International Journal of Grid Distribution Computing
Vol.8, No.1 (2015), pp.67-80
http://dx.doi.org/10.14257/ijgdc.2015.8.1.07

Multi-hop Range-Free Localization Algorithm For Wireless Sensor Network Using Principal Component Regression

Xianghong Tian (a) Wei Zhao (a) Xiaoyong Yan (b)

School of Computer Engineering, Jinling Institute of Technology, Nanjing 211169, China (a)
School of Intelligence Science and Control, Jinling Institute of Technology, Nanjing 211169, China(b)
txh@jit.edu.cn

Abstract

In this paper, a novel approach to multi-hop range-free localization algorithm in wireless sensor network is proposed using principal component regression. The localization problem in the wireless sensor network is formulated as a multiple regression problem, which is resolved by principal component regression. The proposed methods are simple and efficient that no additional hardware is required for the measurements, and only hop-counts information and location information of the beacons are used for the localization. The proposed method consists of two phases: the offline training phase and the online localization phase. In offline training phase, the real distances and the hop-counts among sensor nodes are collected to build localization model. In online localization phase, each unknown sensor node finds its own location using the localization model. The experimental results show that compared with previous localization methods, the proposed method exhibits excellent and robust performances not only in the isotropic sensor networks but also in the anisotropic sensor networks.

Keywords: *Wireless sensor network, Multi-hop Range-Free Localization, Principal component regression*

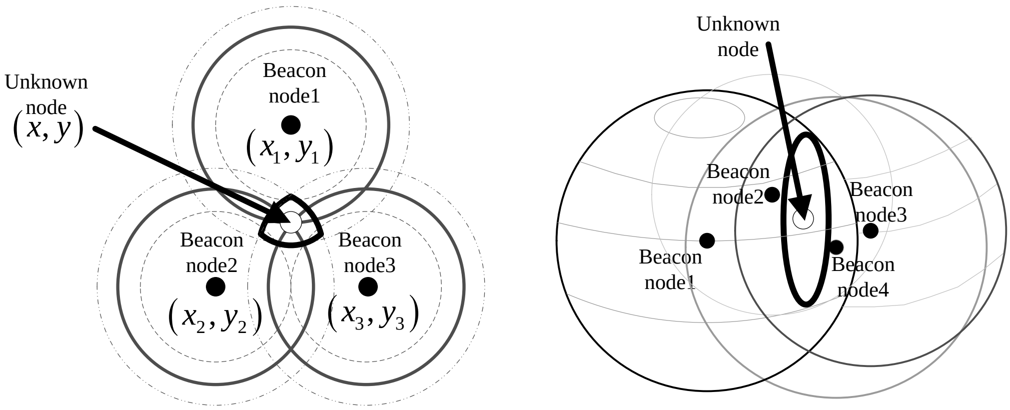
1. Introduction

With the fast development of micro-electro-mechanical systems (MEMS) technology, as well as the advancement in digital electronics and wireless communications, wireless sensor network is being increasingly applied in many fields[1-3], such as public safety and military affairs, environment inspection, health applications, home Applications, traffic management, long-distance control of dangerous region, and so on. In numerous applications, an important problem after event detection is occurrence position of the event. Drawing support from GPS (Global Positioning System) or BDS (BeiDou Navigation Satellite System) device installed in the sensor node, position information of sensor node may be easily acquired[4]. However, this is only applies to unobstructed outdoor environment. Limited by GPS/BDS satellites in space, during war time, the technology may be not always possible. In addition, GPS/BDS requires expensive and high energy consumption electric equipment to control accurate satellite synchronization. Thus, it would be impossible to add every sensor node with GPS/BDS device to collect sensor node's position. For this reason, within the monitoring area of wireless sensor network, sensor node's position is normally figured out with certain estimation algorithm and based on the position of known sensor nodes.

So far, the most existing localization technology of wireless sensor network is roughly divided into two categories[5; 6]: range-based and range-free depending on whether an algorithm uses distance estimation or some other information for estimating sensor node locations. Range-based algorithms usually use distance or angle datum estimates in their

ISSN: 2005-4262 IJGDC
Copyright ⓒ 2015 SERSC

locations estimations, while range-free algorithms only use connectivity information between unknown sensor nodes and beacon nodes. For node localization in 2-dimensional environment, precise distance measurements from at least three beacon nodes are required and we use trilateration for the location estimation of an unknown node. The intersection of three circles around three beacon nodes gives a single point as the location of the unknown node, as showed in Figure 1(a).The same algorithm can be extended to three-dimensional environment by the addition of a fourth beacon node. The intersection of four spherical surfaces around four beacon nodes gives a single point as the location of the unknown node, as showed in Figure 1(b). Range-based approaches have used Received Signal Strength Indicator(RSSI)[7], Time of Arrival (TOA) [8], Time Difference of Arrival (TDoA)[9], and Angle of Arrival (AoA)[10]. Beacon nodes can obtain their own location information in advance by GPS/BDS devices or the artificial deployment information. Generally speaking, range-based technologies can obtain higher localization accuracy, but their performance is limited by the high hardware cost and the heavy power consumption, so they are generally not suitable for large-scale scenarios.



