WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2010; **10**:1045–1054
Published online 23 June 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.820

Sensor network localization using kernel spectral regression

Chengqun Wang, Jiming Chen*,† and Youxian Sun

State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

Summary

This paper addresses the localization problem in wireless sensor networks using signal strength. We use a kernel function to measure the similarities between sensor nodes. The kernel matrix can be naturally defined in terms of the signal strength matrix. We show that the relative locations of sensor nodes can be obtained by solving a dimension reduction problem. To capture the structure of the whole network, we use the kernel spectral regression (KSR) method to estimate the relative locations of the sensor nodes. Given sufficient anchor nodes, the relative locations can be aligned to global locations. The key benefits of adopting KSR are that it allows us to define a graph to optimally preserve the topological structure of the sensor network, and a kernel function can capture the nonlinear relationship in the signal space. Simulation results show that we can achieve small average location error with a small number of anchors. We also compare our method with several related methods, and the results show that KSR is more efficient than the others in our simulated sensor networks. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: graph embedding; kernel spectral regression; localization; sensor network

1. Introduction

Wireless sensor networks (WSNs) are usually implemented with a number of wireless sensor nodes that communicate over multiple hops with one or more base stations. It allows for inexpensive, and high quality monitoring of large geographical areas, e.g., wildlife habitat monitoring [1], manufacturing [2], remote health-care monitoring [3]. Many of these WSNs applications either require or benefit from the localization service which provides the location of each sensor node. For example, given a network of sensors deployed across a large area for environmental monitoring, such as soil moisture monitoring or water quality surveillance, the main task of the nodes is to send the measured data to a sink at a given location. Without knowing the locations from where the data are obtained, the sink node would deem the data meaningless.

The simplest way of providing locations is to equip every sensor node with a GPS receiver. However, GPS receivers are relatively expensive, power-consuming, and cannot work indoors, while wireless sensor nodes should be inexpensive and low powered [4]. It is inadvisable and unnecessary to equip all of the nodes with GPS receivers. In most cases, there are only a small fraction of all the sensor nodes, whose locations are obtained by GPS-like tools, or by deploying them at points with known locations. These sensor nodes are called *anchors*. Most of the sensor nodes are

*Correspondence to: Jiming Chen, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China.
†E-mail: jmchen@ieee.org

randomly placed in the region of interest (ROI) and do not know their absolute locations. The sensor nodes with unknown locations are called *non-anchor* nodes and their locations need to be estimated using the localization method.

There have been significant research activities for the localization problem, and various localization methods have been proposed in the literature. In general, these methods can be categorized into *range-based* and *range-free* methods. The range-based methods are defined by protocols that use absolute pairwise distance estimation (ranging) or angle estimation for calculating locations [5]. For example, RADAR [6] is a typical range-based localization system that measures the signal strength at all locations in the ROI, and records the signal strength in a database during the calibration phase. In the localization phase, the location of the object is estimated by finding the best match of the measured signal strengths in the database. The *range-free* methods make no assumptions about the availability of the distance or angle estimation. For example, DV-Hop method proposed in Reference [7] uses offline hop-distance estimation to improve location estimation via exchanging neighboring data. The approximate point in triangle (APIT) [8] uses a small fraction of anchors to divide the ROI into overlapping triangular regions, and then estimates the location of a sensor node via the aggregation of the two different triangular regions. Finally, the location of a sensor node can be estimated by the center of the intersection of the triangulars.

In recent years, kernel-based range-free methods have attracted considerable research interests for localizing sensor nodes. These kinds of methods employ an appropriate kernel function to measure the similarities between sensor nodes, and then formulate the localization problem as classification [9], regression [10], or manifold learning problem [11,12]. For example, Nguyen *et al*. [13] divide the ROI into overlapping subregions, and then use Support Vector Machine (SVM) to select a set of subregions where the sensor node is likely to be located. The estimated location can be calculated as the center of the selected subregions. Kuh *et al*. [14] use a kernel matrix to store the signal strength information, and then employ a least squares regression method to estimate the non-anchors' locations. These two kernel-based localization methods work well in densely distributed sensor networks, especially when the number of anchors is large. Honeine *et al*. [15] seek a mapping function between the inner products of the anchors' locations and their corresponding kernel matrix, and then estimate the inner products between the locations of anchors and the locations of non-anchors using a kernel matrix regression method. The locations of non-anchors can be retrieved from these estimated inner products. The main advantage of this method is that the kernel selection procedure can be avoided.

In this paper, we consider the WSN localization problem using signal strength. We aim at accurately estimating the locations of non-anchors with a small quantity of anchors. We bypass the ranging procedure, and formulate the localization problem as a dimension reduction problem. The kernel spectral regression (KSR) is then chosen to estimate the relative locations of sensor nodes. As we know, the signal strength attenuation s satisfies the following formulation in the ideal space:

$$s \propto Pd^{-\nu} \tag{1}$$

where d is the distance between the transmitter and receiver, ν is a constant, and P is the transmitting power [16]. This provides an efficient way to estimate the pairwise distance. Unfortunately, due to multi-path effects and noise interferences in real environments, the received signal strength is highly inaccurate and usually attenuates in a way that is highly nonlinear and uncertain, especially in a complex environment [17]. To accurately estimate the location of the non-anchors based solely on the inaccurate pairwise signal strengths, we consider each sensor node as an independent device, and adopt an appropriate kernel function to measure the similarities between these nodes. The kernel matrix can be naturally defined based on the pairwise signal strengths. The main advantage of choosing kernel functions is that they can better capture the nonlinear relationship in the signal strength space. Consequently, we formulate the relative location estimation problem as a dimension reduction problem. To alleviate the influence of the number of anchors, we construct a graph to represent the topological structure of the network, and use KSR to estimate the relative locations of non-anchors. When there are sufficient anchor nodes (3 or more for the two-dimension (2D) case), we can construct a coordinate transformation system to align the relative locations to global locations. The transformation matrix can be estimated by solving a set of linear equations. Compared with the related kernel-based localization methods [9,14,15], KSR is a kernel graph embedding method and allows us to construct a graph to optimally preserve the topological structure of the WSNs. This allows us to accurately estimate the locations of non-anchors using a small amount of anchors. Simulation results also show that this

DOI: 10.1002/wcm

method is not significantly sensitive to the number of anchors.

