinate may not be the efficient estimator, even not the asymptotically efficient estimator. In order to correct the adverse effect caused by the heteroscedasticity, Meesookho et al. [9] proposed the weighted least squares (WLS); they used the reciprocal of error variance as the weighting of weight to restrain the error propagation. WLS is considered to be the improved OLS method; similar to OLS, the residual sum of squares is solved firstly by WLS and followed by the minimum value. However, in the process of finding the residual sum of squares, unlike OLS, WLS considers the influence of heteroscedasticity. In view of considering the heteroscedasticity based on location estimation by WLS, the location accuracy is improved through corresponding different weights with different data. Subsequently, Xiong et al. [10] proposed a kind of incremental node localization method with the optimal weighted least squares on the basis of WLS; this method is based on the obtained optimal weighted least squares when the error variance matrix is estimated as the minimal. Ji and Liu [5] proposed another strategic improved incremental localization approach (IILA) with the hypothesis that previous localization accuracy is greater than next one during incremental localization, this means that the estimation of location nearer to original beacon nodes is more accurate; on the basis of this assumption, with estimated distance of previous location as a constraint condition, the localization problem is converted into trust region sequences that can be solved by sequential quadratic programming (SQP) method. However, IILA did not consider sensor network as a kind of multihop network that there are many paths to certain node, and neglected complexity of deployment environment but only assumed errors during localization process that definitely increase with increasing hop counts; that is, heteroscedasticity of localization process is only monotonically increasing. In complex monitoring environment, variation tendency of heteroscedasticity is difficult to predict, is not necessarily monotonically increasing, but also is possibly decreasing or concurrently increasing and decreasing. For example, as shown in Figure 1, in monitoring area, there are many paths from node A to unknown node D, such as $A \rightarrow B \rightarrow C \rightarrow D$ and $A \rightarrow E \rightarrow D$ due to environment complexity of monitoring area, barrier or interference sources exist between node A and node E, and accuracy of measurement from node A to Node E is far less than that of other nodes; as a result, it is not appropriate to estimate ED distance if AE distance acts as the constraint condition.

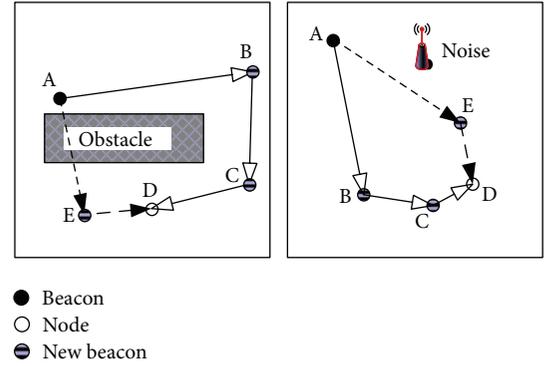


FIGURE 1: Localization in complex environment.

In addition, there is a problem that error of locations through estimation has directivity, for example, in Figure 1, errors between node A and node E could along direction of \vec{AE} also could along \overleftarrow{AE} ; similarly, errors between node E and node D also have directivity; if the direction of errors between node A and node E is in opposite direction of errors between node E and node D, errors between node E and node D is possibly less than those between node A and node E therefore, the assumption of IILA would not be valid any more.

In pervious incremental algorithms, most incremental localization algorithms are used to adjust heteroscedasticity during localization process and assume that heteroscedasticity is only monotonically increasing but fail to take deployment environment and networking features of sensor network into account. Sensor network is a kind of multi-hop network with relatively worse deployment environment, and incremental pattern of its heteroscedasticity is complicated and diversified for incremental localization algorithm. Furthermore, as same as concurrent localization algorithm, the accuracy of incremental localization is influenced by topology quality of beacon nodes also, and multicollinearity problem [11, 12] caused by topological relations among beacon nodes is not considered by previous incremental algorithms. For these reasons, we will propose a feasible incremental localization algorithm, (Location Estimation-FWLS-CCR) LE-FWLS-CCR, which uses less beacon nodes and considers multi-hops features of sensor network, error accumulation, heteroscedasticity and multicollinearity, and other problems. This algorithm will resolve heteroscedasticity problem through feasible weighted least squares (FWLS) [13] iterative computation mode. The iterative process is more proper for multi-hops features, while in estimation process, we adopts Canonical Correlation Regression (CCR) [12, 14–16] multivariate analysis to solve errors exist in newly added beacon nodes as well as topological relations between newly added and original beacon nodes.

2. Relevant Knowledge Review

2.1. Feasible Weighed Least Squares. In a concurrent localization process, distance-coordinates formula is generally transformed into a form of $\mathbf{Ax} = \mathbf{b} + \xi$ [11], in which ξ

is distance-measuring error term. The hypothesis to deduce locations of unknown nodes by distance-coordinates formula is variance of distance-measuring errors, $\text{Var}(\xi) = \sigma^2 \mathbf{I}_n$ (σ^2 is a constant, \mathbf{I}_n is unit matrix), and in estimation algorithm, constant variance of errors is also known as homoscedasticity [7, 8]. While in location estimation by incremental algorithm, distance-measuring error is not inevitable, so variance of distance-measuring errors ($\text{Var}(\xi) = \sigma^2 \Omega \neq \sigma^2 \mathbf{I}_n$) in step-by-step localization process would not be a constant, and this manifestation is called as heteroscedasticity. Because incremental localization process is complicated, there are abundant heteroscedasticity problems in it. Due to existence of heteroscedasticity, the results of typical location estimation models are not accurate and effective.

In the presence of heteroscedasticity, the positions of localization data in location estimation are different; smaller variance of error term of data means higher confidence level of residuals, while bigger variance of error term means lower confidence level of residuals. Therefore, as to the estimation of location in coordinates under the circumstance that heteroscedasticity exists, it usually uses weighted least squares method to discriminate different residuals [17], namely, to pay adequate attention to data terms with relatively smaller residuals and so assign larger weights on them and smaller weights on data term with larger residual in order to adjust the effect of various data items on estimation computation, and therefore to obtain an effective estimation value in localization process.

So, variance of error term of distance-coordinates formula exhibits heteroscedasticity which is expressed as

$$\text{Var}(\xi) = \sigma^2 \Omega, \quad (1)$$

among which σ^2 represents a constant; Ω denotes the n -order symmetric positive definite matrices. It is easy to understand that a n -order invertible matrix must exist, so as to make the following formula true.

$$\Omega = \mathbf{D}\mathbf{D}^T \implies \mathbf{D}^{-1}\Omega(\mathbf{D}^T)^{-1} = \mathbf{I}_n. \quad (2)$$

\mathbf{D}^{-1} is multiplied at both sides of the equation $\mathbf{A}\mathbf{x} = \mathbf{b} + \xi$

$$\mathbf{D}^{-1}\mathbf{A}\mathbf{x} + \mathbf{D}^{-1}\xi = \mathbf{D}^{-1}\mathbf{b}. \quad (3)$$

Assume, $\mathbf{b}^* = \mathbf{D}^{-1}\mathbf{b}$, $\mathbf{A}^* = \mathbf{D}^{-1}\mathbf{A}$, $\xi^* = \mathbf{D}^{-1}\xi$, then (3) can be converted into

$$\mathbf{b}^* + \xi^* = \mathbf{A}^* \mathbf{x} \quad (4)$$

Then the variance of the error term is

$$\begin{aligned} \text{Var}(\xi^*) &= E[\xi^*(\xi^*)^T] = E[\mathbf{D}^{-1}\xi(\mathbf{D}^{-1}\xi)^T] \\ &= E[\mathbf{D}^{-1}\xi\xi^T(\mathbf{D}^{-1})^T] \\ &= \mathbf{D}^{-1}E[\xi\xi^T](\mathbf{D}^{-1})^T \\ &= \mathbf{D}^{-1}\sigma^2\Omega(\mathbf{D}^{-1})^T \\ &= \sigma^2\mathbf{D}^{-1}\Omega(\mathbf{D}^{-1})^T \\ &= \sigma^2\mathbf{I}_n. \end{aligned} \quad (5)$$