(a)Two-dimensional Location Estimation (b)Three-dimensional Location Estimation

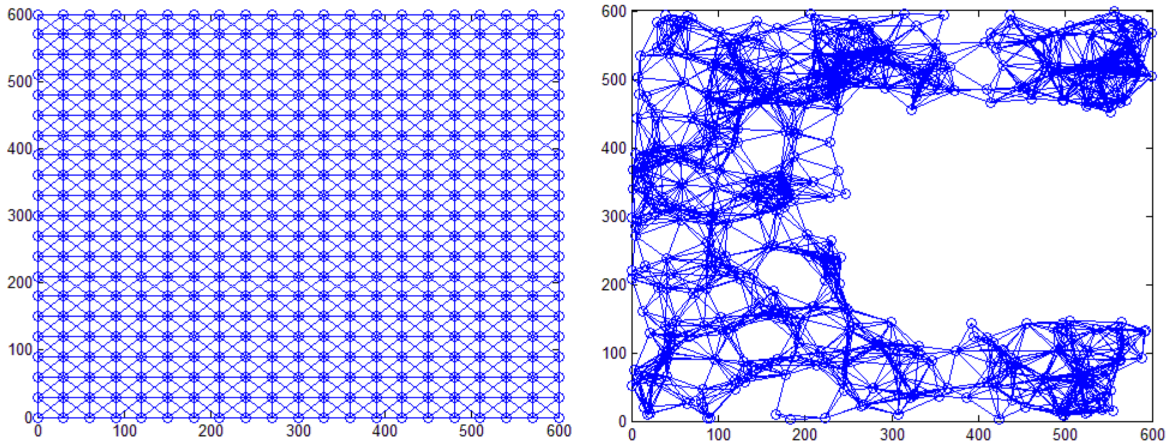
Figure 1. Sensor Node Localization Methods

Range-free localization algorithm calculates location without needing to find the distance between the sensor nodes. Range-free method normally obtains hop-counts (proximity) between sensor nodes from the connection relation between sensor nodes. On this basis, hop-distances(average physical distance of one hop multiplies hop-counts) will be used to replace the real distance between sensor nodes, so as to figure out the estimated position of node. As for this, this algorithm is quite suitable for localization application under large-scale situation. However, influenced by limited hardware of error, lack of energy, severe weather and other practical environmental factors, sensor network often takes on the feature of anisotropy. In anisotropic networks, using hop-distances to replace real distances may lead to significant error. In order to accurately find the relationship between hop-counts and real distances in general network topology and anisotropic network, so as to design a popular positioning algorithm, the paper proposes a kind of more pervasive location estimation method: principal component regression-range-free positioning method, *i.e.* PCR-RF. The algorithm to construct the mapping relationship between hop-counts and real distances via principal component regression, so as to calculate corresponding estimated distances after getting the hop-counts from unknown sensor nodes to know sensor nodes. Eventually, the coordinate position of unknown sensor nodes may be estimated by trilateration or multilateration methods.

 Copyright ⓒ 2015 SERSC

2. Background

Assuming a wireless sensor network, it is assumed that there is a mapping relationship $f:\mathbb{R}^{2d}\to\mathbb{R}$, which describing the real distances and hop-counts mapping relationship between sensor node pairs. Assuming that the measured hop-count from sensor node x_i to sensor node x_j is described as $h_{ij}=f(x_i,x_j)$, while the real distances between sensor nodes is d_{ij} , if $h_{ij}=f(x_i,x_j)=g(d_{ij})$, where $g:\mathbb{R}\to\mathbb{R}$ then such wireless sensor network is referred to as isotropy; or else, it is called anisotropy[11]. Isotropic network is quite rare in practice environment. Instead, most existing sensor network are anisotropic[12]. Anisotropic network is mainly resulted by barriers, uneven distribution, or failure of certain sensor nodes, which lead to significant holes in sensor node deployment area. Figure.2 shows isotropic network and anisotropic network .



(a) Isotropic Sensor Network (b) Anisotropic Sensor Network

Figure 2. Two Kinds of Sensor Network

DV-hop [13]localization method is the typical implementation of the multihop-based localization designs, which measure the internode hop-counts and linearly converts them to the real distances estimates by computing average hop-distances. DV-Hop localization algorithm has good distributiveness and expandability. DV-hop follows three steps:

1. First, each sensor node estimates the minimum hop-counts to each beacon and maintains a data table $[x_i,y_i,h_i]$ and exchanges updates only with its neighbor nodes. In which $[x_i, y_i]^T$ denotes the real coordinate of beacon node i , h_i is a counter to record the hop-counts to beacon node. This step is the classical Bellman-Ford distributed shortest path algorithm.

2. In the second step ,beacon nodes cooperatively estimate the mean distance of each hop in the deployment area. Once an beacon node j gets the hop-count h_i to beacon node i, beacon node j reports the value of h_i to beacon node i. After collecting these datum from all other beacon nodes, beacon node i (coordinate at $[x_i, y_i]^T$) estimates the average distance of each hop in the deployment area, employs it as an modified value and broadcasts it to the whole network. The average distance of each hop can be expressed by the following equation:

$$HopSize_i=\frac{\sum_{i\neq j}\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}{\sum_{i\neq j}h_i}$$

3. After receiving the adjusted value, an unknown sensor node may estimate distances to the rest of the beacon nodes. Assuming that an unknown node receives the messages from more than three beacon nodes, *e.g.* beacon nodes i, j, and k. It uses the three distance estimates ($HopSize_i \times h_i$, $HopSize_j \times h_j$, and $HopSize_k \times h_k$) to determine its coordinate location by trilateration method.

The positioning accuracy of DV-Hop algorithm mainly relies on the precision of estimated hop-distances[14]. However, there exists certain error between hop-distance and real-distances between sensor nodes. What's more, topological structure of the network also influences positioning performance. Hereby, normally, DV-Hop algorithm is only suitable for isotropic and dense node distribution network where hop length is uniform across all segments of the network. DV-Hop algorithm may not get satisfactory results for anisotropic networks, where node density and hop-counts in certain parts of the network do not correspond well to the correction factor calculated by the beacon nodes.

