Efficient Kernel Based 3-D Source Localization via Tensor Completion

Shan Lu[1], Jun Zhang[1], Xianmin Ma[2] and Changju Kan[3*]
[1] Information College, Capital University of Economics and Business
Beijing, China
[e-mail : lushan@cueb.edu.cn, zhangjun@cueb.edu.cn]
[2] Heilongjiang International University
Heilongjiang, China
[e-mail : hljmxm@163.com]
[3] College of Communications Engineering, Army Engineering University of PLA
Nanjing, China
[e-mail : kanchangju@163.com]
*Corresponding author: Changju Kan

Abstract

Source localization in three-dimensional (3-D) wireless sensor networks (WSNs) is becoming a major research focus. Due to the complicated air-ground environments in 3-D positioning, many of the traditional localization methods, such as received signal strength (RSS) may have relatively poor accuracy performance. Benefit from prior learning mechanisms, fingerprinting-based localization methods are less sensitive to complex conditions and can provide relatively accurate localization performance. However, fingerprinting-based methods require training data at each grid point for constructing the fingerprint database, the overhead of which is very high, particularly for 3-D localization. Also, some of measured data may be unavailable due to the interference of a complicated environment. In this paper, we propose an efficient kernel based 3-D localization algorithm via tensor completion. We first exploit the spatial correlation of the RSS data and demonstrate the low rank property of the RSS data matrix. Based on this, a new training scheme is proposed that uses tensor completion to recover the missing data of the fingerprint database. Finally, we propose a kernel based learning technique in the matching phase to improve the sensitivity and accuracy in the final source position estimation. Simulation results show that our new method can effectively eliminate the impairment caused by incomplete sensing data to improve the localization performance.

---

1. Introduction

Recently, the problem of the source localization in wireless sensor networks (WSNs) has received considerable attention due to its increasingly widespread application, such as in smart transportation system, monitoring systems, and intrusion target detection systems [1]-[4]. Among the detection area, there are a number of sensors called sensing nodes which are deployed for localization problems in WSNs. Based on the measured data of sensing nodes, the final goal of the localization problem is to estimate the position of an unknown source. Considering the scenarios, most of the current localization methods were designed for two-dimensional (2-D) problems. However, three-dimentional (3-D) localization is also needed in some cases, such as the localization problem of unmanned aerial vehicles (UAVs) in some special environments, and the target sourc localization in vocal concert environments [5]-[7].

Many kinds of the localization methods such like received signal strength (RSS) based algorithms or their improved methods have been investigated which are based on the measured energy data [8]–[11]. Most of these methods were designed in 2-D and can be extended to 3-D cases. However, one problem with these distance estimated localization techniques is that the signal propagation model should be correctly and concisely described. Due to the varying air-ground topography, wireless radio channels are more seriously suffered by noise and other interference, the signal model can not be clearly depicted. Hence, the final measurements are deviating from the real data, which further decrease the final estimated source position precision [12]–[13].

The fingerprint localization technique, which uses a radio map to determine user locations, has received considerable attention [14]–[16]. Different from determining distances, this method first constructs the fingerprint database by measuring the source's RSS values at each point during an offline phase. Whenever the source requests services, the position can be quickly estimated by matching the sensing data with the previously constructed fingerprint database and finding the best item during the online phase. Due to the learning mechanism used, the fingerprinting method is less sensitive to complex conditions and can provide relatively accurate localization performance [17]–[19].

As described above, an offline fingerprint database must be constructed for the training phase of the fingerprint based localization, which is composed of the RSS value collected at each grid point of the testing area. However, due to various obstacles, quickly building a complete fingerprint database may not be possible. Firstly, some measured data may not be available given the difficulty of a complicated environment, such as some remote corners, which leads to an incomplete fingerprint database. Secondly, the number of grid points to be trained in 3-D localization are increasing when compared to 2-D scenes, which results in a high computation requirement. Also, to more precisely estimat a position, a high density of grid points is prefered, which also increases the training burden. Thirdly, to strive for an accurate estimation, a recalibration procedure is needed whenever the environmental changes occur in the measured area, which will further increase the training effort [20]. To complete the data matrix and reduce the data collecting overhead, some related works based on signature map construction were investigated [21]-[26]. To data, this idea have mainly been applied in the field of data processing, which shows a bright future in the localization problem.

In this paper, an efficient source localization technique is developed to solve the above mentioned problems. By analyzing the problems in the completion of the fingerprint and the

reduction of the sensing overhead in the training phase, we find that the core issue involves using less measurement data to complete the unknown data of other grid points. Aimed at the issue, we first exploit the RSS spatial correlation and show the low rank property of the RSS data matrix measured in the localization area. By using this property, a tensor completion scheme is then proposed to effectively recover the incomplete fingerprint database in the training phase. Furthermore, we propose a kernel based learning technique in the matching process that can simplify the localization problem and improve the sensitivity and accuracy in fingerprint matching.

Specifically, the contributions of this paper are summarized as follows:

- We exploit the spatial correlation of the RSS data and propose a novel training scheme based on tensor completion. Since the tensor completion can efficiently generate a complete fingerprint database, the effort and time spent during training phase can be substantially reduced by the proposed training scheme and the integrity of the fingerprint database is also improved.

- We introduce a kernel based learning method in the matching phase. The kernel based method simplifies source localization by clarifying the complicated relationship between the fingerprint and the corresponding position and identifying the source position more effectively, which further improves the localization precision.

- Finally, we present some deeper insights into the proposed scheme. By comparison, the simulation results show the effectiveness of the proposed method.

The rest of this paper is structured as follows. In Section 2, the preliminaries of the our workis described, which conclude the basic signal model and the RSS spatial correlation property. Then, the efficient kernel based 3-D localization algorithm via tensor completion is proposed in Section 3. In Section 4, we provide the simulation results which to demonstrate the effectiveness of the new proposed algorithm. Finally, the main conclusions are summarized in Section 5.

