Computer Networks 106 (2016) 151–160

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Wireless sensor networks localization based on graph embedding with polynomial mapping☆

Hao Xu[a], Huafei Sun[a,*], Yongqiang Cheng[b], Hao Liu[c]

[a] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing (100081), China
[b] School of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan (410000), China
[c] School of Science, University of Science and Technology Liaoning, Anshan (114051), China

ARTICLE INFO

Article history:
Received 20 December 2015
Revised 5 June 2016
Accepted 24 June 2016
Available online 25 June 2016

Keywords:
Graph embedding
Polynomial mapping
Pair-wise distance
Relative locations
Coordinate transformation

ABSTRACT

Localization of unknown nodes in wireless sensor networks, especially for new coming nodes, is an important area and attracts considerable research interests because many applications need to locate the source of incoming measurements as precise as possible. In this paper, in order to estimate the geographic locations of nodes in the wireless sensor networks where most sensors are without an effective self-positioning functionality, a new graph embedding method is presented based on polynomial mapping. The algorithm is used to compute an explicit subspace mapping function between the signal space and the physical space by a small amount of labeled data and a large amount of unlabeled data. To alleviate the inaccurate measurement in the complicated environment and obtain the high dimensional localization data, we view the wireless sensor nodes as a group of distributed devices and use the geodesic distance to measure the dissimilarity between every two sensor nodes. Then employing the polynomial mapping algorithm, the relative locations of sensor nodes are determined and aligned to physical locations by using coordinate transformation with sufficient anchors. In addition, the physical location of a new coming unknown node is easily obtained by the sparse preserving ability of the polynomial embedding manifold. At last, compared with several existing approaches, the performances of the presented algorithm are analyzed under various network topology, communication range and signal noise. The simulation results show the high efficiency of the proposed algorithm in terms of location estimation error.

1. Introduction

Wireless sensor networks (WSN) have received extensive interest lately as a promising technology in many applications of wireless communications, containing manufacturing [1], health caring [2], environment monitoring and forecasting [3], habitat monitoring and tracking. The location of nodes in WSNs plays an important role in most application fields. In addition, knowing the relative locations of sensors makes use of location-based addressing and routing protocols, which can improve network robustness and energy-efficiency effectively. Therefore, sensor nodes localization is one of the fundamental issues in the implementation of WSNs. In the large scale of sensor networks, even though some sensor nodes could be equipped with a global positioning system(GPS) to provide them with their absolute position, this is currently a costly solution or impossible solution to some indoor cases. Therefore, it is often assumed that the positions of some nodes are known exactly, so that it is possible to find the absolute positions of the remaining nodes in the WSNs through the known locations of sensor nodes and the measurement data. The main task of WSNs localization algorithm is to determine the positions of sensor nodes in a network given incomplete and disturbed by noise. Locating the unknown nodes in a wireless system involves the collection of location information from radio signals traveling between the unknown nodes and a number of reference anchor nodes(anchors). There are many classical positioning techniques, including the angle of arrival(AOA) [4–6], the received signal strength(RSS) [7–9], or time of arrival(TOA) [10–12], which can be all used to determine the location of unknown nodes. The AOA technique uses the angles between the unknown node and a number of anchors to estimate the location of the unknown nodes, the RSS estimates the received signal strength, and the time-based approaches measure the arrival time of the received signal, respectively.

☆ This subject is supported by the National Natural Science Foundations of China (No. 61179031, No. 61302149).

* Corresponding author.
E-mail addresses: xhheqiao@126.com (H. Xu), huafeisun@bit.edu.cn (H. Sun), nudtyqcheng@gmail.com (Y. Cheng), liuhao123@ustl.edu.cn (H. Liu).

http://dx.doi.org/10.1016/j.comnet.2016.06.032
1389-1286/

Though many traditional localization methods can reach a high localization accuracy, they are not suitable for dealing with large scale sensor networks due to lot of time and material cost. Recently, in order to handle these noisy measurements and time consumption, several algorithms making use of kernel-based machine learning have been proposed for locating sensor nodes with inaccurate measurement in WSNs. Some different measurement methods about the sensor communication with each other are presented in [7]. A graph embedding localization algorithm for WSNs is introduced in [13], which views the sensor nodes as a group of distributed devices to construct a graph to preserve the topological structure of the sensor networks, and employs an appropriate kernel function to measure the dissimilarity between sensors. A kernel isometric mapping(KIsomap) algorithm is used to determine the relative locations of sensors based on geodesic distance [14] and a semi-supervised Laplacian regularized least squares algorithm is presented in [15] using the alignment criterion to learn an appropriate kernel function. In addition, a semi-supervised manifold learning is used to estimate the location of mobile nodes in the WSNs [16].

Most kernel-based localization algorithms can capture the nonlinearity of the measured data due to the nonlinear property of the kernel function. However, they cannot deal with the sensibility of measurement parameters. In order to overcome the shortcoming of the sensitive to the neighborhood parameters, a robust localization method with ensemble-based manifold learning is introduced in [17], and a local patches alignment embedding localization is proposed in [18]. However, a main drawback of many manifold learning methods is that they learn the low-dimensional output data samples implicitly. We cannot obtain an explicit mapping relationship from the input data manifold to the output embedding after the training process. Therefore, in order to obtain the physical locations of new coming samples, the learning procedure has to be extremely time consuming for sequentially arrived localization data. Although some methods based on linear projection have been proposed to get an explicit mapping, the linearity assumption may still be too restrictive. Meanwhile, the kernel embedding methods have been also proposed to give nonlinear but implicit mappings for manifold learning localization [19–21]. In addition, because these mappings are computed within a subset of the feature space rather than the whole feature space itself and are given in terms of the kernel and the training data samples explicitly, their computational process would be quite complicated for large-scaled sensor networks.

In this paper, a new graph embedding algorithm with polynomial mapping (GEPM) is proposed for locating unknown nodes in WSNs, especially to the new coming nodes. Firstly, location data based on pair-wise distance is obtained by the geodesic distance measurement. Consequently, in order to solve the problem of the sparse structure of the sensor networks and the sensitive to neighborhood parameters easily, we can construct a graph to represent the topological structure of the sensor networks and calculate the weight matrix and the sparse preserving matrix. And then, an eigenvalue function can be obtained by the polynomial mapping based on spectral embedding optimization problem. Letting some eigenvectors corresponding to the smallest eigenvalues of function be the polynomial coefficients, the relative locations of all sensor nodes can be obtained based on Rayleigh-Ritz theorem. At last, by the coordinate transformation, we can get final physical locations of all unknown nodes. Meanwhile, for the new coming unknown nodes, we can easily obtain the geodesic distance between them with the datum nodes. Based on the graph theorem, the relative coordinates and physical coordinates can be calculated simply. Compared with linear projection methods and kernel-based nonlinear mapping methods, the presented algorithm gives more accurate embedding and out-of-sample extension results and, meanwhile, is very fast in locating new coming samples.

The main features of the proposed locating method are given as follows.