The remainder of the paper is organized as follows. Section 2 introduces the sensor network model. Section 3 gives a detailed description of WSNs localization using KSR. Section 4 illustrates the simulation results. Finally, Section 5 concludes the paper.

2. Sensor Network Model

Consider a WSN of N sensor nodes deployed in a 2D ROI $\mathcal{C} \subseteq \mathbb{R}^2$[‡]. An integer from 1 to N denotes each node as its ID. The set of all nodes in the network is denoted by $\mathcal{N} = (1, \cdots, N)$. The location of node $i \in \mathcal{N}$ is denoted by $\mathbf{x}_i$. Without loss of generality, we suppose that the first M ($M \ll N$) nodes are anchors and that their locations are known. The task of the localization problem is to estimate the locations of the other $N - M$ non-anchor nodes.

We assume that each node is capable of transmitting signal strength to each of its neighbors, up to a determined communication range. As analyzed in Reference [16], the signal strength fades along the transmitting direction and is often sensitive to noise. In this paper, we choose the following signal model (Equation (2)) to simulate the received signal strength between sensor node i and j when node i is among the communication range of node j or node j is deployed in the communication range of node i,

$$s(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-w_s \|\mathbf{x}_i - \mathbf{x}_j\|^2 + N(0, \sigma^2)\right) \quad (2)$$

where w_s is a pre-defined parameter, which implicitly determines the fading ratio of the signal strength, and $N(0, \sigma^2)$ denotes an independent normal random variable with standard deviation σ, which corresponds to the level of noise of the WSNs environment. We can see that the value of signal-to-noise ratio (SNR) is relatively low when the value of σ is large. This signal model is widely used in previous works [13,14], and is used merely to generate the signal strength between pair-wise sensor nodes in this paper. There is one issue that must be mentioned. When the sensor nodes are deployed in a large region, most of the simulated value of the signal strength is very close to zero, even for the neighborhood nodes. To overcome this problem, we suppose the location $\mathbf{x}_i$ is scaled in [0, 1].

[‡]Extension to the three-dimensional (or even higher dimensional) environment is straight forward.

3. Localization with Kernel Spectral Regression (KSR)

KSR is a graph embedding technique, which was first proposed for dimension reduction problem. It casts the learning of an embedding function problem into a regression framework, which avoids the generalized eigen-decomposition of dense matrices [18,19]. In this section, we will adopt KSR to solve the localization problem in WSNs with the goal of achieving relatively high localization accuracy using a small amount of anchors.

3.1. Kernel Spectral Regression (KSR)

Consider the sensor network with N sensor nodes, as described in Section 2. We use s_{ij} to denote the abbreviation of $s(\mathbf{x}_i, \mathbf{x}_j)$. We set $s_{ii} = 1$ for all $i = 1, \cdots, N$. When the sensor node i is out of the communication range of sensor node j or j is out of the communication range of node i, we simply set $s_{ij} = 0$.

We first construct the adjacent graph, which represents the topological structure of the whole sensor network. Let G denote a graph with N vertices. The i-th vertex corresponds to the sensor node i. We put an edge between vertices i and j, if $s_{ij} \neq 0$ or $s_{ji} \neq 0$. Let W be the weight matrix of graph G. W is a sparse symmetric $N \times N$ matrix with W_{ij} denoting the weight of the edge between vertices i and j.

- If there is no edge between vertices i and j, we set $W_{ij} = 0$.
- Otherwise, we set

$$W_{ij} = \begin{cases} \max\{s_{ij}, s_{ji}\}, & \text{if } s_{ij} = 0 \text{ or } s_{ji} = 0 \\ \dfrac{s_{ij} + s_{ji}}{2}, & \text{otherwise} \end{cases}$$

The localization problem of finding the relative locations of the sensor nodes becomes the problem of constructing a 2D embedding graph. And the purpose is to optimally preserve the topological structure of the sensor networks by representing each node of the graph as a 2D vector, and to optimally preserving the similarities between sensors, where the similarity is measured by the edge weight. Suppose $\mathbf{v} = [v_1, \cdots, v_N]^T$ be a map from the graph G to a real line. The optimal $\mathbf{v}$ tries to minimize

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \left(v_i - v_j\right)^2 W_{ij} \quad (3)$$

DOI: 10.1002/wcm

subjecting to the constraint $\mathbf{v}^T D\mathbf{v} = 1$, where D is a diagonal degree matrix with $D_{ii} = \sum_{j=1}^{N} W_{ij}$ [20]. D_{ii} reflects the degree of sensor node i. With some algebraic formulations, the minimization problem reduces to

$$\mathbf{v}^* = \underset{\mathbf{v}^T D\mathbf{v}=1}{\operatorname{argmin}} \mathbf{v}^T L\mathbf{v} = \operatorname{argmin} \frac{\mathbf{v}^T L\mathbf{v}}{\mathbf{v}^T D\mathbf{v}} = \operatorname{argmin} \frac{\mathbf{v}^T W\mathbf{v}}{\mathbf{v}^T D\mathbf{v}} \tag{4}$$

where $L = D - W$ is the graph Laplacian matrix [20]. The optimal $\mathbf{v}^*$ is the maximum eigenvector of the following eigen-problem,