2. Preliminaries

2.1 Signal Model

We considered the fundamental signal model of the localization problem. In the measured environment, we put M sensing nodes for localization, the positions of these nodes are known and the sensing nodes are gridding distributed. The target source performs random distribution and its position is unknown. The estimate the source position, we should firstly measured the received power of each sensing node from the unknown source. By using the radio propagation path loss model, the average received power P_i of the ith sensing node can be written as [11]

$$P_i\left(\mathrm{dB}\right) - P_t\left(\mathrm{dB}\right) = K - 10\gamma \log_{10}\left(\frac{d_i}{d_0}\right) + n_i, \tag{1}$$

where P_t is the transmit power of the unknown source; d_0 is the reference distance which is normally taken as 1 m; d_i is the real distance of the ith sensing node and source; and n_i is the Gaussian distributed measurement noise; K is a constant value and γ is the pass loss exponent.

2.2 RSS Correlation Property

In this part, the spatial correlation of the RSS measurements received from the sensing nodes is exploited. Since that the limited number of freedom degrees means that the rank of the measurement data matrix is much lower than its dimension. The correlations of the measured data can also be implied by the lower freedom degrees of the data matrix [20]. The RSS data correlation as well as thelow rank nature of the data matrix which showed by the normalized singular values can be observed in **Fig. 1**.
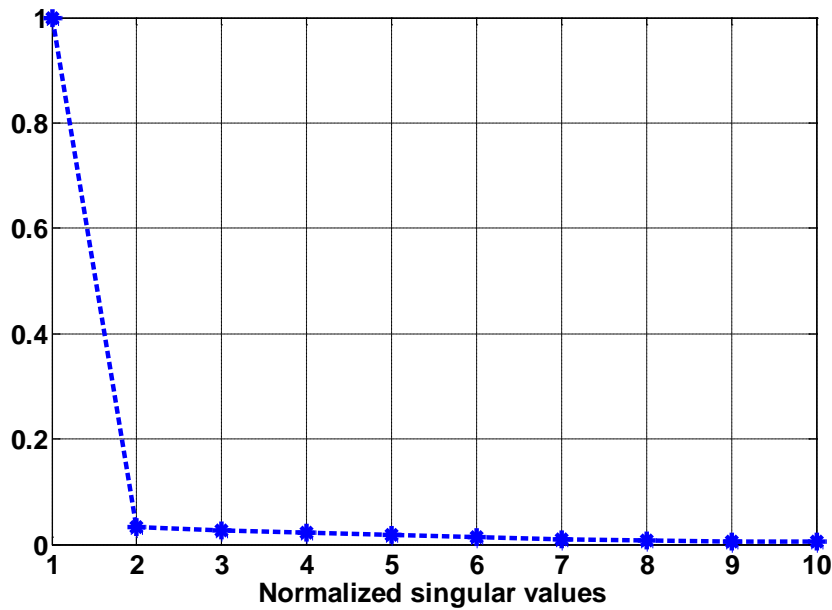

Fig. 1. Normalized singular values of the data matrix in the simulation model.

The figure can clearly show that the first few singular values are containing most of the energy, whereas additional singular values which caused by noise-like phenomena that exhibit much lower energy. The exploited low rank nature of the sensing data matrix plays a pivotal role in our proposed localization technique.

3. Efficient Kernel Based 3-D Source Localization via Tensor Completion

In this part, we propose our novel efficient kernel based localization technique. The new algorithm exploits the spatial correlation of the RSS measurments in the sensing data matrix. By taking advantage of this property, the localization overhead in the training procedure is further reduced. The proposed algorithm is divided into two phases: the training phase, which provides a subset of the RSS measurements for some grid points and then recovers the grid points to build the complete reference fingerprint database; and the matching phase, which uses kernel learning to find the best estimation for the source position. In the following, we describe these two phases in detail.

3.1 Training Phase

The core task of the training phase involves building the fingerprint database. Firstly, in some grid points among the measurement area that were chosen randomly, we record all the

measured RSS data received by sensing nodes from the constructing node. Since only part of the grid points are used for data collecting, tensor completion was used to complete the entire data matrix based on the incomplete data. Finally, all RSS vector data of all the grid points formed the entire fingerprint database.

Specifically, for a random sensing node m, the recorded data matrix can be written as

$$\mathbf{\Phi}_{\Omega} = \begin{cases} \mathbf{\Phi}, & i \in \Omega \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where Ω is the subset of the grid points that are used for data collecting, i is the corresponding serial number, and $\mathbf{\Phi}$ is the measured data matrix we want to recover as precisely as possible.

After that, an optimization problem can be modeled by the tensor completion for the missing RSS value, which is try to complete the whole data by using the known subset. The problem can be which is written as

$$\begin{aligned} \min_{\mathbf{\Phi}} & \|\mathbf{\Phi}\|_{*} \\ s.t. & \mathbf{\Phi}_{\Omega} = \mathbf{\Gamma}_{\Omega} \end{aligned}, \tag{3}$$

where $\| \ \|_{*}$ denotes the tightest convex envelop for the rank of matrix. $\mathbf{\Phi}$ is the incomplete tensor to be completed. Since the sensing data are a 3-D matrix, $\mathbf{\Phi}_{\Omega}$ and $\mathbf{\Gamma}_{\Omega}$ are three-mode tensors and the size in each mode are the same. Entries of $\mathbf{\Gamma}_{\Omega}$ from the set Ω are known whereas the remaining entries are missing. Since that computing the rank of a tensor is an NP hard problem which can not be directly solved, we introduce the following definition for the tensor trace norm [23]:

$$\|\mathbf{\Phi}\|_{*} := \sum_{i=1}^{3} \delta_i \left\|\mathbf{\Phi}_{(i)}\right\|_{*}, \tag{4}$$

where $\sum_{i=1}^{3} \delta_i = 1$ is content and $\delta_i \geq 0$. $\left\|\mathbf{\Phi}_{(i)}\right\|_{*}$ denotes the norm of unfolded matrix $\mathbf{\Phi}_{(i)}$. However, a tensor's trace norm is consistent with all the matrices unfolded along each mode. Based on this, the above mentioned problem is further written as:

$$\begin{aligned} \min_{\mathbf{\Phi}} & \sum_{i=1}^{3} \delta_i \left\|\mathbf{\Phi}_{(i)}\right\|_{*}. \\ s.t. & \mathbf{\Phi}_{\Omega} = \mathbf{\Gamma}_{\Omega} \end{aligned} \tag{5}$$