- The geodesic distance matrix is used to construct the neighbor graph of the WSNs, which reduces the influence of noise in the measurement and receives high accurate location estimation, even with error-prone distance information. In addition, the datum nodes are selected to determine the geodesic distance vector, i.e., the high dimensional localization data, based on the geodesic distance matrix. Comparing with some previous manifold learning localization methods, it is to further reduce the computational complexity and enhance the precision of the solution in our approach. The detailed analysis is given in Section 4.4.
- The embedding mapping is nonlinear and the high dimensional data space is considered as a nonlinear manifold. Comparing with the linear projection-based methods, the proposed mapping provides a nonlinear polynomial mapping from the high dimensional localization data space to the low dimensional representation space. Therefore, it is more reasonable to use a polynomial mapping to handle data samples lying on nonlinear manifolds. Meanwhile, the proposed localization method is suitable to the large scale WSNs and has high localization accuracy.
- The mapping can be straightforward to locate any new sensor nodes participating the WSNs. It is different from the traditional manifold learning localization methods such as MDS, SDP, SVM, and ISOMAP, which are not clear how new localization data sample can be embedded in the low dimensional space for the implicit mapping. Meanwhile, in contrast to the explicit manifold learning algorithm such as KLPP and S^2LapRLS, the proposed localization method has a lower computational complexity and is fast and efficient in finding the low dimensional representations of new localization data even for vary lager data sets.
- The influence of the datum nodes to the localization accuracy is analyzed in detail based on the RMS error under different anchors, communication range and noise standard deviation for WSNs. Finally, the optimal selection of the datum nodes is first given by $m_{op} = [-\frac{3}{2} + \sqrt{2N + \frac{9}{4}}]$ with $q = 2$ for different sensor nodes and interested localization area. The detailed analysis and experiments are given in Section 4.1.

The rest of this paper is organized as follows. The localization problems, containing graph construction, original nodes location and new coming data calculation, are presented in Section 2. Section 3 presents the main contribution of this paper about the sensor nodes location estimation algorithm based on the graph embedding theorem with polynomial mapping. The extensive simulation results are described in Section 4. Finally, some concluding remarks and future works are given in Section 5.

2. Localization problem statement

Let us consider a p-dimensional(pD) localization problem in WSNs consisting of N sensor nodes in an interesting area $C \subseteq \mathcal{R}^p (p = 2 \text{ or } 3)$. Without loss of generality, we assume that the first n sensors are anchors, and the rests are unknown nodes. At the same time, we select the first $m (p < m)$ sensor nodes as datum nodes. The locations of anchors can be obtained by location equipments, e.g., global positioning system, while the unknown nodes need to be measured by the anchors and localization data. Let $\mathbf{p}_i$ be the ith sensor node whose coordinate is expressed as $(\mathbf{p}_{1i}, \ldots, \mathbf{p}_{pi})^T, i = 1, 2, \ldots, N,$ and every sensor node can transmit

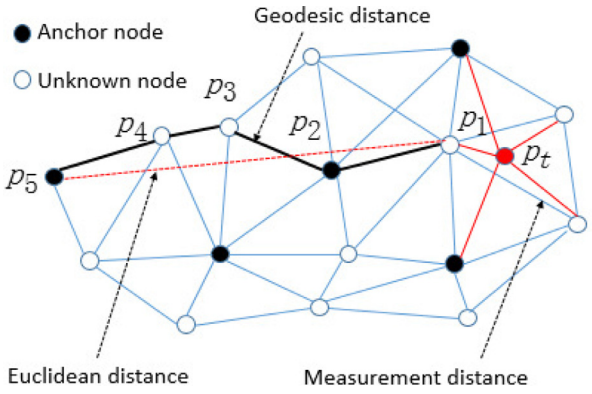

Fig. 1. Distance between the nodes.

localization data to each of its neighbors upon some communication range d. Our goal is to determine the locations $\{\mathbf{p}_i\}_{i=n+1}^{N}$ of the unknown nodes by the anchors locations and the localization data with high localization accuracy and low complexity. Here, we use the measurement pair-wise distance to calculate the localization data which can be obtained by the following description.

For every pair of nodes $\mathbf{p}_i$ and $\mathbf{p}_j$, their measured distance denoted as x_{ij} can be obtained by several measurement techniques which contains time of arrival (TOA), round trip time of arrival (RTTOA), time difference of arrival (TDOA), received signal strength(RSS), connectivity and so on. As analyzed in many literatures such as [14,15], the pair-wise distance x_{ij} can be replaced by the geodesic distance g_{ij} which contains the shortest path between nodes $\mathbf{p}_i$ and $\mathbf{p}_j$ where the geodesic distance matrix $G = [g_{ij}]_{N\times N}$ can be obtained by the following steps. First, we set $g_{ij} = x_{ij}$ if node $\mathbf{p}_i$ is within the communication range of node $\mathbf{p}_j$; $g_{ij} = \infty$ otherwise. Second, for each value of $l = 1, \ldots, N$, replace all entries g_{ij} in turn by $\min\{g_{ij}, g_{il} + g_{lj}\}$. As shown in Fig. 1, where the black circles represent anchors and the circles denote the unknown nodes; the blue lines between the nodes represent the measurement distance through the Euclidean distance. When the nodes are not in each other's communication radius, such as nodes $\mathbf{p}_1$ and $\mathbf{p}_5$, we cannot estimate the Euclidean distance between them(the dotted line); consequently, the geodesic distance(the heavy line) can be obtained, which represents the shortest path between them in the neighborhood.

According to the above criteria, the final matrix G is the geodesic distance matrix which can be seen as a neighbor graph about the WSNs. Therefore, the topological structure of the WSNs can be represented by a neighbor graph based on the geodesic distance, where each vertex represents a sensor node. We denote the geodesic distance vector between the ith sensor and the m datum nodes in the sensor networks as

$$\mathbf{v}_i = (g_{1i}, \ldots, g_{mi})^T, i = 1, \ldots, N, \tag{1}$$

which is also called the high dimensional localization data.

In addition, for a new coming sensor node $\mathbf{p}_t$(the red circle) as shown in Fig. 1, we can firstly obtain the nearest neighboring $\{\mathbf{p}_j\}_{j=1}^{r}$ and the measurement distance between them. Then the geodesic distance between $\mathbf{p}_t$ and the datum node $\mathbf{p}_5$ can be calculated by $\min\{x_{jt} + g_{j5}\}$, for $j = 1, \ldots, r$, where g_{5j} is the geodesic distance between the anchor node $\mathbf{p}_5$ and the jth sensor node $\mathbf{p}_j$ of the nearest neighboring of $\mathbf{p}_t$. Therefore, we can easily get the corresponding geodesic distance vector $\mathbf{v}_t$ by (1). Due to the above processing of the measurement information where each node should communicate measurement data to the center node or the based station, the proposed localization method is a centralized algorithm for wireless sensor networks localization problem [22].

3. Localization algorithm

In this section, we attempt to learn an explicit nonlinear mapping from the localization data space to the physical space for sensor nodes localization. As well known that, the closer the locations of some sensor nodes are, the more similar their localization data are, i.e., the high dimensional localization data lie on a low dimensional embedding manifold determined by the physical location space. In other words, when the locations of some sensor nodes are known, the unknown nodes can be ground by exploiting the geometry of the distribution of localization data, assuming their conditional distribution are similar. After constructing the neighbor graph, we can know that the topological structure of the WSNs can be depicted by the neighbor graph G with a suitable similarity between every two sensors. Hence, the localization problem can be viewed as a graph embedding problem.