In recent years, statistical learning method [15; 16]is adopted to excavate knowledge concealed in data collection, so as to figure out the localization model. This has become a development tendency of researches on localization of sensor network. For range-free method, learning machine method may be used to calculate the similarity between nodes based on their connectivity, *i.e.* training a prediction model according to the similarity or dissimilarity between nodes. On this basis, nodes' relative coordinates or absolute coordinates may be predicted with the prediction model. We can obtain the optimization transformation matrix through statistical learning method. The matrix can compensate the error caused by the distance estimation of the anisotropic problem, and obtain more accurate mapping relationship between hop-counts and real distances. It makes the localization algorithm can adapt to different complex environments. Shang *et al.* [17]puts forward the MDS-MAP positioning algorithm based on multidimensional scaling method on the basis of connectivity between sensor nodes. Multidimensional scaling algorithm[18] is a case in statistical learning method which is originally developed for mathematical psychology. The MDS-MAP positioning algorithm is based on the assumption that there are no isolated sensor nodes in the sensor network and uses the minimum hop-counts between any pairs-nodes as its distance, which transfers the localization method into the dimension reduction method. The classical MDS-MAP localization algorithm has three phases:

1. Generate a $n \times n$ squares distance matrix **D**, which is composed of the estimated distance between sensor node i and sensor node j. The matrix **D** as follows:

$$\mathrm{D} = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{pmatrix}$$

Where, d_{ij} is the estimated distance between sensor nodes i and j, which is given by

$$d_{ij} = \sqrt{\sum_{k=1}^{2or3} (s_{ik} - s_{jk})^2}$$

Here, s_{ik} is location of sensor node i.

2. Apply the classical metric MDS method for matrix **D** to determine a map that gives the relative locations of all sensor nodes in space.

3. Transform the relative location to the absolute location based on the given beacon sensors.

It is noted that MDS-MAP is a centralized method and is not much suitable for complicated environment of wireless sensor networks. It has a high requirement for global

 Copyright ⓒ 2015 SERSC

communication network of sensor nodes, however, under the anisotropic network; the hop-distances are markedly different from the real distances between sensor nodes, so the positioning accuracy is decreased significantly.

Lim *et al.* [19]proposed a PDM(proximity distance map) algorithm based on linear TSVD(truncated singular value decomposition)technique[20]. PDM describes the optimal linear transformations between the hop-counts and the real distances under the least-squares metric. With the help of PDM, an arbitrary unknown sensor node gets more accurate distance translation, thus to gain a better position estimation. Firstly, the PDM method make uses of matrices to express the collected real distances and the hop-counts between known sensor nodes; secondly, linear TSVD technique is used to conduct linear transformation of two matrices to obtain an optimum linear transformation model; lastly, the hop-counts from the unknown sensor nodes to the known sensor nodes will be applied to this transformation model to estimate the real distances between the unknown sensor nodes and the known sensor node. In essence, TSVD is a multivariable linear regularization learning method, the estimated real distances obtained through method is actually the weighted sum of the estimated distances of other known sensor nodes in the monitoring region, and therefore, the obtained estimated distances are close to the real distances. In addition, the linear TSVD method has abandoned the small singular values, which can to a certain extent reduce the impact of noise during the transformation process, so the ill-posed problem [21-23]during the node localization process can be avoided, and the stability of algorithm can be increased. All these have caused the algorithm to have a low requirement for the deployment of sensor nodes, connection and signal attenuation method, which more benefits its use in complex real environments. The process of PDM localization algorithm can be implemented in a distributed way as follows:

1. Suppose there are m beacon nodes. The hop-counts measured from a (beacon or non-beacon) sensor node to beacon nodes define its coordinate in a linear system. The location of the i^{th} beacon is represented by the proximity vector.

   $$\mathrm{h}_i = [h_{i1}, \cdots, h_{im}]^T$$

   Where hij is the hop-count between beacon node *i* and beacon node *j* and h_{ii}=0. After communication, the overall region can be represented by an $m \times m$ proximity matrix **H**, whose i^{th} column is the coordinate of the i^{th} beacon:

   $$\mathrm{H} = [\mathrm{h}_1, \cdots, \mathrm{h}_m]$$

   Here, **H** is a square matrix with zero diagonal elements.
   Similarly, the real distance vector and matrix between beacon nodes are defined as:

   $\mathrm{d}_i = [d_{i1}, \cdots, d_{im}]^T$ and $\mathrm{D} = [\mathrm{d}_1, \cdots, \mathrm{d}_m]$

   where d^{ij} is the distance of beacon node *i* and beacon node *j*, which can be obtained according to the coordinates of the two beacon nodes. The real distance matrix D is an $m \times m$ symmetric square matrix with zero diagonal elements.

2. TSVD method is used to execute a linear transformation of real distance matrix **D** and proximity matrix **H** to gain an optimum linear transformation **T.** Note that **T** is an $m \times m$ square matrix. Each row vector $\mathbf{t}_i$ of **T** can be obtained by minimizing the following square error:

   $$\varepsilon_i = \sum_{k=1}^{m} (d_{ik} - \mathrm{t}_i \mathrm{h}_k)^2 = \|\mathrm{d}_i - \mathrm{t}_i \mathrm{H}\|^2$$

   The least square solutions of the row vector t_i is

   $$\mathrm{t}_i = \mathrm{d}_i^T \mathrm{H}^T (\mathrm{H}\mathrm{H}^T)^{-1}$$

   As a result, the transformation model of PDM localization is defined as

   $$\mathrm{T} = \mathrm{D}\mathrm{H}^T (\mathrm{H}\mathrm{H}^T)^{-1}$$

3. An unknown node S_u can obtain its measured hop-count vector $\mathbf{h}_u$ by counting the hop counts for all other beacon nodes. It then obtains the estimate of its real distances to all other remaining beacon nodes by multiplying $\mathbf{h}_u$ with PDM:

$$\hat{\mathrm{d}}_u = \mathrm{Th}_u$$

Finally, integrating coordinates and estimation real distances of known nodes, trilateration or maximum likelihood estimation method may be used to estimate unknown sensor nodes, so as to obtain the estimation coordinates.