Due to the interdependence among the matrix trace norm terms, the problem in Equation (5) is challenging to solve. Hence, we introduced the additional matrices $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\}$ to split these interdependent terms. The problem in Equation (5) can be simplified as follows:

$$\begin{aligned} \min_{\mathbf{\Phi}, \mathbf{T}_i} & \sum_{i=1}^{3} \delta_i \left\|\mathbf{T}_i\right\|_{*} \\ s.t. & \mathbf{\Phi}_{(i)} = \mathbf{T}_i \qquad i = 1, 2, 3 . \\ & \mathbf{\Phi}_{\Omega} = \mathbf{\Gamma}_{\Omega} \end{aligned} \tag{6}$$

In order to solve this convex optimization problem, we are trying to use the alternating direction method of multipliers (ADMM). Firstly, we use the tensor versions $\boldsymbol{\beta}_i$ to replace the corresponding matrix $\mathbf{T}_i$:

$$\min_{\boldsymbol{\Phi}, \beta_1, \beta_2, \beta_3} \sum_{i=1}^{3} \delta_i \left\| \boldsymbol{\beta}_{i(i)} \right\|_*$$
$$s.t. \boldsymbol{\Phi} = \boldsymbol{\beta}_i \qquad i = 1, 2, 3 .$$
$$\boldsymbol{\Phi}_{\Omega} = \boldsymbol{\Gamma}_{\Omega} \tag{7}$$

Finally, we define the augmented Lagrangian function as follows:

$$L_\rho \left(\boldsymbol{\Phi}, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i \right) = \sum_{i=1}^{n} \delta_i \left\| \boldsymbol{\beta}_{i(i)} \right\|_* + \left\langle \boldsymbol{\Phi} - \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i \right\rangle + \frac{\rho}{2} \left\| \boldsymbol{\Phi} - \boldsymbol{\beta}_i \right\|_F^2 . \tag{8}$$

According to the ADMM framework, the augmented Lagrangian function is applied to update $\boldsymbol{\beta}_i$, $\boldsymbol{\Phi}$, and $\boldsymbol{\gamma}_i$ iteratively.

After the iteration, the final recovered data matrix of the mth sensing node is achieved. The same procedure can be applied to the other sensing nodes for their data matrix completion, $\left\{ \boldsymbol{\Phi}^1, \boldsymbol{\Phi}^2, \cdots, \boldsymbol{\Phi}^M \right\}$.

To facilitate the following analysis, we further translated the above data matrices into a new matrix form. Specifically, for a fixed grid p, the fingerprint vector can be written as

$$\vec{F}_{R,p} = \left[P_{R,p}^1, P_{R,p}^2, P_{R,p}^3, \cdots, P_{R,p}^M \right] . \tag{9}$$

The entire fingerprint database is:

$$\mathbf{D} = \left[\vec{F}_{R,1}, \vec{F}_{R,2}, \vec{F}_{R,3}, \cdots, \vec{F}_{R,N} \right]^T . \tag{10}$$

3.2 Matching Phase

When the unknown source is set in the area, the fingerprint vector can also be achieved, which is written as

$$\vec{F}_S = \left[P_{r,S}^1, P_{r,S}^2, P_{r,S}^3, \cdots, P_{r,S}^M \right] . \tag{11}$$

In the matching phase, by defining a model $\boldsymbol{\Psi}(\ )$ to estimate the source position, the estimated coordinates can be written as

$$\hat{\mathbf{x}} = \boldsymbol{\Psi} \left(\vec{F}_S \right), \tag{12}$$

where $\vec{F}_S$ is an input entry of the defined model $\boldsymbol{\Psi}(\ )$ and $\hat{\mathbf{x}}$ is the corresponding estimated position.

Hence, the objective was to find the model $\boldsymbol{\Psi}(\ )$ associated with the fingerprint $\vec{F}_{R,p}$ and the corresponding output $\hat{\mathbf{y}}_p$. Since that this is a 3-D problem, let $\hat{\mathbf{y}}_p = \left(\hat{y}_{p,1}, \hat{y}_{p,2}, \hat{y}_{p,3} \right)$ be the estimated position of the reference node in point p. Also, $\boldsymbol{\Psi}(\ )$ can be written as

$\Psi(\ )=\left(\Psi_1(\ ),\Psi_2(\ ),\Psi_3(\ )\right)$, where $\Psi_d(\ )$ is responsible for the estimation of the d th coordinate and d is an indicator for a random dimension.