Here, we propose an explicit nonlinear mapping for manifold learning with the assumption that there exists a polynomial mapping between high dimensional localization data space and low dimensional representation space. Comparing with the linear projection assumption previously used, the polynomial mapping is more accurate for localization data lying on nonlinear manifolds due to providing a high order approximation to the unknown nonlinear mapping. Precisely, we can assume that the kth component $y_{ki}(k = 1, \ldots, p)$ of the low-dimensional representation $\mathbf{y}_i$ corresponding to the ith sensor node $\mathbf{p}_i$ is a polynomial of degree q in $\mathbf{v}_i$ as the following manner

$$y_{ki} = \sum_{1\le l_1+l_2+\ldots+l_m\le q} \alpha_k^{\mathbf{l}} g_{1i}^{l_1} g_{2i}^{l_2} \ldots g_{mi}^{l_m}, \tag{2}$$

where $l_1, l_2, \ldots, l_m$ are nonnegative integers and $\mathbf{v}_i = (g_{1i}, g_{2i}, \ldots, g_{mi})^T$ is the localization data which have been obtained based on the pair-wise distance measurement by (1). It should be noted that there is not similar terms in the polynomial expression (2) such as $g_{1i}g_{2i}$ and $g_{2i}g_{1i}$, and we only select the former to participate the localization. Therefore, the superscript $\mathbf{l}$ stands for the m-tuple indexing array $(l_1, l_2, \ldots, l_m)$, and the vector of polynomial coefficients α_k can be defined as

$$\alpha_k = \begin{pmatrix} \alpha_k^{\mathbf{l}}|_{l_1=q,l_2=0,\ldots,l_m=0} \\ \alpha_k^{\mathbf{l}}|_{l_1=q-1,l_2=1,\ldots,l_m=0} \\ \vdots \\ \alpha_k^{\mathbf{l}}|_{l_1=0,l_2=q,\ldots,l_m=0} \\ \alpha_k^{\mathbf{l}}|_{l_1=0,l_2=q-1,\ldots,l_m=0} \\ \vdots \\ \alpha_k^{\mathbf{l}}|_{l_1=1,l_2=0,\ldots,l_m=0} \\ \vdots \\ \alpha_k^{\mathbf{l}}|_{l_1=0,l_2=0,\ldots,l_m=1} \end{pmatrix}. \tag{3}$$

For a given localization data sample $\mathbf{v}_i$, we define $V_q^{(i)}$ as

$$V_q^{(i)} = \begin{pmatrix} \overbrace{\mathbf{v}_i \otimes (\mathbf{v}_i \otimes \ldots (\mathbf{v}_i \otimes \mathbf{v}_i))}^{q} \\ \vdots \\ \mathbf{v}_i \otimes \mathbf{v}_i \\ \mathbf{v}_i \end{pmatrix}, \tag{4}$$

where $\otimes$ denotes the Kronecker-like product defined on matrices, i.e., for a vector $\mathbf{v}_i = (g_{1i}, g_{2i}, \ldots, g_{mi})^T$, $\mathbf{v}_i \otimes \overbrace{(\mathbf{v}_i \otimes \ldots (\mathbf{v}_i \otimes \mathbf{v}_i))}^{q-1}$ is a block column vector whose jth block is denoted as $\mathbf{v}_{ji}^q = g_{ji}\mathbf{V}_{ji}^{q-1}$ with $\mathbf{V}_{ji}^{q-1} = (\mathbf{v}_{ji}^{q-1}, \ldots, \mathbf{v}_{mi}^{q-1})^T$ and $\mathbf{v}_{ji}^1 = g_{ji}$. Then, the low-dimensional representation (2) can be rewritten as

$$y_{ki} = \alpha_k^T V_q^{(i)}. \tag{5}$$

Next, we should find the polynomial coefficients α_k to determine an explicit mapping relationship, i.e., the unknown mapping from the high-dimensional localization data samples into their low-dimensional embedding space. It is well known that many manifold learning methods, including Laplacian eigenmap, locally linear embedding and isometric mapping, all can be placed into the framework of spectral embedding [23]. Under this framework, the localization problem is reduced to solve the following optimization problem

$$\min_{\mathbf{y}_i} \sum_{i,j=1}^{N} \frac{1}{2} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2, \tag{6}$$

$$\text{s.t.} \sum_{i=1}^{N} \lambda_i \mathbf{y}_i \mathbf{y}_i^T = I_p, \tag{7}$$

where $W = [w_{ij}]_{N\times N}$ defined by the localization data is symmetry and positive weights matrix with w_{ij} being the weight of the edge between $\mathbf{p}_i$ and $\mathbf{p}_j$, while $\lambda_i = \sum_{j=1}^{N} w_{ij}$. By the pair-wise distance measurement data, for a given communication range d, we can define the weights matrix W with the heat kernel function

$$w_{ij} = \begin{cases} \exp(-w_g g_{ij}^2), & \text{if } g_{ij} \le d; \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where w_g is the user-defined parameter which is usually given by $w_g = \frac{1}{\max\{g_{ij}^2\}}$.

Therefore, substituting (5) and (8) into (6) and (7) with a simply algebraic calculation, the optimization problem can be reduced to

$$\min_{\alpha_k} \sum_{k}^{p} \alpha_k^T V_q U V_q^T \alpha_k, \tag{9}$$

$$\text{s.t.} \quad \alpha_j^T V_q D V_q^T \alpha_k = \delta_{jk}, \tag{10}$$

where $V_q = [V_q^{(1)}, V_q^{(2)}, \ldots, V_q^{(N)}]_{M\times N}$ is called the polynomial kernel matrix with $M = \sum_{i=1}^{q} C_{m+i-1}^{m-1}$, $U = D - W$ is the Laplacian matrix and D is a positive diagonal matrix with $D_{ii} = \sum_j^N w_{ij}$ which represents the degree of sensor networks.

In general, because $\mathbf{v}_i \neq \mathbf{v}_j$ with $i \neq j$ and $v_{ij} \geq 0$, then we can know that $V_q D V_q^T$ is $M \times M$ positive matrix if $M \leq N$, otherwise not. Therefore, based on the Rayleigh–Ritz theorem [24], the optimal solutions α_k can be obtained by solving the following generalized eigenvalue problem

$$V_q U V_q^T \alpha_k = \mu V_q D V_q^T \alpha_k, \alpha_k^T V_q D V_q^T \alpha_j = \delta_{kj}. \tag{11}$$

Then, we can get the column vectors, the solution of the above equations, as $\alpha_1, \alpha_2, \ldots, \alpha_p$, ordered according to their p smallest non-zero eigenvalue $\mu_1 \le \mu_2 \le \ldots \le \mu_p$. Thus, for the original training points, the p-dimensional mapping can be represented as

$$Y = AV_q, \tag{12}$$

where $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$ is the relative locations of sensor nodes and $A = [\alpha_1, \alpha_2, \ldots, \alpha_p]^T$ is the coefficient matrix. For a new coming sample $\mathbf{v}_t$, its location in the low-dimensional embedding manifold can be simply obtained by

$$\mathbf{y}_t = (\alpha_1^T V_q^{(t)}, \alpha_2^T V_q^{(t)}, \ldots, \alpha_p^T V_q^{(t)}), \tag{13}$$

where $V_q^{(t)}$ can be calculated by (1) and (4). In addition, based on the property of manifold learning, the more the datum nodes, the closer the obtained high-dimensional data manifold to the real data space of the WSNs in general. And from (11), we can know that it should be $M \le N$ for the Rayleigh-Ritz theorem problem. Thus, for the given q and N, the optimal datum nodes m_{op} is given by

$$m_{op} = \max\{m | \sum_{i=1}^{q} C_{m+i-1}^{m-1} \le N, m \in Z^+\}. \tag{14}$$

For example, when $q = 2$, there is $\frac{m(m+3)}{2} \le N$, then the optimal datum nodes $m_{op} = [-\frac{3}{2} + \sqrt{2N + \frac{9}{4}}]$.