In a certain degree, PDM method can solve some problems of range-free method under anisotropic network, but the article and experiment show that the PDM method only works under certain monitoring areas, and when the beacon nodes are sparse or serious interference, the positioning accuracy of PDM method will sharply decrease. The major drawback of PDM is that it need to set a threshold parameter *k*. TSVD technique directly sets the singular values smaller than the threshold parameter *k* as zero, and if k is properly chosen, the solution of TSVD is stable, otherwise, it will reduce the algorithm's positioning accuracy. In addition, the PDM method has not conducted standard processing to the hop-count matrix and real distance matrix, and different sizes have caused a certain degree of data from submergence.

Inspired by PDM localization method, PCR-RF method proposed in this paper reforms traditional range-free localization method, which makes use of the maximum variance to construct the mapping relationship from hop-counts to real-distances. Moreover, a specified proportion accumulative contribution rate is set to reduce the influence of multiple correlations. As for this, PCR-based nature improves the estimation precision and performance of range-free positioning algorithm.

3. Sensor Node Location Estimation

3.1. Problem Statement

Consider a plane localization problem. Suppose there are n sensor nodes $\{S_i\}_{i=1}^{n}$ placed in a 2D geographical region $C \subseteq \mathbb{R}^2$ .Let $\mathrm{x}_i \in \mathbb{R}^2$ denote the coordinate location of the sensor node S_i.Without loss of generability, let the first $m(m << n)$ sensor nodes be beacon nodes, whose coordinate locations are known. For every pair of sensors S_i and S_j , h_{ij} denotes shortest hop-count and d_{ij} denotes corresponding real distance that sensor node S_i receives from sensor node S_j. After running for a period of time, we can obtain two kinds of locations data, *i.e.*, the shortest hop-counts matrix $\mathbf{H}=[\mathbf{h}_1,\ldots,\mathbf{h}_m]$ and the physical distance matrix $\mathbf{D}=[\mathbf{d}_1,\ldots,\mathbf{d}_m]$, where $\mathbf{h}_i=[h_{i1},\ldots,h_{im}]^T$, $\mathbf{d}_i=[d_{i1},\ldots, d_{im}]^T$, h_{ij} is the hop-count between the i^{th} beacon node to the j^{th} beacon node and $h_{ii}=0$, Similarly, d_{ij} is the corresponding real distance of beacon nodes i and j, which can be calculated according to the locations of the two beacon nodes. Our objective is to determine the coordinate locations $\{\mathbf{x}_i\}_{i=m+1}^{n}$ of the n-m non-beacons based upon the locations of beacon nodes $\{\mathbf{x}_i\}_{i=1}^{m}$ and other localization information (hop-counts $\{h_{ij}\}_{i,j=1}^{n}$).

3.2. Localization Algorithm

In general, PCR-based location estimation systems have two phases: an offline training phase and an online localization phase. In the offline training phase, a mapping model is under construction for location estimation. In the online localization phase, the mapping model is used to estimate the location of unknown nodes.

 Copyright ⓒ 2015 SERSC

Assuming that, in training phase, after receiving useful data value for a period of time, beacon nodes will get hop-count matrix **H** and corresponding real distance matrix **D**. Considering the relationship between real distances and hop-counts, and according to the principal component regression theory[24], we can obtain an equation, which can be expressed as :

$$\mathbf{D} = \mathbf{H}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{1}$$

Where, $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_m)^T$ is the regression coefficient vector, $\boldsymbol{\varepsilon}$ is the errors vector.

In offline training phase, we firstly calculate the maximum variance direction of hop-counts matrix. The minimum hop-count matrix **H** and its mean vector is set as $\bar{\mathbf{h}}$. Hereby, the covariance matrix can be described as follows:

$$Cov = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{h}_i - \bar{\mathbf{h}})(\mathbf{h}_i - \bar{\mathbf{h}})^T \tag{2}$$

Setting that $\tilde{\mathbf{H}} = (\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_m)$, where $\tilde{\mathbf{h}}_i = \mathbf{h}_i - \bar{\mathbf{h}}$. Then Equation (2)can be transformed into:

$$S_t = \frac{1}{m}\sum_{i=1}^{m}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T \tag{3}$$

Extracting equation (3) the first k orthonormal eigenvectors corresponding to eigenvalues $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$, so as to maximize the general divergence of features when criterion function (4) reaches the projection; the target function may be maximized:

$$J(\mathbf{u}) = \mathbf{u}^T S_t \mathbf{u} \tag{4}$$

In practice, in order to computational efficiency, the eigenvector of covariance is resolved by the singular value decomposition (SVD). Details are as follows:

Building a matrix: $Q = \frac{1}{m}\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}$, where, the dimensionality of Q is $m \times m$. According to the singular value decomposition [25]theorem, the first k non-zero feature values $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_k$ of Q and S_t are exactly the same. Thus, firstly, by drawing support from the matrix Q, we are to figure out its features values $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_k$, as well as corresponding standard orthogonal feature vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$. Then, according to function (5), we may figure out the standard orthogonal feature vectors corresponding to the first k maximum non-zero feature values of S_t,which are $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$.

$$\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}}\tilde{\mathbf{h}}\mathbf{v}_i,\ i = 1, \cdots, k \tag{5}$$

Thus, for any hop-count $\mathbf{h}$, the principal component feature shall be:

$$\breve{\mathbf{h}} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)^T\mathbf{h} \tag{6}$$

So, we can get the mapping model of hop-counts and physical distances:

$$\hat{\boldsymbol{\mu}} = diag(1/m) \times \breve{\mathbf{H}}^T \times \mathbf{D} \tag{7}$$

At this time, unknown nodes $\{S_i\}_{i=m+1}^{n}$ can obtain theirs hop-count matrix $\mathbf{H}_t$ by counting the hop-counts to all beacons.it then obtains the estimate of theirs physical distances to all beacons by multiplying $\hat{\boldsymbol{\mu}}$. Finally, we use the trilateration or multilateration method to estimate location of unknown nodes.