Table 1. Commonly used kernels with parameters $c > 0, d \in N_+, \sigma > 0$

Kernel	Expression
Monomial	$\left(\left\langle \vec{F}_i, \vec{F}_j \right\rangle\right)^d$
Exponential	$\exp\left(\left\langle \vec{F}_i, \vec{F}_j \right\rangle \big/ 2\sigma^2\right)$
Gaussian	$\exp\left(-\left\| \vec{F}_i, \vec{F}_j \right\|_2^2 \big/ 2\sigma^2\right)$
Multiquadratic	$\sqrt{\left\| \vec{F}_i, \vec{F}_j \right\|_2^2 + c}$

In this part, the functions $\Psi_d(\ )$ is determined by the kernel based learning method. Hence, a reproducing kernel $\mathbf{K}$ is considered and υ is its Reproducing Kernel Hilbert Space (RKHS) with inner product $\langle , \rangle_\upsilon$ [27]. Here some of the commonly used kernel functions are showed in **Table 1**. Then, the function $\Psi_d(\ )$ is obtained by minimizing the following regularized empirical risk:

$$\xi\left(\left(p_{1,d}, \Psi_d\left(\vec{F}_{R,1}\right)\right), \ldots, \left(p_{N,d}, \Psi_d\left(\vec{F}_{R,N}\right)\right)\right) + \eta\zeta\left(\left\|\Psi_d\right\|_\upsilon^2\right), \tag{13}$$

where ξ is an arbitrary cost function, for example, the mean squared error. ζ is defined as a real-valued function which is strictly monotonically increased. $\| \ \|_\upsilon$ is the norm in the RKHS. η is set to be a positive tunable parameter, which is used to control the trade-off between the fitness error and the solution complexity as measured by the norm in the RKHS.

Specifically, the first term in Equation (13) is taken using the mean squared error of the estimated position $\Psi_d\left(\vec{F}_{R,p}\right)$ and the real position $y_{p,d}$,

$$\xi\left(\left(p_{1,d}, \Psi_d\left(\vec{F}_{R,1}\right)\right), \ldots, \left(p_{N,d}, \Psi_d\left(\vec{F}_{R,N}\right)\right)\right) = \frac{1}{N}\sum_{p=1}^{N}\left(y_{p,d} - \Psi_d\left(\vec{F}_{R,p}\right)\right)^2. \tag{14}$$

The regularization term in Equation (13) is changed in its simplest form as $\left\|\Psi_d\right\|_\upsilon^2$. Hence, the optimization problem is finally defined, which is written as [28]:

$$\min \frac{1}{N}\sum_{p=1}^{N}\left(y_{p,d} - \Psi_d\left(\vec{F}_{R,p}\right)\right)^2 + \eta\left\|\Psi_d\right\|_\upsilon^2. \tag{15}$$

By determining the $\alpha_{p,d}$ and transforming it into a finite dimension, we express the minimizer $\Psi_d(\ )$ as a finite linear combination of the kernels. Hence, the problem of the minimizer of the regularized empirical risk in Equation (13) is described in a simpler form, as follows:

$$\Psi_d\left(*\right) = \sum_{p=1}^{N} \alpha_{p,d} \mathbf{K}\left(\vec{F}_{R,p}, *\right). \tag{16}$$

By combining Equation (15) and (16), we obtain the following optimization problem in terms of $\boldsymbol{\alpha}_{*,d}$:

$$\min\left(\mathbf{P}_{*,d} - \mathbf{K}\boldsymbol{\alpha}_{*,d}\right)^T \left(\mathbf{P}_{*,d} - \mathbf{K}\boldsymbol{\alpha}_{*,d}\right) + \eta N \boldsymbol{\alpha}_{*,d}^T \mathbf{K}\boldsymbol{\alpha}_{*,d}, \tag{17}$$

where $\mathbf{P}_{*,d}$ is the dth coordinate of all the grid points, which is $\mathbf{P}_{*,d} = \left[y_{1,d}, \cdots, y_{p,d}, \cdots, y_{N,d}\right]$. $\mathbf{K}$ is the $N \times N$ matrix and the (i,j)th entry is $\upsilon\langle\rho_i, \rho_j\rangle$, for $i, j \in \{1, 2, \cdots, N\}$. Also, $\boldsymbol{\alpha}$ is defined as a $N \times N$ matrix, which the dth column of $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}_{*,d}$ and the pth line is $\boldsymbol{\alpha}_{p,*}$.

This is a classical quadratic regression problem, we further transform it into the following forms:

$$\mathbf{K}\mathbf{P}_{*,d} + \mathbf{K}^2\boldsymbol{\alpha}_{*,d} + \eta N \mathbf{K}\boldsymbol{\alpha}_{*,d} = \mathbf{0}. \tag{18}$$

Thus, we obtain the following form of the solution:

$$\boldsymbol{\alpha}_{*,d} = \left(\mathbf{K} + \eta N \mathbf{I}\right)^{-1} \mathbf{P}_{*,d}, \tag{19}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

After calculating the estimation of all the dimensions, the whole formulation can be written as:

$$\boldsymbol{\alpha} = \left(\mathbf{K} + \eta N \mathbf{I}\right)^{-1} \mathbf{P} \tag{20}$$

Using Equation (16) and the definition of the vector of functions $\boldsymbol{\Psi}(\ )$, we now define a model that allows us to estimate all three coordinates of the unknown source at once, as follows:

$$\boldsymbol{\Psi}\left(*\right) = \sum_{p=1}^{N} \boldsymbol{\alpha}_{p,*} \mathbf{K}\left(\vec{F}_{R,p}, *\right) \tag{21}$$

Therefore, by using the source fingerprint $\vec{F}_S$ as the input entry, the final estimated position $\hat{\mathbf{x}}$ is calculated.

The overall efficient source localization algorithm is summarized in Algorithm 1.