Then, the heteroscedasticity of the error term is eliminated, and it is easy to learn that $E(\xi^*) = 0$. Obviously, the error term ξ^* in (5) meets the assumption of the least squares model; therefore, there is the loss equation $S(\mathbf{x})$:

$$\begin{aligned} S(\mathbf{x}) &= (\xi^*)^T \xi^* \\ &= (\mathbf{b}^* - \mathbf{A}^* \mathbf{x})^T (\mathbf{b}^* - \mathbf{A}^* \mathbf{x}) \\ &= (\mathbf{b} - \mathbf{A}\mathbf{x})^T \Omega^{-1} (\mathbf{b} - \mathbf{A}\mathbf{x}). \end{aligned} \quad (6)$$

In order to obtain the optimal solution, we must make

$$\min (\mathbf{b} - \mathbf{A}\mathbf{x})^T \Omega^{-1} (\mathbf{b} - \mathbf{A}\mathbf{x}). \quad (7)$$

It is assumed that $\hat{\mathbf{x}}_{\text{WLS}}$ is the minimized optimal solution. Therefore, $\hat{\mathbf{x}}_{\text{WLS}}$ meets the minimal least squares equation solution as follows:

$$(\mathbf{A}^T \Omega^{-1} \mathbf{A}) \hat{\mathbf{x}}_{\text{WLS}} = \mathbf{A}^T \Omega^{-1} \mathbf{b}. \quad (8)$$

Obviously, if the row vector of \mathbf{A} is linearly independent, then the row vector of \mathbf{A}^* is linearly independent. So, $(\mathbf{A}^*)^T \mathbf{A}^* = (\mathbf{A}^*)^T \Omega^{-1} \mathbf{A}^*$ is reversible; thus, the optimal solution of (8) is expressed as

$$\hat{\mathbf{x}}_{\text{WLS}} = (\mathbf{A}^T \Omega^{-1} \mathbf{A})^{-1} \mathbf{A}^T \Omega^{-1} \mathbf{b}. \quad (9)$$

Through Schwarz inequality [17, 18], it is proved that when the matrix Ω is the reciprocal of the variance matrix of the range error under the condition that the ratio of range error to the distance is independent Gaussian random variables, the error variance by WLS is minimal. But in reality, the variance of the error term is unknown; therefore, if WLS is solved, the weight needs to be taken according to the actual situation.

FWLS is a feasible method which is able to overcome the problem that cannot be implemented by WLS due to the unavailable weight. FWLS uses residuals attained at each computation as weight matrix; therefore, real weight values can be acquired in computation process, and procedure of FWLS algorithm is shown in Algorithm 1.

It can be noted that the FWLS algorithm is in marching iteration; the derivation of the optimal estimated value $\hat{\mathbf{x}}_i$ in

- (1) Firstly, it is essential to estimate the model through OLS method and obtain the estimated value $\hat{\mathbf{x}}$; then substitute it into the equation, and obtain the residual error $\hat{\mathbf{u}}_0 = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ accordingly.
- (2) Utilize the square of the residual term as the $\mathbf{\Omega}$ matrix, for example, $\hat{\mathbf{\Omega}}_i = \text{diag}(\hat{u}_{i,0}^2, \hat{u}_{i,1}^2, \dots, \hat{u}_{i,n-1}^2)$
- (3) Obtain the next-order estimated value and residual value by WLS

$$\hat{\mathbf{x}}_{i+1} = (\mathbf{A}^T \hat{\mathbf{\Omega}}_i^{-1} \mathbf{A})^{-1} \mathbf{A}^T \hat{\mathbf{\Omega}}_i^{-1} \mathbf{b}$$

$$\hat{\mathbf{u}}_{i+1} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{i+1}$$
- (4) Go back to Step 2, until the number of iterative times meet the number of times according to the algorithm requirements.

ALGORITHM 1: FWLS algorithm.

each step is based on the assumption of nonexistent multicollinearity problem in $\mathbf{A}^T \hat{\mathbf{\Omega}}_i^{-1} \mathbf{A}$; unfortunately, by virtue of FWLS, it is feasible to eliminate the interference of heteroscedasticity, but the multicollinearity interference cannot be sure to eliminate. Therefore, in the process of iteration, it is essential to make corresponding strategies to avoid the algorithm insolubility caused by the multicollinearity.

2.2. Canonical Correlation Regression. In concurrent localization algorithm, beacon nodes have a very large influence on final location estimation and possibly cause significant errors when beacon nodes relations are collinear or approximately collinear [11]. Principal component analysis (PCA) in multivariate analysis will remove partial information through recombination of coordinate information of beacon nodes in order to reduce noise and effects of multicollinearity.

As to incremental localization algorithm, it can use PCA [19] method in the first estimation of coordinates of unknown nodes to avoid problems caused by multicollinearity in location estimation, but in incremental algorithm, measuring errors cannot be absolutely avoided or eliminated, which indicates that there would be always errors in locations of newly added beacon nodes in coordinates; therefore, it requires processing error information in output message. PCA only processes input variables; for incremental localization algorithm, its output variables act as locations of newly added beacon nodes, and errors contained in them shall be preprocessed to a certain extent.

Canonical correlation analysis (CCA) is a kind feasible and powerful multivariate analysis method especially appropriate for processing and analysis of two correlated data. At the same time, it is a kind of descending dimension method similar to PCA and is also able to remove some noise information that contains collinear information through recombination of data like PCA. CCA pays more attention to data processing and analysis of correlated data; for this reason, it is more proper for regression algorithm and has higher regression accuracy than PCA.

For the equation $\mathbf{Ax} = \mathbf{b} + \xi$, the solution procedure of CCA is as follows

Suppose that there are two groups of data, \mathbf{A} and \mathbf{b} , which have been processed with centralization, $\mathbf{A} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$, CCA is mainly used to seek linear combination of \mathbf{A} and \mathbf{b} , $\mathbf{w}_A^T \mathbf{A}$ and $\mathbf{w}_b^T \mathbf{b}$, respectively, making them correlate with each

other maximally, that is to say, to find the maximum solution of following equation:

$$\rho = \frac{E[\mathbf{w}_A^T \mathbf{A} \mathbf{b}^T \mathbf{w}_b]}{\sqrt{E[\mathbf{w}_A^T \mathbf{A} \mathbf{A}^T \mathbf{w}_A] E[\mathbf{w}_b^T \mathbf{b} \mathbf{b}^T \mathbf{w}_b]}} \quad (10)$$

$$= \frac{\mathbf{w}_A^T \mathbf{C}_{Ab} \mathbf{w}_b}{\sqrt{\mathbf{w}_A^T \mathbf{C}_{AA} \mathbf{w}_A \mathbf{w}_b^T \mathbf{C}_{bb} \mathbf{w}_b}},$$

in which $\mathbf{C}_{AA} \in \mathbb{R}^{p \times p}$, $\mathbf{C}_{bb} \in \mathbb{R}^{q \times q}$ is within-set covariance matrix of variable \mathbf{A} and \mathbf{b} , respectively, $\mathbf{C}_{Ab} \in \mathbb{R}^{p \times q}$ means between-set covariance matrix, and moreover $\mathbf{C}_{Ab} = \mathbf{C}_{Ab}^T \in \mathbb{R}^{p \times q}$.

The correlation function ρ is independent of scales of \mathbf{w}_A and \mathbf{w}_b ; by constraining respective within-set covariance \mathbf{C}_{AA} and \mathbf{C}_{bb} of \mathbf{A} and \mathbf{b} , CCA can be formulated as solution of optimization problem of the following equation:

$$\begin{aligned} & \max_{\mathbf{w}_A, \mathbf{w}_b} \mathbf{w}_A^T \mathbf{C}_{Ab} \mathbf{w}_b \\ \text{s.t. } & \mathbf{w}_A^T \mathbf{C}_{AA} \mathbf{w}_A = 1, \quad \mathbf{w}_b^T \mathbf{C}_{bb} \mathbf{w}_b = 1. \end{aligned} \quad (11)$$

To solve this optimization problem of (11), we can build a Lagrange equation to obtain the optimal solution; that is,

$$\begin{aligned} L(\mathbf{w}_A, \mathbf{w}_b, \lambda_1, \lambda_2) &= \mathbf{w}_A^T \mathbf{C}_{Ab} \mathbf{w}_b \\ &+ \frac{1}{2} \lambda_1 (1 - \mathbf{w}_A^T \mathbf{C}_{AA} \mathbf{w}_A) \\ &+ \frac{1}{2} \lambda_2 (1 - \mathbf{w}_b^T \mathbf{C}_{bb} \mathbf{w}_b). \end{aligned} \quad (12)$$