4. Performance Evaluation

One of the notable features of multi-hop range-free localization method is that, it is quite fit for large-scale deployment. This requires thousands of sensor nodes, while in labs,

Copyright ⓒ 2015 SERSC

it is difficult to realize such large-scale real network. As for this, in researches on large-scale multi-hop range-free node localization algorithm, software simulation is often asked to evaluate the performance of the proposed method with the performance of previous methods.

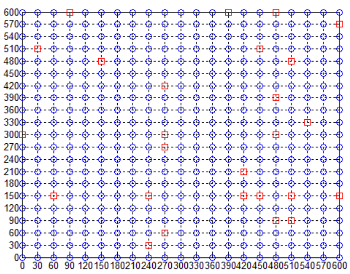
4.1 Simulation Setting

In this section, the performance of PCR-RF method is intended to be analysed and assessed via simulation experiment. Matlab2013b software is employed to analyse and compare methods set out in the present paper. In the experiment, all sensor nodes are evenly distributed in the two-dimensional plane. In order to reduce the one-sidedness of a single experiment, each deployment goes through 50 simulations while nodes in each experiment are randomly re-distributed in the experimental area. Mean value of 50 ALE(Average Localization Error)[26] is taken as the assessment basis.

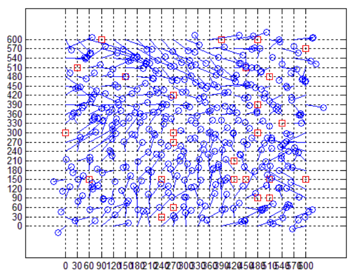
In order to assess the performance of methods proposed in this paper, the sensor nodes are deemed to be randomly or regularly deployed in the deployment area. In addition, in order to assess the adaptability of the proposed methods to network topology anisotropy, an obstruction is added in the aforementioned two deployment strategies, *i.e.* assuming that there is a large obstruction in the deployment area, impeding the direct communication between nodes. Such this area is of C-shape. In allusion to different network topology structure, nodes are re-deployed in the same area for several times, assessing the average localization error. For a fair comparison, this experiment also compare our method with three previous methods: (1)The classic DV-Hop method proposed in[13]; (2) PDM proposed in[19]in two group experiments. Furthermore, for fairness, in PDM localization, we denoted abandoning threshold in TSVD for 3, *i.e.* abandoning eigenvectors with eigenvalue less than or equal to 3.

4.2 Regularly Deployed Sensors

In this set of experiments, 441 and 333 sensor nodes are regularly deployed in the area of square or C-Shape region. The number of beacon nodes was gradually increased from 22 to 32, with a step size of 2.The beacon nodes are randomly placed. Figure 3 shows the result of DV-HOP,PDM and PCR-RF with 26 beacon nodes. In Figure 3, the blue circles (○) denotes the sensor nodes while the red rectangles (□) denote the beacon nodes ,the line connects the actual coordinate and estimated coordinate of unknown nodes, and the longer the line, the more the estimated value deviates from the actual location. In Figure 3, the average localization error(ALE) for DV-Hop, PDM and PCR-RF is about 53.8%,50.9% and 37.2%,respectively.

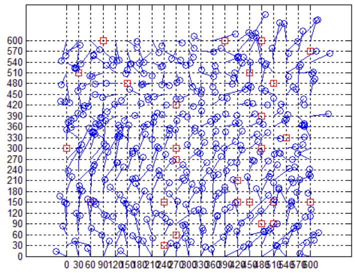


(a) Node Distribution

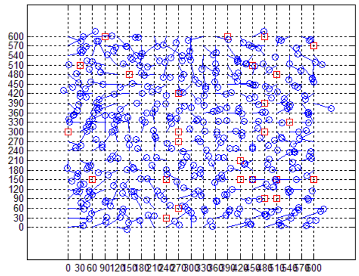


(b) The Positioning Results of the DV-Hop Algorithm

 Copyright ⓒ 2015 SERSC



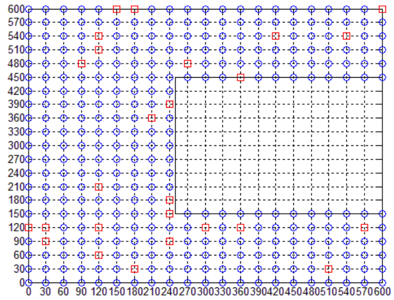
(c) The Positioning Results of the PDM Algorithm



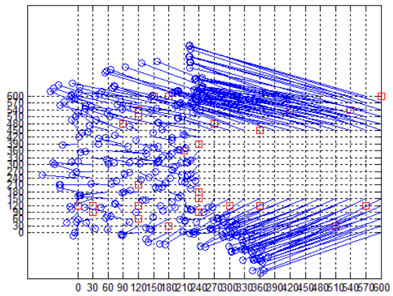
(d) The Positioning Results of the PCR-RF Algorithm

Figure 3. Estimation for the Regular Deployment Network with 26 Beacon Nodes in Square Area

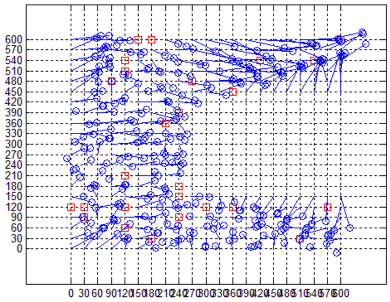
Figure 4 respectively shows the positioning results of DV-Hop, PDM and PCR-RF with 26 beacon nodes. Figure 4(a) shows that the ALE is about 192.01%; Figure 4(b)shows that the ALE is about 66.4%; Figure 4(b)shows that the ALE is about 53.3%.