Algorithm 1: Efficient Kernel-based Localization via Tensor Completion

---

Input $M, N, K, \rho, \mathbf{P}$

Phase I : Training Phase

for m = 1 to M **do**

    collecting subset data matrix Υ_Ω

    Set $\Phi_\Omega = \Gamma_\Omega$ and $\Phi_{\hat{\Omega}} = 0$

    for k = 1 to K **do**

        for i = 1 to 3 **do**

$$\beta_i = \text{fold}_i\left[D_{\frac{\delta_i}{\rho}}\left(\Phi_i + \frac{1}{\rho}\gamma_i(i) \right) \right]$$

        end for

$$\Phi_\Omega = \frac{1}{3}\left(\sum_{i=1}^{3} \beta_i - \frac{1}{\rho}\gamma_i \right)_{\hat{\Omega}}$$

$$\gamma_i = \gamma_i - \rho(\beta_i - \Phi)$$

    end for

end for

$\{\Phi^1, \Phi^2, \cdots, \Phi^M\} \rightarrow \mathbf{D}$

return $\mathbf{D}$; //completed fingerprint database

Phase II : Matching Phase

Collecting source fingerprint $\vec{F}_S$

for i=1 to N **do**

    for j=1 to N **do**

$$\mathbf{K} = \exp\left(-\left\| \vec{F}_{R,i}, \vec{F}_{R,j} \right\|_2^2 \big/ 2\sigma^2 \right)$$

    end for

end for

$\alpha = (\mathbf{K} + \eta N\mathbf{I})^{-1}\mathbf{P}$

$$\Psi(\vec{F}_S) = \sum_{p=1}^{N} \alpha_{p,*}\mathbf{K}(\vec{F}_{R,p}, \vec{F}_S)$$

Output $\hat{\mathbf{x}}$

---

4. Performance Evaluation

4.1 Simulation Setup

Fig. 2 shows the fundamental simulation scenario. In this scenario, the measured area was a $20 \times 20 \times 20$ m cube and nine sensing nodes were deployed at the known coordinates (0,0), (0,10), (0,20), (10,0), (10,10), (10,20), (20,0), (20,10), (20,20) in meters. The unknown source was distributed randomly in the area. The measured area was divided into $10 \times 10 \times 10$ grids, thus, each grid was about a $2 \times 2 \times 2$ m cube area. For signal propagation modeling, the transmitting power of the source is $P_S = 0\text{dBm}$. The pass loss exponent is $\gamma = 3$. The AWGN

is with zero mean and a variance $\sigma = 1$. To reduce the overhead, only a part of the grid are used to collect the RSS data in the training process. The sampling rate is set to be 0.5 and the samples are chosen randomly.

Unknown Source   **Sensing Node**
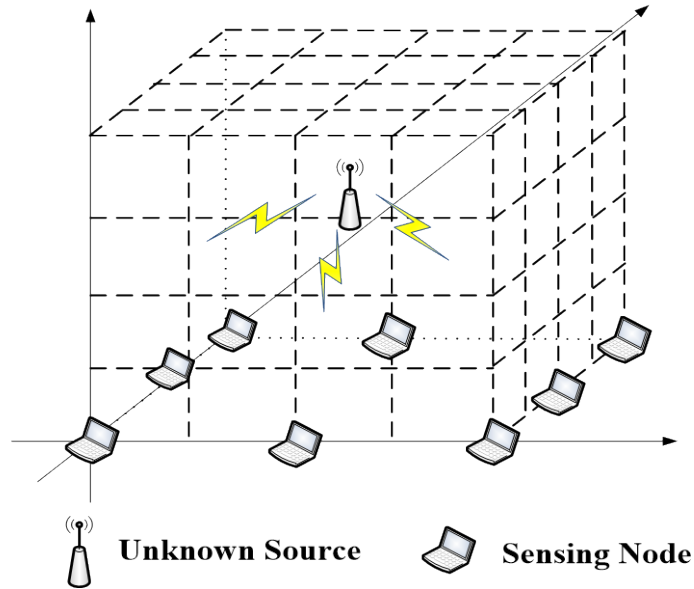
Fig. 2. The configuration of the simulation scenario.

4.2 Effectiveness analysis of the proposed localization algorithm

In the following analysis, the root mean square error (RMSE) metric is used to compare the position estimation performance of different algorithms. The RMSE can clearly show the performance variation under each condition.

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^{T} \left\| \hat{\mathbf{x}}_i - \mathbf{x}_i \right\|^2} \tag{22}$$

where $\hat{\mathbf{x}}$ is the estimated position of the unknown source and $\mathbf{x}$ is the correct position. T is set to be the simulation time.

Also, we have performed several localization schemes under different conditions for comparison, which is to demonstrate the effectiveness of the proposed novel localization algorithm. The schemes are listed as follows.

- A baseline scheme called "Original Data + Kernel" (OD-Kernel scheme), which directly uses the complete sensing data to perform the position estimation. Meanwhile, the kernel learning method is used for the matching phase.

- A baseline scheme called "Tensor Completion + Kernel" (TC-Kernel scheme), which uses the incomplete data and processed by tensor completion method for the position estimation. In the matching phase, the kernel learning method is used to find the best point. Note that this scheme is also our proposed algorithm.

- A baseline scheme called "Interpolation Completion + Kernel" (IC-Kernel scheme), which uses the incomplete data and processed by an interpolation completion method in [29] for the final position estimation. The kernel learning method is used to find the best point in the matching phase.

- A baseline scheme called "Tensor Completion + KNN" (TC-KNN scheme), which uses the incomplete data and processed by tensor completion method for the final position estimation. However, the matching phase uses the K-nearest neighbor (KNN) method to find the best matching point.

- A baseline scheme called "Interpolation Completion + KNN" (IC-KNN scheme), which uses the incomplete data and processed by an interpolation completion method for the final position estimation. Also, the KNN method is used to find the best matching point.