$$W\mathbf{v} = \lambda D\mathbf{v} \tag{5}$$

where λ is the eigenvalue. Let $\mathbf{v}_0$, $\mathbf{v}_1$ and $\mathbf{v}_2$ be the three largest generalized eigenvectors of above eigen-problem. It is easy to see that $\mathbf{v}_0$ is a vector of ones with eigenvalue $\lambda = 1$. So we use $\mathbf{v}_1$, $\mathbf{v}_2$ for the graph embedding. If we choose a linear function, i.e., $v_i^k = f(\mathbf{x}_i) = \alpha_k^T \mathbf{x}_i$, where v_i^k is the i-th element of $\mathbf{v}_k$, $k = 1, 2$, a popular way is to find α_k which can best fit the equation in the regularized least squares sense,

$$\alpha_k^* = \underset{\alpha}{\operatorname{argmin}} \left(\sum_{i=1}^{N} (\alpha^T \mathbf{x}_i - \mathbf{v}_i^k)^2 + \gamma \|\alpha\|^2 \right) \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm, and γ is a user-defined tradeoff parameter. The solution of above optimization problem can be reduced to solve a set of linear equations,

$$(XX^T + \gamma I)\alpha_k = X\mathbf{v}_k \tag{7}$$

where I is the $N \times N$ identity matrix, and $X = [\mathbf{x}_1, \cdots, \mathbf{x}_N]^T$ is the location matrix of the sensor nodes. Only the first M anchors' locations are known, so the matrix X is incomplete. Therefore, we choose a kernel function instead of the location matrix. Using the *Representor Theorem* [21] in reproducing Hilbert kernel space (RKHS), we have $v_i^k = f(\mathbf{x}_i) = \sum_{j=1}^{N} \alpha_j^k \kappa(\mathbf{x}_i, \mathbf{x}_j)$, α_j^k is the j-th element of α_k, $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ is the Mercer kernel of RKHS $\mathcal{H}_\kappa$. Equation (6) can be rewritten as

$$\alpha_k^* = \underset{f \in \mathcal{H}_\kappa}{\operatorname{argmin}} \left(\sum_{i=1}^{N} \left(f(\mathbf{x}_i) - \mathbf{v}_i^k \right)^2 + \gamma \|f\|_{\mathcal{H}_\kappa}^2 \right) \tag{8}$$

and Equation (7) becomes

$$(K + \gamma I)\boldsymbol{\alpha}_k = \mathbf{v}_k \quad k = 1, 2 \tag{9}$$

where K is an $N \times N$ Gram matrix with $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

Let $\mathbf{y}_i \in \mathbb{R}^2$ be the relative location of sensor node i, and $Y = [\mathbf{y}_1, \cdots, \mathbf{y}_N]^T$ be the relative location matrix of all sensor nodes. The relative location matrix Y can be estimated by the following equation,

$$Y = KA \tag{10}$$

where $A = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2]$.

3.2. Sensor Node Location Estimation

As mentioned in Section 2, our task is to estimate the locations of the $N - M$ non-anchors based on the signal strength matrix. We choose KSR to estimate the relative locations of sensors. The choice of the kernel function is highly dependent on the nonlinear and noisy characteristics of the localization problem. The signal strength matrix would be a positive semidefinite matrix in the ideal environment. The signal strength matrix itself can be a kernel matrix. More generally, to capture the nonlinear relationship in the signal space, the kernel function must be redefined. In particular, the radio signal model, described by Equation (1), has a form similar to the polynomial kernel function. The acoustic energy model has the form of the Gaussian kernel function [22]. As argued in Reference [23], Gaussian kernel has the ability of universally approximating any distribution, including some kind of noisy tasks. Hence, we choose Gaussian kernel function in this paper.

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-w_G \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|^2\right) = \exp\left(-\sum_{l=1}^{N} w_G \left(\mathbf{s}_{il} - \mathbf{s}_{jl}\right)^2\right) \tag{11}$$

where w_G is a user-defined band-width parameter, which reflects the smoothness of the nonlinear kernel map. There are methods for choosing the parameter w_G based on the locations of anchors [24].

After employing KSR to estimate the relative locations of all sensors, the resulting coordinates are determined in the space defined by the aforementioned eigenvectors. Thus we need to conduct a final step of aligning the relative location matrix Y into the initially

DOI: 10.1002/wcm

absolute location space using a linear transformation. Such step is commonly used in MDS-liked techniques [25].

For a given anchor i, we set $\Delta X_a = [\mathbf{x}_1 - \mathbf{x}_i, \cdots, \mathbf{x}_{i-1} - \mathbf{x}_i, \mathbf{x}_{i+1} - \mathbf{x}_i, \cdots, \mathbf{x}_M - \mathbf{x}_i]$ and $\Delta Y_a = [\mathbf{y}_1 - \mathbf{y}_i, \cdots, \mathbf{y}_{i-1} - \mathbf{y}_i, \mathbf{y}_{i+1} - \mathbf{y}_i, \cdots, \mathbf{y}_M - \mathbf{y}_i]$. Let $A = \Delta X_a \Delta Y_a^T (\Delta Y_a \Delta Y_a^T)^{-1}$. The absolute locations of non-anchors can be estimated by

$$\tilde{\mathbf{x}}_j = A(\mathbf{y}_j - \mathbf{y}_i) + \mathbf{x}_i, \quad j = M+1, \cdots, N \tag{12}$$

where $\tilde{\mathbf{x}}_j$ is the estimation of $\mathbf{x}_j$. These equations can be rewritten in matrix manner,

$$\tilde{X}_n = A(Y_n - \mathbf{y}_i \mathbf{e}_n^T) + \mathbf{x}_i \mathbf{e}_n^T \tag{13}$$

where $\tilde{X}_n = [\tilde{\mathbf{x}}_{M+1}, \cdots, \tilde{\mathbf{x}}_N]$, $Y_n = [\mathbf{y}_{M+1}, \cdots, \mathbf{y}_N]$, and $\mathbf{e}_n$ denotes the $(N - M)$-dimensional column vector of ones. A summary of the method is provided in Algorithm 1.

Algorithm 1 Sensor node location estimation method

Require: $\{s_{ij}\}_{i,j=1}^N$: signal strength information, $X_a = [\mathbf{x}_1, \cdots, \mathbf{x}_M]$: optional location matrix of the anchors.