Differentiating (12) with \mathbf{w}_A and \mathbf{w}_b , partial derivatives are as follows:

$$\frac{\partial L}{\partial \mathbf{w}_A} = \mathbf{C}_{Ab} \mathbf{w}_b - \lambda_1 \mathbf{C}_{AA} \mathbf{w}_A \quad (13a)$$

$$\frac{\partial L}{\partial \mathbf{w}_b} = \mathbf{C}_{bA} \mathbf{w}_A - \lambda_2 \mathbf{C}_{bb} \mathbf{w}_b. \quad (13b)$$

To obtain the optimal solution, let (13a), (13b), (14a), and (14b) equal to zero, then

$$\mathbf{C}_{Ab}\mathbf{w}_b = \lambda_1 \mathbf{C}_{AA}\mathbf{w}_A, \quad (14a)$$

$$\mathbf{C}_{bA}\mathbf{w}_A = \lambda_2 \mathbf{C}_{bb}\mathbf{w}_b. \quad (14b)$$

Multiply both sides of formula (14a) and (14b) by \mathbf{w}_A and \mathbf{w}_b from left, respectively, easily obtain $\lambda_1 = \lambda_2$, and take $\lambda_1 = \lambda_2 = \lambda$, then the above formula can be simplified into

$$\mathbf{C}_{Ab}\mathbf{w}_b = \lambda \mathbf{C}_{AA}\mathbf{w}_A, \quad (15a)$$

$$\mathbf{C}_{bA}\mathbf{w}_A = \lambda \mathbf{C}_{bb}\mathbf{w}_b. \quad (15b)$$

Given \mathbf{C}_{bb} is reversible, from (15b), obtain $\mathbf{w}_b = (1/\lambda)\mathbf{C}_{bb}^{-1}\mathbf{C}_{bA}\mathbf{w}_A$, and then substitute it into (15a), and reorganize them into

$$\mathbf{C}_{Ab}\mathbf{C}_{bb}^{-1}\mathbf{C}_{bA}\mathbf{w}_A = \lambda^2 \mathbf{C}_{AA}\mathbf{w}_A, \quad (16a)$$

$$\mathbf{C}_{bA}\mathbf{C}_{AA}^{-1}\mathbf{C}_{Ab}\mathbf{w}_b = \lambda^2 \mathbf{C}_{bb}\mathbf{w}_b. \quad (16b)$$

Here, the solution of CCA was translated into generalized eigenvalue-eigenvector problem of two matrixes whose scales are $p \times p$, and $q \times q$ respectively. And then CCA problem is equally described as generalized eigenvalue problem of formula (17):

$$\begin{pmatrix} \mathbf{A}\mathbf{b}^T \\ \mathbf{b}\mathbf{A}^T \end{pmatrix} \begin{pmatrix} \mathbf{w}_A \\ \mathbf{w}_b \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{A}\mathbf{A}^T & \\ & \mathbf{b}\mathbf{b}^T \end{pmatrix} \begin{pmatrix} \mathbf{w}_A \\ \mathbf{w}_b \end{pmatrix}. \quad (17)$$

Formula (17) can be abridged into $\mathbf{X}\mathbf{w} = \lambda\mathbf{Y}\mathbf{w}$, in which \mathbf{X} , \mathbf{Y} , respectively, corresponds to left and right matrix of previous formula, $\mathbf{w} = [\mathbf{w}_A^T, \mathbf{w}_b^T]^T$ therefore, \mathbf{w}_A and \mathbf{w}_b are eigenvectors of $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T\mathbf{A}$ and $(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, respectively.

The literature [12, 15] proposed a regression method based on canonical correlation analysis, namely, canonical correlation regression (CCR). CCR combines least squares with canonical correlation analysis with the purposes of optimizing solution of regression coefficient under most relevant significance. CCA-based regression method, to some extent, avoids interferences of multicollinearity of samples through utilization of components that have been features-extracted for regression; in addition, CCA considers correlation between output and input invariables and so can be regarded as an advanced regression between two multivariates, an extension of multiple linear regression (MLR), and known as a ‘‘multi-to-multi’’ regression method. Regression coefficient of canonical correlation regression can be computed out by following equation:

$$\hat{\mathbf{x}}_{\text{CCR}} = (\mathbf{W}_k \mathbf{W}_k^T)^{-1} \mathbf{W}_k^T \mathbf{b}, \quad (18)$$

in which, $\mathbf{W}_k = [\mathbf{w}_A^1, \mathbf{w}_A^2, \dots, \mathbf{w}_A^k]$ is composed of first k of maximum eigenvector.

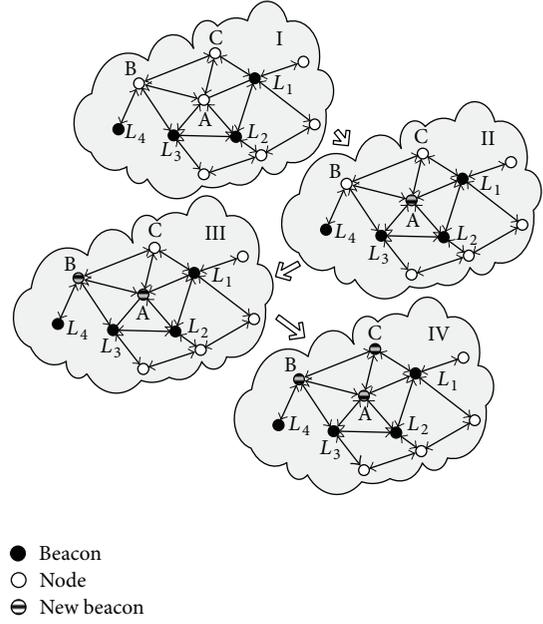


FIGURE 2: Example and phases of LE FWLS-CCR.

3. Methodology

3.1. The LE-FWLS-CCR Algorithm. The existence of multicollinearity usually brings seriously adverse effects on model estimation, testing, and prediction. During localization estimation, multicollinearity not only exists in concurrent localization but also exists in incremental location estimation process. For this reason, we add canonical correlation regression method into FWLS algorithm to acquire optimal prediction direction through correlation analysis of input and output variables and dimension reduction processing then use FWLS method to resolve problems caused by heteroscedasticity. Because the procedure of FWLS-CCR localization algorithm is similar to that of FWLS algorithm, its solution is carried out through iteration, and the algorithm process is shown in Algorithm 2.

Node localization process based on FWLS-CCR is shown in Figure 2. Assume that the monitoring area is deployed with several sensor nodes and L_1, L_2, L_3 , and L_4 are original beacon nodes, which are set to be zero-level beacon nodes. The nodes of A, B, and C are the nodes to be localized. Node A is directly connected with the three original beacon nodes of L_1, L_2 , and L_3 . Node B is connected with L_3 and L_4 , while node C is only connected with L_1 .

Obviously, the coordinates of node A can be calculated and estimated according to the beacon nodes of L_1, L_2 , and L_3 . According to CCR incremental localization algorithm, node A may be updated to a new beacon node after adopting CCR algorithm to calculate its estimated coordinates, and the node may be set as the first level beacon node. By calculating the residual, matrix Ω can be obtained. Assuming that $\mathbf{b}^* = \mathbf{D}^{-1}\mathbf{b}$, $\mathbf{A}^* = \mathbf{D}^{-1}\mathbf{A}$, the location calculation equation is transformed into the form of $\mathbf{b}^* + \xi^* = \mathbf{A}^* \mathbf{x}$. taking original beacon nodes L_3, L_4 , and the newly added beacon node A

Input: Beacon nodes coordinates:

Distance from beacon nodes to unknown nodes $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}, m \geq 3$

Output: Estimated coordinates of unknown nodes: $\{\hat{\mathbf{x}}_{m+1}, \hat{\mathbf{x}}_{m+2}, \dots, \hat{\mathbf{x}}_n\}$

- (1) Beacon nodes deliver their location information outwards through controllable flooding, if unknown nodes acquire more than 3 beacon nodes, it firstly uses CCR to estimate locations of unknown nodes, and will stop if there is no rest unknown nodes, otherwise, will carry out next step.
- (2) Uses estimated location to estimate residual vector by the estimation formula: $\hat{\mathbf{u}}_0 = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}_{\text{CCR}}$;
- (3) Constructs covariance matrix through FWLS method residual vectors, $\hat{\mathbf{\Omega}}_i = \text{diag}(\hat{u}_{i,0}^2, \hat{u}_{i,1}^2, \dots, \hat{u}_{i,m-1}^2)$;
- (4) Uses newly-constructed covariance matrix to rewrite location distance equation, $\mathbf{A}\mathbf{x} = \mathbf{b} + \xi$;
- (5) According to new equation, uses CCR method to estimate locations of secondary beacon nodes.
- (6) If there still are some nodes which locations have not be estimate out in deployment area, skip to Step 2.
- (7) The algorithm will finish if there is no node to be estimated in deployment area, and it will output coordinates of unknown nodes.