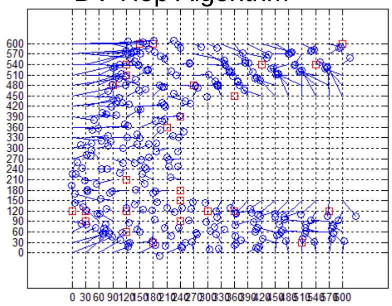
(a) Node Distribution



(b) The Positioning Results of the DV-Hop Algorithm



(c) The Positioning Results of the PDM Algorithm



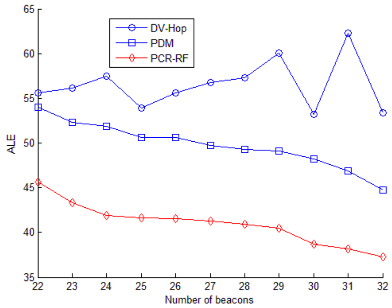
(d) The Positioning Results of the PCR-RF Algorithm

Figure 4. Estimation for the Regular Deployment Network with 26 Beacon Nodes in C-shape Area

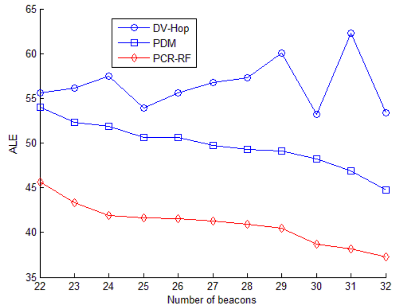
Figure 5 illustrates the curve of average ALE of repeated experiments by three localization algorithms varying with different numbers of beacon nodes in regular deployment scenarios. From Figure 5 we can clearly see that both in n the C-shape area and in the square area, its ALE value is still have ups and downs. The maximum ALE of DV-Hop algorithm in square area and C-shape area of the maximum value is 62.27% and 216% respectively. Since the DV-Hop method using hop-distances instead of real distances in according to distance vector routing theory, so its error is large and the algorithm is not stable. The ALE value of PDM method is obviously smaller than that of DV-Hop algorithm along with the increment of beacon nodes. PDM method eliminates some data of smaller eigenvalue. As the threshold parameter is fix artificially, disregarding factors of hop-counts and real distances dimension, its localization accuracy is only slightly

Copyright ⓒ 2015 SERSC

improved than DV-Hop method. In addition, collinear problem between nodes is further intensify the positioning error. Its average performance index relative to DV-Hop method was increased by 14% and 166% in two scenarios. The PCR-RF method takes full account of the relationship between the hop-counts and real-distances, therefore, its positioning performance than DV-Hop algorithm are improved respectively 38% and 288%. In particular, the positioning performance of PDM and PCR-RF method is improved particularly evident.



(a) Average Localization Error in Square Area

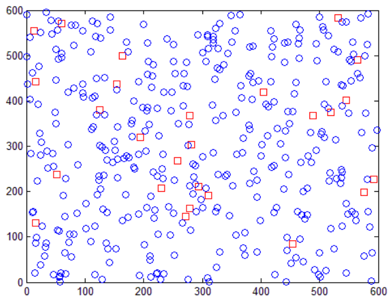


(b) Average Localization Error in C-Shape Area

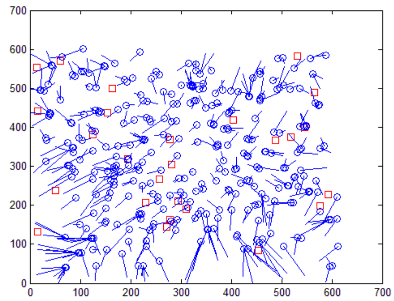
Figure 5. Average Localization Error of Regular Deployment with Different Number of Beacon Nodes

4.3 Random Deployed Sensors

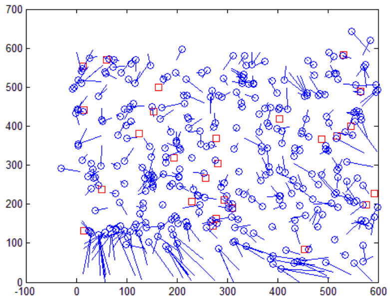
Due to random deployment is more close to the actual situation, the experiment of this section is mainly designed to compare the performance of different algorithms in multiple random deployment environment. Similarly, random deployment experiment also includes two kinds of situations, C-shaped area with obstruction and square area without obstruction. Figure 6 shows the result of DV-HOP,PDM and PCR-RF with 26 beacon nodes in the random deployment scenarios. In Figure 3, the average localization error(ALE) for DV-Hop, PDM and PCR-RF is about 58.6%,52.9% and 47.2%,respectively.