For getting the physical locations of unknown nodes, based on the coordinate transformation theorem, we can assume that there is an affine coordinate transformation for each node as

$$\mathbf{p}_i = B\mathbf{y}_i + \mathbf{c}, i = 1, \ldots, N, \tag{15}$$

where $B = [b_{jk}]_{p\times p}$ and $\mathbf{c} = (c_1, \ldots, c_p)^T$. Given sufficient number of anchors, we can transform the relative locations to physical locations. The goal is to minimize the sum of squares of the errors between the true locations of the anchors and their transformed locations and the bias $\mathbf{c}$ in (15) can be eliminated by considering the relative distance between sensors.

Making use of the anchors $\{\mathbf{p}_i\}_{i=1}^n$ and the corresponding relative locations $\{\mathbf{y}_i\}_{i=1}^n$, for a given anchor node $\mathbf{p}_u$, we set

$$\Delta Y_u = [\Delta\mathbf{y}_1, \Delta\mathbf{y}_2, \ldots, \Delta\mathbf{y}_{u-1}, \Delta\mathbf{y}_{u+1}, \ldots, \Delta\mathbf{y}_n], \tag{16}$$

$$\Delta P_u = [\Delta\mathbf{p}_1, \Delta\mathbf{p}_2, \ldots, \Delta\mathbf{p}_{u-1}, \Delta\mathbf{p}_{u+1}, \ldots, \Delta\mathbf{p}_n], \tag{17}$$

where $\Delta\mathbf{y}_i = \mathbf{y}_i - \mathbf{y}_u$ and $\Delta\mathbf{p}_i = \mathbf{p}_i - \mathbf{p}_u$. And then, we can formulate the alignment problem as a least-squares optimization problem

$$\min_{B,\mathbf{c}} \sum_{i=1, i\neq u}^{n} \|B\Delta\mathbf{y}_i - \Delta\mathbf{p}_i\|_2^2. \tag{18}$$

Through translating into a set of linear equations, it is easy to get the analytical solution of above least-squares problem as

$$B = \Delta P_u \Delta Y_u^T (\Delta Y_u \Delta Y_u^T)^{-1}. \tag{19}$$

Finally, the estimation physical locations $\tilde{\mathbf{p}}_i$ of unknown nodes can be obtained by the following formula as

$$\tilde{\mathbf{p}}_i = B(\mathbf{y}_i - \mathbf{y}_u) + \mathbf{p}_u, i = n+1, \ldots, N. \tag{20}$$

For a new coming sensor node $\mathbf{p}_t$, the estimation physical location can be also simply calculated by

$$\tilde{\mathbf{p}}_t = B(\mathbf{y}_t - \mathbf{y}_u) + \mathbf{p}_u, \tag{21}$$

where $\mathbf{y}_t$ obtained by (13) is the low dimensional representation of $\mathbf{p}_t$ in the embedding manifold.

Remarks:

- This is a typical sensor network self localization problem for a large number of unknown nodes using some reference anchors and localization data via a nonlinear manifold learning method.
- The datum nodes m and polynomial degree q are two important polynomial parameters which should be decided in advance through simulating with empirical data. In general, we select the anchors to be the datum nodes. However, the actual number of anchors can be more than the datum nodes m or not. Obviously, the more the actual anchors are, the better the topological structure of the sensor networks is. Consequently, we can also select the unknown nodes to be the datum nodes.
- There are many different measurement pair-wise distance models which can be selected to take part in the simulation. And because of without the additivity, the proposed algorithm cannot be used to deal with the signal strength measurement for short communication range in general. However, we can select an appropriate datum nodes set or eliminate the no communicating nodes for constructing the neighboring graph to deal with received signal strength in the real world application.

- Many applications only require the relative locations, so that the coordinate transformation is optional.

A summary of our algorithm is provided as follows:

Node localization algorithm based on GEPM

Input: $\{\mathbf{p}_i\}_{i=1}^{n}$: the location date set of the anchors; x_{ij}: measurement pair-wise distance.

Output: The estimation physical locations $\{\bar{\mathbf{p}}_j\}_{j=n+1}^{N}$ of the unknown nodes.

1: Construct the neighbor graph G. Compute the polynomial kernel matrix V_q, the basic weight matrix W with the localization data and the graph Laplacian matrix U with $U = D - W$.
2: Obtain the high-dimensional sample data $\{\mathbf{v}_i\}_{i=1}^{N}$ of the sensor nodes and the new coming localization data $\mathbf{v}_t$ based on measured pair-wise distance.
3: Calculate the relative locations using GEPM to estimate the relationship between the localization data space and relative location space. Solving the eigenvector problem in Eq. (11), the p orthogonal eigenvectors $\alpha_1, \ldots, \alpha_p$ corresponding to the smallest p non-zero eigenvalues can be obtained by Rayleigh–Ritz theorem. Then the p-dimensional relative coordinate matrix Y can be obtained by Eq. (12).
4: Given sufficient anchor nodes ($n \geq 3$ for 2D or $n \geq 4$ for 3D), the estimation physical locations $\{\bar{\mathbf{p}}_i\}_{i=n+1}^{N}$ can be transformed by the relative coordinates $\{\mathbf{y}_i\}_{i=n+1}^{N}$. And for a new coming node $\mathbf{p}_t$, we can simply get its physical coordinate by Eq. (21).

4. Simulation results

To evaluate the performance of the proposed algorithm GEPM, we simulate the localization algorithms with Matlab[1] and suppose 200 sensor nodes including n anchors are randomly placed in 2D environment, which is formed by 10 × 10 unit's square region. The measurement distance x_{ij} between $\mathbf{p}_i$ and $\mathbf{p}_j$ can be represented as $x_{ij} = \tilde{x}_{ij}(1 + N(0, \sigma^2))$, where $\tilde{x}_{ij}$ is the real distance and σ is the standard deviation of noise [15]. In the experiments, the heat kernel parameter is simply set as $w_g = 0.05$ and the detailed analysis of the polynomial parameters selection is given in subsection A. The mean location estimation error, i.e., root-mean-square(RMS) error calculated by $Error = \frac{1}{N-n}\sum_{i=n+1}^{N} \|\bar{\mathbf{p}}_i - \mathbf{p}_i\|_2$, is used to measure the performance of the presented algorithm. To get the optimal estimations and make full use of the localization data, the degree of polynomial mapping is usual set with $q = 2$. Otherwise, there is much more complicated calculation and time costing. Thus, the optimal datum nodes m_{op} is set as 18 with $N = 200$. It should be noted that the measurement data is always affected by kinds of noise. For example, some objects around the sensor nodes, including cars, trees and buildings, may block the sensor nodes, which in turn makes the measurements inaccurate. Consequently, we assume that the standard deviation $\sigma \geq 0.05$ and all of the reported simulation results are the average over 30 trials.