Ensure: Y: relative location matrix, $\tilde{X}$: estimated location matrix of the non-anchors.

1: For $i = 1, \cdots, N$, set $s_{ii} = 1$.
2: Compute the Gram matrix K based upon $\{s_{ij}\}_{i,j=1}^N$, where $K_{i,j} = \exp(-\sum_{l=1}^N w_G(\mathbf{s}_{il} - \mathbf{vs}_{jl})^2)$.
3: Construct the adjacent graph G, define its weight matrix W, and calculate the diagonal degree matrix D, $D_{ii} = \sum_{j=1}^n W_{ij}$.
4: Find $\mathbf{v}_0$, $\mathbf{v}_1$, $\mathbf{v}_2$, the largest 3 eigenvectors of the eigen-problem (Equation (5)).
5: Solving the optimization problem (Equation (8)) for optimum vectors α_1 and α_2.
6: For $i = 1, \cdots, N$, $\mathbf{y}_i = K(i, :)[\alpha_1, \alpha_2]$.
7: if X_a is given, estimate the absolute locations of non-anchors matrix $\tilde{X}$ by Equation (13); otherwise, $\tilde{X} = \mathbf{0}$.

We provide the computational complexity of our location estimation method as below. In step 2, we compute the kernel matrix. Note that the signal strength matrix is sparse as the maximum communication range of sensor node i is fixed and small compared with the scale of the ROI in most real applications. We suppose that the average number of sensor nodes in the communication range is k, then the complexity of calculating K is $O(kN^2)$, while the weight matrix W and degree matrix D determined in step 3 is also calculated based upon $\{s_{ij}\}_{i,j=1}^N$. Their complexity is $O(N^2)$.

Steps 5 and 6 calculate the relative locations of all sensor nodes including anchors. We first solve the eigen problem in Equation (5). Note that the weight matrix W is sparse, so we use a Lanczos method to compute the three largest eigenvectors. Let k_1 be the number of iterations of the Lanczos method. The corresponding time complexity is $O(3k_1N(k + 8))$ [26]. As analyzed in Reference [18], the optimization problem in Equation (8) is a regularized kernel least squares problem which can be quickly solved by the iterative algorithm LSQR with $O(4k_2Nk + 10k_2N)$ operations [27], where k_2 is the number of iterations in LSQR. Finally, we calculate the relative locations of all sensors by Equation (10), which has time complexity $O(2N^2)$ in the 2D case.

The purpose of step 7 is to transform the relative locations into absolute locations with the aid of the locations of anchors. Many previous works have discussed this problem and the linear transformation can be efficiently solved by approximately $O(M^2)$ operations [28]. Note that many WSNs applications only require the relative locations, and so step 7 is optional in some real WSNs applications.

Finally, the performance of the localization method can be measured with the average location estimation error, which is a widely used performance metric in existing works,

$$error = \frac{\sum_{i=M+1}^N \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2}{N - M} \tag{14}$$

where $\|\cdot\|_2$ denotes the 2-norm. A small error indicates good performance of the evaluated method.

4. Simulation Results

In this section, several simulations are conducted to test the performances of our localization method. We consider that 400 sensor nodes are uniformly or randomly deployed in a 50×50 unit 2D ROI. There are $M = l^2$ anchors which are initially deployed at grid point of the $l \times l$ square, as shown in Figure 1. We suppose that the average communication range of each sensor node is 20 units. We choose Equation (2)

DOI: 10.1002/wcm

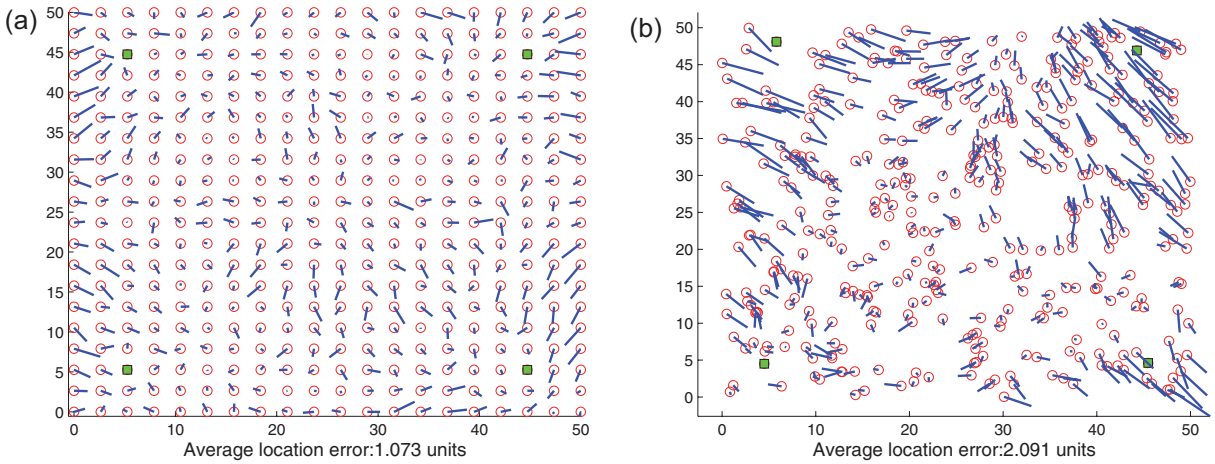

Fig. 1. Estimated locations for the (a) uniformly and (b) randomly deployed sensor network.

to simulate the signal strength between sensor nodes, and we simply set the parameter $w_s = 1$. For the tradeoff parameter used to solve the RKLS problem in Equation (8), we set $\gamma = 0.01$.

As described in Section 3, we calculate the relative locations of all sensor nodes, and then transform the relative locations to absolute locations. Since it is difficult to assess the quality of the relative locations, we focus on evaluating the accuracy of the absolute locations of the non-anchors, and use Equation (14) to estimate the average location error. All of the reported results are averaged over 50 trials.