ALGORITHM 2: LE-FWLS-CCR algorithm.

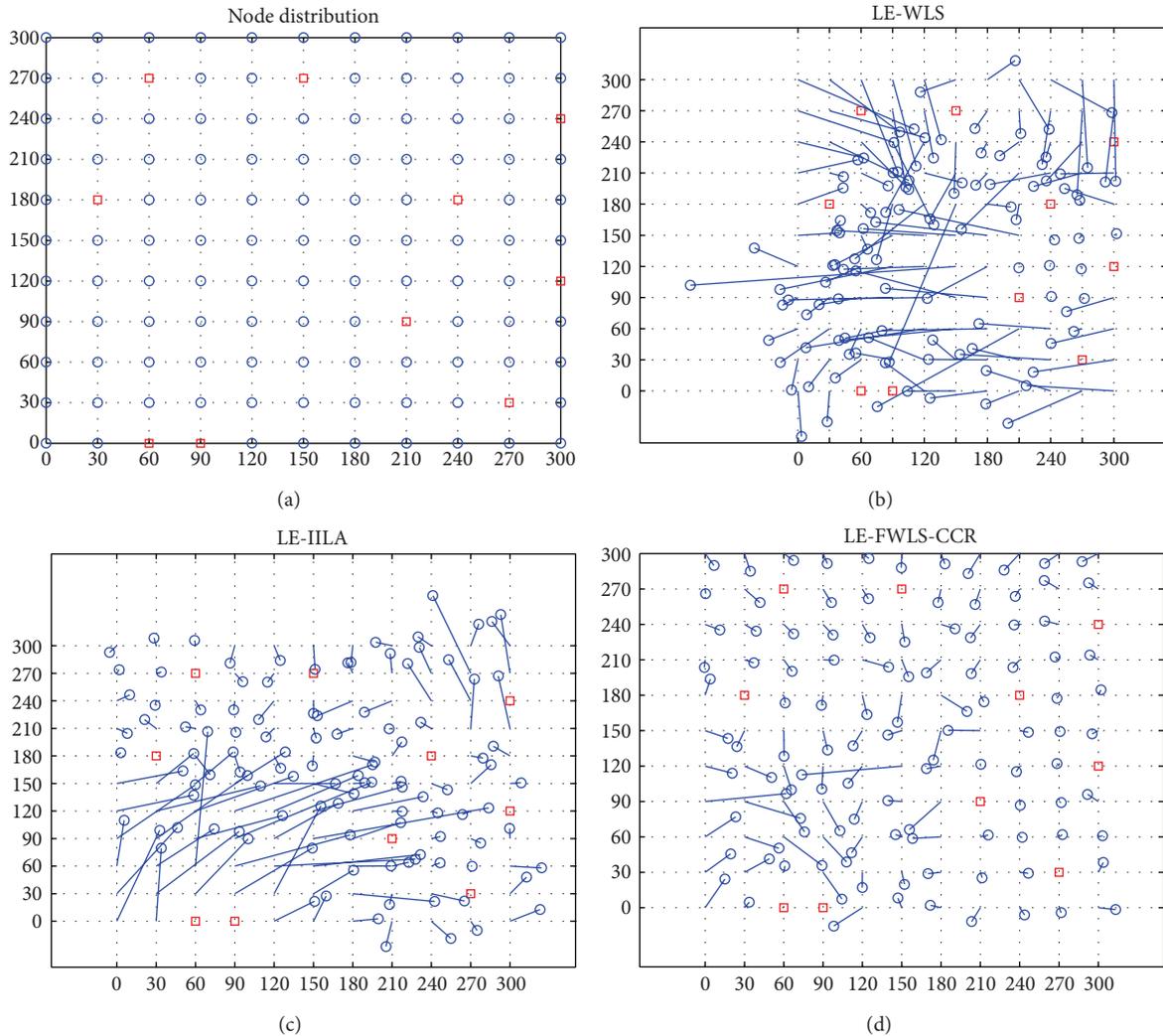


FIGURE 3: Localization results of regular deployment in square area.

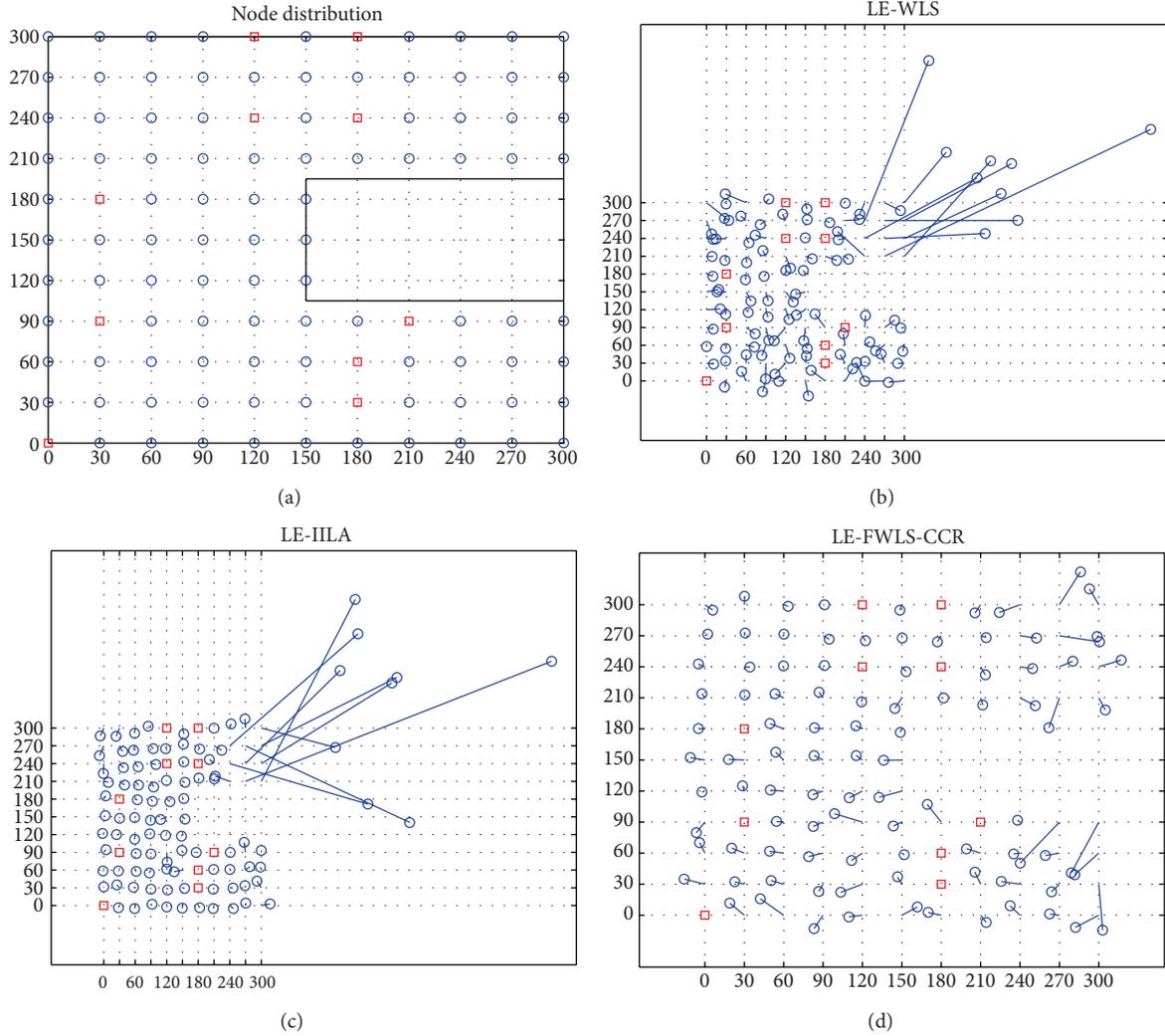


FIGURE 4: Localization results of regular deployment in C-shape area.

as the referential nodes, and then readopt CCR algorithm to figure out the estimated coordinates of node B. After the calculation, node B is added as a new beacon node, with second level. Similarly, the location of node C can be estimated based on beacon L_1 , node A, and node B. Node C is defined to be a third level beacon node.