4.1. The polynomial parameters selection

In this subsection, the influence of the datum nodes m to the localization accuracy is analyzed in detail based on the RMS error under different anchors n, communication range d and noise σ. As shown in Fig. 2 with $d = 5, \sigma = 0.2$ and $n = 10, 30, 50, 70$,

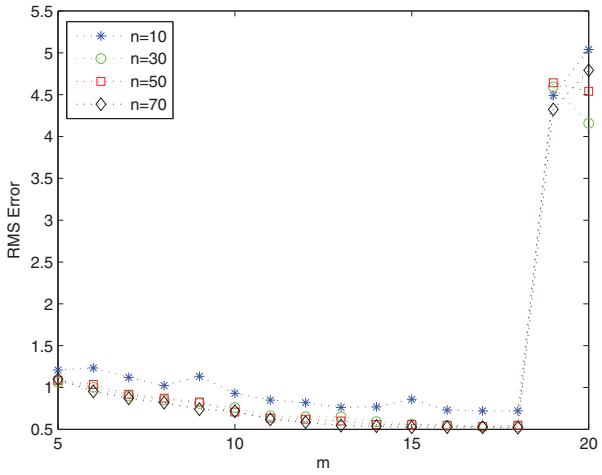

Fig. 2. The performance of datum nodes m under different anchors n.

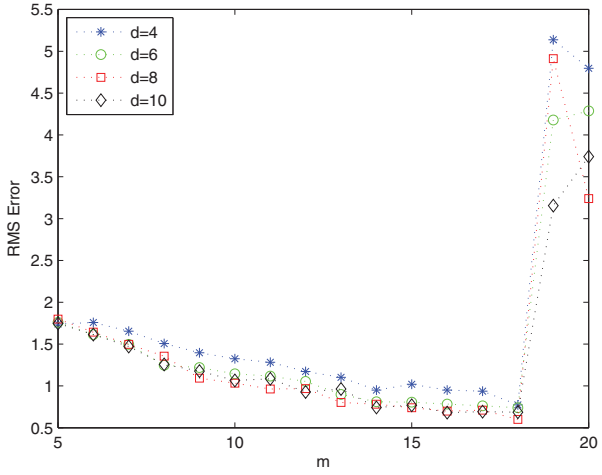

Fig. 3. The performance of datum nodes m under different communication range d.

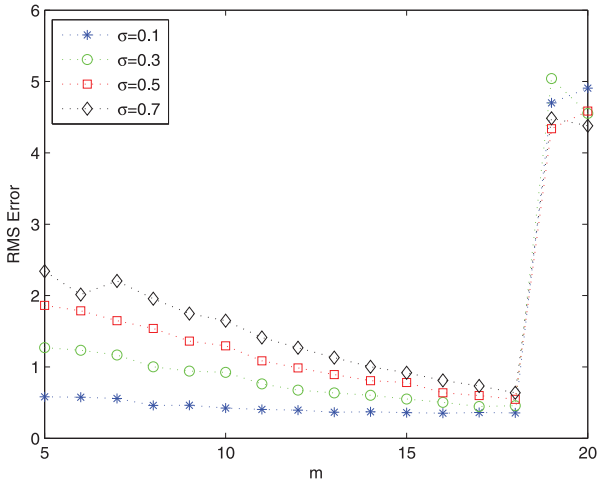

Fig. 4. The performance of datum nodes m under different noise σ.

Fig. 3 with $n = 20, \sigma = 0.4$ and $d = 4, 6, 8, 10$, and Fig. 4 with $n = 40, d = 7$ and $\sigma = 0.1, 0.3, 0.5, 0.7$, respectively, we can know that the RMS error is decreasing slowly with the value of m and reaches the minimal value at $m = 18$. And when the fundamental nodes number $m \geq 19$, the RMS error would reach a big value rapidly. It can be seen that a good localization accuracy can be obtained when $m \leq 18$, especially at $m = 18$, the RMS error can reach the minimal value. Even for $m = 5$, the RMS error is also not

[1] http://cn.mathworks.com/matlabcentral/fileexchange/57394-sensorsourcecode/content/SensorSourceCode.m

Table 1
Parameter selection with different interesting region.

Side length	10	20	30	40	50	60
Com. range	6	12	18	24	30	36
Datum nodes	18	18	18	18	18	18
RMS error	0.45	1.56	2.82	4.05	5.38	6.56

Table 2
Parameter selection with different sensor number.

Sensor nodes	50	100	150	200	250	300
Datum nodes	8	12	15	18	20	23
RMS error	0.65	0.56	0.52	0.45	0.38	0.36
Sensor nodes	350	400	450	500	550	600
Datum nodes	25	26	28	30	31	32
RMS error	0.33	0.32	0.33	0.32	0.32	0.33

vary large. Meanwhile, for the given $N = 200, n = 20$ and $\sigma = 0.1$, we present the localization results under different localization region with different side length and an appropriate communication range(Com. Range) in Table 1. We can see that the optimal datum nodes m is also at $m = 18$. Finally, we can know that the optimal parameter of datum nodes is not effected by anchors n, communication range d and noise standard deviation σ for different interesting region in general.

Furthermore, through sequential experiment with $n = 20, \sigma = 0.1, d = 6$ and $r \in [5,50]$, we can obtain the optimal datum nodes number and the corresponding RMS error as shown in Table 2 under different sensor nodes in the interesting area 10×10, where $N = 50, 100, \ldots, 600$. From Table 2, we can known that the optimal datum nodes m_{op} is consistent with $m_{op} = [-\frac{3}{2} + \sqrt{2N + \frac{9}{4}}]$ and the RMS error is decreasing with the density of the sensor nodes in the localization region. At the same time, the presented algorithm also has a great ability to preserve the sparsity structure of the localization data. Even when $N = 50, m = 8$, the RMS error is only about 0.65 at $d = 6$ and $\sigma = 0.1$.

From the above analysis, we can also know that the selection of datum nodes m is only relative to the sensor nodes N with a given polynomial degree q. In theory, the topological structure of the sensor network is represented by a network graph G with N sensor nodes under a given communication range d, where each vertex represents a sensor node. Because the communication range is not too small and the noise is also not too large comparing the interesting region, otherwise there would be no sense, with the application of geodesic distance, the dimension m of the high-dimensional data manifold is mainly determined by the density of the sensor network, i.e., the number of sensor nodes N, especially for a big change with N.

4.2. The localization result analysis

In this subsection, we present the localization results and make a detailed analysis under different anchors n, communication range d and noise standard deviation σ. There are 200 sensor nodes which are placed randomly in the interesting area mentioned previously and the number of datum nodes is always set as $m = 18$ in the next simulations. The location results of each sensor node are shown in Figs. 5 and 6. The squares are anchors and the circles denote the unknown nodes. Each line connects a true node location and its estimation whose length denotes the localization error. We set $n = 40, d = 4, \sigma = 0.1$, and depict the estimated location of each sensor node in Fig. 5. The RMS error is about 0.6235. We take $n = 20, d = 6, \sigma = 0.4$, and plot the final estimated location in Fig. 6. The final RMS error is about 0.5742.