The effect of the data sampling rate on the RMSE of the source position estimate is presented in **Fig. 3.** Especially, the effectiveness of the tensor completion is shown in this simulation, so that only the localization schemes using kernel learning are compared. From the figure, by using the tensor completion, our proposed TC-Kernel scheme obviously improved the localization precision when compared with the IC-Kernel scheme. This occurred because the tensor completion can efficiently recover the incomplete sensing data matrix, which increased the precision of the final position estimation. When the data sampling rate increased, our proposed TC-Kernel scheme gradually reached a comparative performance to the OD-kernel scheme, while the overhead was almost cut in a half. Also, a phenomenon is seen that our proposed method may have worse performance when the sampling rate is lower, such as 0.2. A rational explanation may be that too little existed sensing data exceed the tolerance of tensor completion, which is more sensitive compared to the IC-Kernel scheme.
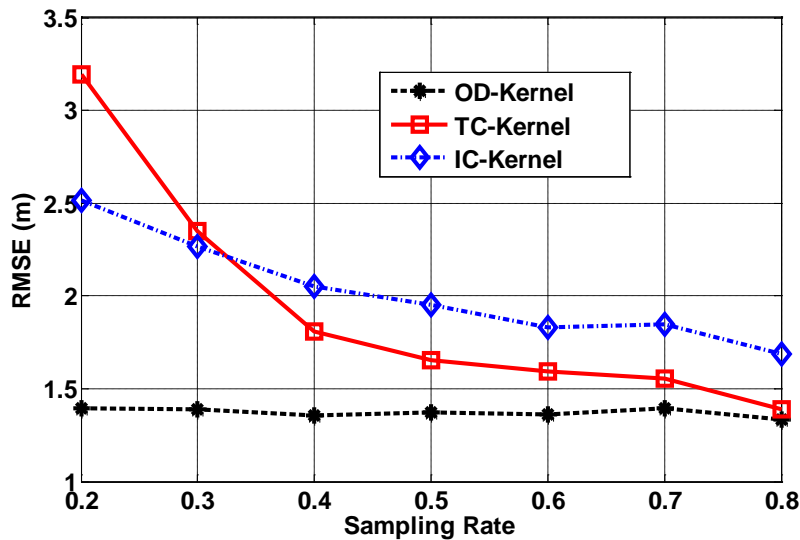

Fig. 3. Effect of the tensor completion on RMSE versus data sampling rate.

Fig. 4 shows the effect of the data sampling rate on the RMSE of several localization schemes. The effectiveness of the kernel method on the localization performance is mainly verified in this part. By using the kernel method, the TC-Kernel scheme and the IC-Kernel scheme are all performed well when compared with the other schemes, which attained a position precision of about 0.8 m, when the sampling rate was greater than 0.5. However, when comparing the TC-Kernel scheme with the IC-Kernel scheme, an interesting phenomenon occurred: the superiority of our proposed scheme disappeared gradually as the sampling rate decreased. A rational explanation for this observation may be that the kernel

method is more sensitive than the KNN method especially in harsh cases. As the sampling rate decreased, the effciency of the tensor completion similarly decreased, the increasing error data generated will definitely impair the finally localization, which influences the kernel learning method more than the KNN method.
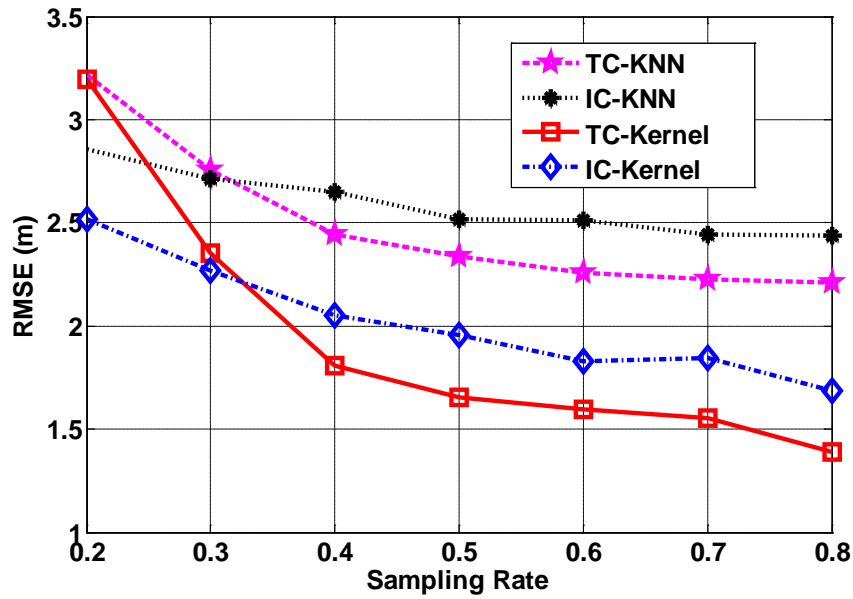

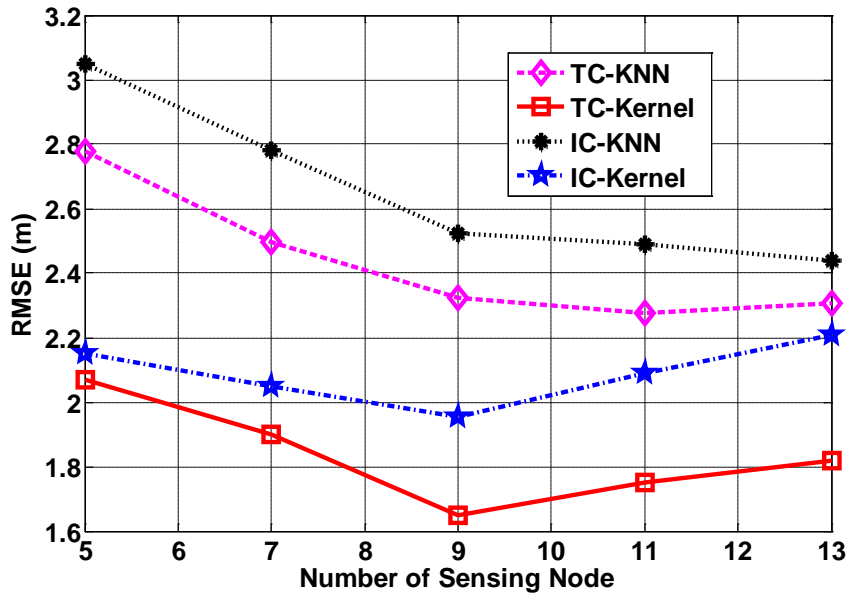**Fig. 4.** Effect of the kernel learning on RMSE versus data sampling rate.