In incremental localization approach, localization of nodes is implemented in batches. Owing to distance measurement error, there is a certain difference between the estimated value and the practical value of first level beacon node. As for second level beacon nodes, their estimated values are as well influenced by measurement error and the intrinsic error of first level beacon nodes. In addition, it is shown in Figure 2 that, when estimating node C, the positions of referential nodes L_1 , A, and B are founded to be approximately collinear, which may lead to even larger error when estimating node C. On this basis, FWLS-CCR method is applied to utilize CCR to reduce the dimensionality of input and output data after the data is analyzed. The objective is to partially eliminate error and multicollinearity.

Meanwhile, attainable residuals are adopted as the weight value, making the algorithm even feasible and applicable to practical environment. Thus, compared with previous algorithms, localization algorithm based on FWLS-CCR is more feasible, with higher adaptability.

3.2. Time Complexity Analysis. In this Section, we compare the time complexities for localization as required by LE-FWLS-CCR and other popular incremental location estimation algorithms, namely, WLS-based location estimation (LE-WLS) proposed in the literature [10] as well as SQP-based location estimation (LE-IILA) proposed in the literature [5]. These algorithms will also be compared experimentally in Section 4.

LE-FWLS-CCR: basically, the complexity of our algorithm is dominated by two parts: canonical correlation regression and feasible weighted least squares. The complexity of computing CCR is mainly determined by the core algorithm CCA, whose complexity is $O(n^3 \log n)$. FWLS method is OLS improvement, which uses residuals attained at

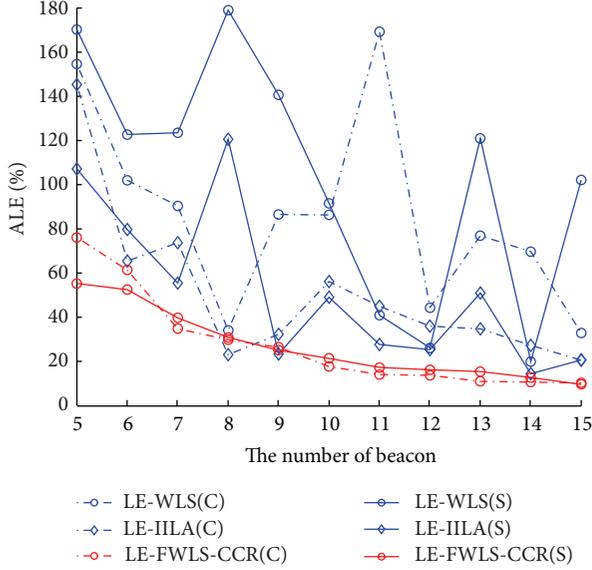


FIGURE 5: Average localization error of regular deployment.

each computation as weight matrix; therefore, complexity of FWLS is $O(n^4)$. Thus, the total complexity of the FWLS-CCR is $O(n^4)$.

LE-WLS: WLS-based location estimation method using the WLS method as the core algorithm, and the complexity of WLS is $O(n^3)$; thus, the computer complexity of LE-WLS method is $O(n^3)$.

LE-IILA: the complexity of SQP method is the main reason in LE-IILA algorithm. If the limited number of iterations, the computational complexity of SQP is $O(k^2 + kn)$, where k is the number of iterations. If $n \gg k$, the complexity of computer is $O(kn)$, otherwise the complexity of computer is $O(k^2)$.

The literature [20] has proved that properly increasing the calculation amount of algorithm will not affect the performance of sensor network. For this reason, it is worthy to improve the localization accuracy based on FWLS-CCR by sacrificing partial calculation volume.

4. Simulation and Experiments

This section will analyze and evaluate LE-FWLS-CCR localization algorithm on Matlab platform. In simulation experiment, it is supposed that nodes are deployed in a two-dimensional monitoring area and adopt transformation of RSSI signals to distance for matrix of distance among nodes. In order to compare impartiality of experimental results, this section adopts signal model proposed in the literature [21] to simulate signal strength among nodes; that is,

$$P_{ij} \sim N(\bar{P}_{ij}, \sigma_{dB}^2), \quad (19)$$

$$\bar{P}_{ij} = P_0 - 10n_p \lg\left(\frac{d_{ij}}{d_0}\right),$$

among which P_{ij} represents the transmitted signal power which is received by node i from the node j , and the unit is dBm; P_0 represents the received signal power corresponding to the point of the reference range d_0 ; d_0 represents the reference range; n_p represents the attenuation coefficient of the wireless transmission and is related to the environment; \bar{P}_{ij} represents the received signal power corresponding to the point of the reference range d_0 (dBm); σ_{dB}^2 represents the shadow variance. n_p uses fitting data from real collection in the literature [20]; as for σ_{dB}^2 , let $\sigma_{dB}^2/n_p = 1.2$ in this experiment.

Due to higher coverage of incremental algorithm, the experiment in this section mainly examines the accuracy of localization of nodes with ALE as evaluation basis, and the definition of ALE is as follows:

$$\text{ALE} = \frac{\sum_{i=1}^n \sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{n \times R} \times 100\%. \quad (20)$$

In the formula, (\hat{x}_i, \hat{y}_i) represents the estimated coordinate location of the i th node, (x_i, y_i) represents the actual coordinate location of the i th node n represents the number of the unknown nodes, and R represents the communication radius. It can be seen from the above formula that ALE refers to the ratio of the average error of the Euclidean distance from the estimation location of all nodes to the real location in the area to the communication radius. ALE can reflect the stability of the localization algorithm and the positioning accuracy; when the communication radius of the node is given, if the average localization error of the node is smaller, then the positioning accuracy of the algorithm is higher and vice versa.

This experiment also compares the method proposed in this paper, the localization algorithm based on FWLS-CCR (LE-FWLS-CCR), with WLS-based location estimation (LE-WLS) proposed in the literature [10], as well as IILA-based location estimation (LE-IILA) proposed in the literature [5]. In addition, the experiment carries out comparison by use of data collected from actual scenes provided in the literature [20].

4.1. Simulation Experiments Based on Distance-Measuring Model. The experiments based on distance-measuring model have set four experimental scenes: random deployment nodes in square area, regular deployment nodes in square area, random deployment nodes in C-shape area and regular deployment nodes in C-shape area, in which C-shape area, is formed because of a bigger barrier, mainly used to evaluate localization performance with lager barrier, that is, in case of non-line-of-sight. In order to decrease the effect of single one experiment, each group of experiments will be repeated for 50 times in each scene, finally the average indicators of the 50 experiments will be reported. The experiments will examine accuracy of final localization results of unknown nodes with incremental quantity of beacon nodes. In these experiments, the valid communication radius of nodes is supposed to be 60 m.