4.1. Quality and Efficiency of the Localization Method

KSR needs a graph that represents the topological structure of the sensor network, and a kernel function to calculate the similarities between sensor nodes. The weight matrix W and the Gaussian kernel matrix K are determined by the noisy signal strength matrix, so the accuracy of the relative locations is related to the standard deviation parameter σ and the user-defined parameter w_G. Meanwhile, the transformation from the relative locations to absolute locations is determined by the locations of anchors, and hence the number of anchors M is related to the estimation of the absolute locations of non-anchors. We present a quantitative analysis of the effect of σ, number of anchors M, the user-defined parameter w_G, and the density of sensors on the localization performance in this subsection.

Figure 1 shows the localization results for each sensor node with the user-defined parameter $w_G = 1$ and the number of anchors $M = 4$. The squares are anchors and the circles denote the non-anchors. Each line connects a true sensor location and its estimation. The length of each line denotes the estimated error. The longer the line, the larger the error is. Figure 1(a) shows the final solution in the uniformly deployed sensor networks, whose average location error is approximately 1.073 units. Figure 1(b) shows the final solution in the randomly deployed sensor networks, whose average location error is about 2.091 units.

Figure 2 shows the effects of σ on the localization performance. We set $w_G = 1$ and choose $l = 2$, 4, 6, 8, respectively. Figure 2(a) shows the average location error (averaged over all trials and over all sensors) in the uniformly deployed sensor network, while Figure 2(b) shows the average location error in the randomly deployed network. It is shown that the localization accuracy monotonically decreases with the increase of the value of σ. With the same value of σ, the worst case is observed when $l = 2$. If σ is small, the average location error will be small, especially for the uniformly deployed networks, e.g., the average location error is less than 1 unit if σ is not larger than 0.2. Compared with the uniformly deployed network, the randomly deployed network is more sensitive with respect to the value of σ, i.e., it is more sensitive with the level of noise.

Figure 3 shows the effects of the number of anchors M on the localization performance. We simply set $w_G = 1$ and choose $\sigma = 0.2, 0.4, 0.6, 0.8$, respectively. Figure 3(a) and (b) show the average location errors in the uniformly deployed network and the randomly deployed network, respectively. We can see that the average location error increases with the decrease of

DOI: 10.1002/wcm

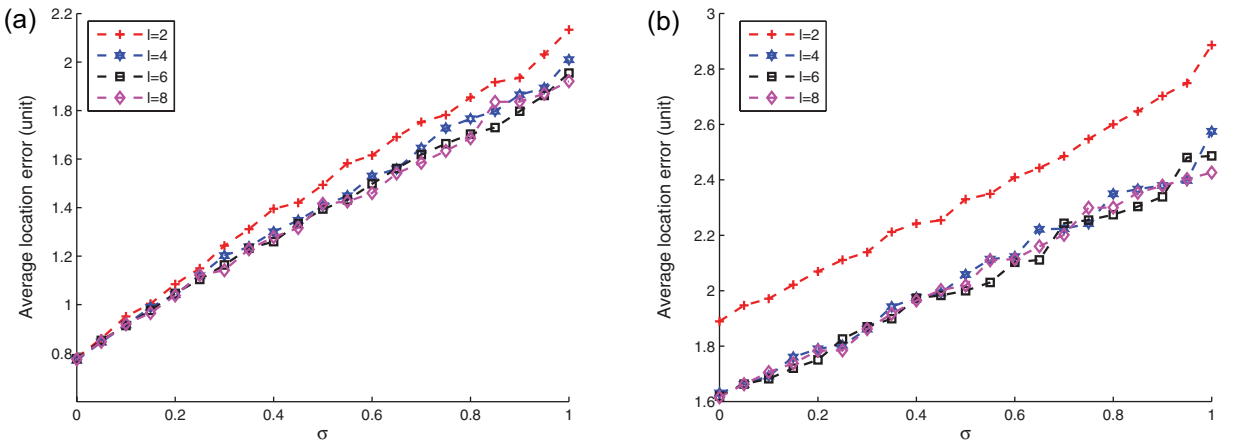

Fig. 2. Simulation results under different value of σ for the (a) uniformly and (b) randomly deployed sensor networks.

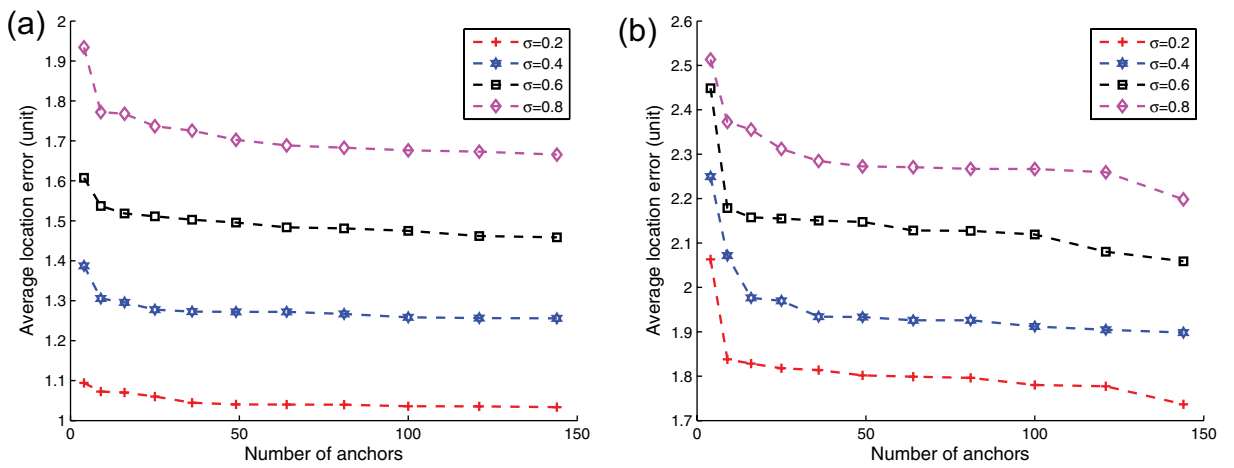

Fig. 3. Average location error with different number of anchors M in the (a) uniformly and (b) randomly deployed sensor networks.