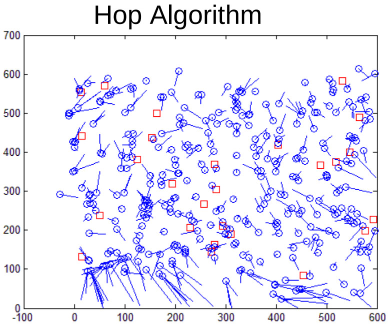
(a) Node Distribution



(b) The Positioning Results of the DV-Hop Algorithm



(c) The Positioning Results of the PDM Algorithm

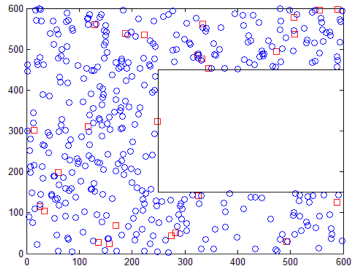


(d) The Positioning Results of the PCR-RF Algorithm

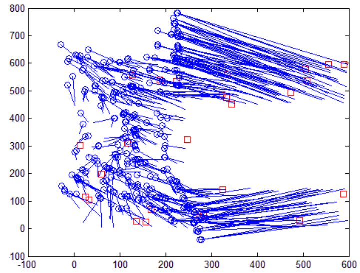
Figure 6. Estimation for the Random Deployment Network with 26 Beacon Nodes in Square Area

Copyright ⓒ 2015 SERSC

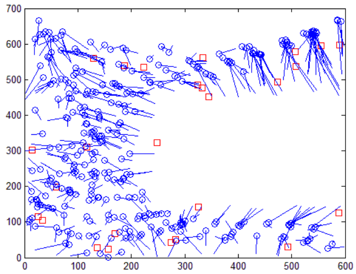
Figure 7(b)-(d) shows the final localization result of three kind of positioning algorithm with 26 beacon nodes in a square area. Figure 7(a) shows deployment of nodes; Figure 7(b) shows localization results of DV-Hop method, and in this figure, ALE = 204.39%; Figure 7(c) shows PDM method proposed by Lim; and in this figure, ALE = 65.7%;Figure 7(d) shows PCR-RF proposed in this paper, and in this figure, ALE = 52.6%.



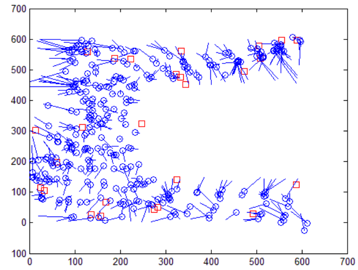
(a) Node Distribution



(b) The Positioning Results of the DV-Hop Algorithm



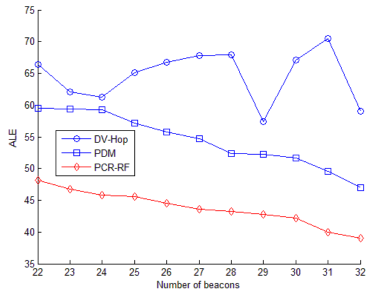
(c) The Positioning Results of the PDM Algorithm



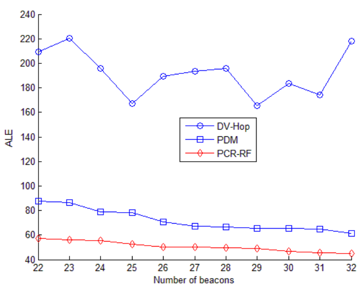
(d) The Positioning Results of the PCR-RF Algorithm

Figure 7 Estimation for the Random Deployment Network with 26 Beacon Nodes in C-shape Area

Fig.8(a-b) shows ALE curve of three localization algorithms, with the number of beacon nodes varying from 22 to 32. Fig.8(a) is ALE curve for random deployment in square area. Fig.8(b) is ALE curve for random deployment in C-shaped area. Similarly, the ambiguous relationship between hop-counts and real distances and beacon node collinearity , so that the localization result of DV-Hop is poor and unstable. Especially in C-shaped area, NLOS problem even enlarged the ALE of DV-Hop. The other methods are more or less improved than DV-Hop algorithm. Compared with DV-Hop method, the performance of PDM was increased by 18.9%and 166%. Similarly, the performance of the PCR method to improve the 47.7% and 279% respectively.



(a) Average Localization Error in Square Area



(b) Average Localization Error in C-Shape Area

Figure 5 Average Localization Error of Random Deployment with Different Number of Beacon Nodes

5. Conclusion

The paper is intended to study PCR-based multi-hop range-free positioning method. Basic ideology of the algorithm is to take PCA as the pioneer of regression. PCA defines the so-called main axis to measure data variance. Such main axis gives the direction of maximum variance in decreasing order of importance. PCA defines a threshold value to eliminate principal directions lower than the threshold value, so as to reduce noise, and to avoid co-linearity between samples. As for this, its regression method may be able to accurately establish the projection from the hop-counts to real distances of beacon nodes. After having figured out the hop count between unknown nodes and known nodes, the hop-counts may be converted into estimated distances, realizing compensation on distance estimation under complicated environment. The method inherits advantages of innovative multi-hop range-free method. On the other hand, it keeps fine positioning performance under complicated anisotropic environment. However, similar to other key technologies, PCR-RF still has plenty of technical issues to be resolved. Especially, when the number of beacon nodes is too small or too large, calculation of PCR may be quite unintelligible. With regard to the problem, in future, improvement shall be made to fit the algorithm for beacon nodes of different quantity.

Acknowledgements

The paper is sponsored by Provincial University Natural Science Research Foundation of Jiangsu Education Department (12KJD510006, 13KJD520004); Natural Science Foundation of Jiangsu(BK2012082); Doctoral Scientific Research Startup Foundation of Jinling Institute of Technology (JIT-B-201411); Educational Reform Foundation of Jinling Institute of Technology(2010JGXM-02-2).