Fig. 5. Effect of the number of sensing node on RMSE for various schemes.

Fig. 5 shows the effect of the number of sensing node on the source position estimation for each schemes. In the figure, using different matching methods have distrinct performance

trends. The KNN method, the TC-KNN scheme and the IC-KNN scheme all demonstrate a monotonically decreasing trend along with the number of sensing nodes, which means that the localization performance improves gradually. For this reason, increasing the number of sensing nodes will enlarge the sensing data matrix and further improve the localization precision. However, when using the kernel based schemes (TC-KNN and IC-KNN), the performance decreased whenzhe number of nodes was nine or less, but as the sensing node number increased above nine, the localization performance decreased. A possible reason for this may be that since the sampling rate was fixed, increasing the number of sensing nodes enlarged the sensing data matrix, but also introduced more missing data that required completion. Since the kernel method is more sensitive than the KNN method, when the sensing node number is much larger, the increasing error in the completed data becomes excessive, so the performance of the kernel based schemes decreased. However, our proposed scheme still outperformed the other schemes despite the variation in the number of the sensing nodes.
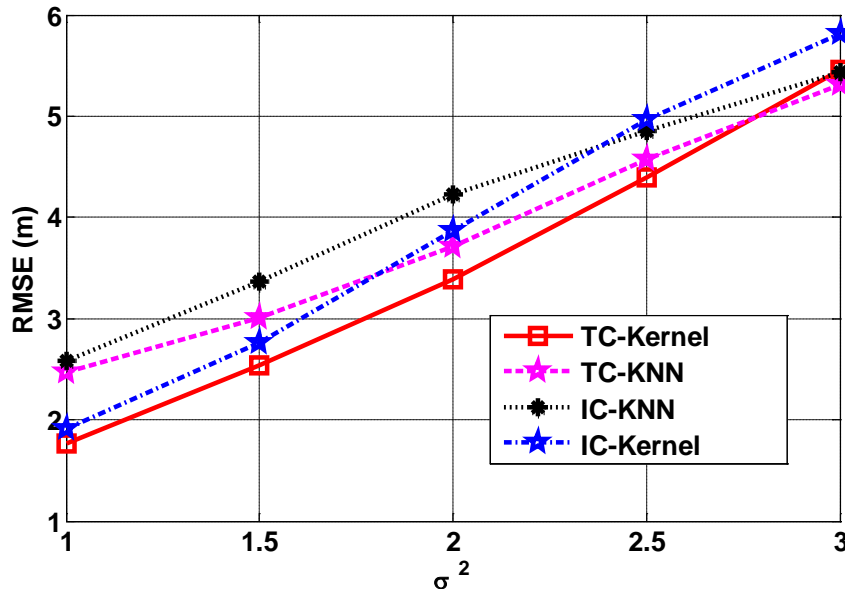
Fig. 6. Effect of the noise variance on RMSE for various schemes.

Fig. 6 illustrates the effect of noise variance on the source position precision for each schemes. The large noise variance resulted in a small signal tonoise ratio (SNR), which resulted in poor performance for all the schemes. From **Fig. 6** we can see that whatever the noise variance is set, the proposed TC-Kernel scheme is always achieves a better performance when compared with other algorithms.

5. Conclusion

In this paper, an efficient kernel based source localization method via tensor completion is proposed. We first exploit the RSS spatial correlation and uses a tensor completion scheme to effectively recover the incomplete fingerprint database. Furthermore, we propose a kernel based learning technique which can simplify the localization problem and improve the sensitivity and accuracy in fingerprint matching. Simulation results showed that our new

localization technique can effectively eliminate the impairment caused by incomplete sensing data and improve the localization performance. For future work, we will consider the real multi-channel data and try to figure out the frequency domain property of the measured data, which will further improve the accuracy of the completed data and increase the precision of the localization estimation.

Acknowledgement

References

[1] I. T. Haque, and C. Assi, "Profiling-based indoor localization schemes," *IEEE Systems Journal*, vol. 9, no. 1, pp. 76-85, Sep. 2013. Article (CrossRef Link)

[2] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205-1211, Mar. 2011. Article (CrossRef Link)

[3] J. Hong, and T. Ohtsuki, "Signal eigenvector-based device-free passive localization using array sensor," *IEEE Trans. Veh. Technol.*, vol. 64, no.4, pp. 1354-1363, Apr. 2015. Article (CrossRef Link)

[4] N. Saeed, and H. Nam, "Robust multidimensional scaling for cognitive radio network localization," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4056-4062, Sep. 2015. Article (CrossRef Link)

[5] F. Lazzari, A. Buffi, P. Nepa, and S. Lazzari, "Numerical Investigation of an UWB Localization Technique for Unmanned Aerial Vehicles in Outdoor Scenarios," *IEEE sensors J.*, vol. 17, no. 9, pp. 2896-2903, May 2017. Article (CrossRef Link)

[6] W. Cui, C. Wu, W. Meng, B. Li, Y. Zhang, and L. Xie, "Dynamic Multidimensional Scaling Algorithm for 3-D Mobile Localization," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 12, pp. 2853-2865, Dec. 2016. Article (CrossRef Link)

[7] Y. Wang, and K. C. Ho, "An Asymptotically Efficient Estimator in Closed-Form for 3D AOA Localization Using a Sensor Network," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6524-6535, Dec. 2015. Article (CrossRef Link)

[8] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1389-1403, Mar. 2013. Article (CrossRef Link)

[9] M. R. Gholami, R. M. Vaghefi, and E. G. Strom, "RSS-based sensor localization in the presence of unknown channel parameters," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3752-3759, Aug. 2013. Article (CrossRef Link)