Firstly, we present the effects of the anchors n with different d and σ on the localization performance as shown in Figs. 7 and 8, corresponding to $d = 6, \sigma = 0.1, 0.3, 0.5$ and $\sigma = 0.1, d = 4, 6, 8$,

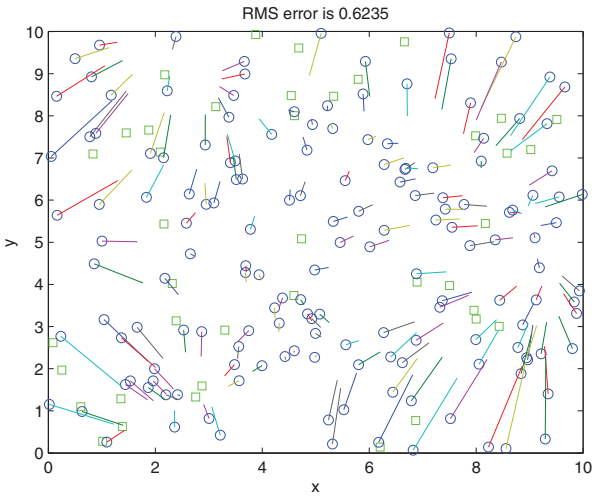

Fig. 5. Localization result with $n = 40, d = 4$ and $\sigma = 0.1$.

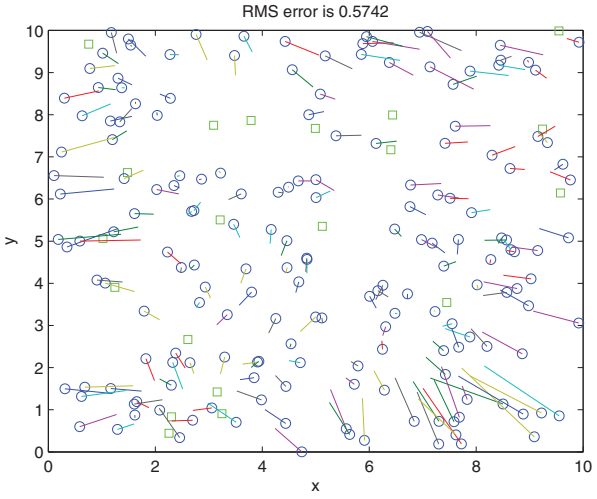

Fig. 6. Localization result with $n = 20, d = 6$ and $\sigma = 0.4$.

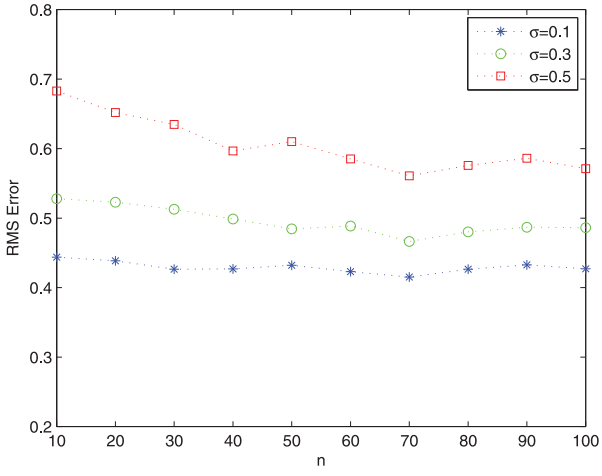

Fig. 7. Root-mean-square error on location with different noise σ and anchors n.

respectively. We can see that the performance is not significantly improved with increasing of n. This demonstrates that the presented algorithm can achieve a good performance with a small number of anchors. Even when $n = 4$, the RMS error is only about 0.56 with $d = 6$ and $\sigma = 0.1$. Consequently, in practical application, if the anchors are very large, we can select appropriate anchors to

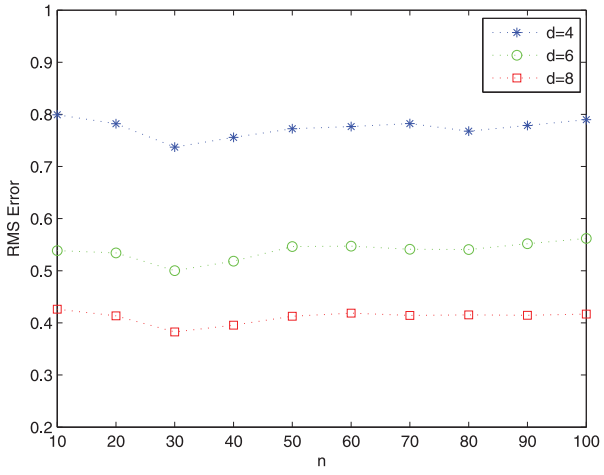

Fig. 8. Root-mean-square error on location with different communication range d and anchors n.

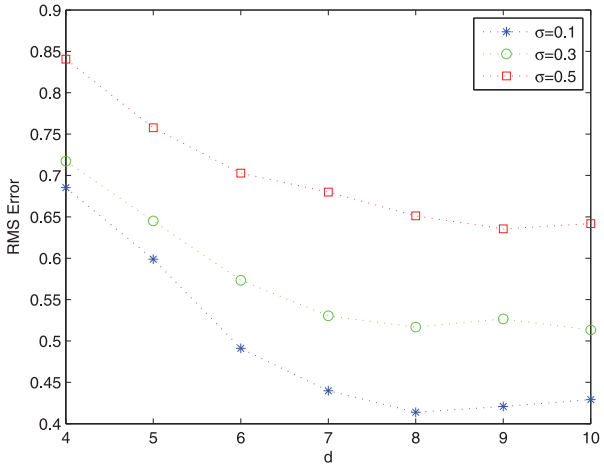

Fig. 9. Root-mean-square error on location with different noise σ and communication range d.

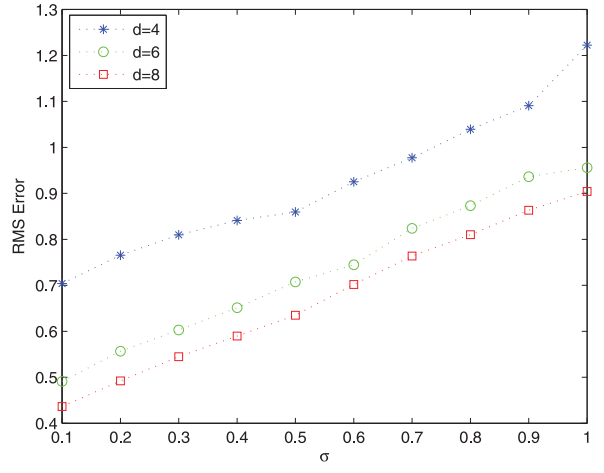

Fig. 10. Root-mean-square error on location with different communication range d and noise σ.

construct the neighbor graph for reducing the computational complexity.