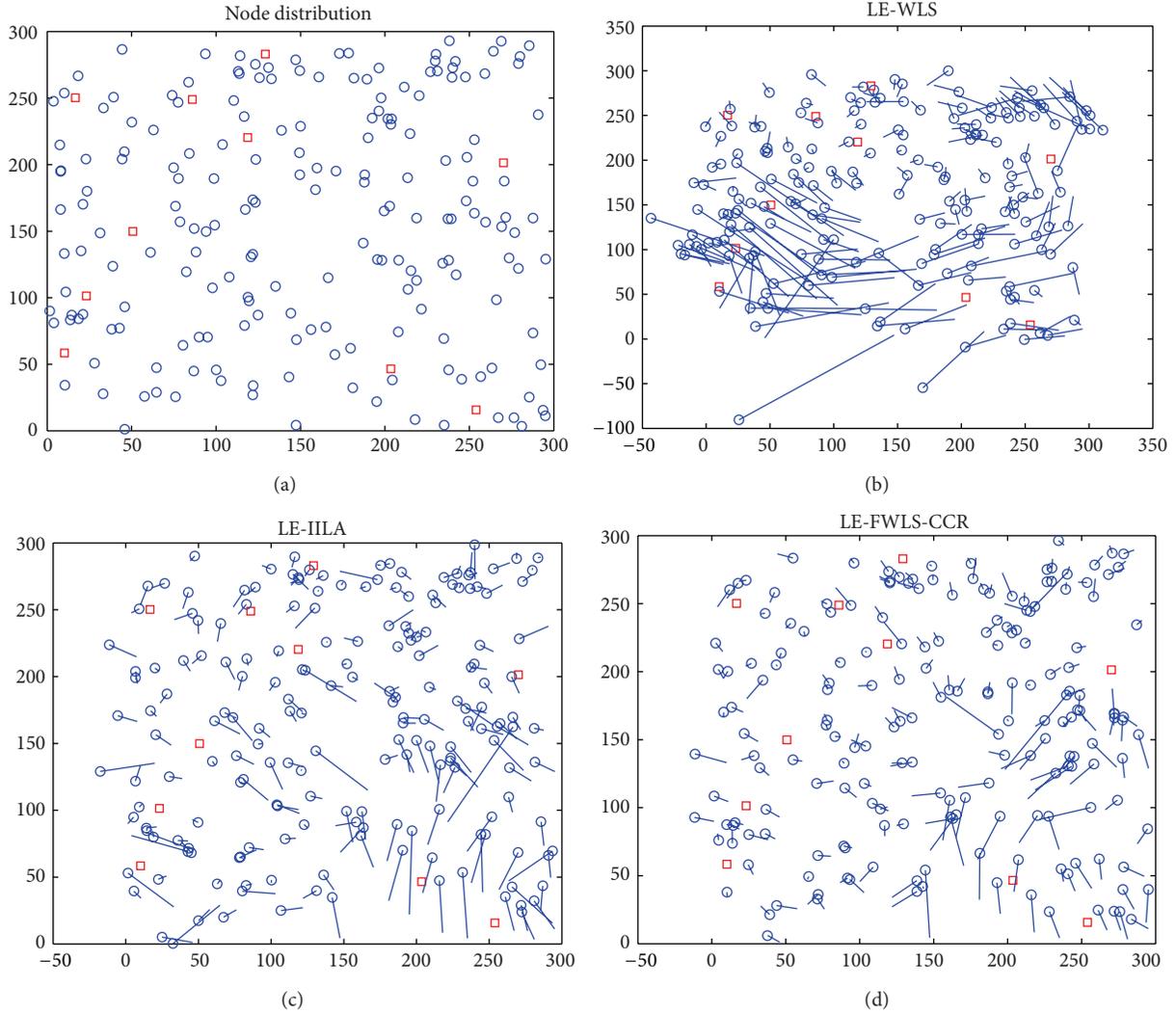


FIGURE 6: Localization results of random deployment in square area.

4.1.1. Rules Deployment. Regular deployment of nodes in monitoring area principally aims to explore effects of collineation of beacon nodes on localization accuracy, while regular deployment in C-shape area is used to observe effects of non-line-of-sight caused by barriers in monitoring area on localization accuracy.

In this group of experiments, regular deployment of nodes is within a $300\text{ m} \times 300\text{ m}$ area, whose side length of grids is 30 m , there are 121 nodes in total without barrier, but a $150\text{ m} \times 90\text{ m}$ barrier was placed in C-shape area, and then the quantity of nodes becomes 106. In these experiments, select 5–15 nodes as beacon nodes, provided that their location information is known. Figure 3 is the final localization result of nodes of certain deployment in square area under the circumstance that there are 10 beacon nodes, in which circle denotes unknown node, box denotes beacon node, and the straight line connects actual coordinates of unknown node with its estimated coordinates. Figure 3(a) shows deployment of nodes; Figure 3(b) shows localization results based on weighted least squares with weights being reciprocals of

variance of error term that is optimal theoretically, and in this figure, $\text{ALE} = 70\%$; Figure 3(c) shows LE-IILA method proposed by Ji, and in this figure, $\text{ALE} = 43.2\%$; Figure 3(d) shows LE-FWLS-CRR proposed in this paper, and in this figure, $\text{ALE} = 18.2\%$.

The circumstance that there is a barrier in regular deployment area was described in Figure 4(b) to Figure 4(d), and this experiment mainly is used to study effects of non-line-of-sight on localization results. Figure 4(a) shows deployment of nodes; Figure 4(b) shows localization results of LE-WLS, $\text{ALE} = 46.8\%$; Figure 4(c) shows LE-IILA method proposed by Ji, $\text{ALE} = 42.2\%$; Figure 4(d) shows LE-FWLS-CRR proposed in this paper, $\text{ALE} = 16.1\%$.

It can be seen from Figures 3 and 4 that LE-WLS and LE-IILA only considered heteroscedasticity but didn't take multicollinearity into account; therefore, only nodes in partially incremental area obtained satisfactory results in experiments, and localization errors in some area are still large. LE-FWLS-CCR method proposed in this paper comprehensively considered heteroscedasticity, error escalation as well as

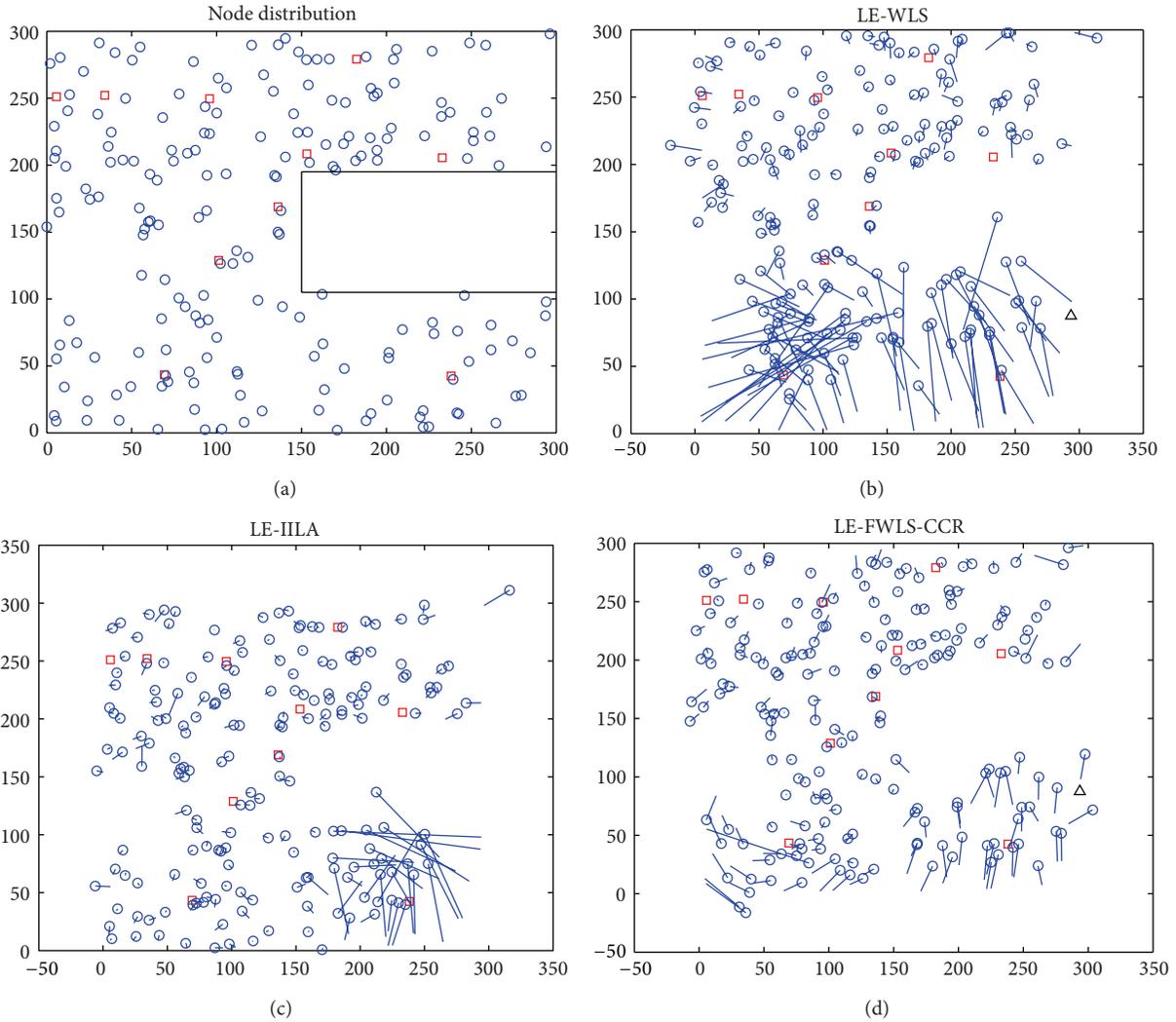


FIGURE 7: Localization results of random deployment in C-shape area.

multicollinearity, and other problems, and so localization results are more effective than those of LE-WLS and LE-IILA methods.