the number of anchors. However this trend is not very significant. The biggest elbow takes place at the point $M = 9$, especially in the randomly deployed network. In Figure 3(b), the average location error decreases about 0.25 units from $M = 4$ to $M = 9$. However, for the cases of $\sigma = 0.2$, 0.4 or 0.6 in the uniformly deployed networks, the influence of the number of anchors is very small. If $M \geq 9$, the decrease of the location error is not significant for both uniformly and randomly deployed sensor networks. Therefore, our localization method can obtain a good performance with a small number of anchors. It is also verified that optimally preserving the topological structure of the sensor networks can alleviate the influence of the number of anchors for localizing sensors.

Figure 4 shows the effects of the value of the bandwidth parameter w_G. We simply set $M = 16$ and choose $\sigma = 0.1$, 0.3, 0.5, 0.7, respectively. Although the errors are different with different values of w_G, the change is very small. The biggest elbow is observed at $w_G = 10^{-3}$ when $\sigma = 0.1$, where the error is less than 0.2 units.

Furthermore, we explore the influence of the sensor density. We assume that there are 36 anchors, and there are 16^2, 18^2, $\cdots$, 40^2 non-anchors uniformly or randomly deployed in the 50×50 square region, respectively. We show the final relationship between the localization accuracy and the number of non-anchors under different value of the noise standard deviation σ in Figure 5. The simulation results show that the localization error monotonically decreases with the increase of the number of non-anchors. When there are 40^2 non-anchors and the noise standard deviation is $\sigma = 0.2$, the average location error is less than 0.8 unit in both uniformly and randomly deployed sensor networks.

DOI: 10.1002/wcm

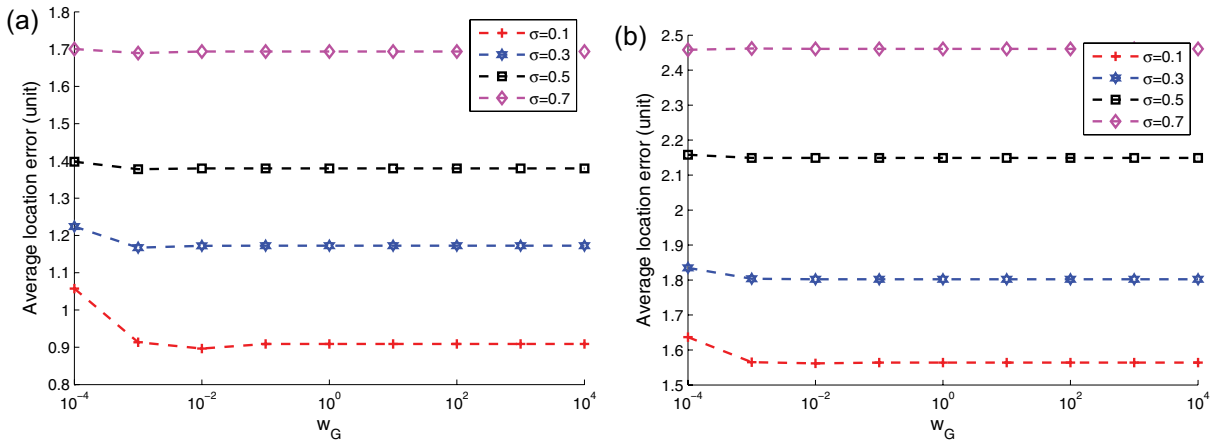

Fig. 4. Simulation results under different band-width parameter w_G in the (a) uniformly and (b) randomly deployed sensor networks. We choose $w_g = \{10^{-4}, 10^{-3}, \cdots, 10^3, 10^4\}$, and report the corresponding average location error under different σ.

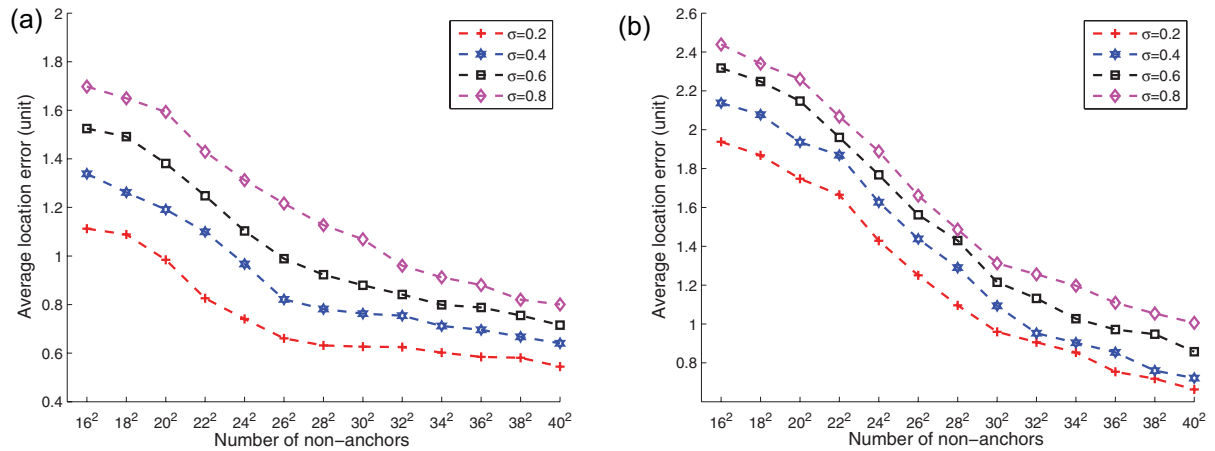

Fig. 5. Simulation results with different number of non-anchors in both (a) uniformly and (b) randomly deployed sensor networks.