Next, we set anchors $n = 20$ in the later simulations and present a quantitative analysis of the effects of d and σ in Figs. 9 and 10. We set $\sigma = 0.1, 0.3, 0.5$, respectively, and plot the RMS error under different values of d in Fig. 9. Meanwhile, we set $d = 4, 6, 8$, respectively, and plot the RMS error under different values of σ in Fig. 10. For a fixed σ, the RMS error is decreasing with the value of d, and when the communication range $d > 7$, the localization accuracy does not increase obviously but decrease slowly on same level. For a given value of d, it can be seen that the RMS error is increasing with the value of σ.

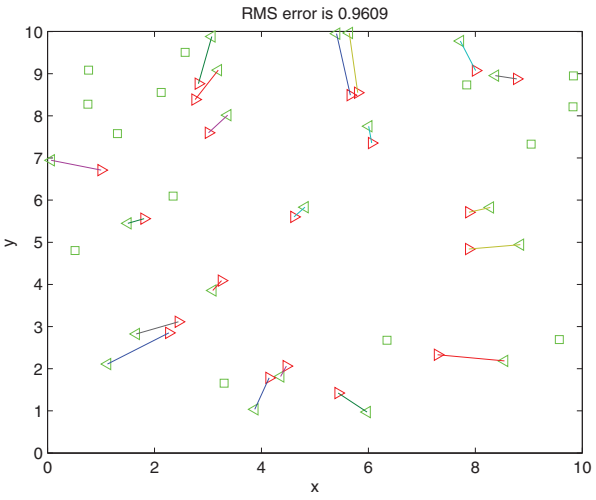

Fig. 11. Localization results with new coming data.

Table 3
The optimal localization results with different sensor number for new coming.

Sensor nodes	50	100	200	300	400	500
RMS error	1.25	1.09	0.96	0.96	0.91	0.86

4.3. The localization for new coming nodes

In this subsection, we assume that 20 new coming unknown nodes are putted into the location region as described in Section 4.1. Then, we can get the geodesic distance between the new coming nodes and the datum nodes by geodesic graph. Under the same condition as Section 4.1, the optimal estimation error with different sensor nodes is shown in Table 3 which indicates that the RMS error is increasing with the sensor nodes due to preserving the topological structure. Meanwhile, the proposed algorithm also has a good preservability, and even for $N = 50$, the RMS error is only about 1.25 with $n = 6, d = 6$ and $\sigma = 0.1$. In addition, we set $N = 200, n = 20, d = 6, \sigma = 0.1$, and plot the localization result in Fig. 11. It can be seen that the new coming nodes at the center of the region have a higher positioning accuracy, especially for the low noise and high density of the localization region.

4.4. Comparison with other kernel learning localization methods

In the simulations, we compare our algorithm with two related rang-based localization algorithms: (1) Isometric mapping(ISOMAP) based on centralized algorithm proposed in [14]; (2) A graph embedding method for wireless sensor network localization based on kernel locality preserving projection(KLPP) [13].

4.4.1. The comparison of computational complexity

For the measurement pair-wise distance, the complexity of computing G and W are $O(N^3)$ and $O(N^2)$, respectively. The complexity of calculating V_2 is $O(\frac{Nm(m+1)}{2})$. The computational complexity of the final generalized eigenvector problem is $O((N+p)[\frac{m(m+3)}{2}]^2)$. The complexity of the coordinate transformation is approximate to $O(m^2)$. Therefore, the total computational complexity of the proposed algorithm is $O(N^3 + N^2 + \frac{Nm(m+1)}{2} + (N+p)[\frac{m(m+3)}{2}]^2 + m^2)$. While the time complexity of the other two localization methods, i.e., ISOMAP and KLPP, are $O(2N^3 + m^2)$ and

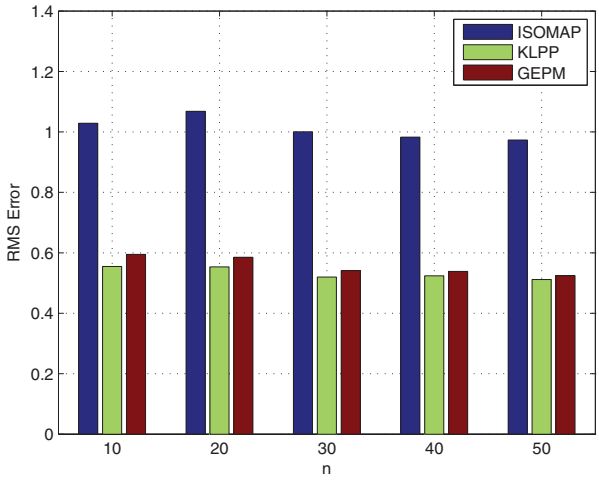

Fig. 12. Performance of diffenrent algorithms with different anchors n.

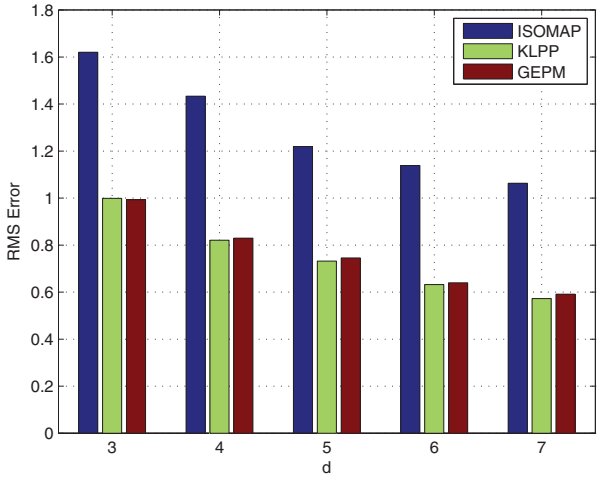

Fig. 13. Performance of diffenrent algorithms with different communication range d.

$O(2N^3 + (p+2)N^2 + m^2)$, respectively [13,14]. Finally, we can easily get that the three algorithms have the similar computational complexity for the localization of sensor nodes in WSNs.

4.4.2. The comparison of localization accuracy

We set $d = 6, \sigma = 0.4$ and plot their RMS error under different anchors n as shown in Fig. 12. We also set $n = 20, \sigma = 0.4$ and plot their RMS error under different d in Fig. 13. While we set $d = 6, n = 20$ and report their RMS error under different σ in Fig. 14. From the simulation results, we can know that GEPM can achieve the same accuracy as KLPP and is better than ISOMAP. Though KLPP can also reach a high localization accuracy, it is more complicated to deal with the new coming nodes than GEPM due to calculating the total geodesic distance.

Based on the all above simulation results for original data and new coming data in WSNs, it can be seen that the presented algorithm can achieve small average location error due to its great ability of preserving the sparsity structure of the localization data and the great ability of dealing with the measurement noise. Especially, it is also effect to the new coming sensor nodes without reconstructing the topological network and reduces a lot of complicated computations. Meanwhile, it is sensitive with the overlarge datum nodes. Therefore, selecting a proper polynomial parameter is very important for different localization environment.