[10] L. Lu, H. Zhang, and H.-C. Wu, "Novel energy-based localization technique for multiple sources," *IEEE Systems J.*, vol. 8, no. 1, pp. 142-150, Mar. 2014. Article (CrossRef Link)

[11] H. Lohrasbipeydeh, T. A. Gulliver, and H. Amindavar, "Blind received signal strength difference based source localization with system parameter errors," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4516-4531, Sep. 2014. Article (CrossRef Link)

[12] C.-Y. Shih and P. J. Marrón, "COLA: Complexity-reduced trilateration approach for 3D localization in wireless sensor networks," in *Proc. of 4th Int. Conf. Sensor Technol. Appl. (SENSORCOMM)*, pp. 24–32, 2010. Article (CrossRef Link)

[13] P. Pivato, L. Palopoli, and D. Petri, "Accuracy of RSS-based centroid localization algorithms in an indoor environment," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3451-3460, Oct. 2011. Article (CrossRef Link)

[14] Q. D. Vo and P. De, "A Survey of Fingerprint-Based Outdoor Localization," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 491-506, 1st Quart. 2016. Article (CrossRef Link)

[15] S. He and S.-H. G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466-490, 1st Quart. 2016. Article (CrossRef Link)

[16] Q. Chen and B. Wang, "FinCCM: Fingerprint Crowdsourcing, Clustering and Matching for Indoor Subarea Localization," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 677-680, Dec. 2015. Article (CrossRef Link)

[17] S.-H. Fang, Y.-T. Hsu, and W.-H. Kuo, "Dynamic Fingerprinting Combination for Improved Mobile Localization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4018-4022, Dec 2011. Article (CrossRef Link)

[18] E. Kupershtein, M. Wax, and I. Cohen, "Single-Site Emitter Localization via Multipath Fingerprinting," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 10-21, Jan. 2013. Article (CrossRef Link)

[19] M. Zhou, Y. Wei, Z. Tian, X. Yang, and L. Li, "Achieving Cost-efficient Indoor Fingerprint Localization on WLAN Platform: A Hypothetical Test Approach," *IEEE Access*, vol. 5, pp. 15865-15874, Aug. 2017. Article (CrossRef Link)

[20] S. Nikitaki, G. Tsagkatakis, and P. Tsakalides, "Efficient Multi-Channel Signal Strength Based Localization via Matrix Completion and Bayesian Sparse Learning," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2244-2256, Nov. 2015. Article (CrossRef Link)

[21] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208-220, Jan. 2013. Article (CrossRef Link)

[22] S. Sorour, Y. Lostanlen, S. Valaee, and K. Majeed, "Joint indoor localization and radio map construction with limited deployment load," *IEEE Trans. Mobile Comput.*, vol. 14, no, 5, pp. 1031-1043, May 2015. Article (CrossRef Link)

[23] M. Tang, G. Ding, Z. Xue, J. Zhang, and H. Zhou, "Multi-dimensional spectrum map construction: A tensor perspective," in *Proc. of Wireless Communications Signal Processing (WCSP), 2016 8th International Conference on*, Oct. 2016. Article (CrossRef Link)

[24] F. Xiao, W. Liu, Z. Li, L. Chen, and R. Wang, "Noise-Tolerant Wireless Sensor Networks Localization via Multi-norms Regularized Matrix Completion," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2409-2419, Mar. 2018. Article (CrossRef Link)

[25] J. Cheng, Z. Song, Q. Ye, and H. Du, "MIL: A mobile indoor localization scheme based on matrix completion," in *Proc. of Communications (ICC), 2016 IEEE International Conference on*, 2016. Article (CrossRef Link)

[26] M. Tang, G. Ding, Q. Wu, Z. Xue, and T. A. Tsiftsis, "A Joint Tensor Completion and Prediction Scheme for Multi-Dimensional Spectrum Map Construction," *IEEE Access*, vol. 4, pp. 8044-8052, Nov. 2016. Article (CrossRef Link)

[27] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, and Y. Chen, "Kernel-Based Learning for Statistical Signal Processing in Cognitive Radio Networks: Theoretical Foundations, Example Applications, and Future Directions," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 126-136, Jul. 2013. Article (CrossRef Link)

[28] S. Mahfouz, F. M.-Chehade, P. Honeine, J. Farah, and H. Snoussi, "Kernel - based machine learning using radio-fingerprints for localization in wsns," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 1324-1336, Apr. 2015. Article (CrossRef Link)

[29] M. Tang, Z. Zheng, G. Ding, and Z. Xue, "Efficient TV White Space Database Construction via Spectrum Sensing and Spatial Inference," in *Proc. of Computing and Communications Conference*, pp. 1-5, 2015. Article (CrossRef Link)

Shan Lu received the master degree in Management Science and Engineering from the Capital University of Economics and Business, China, in 2015. He is the assistant dean of the School of Information of the Capital University of Economics and Business, the chairman of the trade union, and the director of the Computer Public Basic Education and Research Group. His research directions are quantitative finance, enterprise informatization, computer applications.

Jun Zhang, Ph.D. in Computer Management, University of Paris Ninth, France, is the dean of the School of Information, Capital University of Economics and Business. The research field mainly involves the application of management science and information technology in management. Now he is a communication review expert of the National Natural Science Foundation of China.

Xianmin Ma received the master degree in Software Engineering from Harbin Institute of Technology, China, in 2011. She is deputy director of information engineering of Heilongjiang International University. Her research interrests focus on big data and computer application technology.

Changju Kan received his B.S. degree and M.S. degree in communications engineering, from College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2011. He is currently pursuing his B.S. degree at College of Communications Engineering, Army Engineering University of PLA. His research interests are cognitive radio networks and passive source localization.