Because incremental localization method is used to locate nodes through gradually increment, as shown in Figure 3, barrier did not cause localization coverage reduction. While because of “extrusion” of barrier, the ratio of beacon nodes in such deployment area is higher than the same size of deployment area without barrier under the circumstance that the quantity of beacon nodes is equivalent. Ratio of beacon nodes increases; hence, localization accuracy in most of the area is relatively higher as shown in Figure 4. As shown in Figures 3 and 4, LE-WLS and LE-IILA similarly did not consider multicollinearity between original and newly added beacon nodes; as a result, localization errors of nodes in some areas are large and then affect the whole performance of localization.

Figure 5 describes curve of average ALE of repeated experiments by three localization methods varying with the quantity of beacon nodes in regular deployment scenes. It is easily to find ALE of LE-WLS fluctuating as the strongest, and

accuracy is the lowest; LE-IILA is in second place, while LE-FWLS-CCR proposed in this paper is most accurate. This is because original and newly added beacon nodes are being most possibly collinear, furthermore, LE-WLS and LE-IILA algorithms did not consider this and noise was not removed completely, especially LE-WLS method that only considered heteroscedasticity but did not take noise escalation into account; consequently ALE curve waves greatly, sometimes; ALE is approximate to 180%; LE-IILA method only ideally considered noise escalation but did not take multicollinearity into account and so obtain better localization results than LE-WLS method, but the localization results are still not stable, and sometimes ALE is greater than 100%; therefore, the effectiveness of localization by LE-IILA is hard to meet actual needs; the method proposed in this paper considered multiple factors that affect accuracy in incremental localization process and obtained fairly stable localization results and significantly higher accuracy than other incremental localization methods.

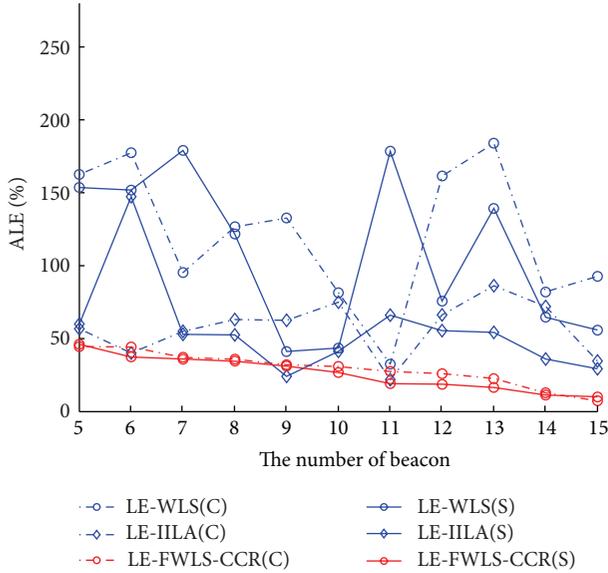


FIGURE 8: Average localization error of random deployment in square area.

4.1.2. Random Deployment. Random deployment is more close to actual situation. The experiments in this scene are used mainly to discuss whether this algorithm is proper for various actual situations or not. In the same way, experiments about random deployment are classified into two groups, in C-shape area with barrier and in square area without barrier. In this group of experiments, there are 200 nodes randomly deployed in a 300 m × 300 m monitoring area, similar with regular deployment, to compare LE-FWLS-CCR with LE-WLS and LE-IILA in order to evaluate changes of ALE with quantity of beacon nodes by two algorithms. In the scene with barrier, a 150 m × 90 m object was placed in deployment area to artificially lead to ineffective communication of nodes in this area. In these experiments, select 5–15 nodes as beacon nodes provided that their location of information is known.

Figure 6 shows final localization result of nodes of certain deployment in square area under the circumstance that there are 10 beacon nodes. Figure 6(b) shows localization results of LE-WLS method, with weights being reciprocals of variance of error term that is optimal theoretically, ALE = 41%; Figure 6(c) shows LE-IILA method proposed by Ji, ALE = 18.5%; Figure 6(d) shows LE-FWLS-CRR proposed in this paper, ALE = 15.2%. From this figure, localization results of unknown nodes surrounding the area crowded with original beacon nodes are better, but with increasing incremental series, effectiveness of localization by LE-WLS is worse; LE-IILA is better than LE-WLS because the former considered error escalation, but in some areas, estimation results of unknown nodes are still far from actual values, for the reason that it did not consider multicollinearity between newly added beacon nodes and original beacon nodes. The results of LE-FWLS-CRR proposed in this paper are still stable and have higher accuracy than the previous two methods.

Figure 7 shows localization result of certain random deployment in C-shape area with barrier. Figure 7(a) shows

nodes distribution diagram of certain localization. Figures (b),(c), and (d) show localization result of LE-WLS, LE-IILA, and LE-FWLS-CCR, respectively, and final ALE of various localization in Figure 7 is 38.6%, 15.6%, and 13.1%, respectively. The figure shows very obvious trace of incremental localization, that is, anticlockwise step-by-step location estimation. Due to existence of barrier, same quantity of beacon nodes accounts for a higher proportion in unit area than random deployment in square area; therefore, if the incremental series is lower, three methods' estimation of accuracy for unknown nodes is high, but with incremental series increasing, advantages and disadvantages of three methods will appear. LE-WLS did not consider error control and multicollinearity, so localization error for higher series is large; although LE-IILA considered error escalation to a certain extent, it did not consider multicollinearity, the localization results were improved but are still poor in some areas. The algorithm proposed in this paper obtained similar localization results with the previous three scenes, and its localization accuracy is still stable and excellent.

Figure 8 describes curve of average ALE of repeated experiments by three localization methods varying with quantity of beacon nodes in random deployment scenes. Curves of LE-WLS and LE-IILA methods still wave ups and downs, because randomness of random deployment is far greater than that of regular deployment, and as a result, maximum ALE of LE-WLS and LE-IILA methods is even approximate to 180%; however, the method proposed in this paper still obtains stable results. Owing to full considerations on adverse factors in localization process, the characteristics of random deployment did not reduce localization accuracy greatly but improved it.

4.2. Simulation Experiment Based on Actually Measured Data. This paper uses actually measured data set provided by Neal Patwari of Utah State University. The experiment was arranged in a standard office area that is a 12 m × 14 m rectangle. There are 44 nodes (in which 4 nodes act as beacon nodes) deployed in it; the communications among nodes adopt direct sequence spread spectrum (DS-SS), and the center frequency of deployment nodes is 2.4 GHz. This paper uses these data and enlarges effective communication radius of nodes to compare LE-WLS, LE-IILA, and LE-FWLS-CCR methods proposed in this paper. Table 1 shows that localization results of LE-FWLS-CCR for four different communication radiuses are better than those of other two localization algorithms. The details are shown in Table 1

Figure 9 shows the localization results of three algorithms under the circumstance that communication radius is 7 m, in which Figure 9(a) shows nodes deployment; Figure 9(b) is localization result of LE-WLS method, ALE = 81.8%; Figure 9(c) is localization result of LE-IILA method, ALE = 39.6%; Figure 9(d) shows localization result of LE-FWLS-CCR method proposed in this paper, ALE = 19.6%. From Figure 9, it can be seen that only localization results of unknown nodes near original beacon nodes are relatively satisfactory in Figure 9(b); in Figure 9(c), localization results of unknown nodes far from original beacon nodes are

TABLE 1: Comparisons of average localization errors based on actual RSSI measurement data.