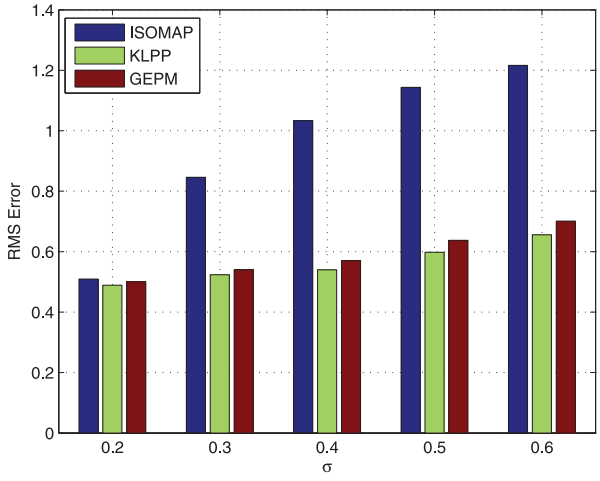

Fig. 14. Performance of diffenrent algorithms with different noise σ.

5. Conclusion

We have presented a graph embedding method for the location estimation problem in WSNs based on measured pair-wise distance. We view the sensors in the network as independently distributed devices and choose a suitable heat kernel function to measure the similarity between each pair of sensor nodes. Then we propose a novel localization algorithm under the manifold learning processes with an explicit nonlinear mapping based on the assumption that there exists a polynomial mapping between the localization data space and their low dimensional representations. For computing the geodesic distance graph, a suitable datum nodes m and polynomial degree q are selected in a right way. And then, the polynomial mapping method is presented and the relative locations are estimated by GEPM method. Finally, the physical coordinates of the sensor nodes can be obtained by the affine coordinate transformation. Experimental results confirm the promising performance of the proposed GEPM. Our future work will focus on the explicit nonlinear mapping of manifold learning localization methods for the target detection and tracking in the wireless sensor networks.

Acknowledgement

The authors wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper.

References

[1] F. Franceschini, A review of localization algorithms for distributed wireless sensor networks in manufacturing, Int. J. Comput. Integrated Manuf. 1 (1) (2007) 1–19.
[2] P. Valdastri, S. Rossi, A. Menciassi, V. Lionetti, F. Bernini, F. Recchia, P. Dario, An implantable ZigBee ready telemetric platform for in vivo monitoring of physiological parameters, Sensors Act. 142 (1) (2008) 369–378.
[3] A. Steed, R. Milton, Using tracked mobile sensors to make maps of environmental effects, Personal Ubiquitous Comput. 12 (2008) 331–342.
[4] A. Kangas, T. Wigren, Angle of arrival localization in LTE using MIMO pre-coder index feedback, IEEE Commun. Lett. 17 (8) (2013) 1584–1587.
[5] X. Li, K. Pahlavan, Super-resolution TOA estimation with diverity for indoor geolocation, IEEE Trans. Wireless Commun. 3 (1) (2004) 224–234.
[6] L. Taponecco, A.A. D'Amico, U. Mengali, Joint TOA and AOA estimation for UWB localization applications, IEEE Trans. Wireless Commun. 10 (7) (2011) 2207–2217.
[7] G. Mao, B. Fidan, B.D.O. Anderson, Wireless sensor network localization techniques, Comput. Netw. 51 (10) (2007) 2529–2553.
[8] C. Meesookho, U. Mitra, S. Narayanan, On energy-based acoustic source localization for sensor network, IEEE Trans. Sig. Process. 56 (1) (2008) 365–377.

[9] H.C. So, L. Lin, Linear least squares approach for accurate received signal strength based source localization, IEEE Trans. Sig. Process. 59 (8) (2011) 4035–4040.
[10] Y. Chan, W. Tsui, H. So, P. Ching, Time-of-arrival based localization under NLOS conditions, IEEE Trans. Veh. Technol. 55 (1) (2006) 12–24.
[11] L. Cong, W. Zhuang, Hybrid TDOA/TOA mobile user location for wideband CDMA cellular systems, IEEE Trans. Wireless Commun. 1 (3) (2002) 439–447.
[12] N. Patwri, A.O. Hero, M. Perkins, N.S. Correal, R.J. O'Dea, Relative location estimation in wireless sensor networks, IEEE Trans. Sig. Process. 51 (8) (2003) 2137–2148.
[13] C. Wang, J. Chen, Y. Sun, X. Shen, A graph embedding method for wireless sensor networks localization, in: IEEE Global Telecommunications Conference (GLOBECOM 2009) Honolulu, Hawaii, USA, 2009a.
[14] C. Wang, J. Chen, Y. Sun, X. Shen, Wireless sensor networks localization with Isomap, in: IEEE International Conference on Communications (ICC 2009), Dresden, Germany, 2009b.
[15] J. Chen, C. Wang, Y. Sun, X. Shen, Semi-supervised Laplacian regularized least squares algorithm for localization in wireless sensor networks, Comput. Netw. 55 (10) (2011) 2481–2491.
[16] B. Yang, J. Xu, J. Yang, M. Li, Localization algorithm in wireless sensor networks based on semi-supervised manifold learning and its application, Cluster Comput. 13 (4) (2010) 435–446.
[17] X. Zeng, S. Tang, S. Li, Ensemble-Based manifold learning methods for localization in wireless sensor networks, in: 2012 Fourth International Conference on Computational and Information Sciences, 2012, pp. 939–942.
[18] Y. Liu, J. Chen, Y. Zhan, Local patches alignment embedding based localization for wireless sensor networks, Wireless Personal Commun. 70 (1) (2013) 373–389.
[19] T. Chin, D. Suter, Out-of-sample extrapolation of learned manifolds, IEEE Trans. Pattern Anal. Mach. Intel. 30 (9) (2008) 1547–1556.
[20] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labelled and unlabelled examples, J. Mach. Learn. Res. 7 (2006) 2399–2434.
[21] Y. Bengio, J.F. Paiement, P. Vincent, O. Delalleau, N.L. Roux, M. Ouimet, Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps and spectral clustering, in Proc. Adv. Neural Inf. Process. Syst. 16 (2003) 177–184.
[22] K. Stone, T. Camp, A survey of distance-based wireless sensor network localization techniques, Int. J. Pervas. Comput. Commun. 8 (2) (2012) 158–183.
[23] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intel. 29 (1) (2007) 40–51.
[24] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrices, Wiley, New York, 1999.

Hao Xu received the B.S. in Neijiang Normal University, Sichuan, China, in 2010, and the M.S. degrees in Hubei University, Wuhan, China, in 2013, respectively. He is currently pursuing the Ph.D. degree from the Beijing Institute of Technology. His research interests lie in the areas of differential geometry, information geometry, statistical signal processing and wireless sensor networks.

Huafei Sun is a full professor of Beijing institute of Technology from 2002. His research interest is information geometry and its applications. He obtained PH.D in 1999 at Tokyo Metropolitan University.

Yongqiang Cheng received the B.S. and M.S. degrees in information and communication engineering from National University of Defense Technology, Changsha, China, in 2005 and 2007, respectively. He is currently pursuing the Ph.D. degree from the National University of Defense Technology. From September 2009 to November 2010, he was a visiting research student with Melbourne Systems Laboratory, University of Melbourne, Australia. His research interests lie in the areas of information geometry and statistical signal processing.

Hao Liu received the Ph.D. degree in applied mathematics from the Beijing Institute of Technology, Beijing, China, 2015. Since 2003, he has been with University of Science and Technology Liaoning, Anshan, China, where he is currently an Assistant Professor with the School of Science. His current research interests include computational intelligence, geometric control and wireless sensor networks.