4.2. Comparison

In this subsection, we conduct a set of simulations to compare KSR with five existing methods: (1) the classic MDS-MAP algorithm proposed in Reference [25]; (2)The Isomap method proposed in Reference [29]; (3) the regularized kernel least squares (RKLS) method proposed in Reference [14]; (4)the kernel matrix regression (KMR) method proposed in Reference [15]; and (5) the support vector regression (SVR) method, which is a simplified variant of a kernel-based method used for location estimation [10]. MDS-MAP and Isomap require the pair-wise distance, which can be estimated from the signal strength in the ranging procedure. For the sake of simplicity, here we directly assume that the measured pair-wise distance is $\tilde{d}(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_i, \mathbf{x}_j)(1 + N(0, \sigma^2))$, where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the real distance and σ is chosen the same as the σ used in our algorithm. As demonstrated in Reference [14], RKLS is sensitive to the number of anchors. Although our method does not significantly vary with the number of anchors, we set $M = 256$ for the sake of fairness. We choose Gaussian kernel function for KSR, RKLS, KMR and SVR, and set the band-width parameter $w_G = 0.002$. Other parameters used in these methods can be found in the caption of Figure 6.

The comparison results are reported in Figure 6. It is found that the performances of the kernel-based methods (KSR,RKLS, KMR and SVR) are better than the range-based methods (ISOMAP and MDSMAP) when $\sigma = 0.1$ and 0.2. When $\sigma = 0.3$ or 0.4, the performances of RKLS, KMR and SVR are low. Although our KSR method does not achieve the lowest

DOI: 10.1002/wcm

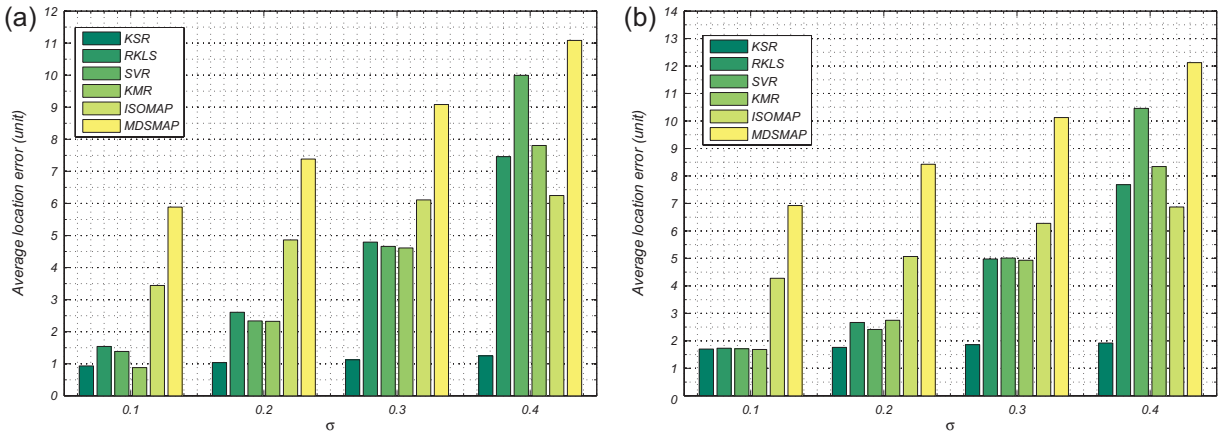

Fig. 6. Performances of different methods for the (a) uniformly and (b) randomly deployed sensor network. We choose the number of anchors $M = 256$, and report the average location error with $\sigma = 0.1$, 0.2, 0.3, and 0.4, respectively. For the ISOMAP, it needs to estimate the geodesic distance between pair-wise sensor nodes, which depends on the k nearest neighbor graph, we set $k = 8$ by default. For the RKLS and SVR, which require a tradeoff parameter C to avoid over-fitting, we simply set $C = 10$.

average location error when $\sigma = 0.1$, it is comparable with the best result (KMR), and the difference is less than 0.1 unit. Furthermore, when $\sigma \geq 0.2$, our method always obtains the best performance in both uniformly and randomly deployed sensor networks. Compared with related methods, the improvement of KSR is significant. This also shows that considering the topological structure of the network is important for localizing sensors in WSNs.

5. Conclusion

In this paper, we study the localization problem in WSNs using signal strength. In specific, we propose a novel localization method based on KSR. We directly adopt a Gaussian kernel function to measure the similarities between sensor nodes. Then we use KSR to estimate the relative locations of all sensor nodes. If there are sufficient anchors, the relative locations can be transformed into absolute locations. Choosing KSR for the localization problem has two advantages. First, the kernel function can capture the nonlinear relationship in the signal strength space. Second, KSR allows us to construct a graph to preserve the topological structure of the sensor networks. The simulation results show that the localization performance can be significantly improved by considering the topological structure of the sensor networks.

In our further work, we hope to evaluate our method in other networks. We would also like to implement this method as a prototype and investigate its performance in real application scenarios.

Acknowledgments

This work was supported in part by NSFC-Guangdong Province joint Project grant no. U0735003; NSFC grant nos. 60604029 and 60736021; 863 High-Tech Project no. 2007AA041201.