References

[1] K. Sohraby, D. Minoli and T. Znati, "Wirelss Sensor Networks Technology, Protocols and Applications", A john wiley & sons, inc., **(2007)**.
[2] I. F. Akyildiz and M. C. Vuran, "Wireless Sensor Networks", A John Wiley and Sons, **(2010)**.
[3] S. Khan, A. S. K. Pathan and N. A. Alrajeh, "Wireless Sensor Networks Current Status and Future Trends", CRC Press, **(2013)**.
[4] S. Sand, A. Dammann and C. Mensing, "Positioning in Wireless Communications Systems", Wiley, **(2014)**.
[5] Y. Liu and Z. Yang, "Location, Localization,and Localizability Location-awareness Technology for Wireless Networks", Springer, **(2011)**.
[6] Z. Chen, F. Xia and T. Huang, "A localization method for the Internet of Things", The Journal of Supercomputing, vol. 63, no. 3, **(2013)**, pp. 657-674.
[7] B. Wei, "Algorithms for wireless sensor network localization based on received signal strength indication", **(2013)**.
[8] A. Giorgetti and M. Chiani, "Time-of-Arrival Estimation Based on Information Theoretic Criteria", Signal Processing, IEEE Transactions,
vol. 61, no. 8, **(2013)**,pp. 1869-1879.
[9] K. H. Choi, W. S. Ra and J. B. Park, "Compensated robust least-squares estimator for target localisation in sensor network using time difference of arrival measurements", IET Signal Processing vol. 7, no. 8, **(2013)**, pp. 664-673.
[10] R. Peng and M. L. Sichitiu, "Angle of Arrival Localization for Wireless Sensor Networks", 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, **(2006)**.
[11] S. Nikoletseas and J. D. P. Rolim, "Theoretical Aspects of Distributed Computing in Sensor Networks", **(2011)**.
[12] G. Zhou, T. He and S. Krishnamurthy, "Models and Solutions for Radio Irregularity in Wireless Sensor Networks", ACM Transactions on Sensor Networks (TOSN), vol. 2, no. 2, **(2006)**, pp. 221-262.
[13] D. Niculescu and B. Nath, "DV Based Positioning in Ad Hoc Networks", Telecommunication Systems, vol. 22, no. 1-4, **(2003)**, pp. 267-280.
[14] G. Wu, S. Wang and B. Wang, "A novel range-free localization based on regulated neighborhood distance for wireless ad hoc and sensor networks", Computer Networks, vol. 56, **(2012)**, pp. 3581-3593.

 Copyright ⓒ 2015 SERSC

[15] M. A. Alsheikh, S. Lin and D. Niyato, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications", Communications Surveys & Tutorials, vol. 99, **(2014)**, pp. 1-24.
[16] S. Afzal and H. Beigy, "A localization algorithm for large scale mobile wireless sensor networks: a learning approach", The Journal of Supercomputing, **(2014)**.
[17] Y. Shang, W. Ruml and Y. Zhang, "Localization from mere connectivity", MobiHoc Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, **(2003)**.
[18] F. W. Young, "Multidimensional Scaling: History, Theory, and Applications", **(1987)**.
[19] H. Lim and J. C. Hou, "Distributed localization for anisotropic sensor networks", ACM Transactions on Sensor Networks (TOSN), vol. 5, no. 2, **(2009)**, pp. 1-26.
[20] L. Hogben, "Handbook of Linear Algebra", United Kingdom: CRC Press, **(2007)**.
[21] D. Kotzor, W. Utschick and S. Member, "Kernel Methods for Ill-Posed Range-Based Localization Problems", IEEE Transactions on Signal Processing, vol. 60,no. 8, **(2012)**, pp. 4151-4162.
[22] K. Bu, Q. Xiao and Z. Sun, "Toward collinearity-aware and conflict-friendly localization for wireless sensor networks", Computer Communications, vol. 35, **(2012)**, **(2012)**, pp. 1549-1560.
[23] Y. X. Yong, Q. H. Yan and G. D. Min, "Multicollinearity-based 3D DV-Hop Localization Algorithm", Computer Science 2011, vol. 38, no. 5, **(2011)**, pp. 37-40.
[24] S. Chattefuee and A. S. Had1 Regression, "Analysis by Example", John Wiley, **(2007)**.
[25] H. Abdi and L. J. William, "Wiley Interdisciplinary Reviews: Computational Statistics", Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, no. 4, **(2010)**, pp. 433-459.
[26] X. Yan, H. Qian and J. Yu, "A Localization Method Based on Principal Component Analysis", Journal of Computational Information Systems, vol. 8, no. 22, **(2012)**, pp. 9425-9432.

Authors

XiangHong Tian. He received the Bachelor Degree from Xi'an Shiyou University in 1994 and Master degree in Computer Application Technology from Nanjing University of Science and Technology in 2008. Since 2010, he is an associate professor at the School of Computer Engineering, Jinling Institute of Technology. His current research interests include wireless sensor network and computer network.

Wei Zhao. In 2001, from Nanjing University of Science and Technology, he received the Bachelor Degree in Optoelectronic Technology and Optical Instrument. From the same university, he received the Master Degree in Computer Software Engineering in 2006, and Ph. D Degree in Computer Application Technology in 2014. Now, he works at the School of Computer Engineering in Jinling Institute of Technology. His major research interests include routing and location technology in Wireless Sensor Networks and Machine learning.

Xiaoyong Yan. He is a lecture at the School of Intelligence Science and Control, Jinling Insitute of Technology, China. He received the B.S. degree in mechanical electronic engineering from Nanjing Agricultural University in 2000 and received the M.S. and Ph.D. Degree in computer application technology from Nanjing University of Science and Technology (NUST) in 2006 and 2013, respectively. Now he is a Postdoctoral Fellow at School of Instrument Science and Engineering, Southeast University, China. His current interests are in the areas of wireless sensor networks, pattern recognition and machine learning.

Copyright ⓒ 2015 SERSC

 Copyright ⓒ 2015 SERSC