Wireless communication radius (m)	LE-WLS ALE	LE-IILA ALE	LE-FWLS-CCR ALE
6.5	99.1%	46.6%	20.8%
7	81.8%	39.6%	19.6%
7.5	70.23%	41.3%	18.4%
8	83.51%	49.12%	15.51%

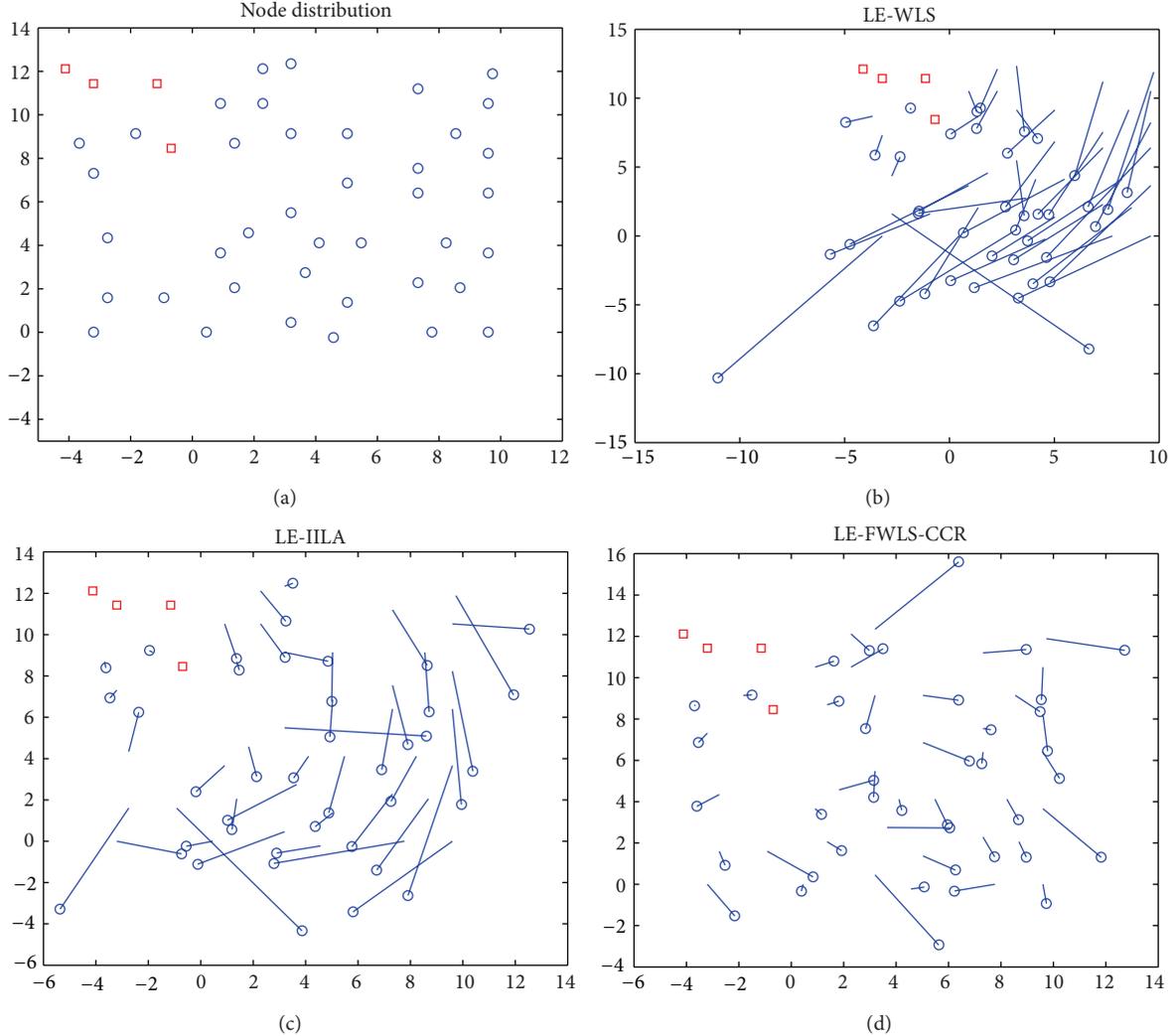


FIGURE 9: Localization results of actually measured data.

better than those in Figure 9(b); while localization result in Figure 9(d) is the best in the three.

5. Conclusion

This paper combines FWLS method and CCR method in a localization process that uses FLWS method to solve problems in location estimation caused by heteroscedasticity and uses dimensionality reduction algorithm CCR in multivariate analysis to deal with topology between original and newly added beacon nodes and error accumulation problems. The results of many groups of experiments indicate that

the method proposed in this paper can effectively resolve heteroscedasticity, accumulative error, and multicollinearity problems, and its localization results are stable and have higher accuracy than previous incremental localization methods.

Acknowledgments

The paper is sponsored by Natural Science Foundation of China (61005008); Provincial University Natural Science Research Foundation of Jiangsu Education Department

(11KJD510002, 12KJD510006); Natural Science Foundation of Jiangsu (BK2012082).

References

- [1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
- [2] G. Cabri, L. Leonardi, M. Mamei, and F. Zambonelli, "Location-dependent services for mobile users," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 33, no. 6, pp. 667–681, 2003.
- [3] J. Zheng and A. Jamalipour, *Wireless Sensor Networks a Networking Perspective*, John Wiley and Sons, New York, NY, USA, 2009.
- [4] N. B. Priyantha, H. Balakrishnan, E. D. Demaine, and S. Teller, "Mobile-assisted localization in wireless sensor networks," in *Proceedings of the IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '05)*, pp. 172–183, Miami, Fla, USA, March 2005.
- [5] W. Ji and Z. Liu, "Accumulative error analysis of incremental node localization approach and its improvement in wireless sensor network," *Journal of Nanjing University of Science and Technology*, vol. 32, no. 4, pp. 496–501, 2008.
- [6] F. Wang, L. Shi, and F. Ren, "Self-localization systems and algorithms for wireless sensor networks," *Journal of Software*, vol. 16, no. 5, pp. 857–868, 2005.
- [7] C. Lim, P. K. Sen, and S. D. Peddada, "Accounting for uncertainty in heteroscedasticity in nonlinear regression," *Journal of Statistical Planning and Inference*, vol. 142, no. 5, pp. 1047–1062, 2012.
- [8] F. Cribari-Neto and W. B. da Silva, "A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model," *ASta Advances in Statistical Analysis*, vol. 95, no. 2, pp. 129–146, 2011.
- [9] C. Meesookho, U. Mitra, and S. Narayanan, "On energy-based acoustic source localization for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 365–377, 2008.
- [10] W. Xiong, M. Tang, and B. Xu, "Incremental node localization approach and its improvement in wireless sensor network," *Chinese Journal of Sensors and Actuators*, vol. 24, no. 4, pp. 576–580, 2011.
- [11] X. Yan, H. Qian, and J. Yu, "A localization method based on principal component analysis," *Journal of Computational Information Systems*, vol. 8, no. 22, pp. 9425–9432, 2012.
- [12] H. Hyötyniemi, "Multivariate Regression Techniques and tools," <http://autsys.aalto.fi/en/Publications/25041>, 2013.
- [13] C. Heij, P. de Boer, P. H. Franses, T. Kloek, and H. K. van Dijk, *Econometric Methods with Applications in Business and Economics*, Oxford University Press, Oxford, UK, 2004.
- [14] B. Abraham and G. Merola, "Dimensionality reduction approach to multivariate prediction," *Computational Statistics and Data Analysis*, vol. 48, no. 1, pp. 5–16, 2005.
- [15] M. Reiter, "Enhanced Multiple Output Regression based on Canonical Correlation Analysis with Applications in Computer Vision," 2010.
- [16] L. Sun, S. Ji, and J. Ye, "Canonical correlation analysis for multi-label classification: a least-squares formulation, extensions, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 194–200, 2011.
- [17] A. Schick, "Weighted least squares estimation with missing responses: an empirical likelihood approach," *Electronic Journal of Statistics*, vol. 7, pp. 932–945, 2013.
- [18] S. Bidwell, M. E. Hassell, and C. R. Westphal, "A weighted least squares finite element method for elliptic problems with degenerate and singular coefficients," *Mathematics of Computation*, vol. 82, no. 282, pp. 673–688, 2013.
- [19] I. Jolliffe, *Principle Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.
- [20] N. B. Priyantha, *The Cricket Indoor Location System*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2005.
- [21] "Measured Channel Impulse Response Data Set," <http://span.ece.utah.edu/pmwiki/pmwiki.php?n=Main>, 2013 .MeasuredCIRDataSet,2013.



Hindawi

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