References

1. Szewczyk R, Osterweil E, Polastre J, Hamilton M, Mainwaring A, Estrin D. Habitat monitoring with sensor networks. *Communications of the ACM* 2004; **47**(6): 34–40.
2. Franceschini F. A review of localization algorithms for distributed wireless sensor networks in manufacturing. *International Journal of Computer Integrated Manufacturing* 2007; **1**(1): 1–19.
3. Hu F, Jiang M, Xiao Y. Low-cost wireless sensor networks for remote cardiac patients monitoring applications. *Wireless Communications and Mobile Computing* 2008; **8**(4): 513–529.
4. Du X, Xiao Y, Chen H, Wu Q. Secure cell relay routing protocol for sensor networks. *Wireless Communications and Mobile Computing* 2006; **6**(3): 375.
5. Picard J, Weiss A. Localization of networks using various ranging bias models. *Wireless Communications and Mobile Computing* 2008; **8**(5): 553–562.
6. Bahl P, Padmanabhan V. RADAR: an in-building RF-based user location and tracking system. *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies* 2000; **2**: 775–784.
7. Niculescu D, Nath B. DV based positioning in ad hoc networks. *Telecommunication Systems* 2003; **22**(1): 267–280.
8. He T, Huang C, Blum B, Stankovic J, Abdelzaher T. Range-free localization schemes for large scale sensor networks. *Technical Report CS-2003-06*, Department of Computer Science, University of Virginia, 2003.
9. Tran D, Nguyen T. Localization in wireless sensor networks based on support vector machines. *IEEE Transactions on Parallel and Distributed Systems* 2008; **19**(7): 981–994.

DOI: 10.1002/wcm

10. Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks* 2005; **47**(6): 825–845.
11. Liu H, Luo X, Yao Y. Two manifold learning techniques for sensor localization. *IEEE International Conference on Systems, Man and Cybernetics* 2007; **14**: 2114–2118.
12. Wang C, Chen J, Sun Y, Shen X. Wireless sensor networks localization with Isomap. *IEEE International Conference on Communications*, 2009.
13. Nguyen X, Jordan M, Sinopoli B. A kernel-based learning approach to ad hoc sensor network localization. *ACM Transactions on Sensor Networks* 2005; **1**(1): 134–152.
14. Kuh A, Zhu C, Mandic D. Sensor network localization using least squares kernel regression. *Lectoure Notes in Computer Science* 2006; **4253**: 1280–1288.
15. Honeine P, Richard C, Essoloh M, Snoussi H. Localization in sensor networks—a matrix regression approach. *IEEE Sensor Array and Multichannel Signal Processing Workshop* 2008; **14**: 284–287.
16. Seidel S, Rappaport T. 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation* 1992; **40**(2): 207–217.
17. Pan R, Zhao J, Zheng V, *et al*. Domain-constrained semi-supervised mining of tracking models in sensor networks. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, USA, 2007; 1023–1027.
18. Cai D, He X, Han J. Efficient kernel discriminant analysis via spectral regression. *Technical Report UIUCDCS-R-2007-2888*, University of Illinois at Urbana-Champaign, 2007.
19. Cai D, He X, Han J. Spectral regression for dimensionality reduction. *Technical Report UIUCDCS-R-2007-2856*, University of Illinois at Urbana-Champaign, 2007.
20. West D. *Introduction to Graph Theory*. Prentice Hall: Upper Saddle River, NJ, 2001.
21. Schölkopf B, Smola A. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press: Cambridge, MA, 2001.
22. Sheng X, Hu Y. Energy based acoustic source localization. *Lecture Notes in Computer Science* 2003; **2643**: 285–300.
23. Steinwart I. On the influence of the kernel on the consistency of support vector machines. *The Journal of Machine Learning Research* 2002; **2**: 67–93.
24. Essoloh M, Richard C, Snoussi H, Honeine P. Distributed localization in wireless sensor networka as a pre-image problem in a reproducing kernel hibert space. *Sixteenth European Signal Processing Conference*, 2008.
25. Shang Y, Ruml W, Zhang Y, Fromherz M. Localization from mere connectivity. *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing*, ACM Press, New York, USA, 2003; 201–212.
26. Stewart G, Mahajan A. *Matrix Algorithms, Volume II: Eigensystems*, Vol. 56. ASME, Philadelphia, PA, USA, 2003.
27. Paige C, Saunders M. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software* 1982; **8**(1): 43–71.
28. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press: New York, 2004.
29. Patwari N, Hero A, III. Manifold learning algorithms for localization in wireless sensor networks. *IEEE International Conference on Acoustics, Speech, and Signal Processing* 2004; **3**: 857–860.

Authors' Biographies

Chengqun Wang received the BE degree in Electronic Engineering from Harbin Institute of Technology, China, in 2003. And now he is a PhD candidate in Department of Control Science and Engineering, Zhejiang University. He is a member of the Group of Networked Sensing and Control (IIPC-nesC) in State Key Laboratory of Industrial Control Technology, China. His research interests include kernel-based learning, convex and concave optimization, wireless sensor networks, and content delivery networks.

Jiming Chen received PhD in Control Science and Engineering. He was a visiting scholar at University of Waterloo, INRIA, NUS, Simula Research Laboratory. He is currently an associate professor with Institute of Industrial Process Control, State Key lab of Industrial Control Technology, Zhejiang University, China. He leads a group of Networked Sensing and Control at Zhejiang University. He has published over 50 different papers, and edited one book on RFID and Sensor Network. He is currently editors of several journals, such as *Ad Hoc & Sensor Wireless Networks, an International Journal (AHSWN)*, *Journal of Computer*, etc. and guest editor of Wireless communication and mobile computing (Wiley). He is a general Symposia Co-chair of IWCMC 2009. He served in over 40 program committee member since 2005, including IEEE PIMRC 2008, 2009 and Globecom 2008, 2009.

Youxian Sun received the Diploma from the Department of Chemical Engineering, Zhejiang University, China, in 1964. He joined the Department of Chemical Engineering, Zhejiang University, in 1964. From 1984 to 1987, he was an Alexander Von Humboldt Research Fellow, and Visiting Associate Professor at University of Stuttgart, Germany. He has been a full-time professor at Zhejiang University since 1988. In 1995, he was elected to be an Academician of Chinese Academy of Engineering. His current research interests include modeling, control and optimization of complex systems, and robust control design and its application. He is author and co-author of more than 450 journal and conference papers. He is currently the director of institute of industrial process control and national engineering research center of industrial automation, Zhejiang University. He also serves as Vice-Chairman of IFAC Pulp and Paper Committee, Vice-President of Chinese Association of Automation, and Vice-President of China Instrument and Control Society.

DOI: 10.1002/wcm