This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

Robust Relative Fingerprinting-Based Passive Source Localization via Data Cleansing

Changju Kan, Guoru Ding, *Senior Member, IEEE,* Qihui Wu, *Senior Member, IEEE,* Rongpeng Li, and Fei Song

***Abstract*—Recently, source localization is becoming a major research focus. Majority of the existing studies focus on the design of received signal strength (RSS) based localization methods. However, when in the face of the complicated environments with severe fading, RSS based localization methods achieve relatively inferior accuracy performance, compared with fingerprinting-based localization methods. Nevertheless, traditional fingerprinting-based localization methods subject to the condition that the source transmit power is known, which cannot be directly used in passive localization cases where the sensing nodes do not have the prior information on the source. Additionally, the received sensing data may contain errors and then affect the location precision due to various abnormal conditions such as device failure and malicious cases. In this paper, we propose a novel robust relative fingerprinting-based passive localization algorithm via a data cleansing approach. First, we figure out the fingerprint correlations property and introduce a new relative fingerprint framework. The key idea is that by exploring the correlations between the source fingerprint and the reference fingerprint database, the correction factors can be achieved to apply the fingerprint idea into the passive localization case. Second, we formulate a generalized modeling of the abnormal data in localization problem and propose a data cleansing approach which utilizes the sparse property of the abnormal data. Based on this, the negative influence of abnormal data can be further eliminated. Third, considering the sparse property of the source position, we use the sparse Bayesian learning in the matching process for the purpose of achieving more precise estimated source position. Simulation results demonstrate that the proposed algorithm achieves higher accuracy performance in passive source localization in terms of eliminating the abnormal data impairment.**

***Index Terms*—Passive localization, robust localization, fingerprint, correlation, data cleansing.**

I. Introduction

A. Background and Motivation

NOWADAYS, source localization has received increasing research attention, as position information is important for a variety of applications, such as emergency services, mobile communications, public safety and intelligent transportation, to mention just a few [1]–[4]. In many practical cases, while the source devices are requested for their positions, it is not possible to have these sources taken part in the localization process. Thus, the passive or non-cooperative localization issue, of which the basic idea is that the sensing nodes do not have the prior information of the source transmit power, is becoming a great interest. On one hand, many devices are designed simply, which do not support the processing module, or have limited use due to the constrained battery. On the other hand, some devices do not wish to be discovered, e.g., the tapping phone in the conference room, the pseudo station in the building, and the interference source in a military fight, etc. Moreover, due to the varying topography, the wireless radio channel is susceptible to noise, interference, and other channel impediments. Also, as the society develops, various installations have been established, which made the surface surroundings becoming much more complicated. In addition, as a large part of people's daily lives is room based, such as schools, office buildings, shopping malls, or residential area, the communications are suffering great shadowing by kinds of obstacles such as walls and floors [5], [6]. All these complex environments make the localization analysis more difficult and unquestionably decrease the precision of location estimation.

Furthermore, since the sensing devices are getting increasingly integrated and portable, all kinds of devices could be used as the sensing nodes, which are an important component of the passive source localization. Hence, various unexpected events such as node equipment failures and malicious behaviors may inevitably occur during the whole localization process, which can be serious threats of degradation in the estimation accuracy [7]. The above observations motivate us in this paper to investigate the interesting but critical issue of robust passive source localization.

B. Related Work

Recently, various passive localization techniques have been investigated which based on the received data measurement in [8]–[18]. Due to the low cost and easy deployment characteristics, energy based techniques have been widely employed such as received signal strength (RSS) [9]–[12] and received signal strength difference (RSSD) based approaches [13]. However, the localization performance may suffer degradation since these methods are affected by the environments more heavily. For another, the authors in [14]–[18] proposed some classical techniques that rely on distance estimation using

This work is supported by the National Natural Science Foundation of China (Grant No. 61501510 and Grant No. 61631020), Natural Science Foundation of Jiangsu Province (Grant No. BK20150717), China Postdoctoral Science Funded Project (Grant No. 2016M590398).

C. Kan, G. Ding, F. Song are with the College of Communications Engineering, Army Engineering University of PLA , Nanjing 210007, China (Email: kanchangju@163.com, dr.guoru.ding@ieee.org, aroucian@163.com). G. Ding is also with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China.

Q. Wu is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China (Email: wuqihui2014@sina.com).

R. Li is with Zhejiang University, Hangzhou 310027, China (email: lirongpeng@zju.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

time of arrival (TOA), time difference of arrival (TDOA), or improved methods. Nevertheless, these methods require precise timing and synchronization, which necessitate the use of very accurate clocks, thus increasing the cost and complexity of implementation. Notably, one severe problem of all the above techniques is that they rely heavily on the signal propagation model. The localization performance may suffers once the real received signal deviates from the assumed signal model too much, which is a general phenomenon in the real world [19], [20].

Fingerprinting-based localization method, which leverages a data mining/learning mechanism to determine user locations, can reduce the dependence of signal model and provide relatively higher accuracy in localization performance (see, e.g., [21]–[27]). Generally, the fingerprinting-based method contains two parts: a training phase and a matching phase. In the training phase, a reference node is used in each reference position to transmit signals and several sensing nodes receive and record the signals. When all reference points have finished collecting data, the reference fingerprint database is created. Then, in the matching phase, a learning mechanism is used to find the best matched item of the source fingerprint in the fingerprint database, which is considered to be the estimated source position. Due to the preconstruction of the reference fingerprint database, fingerprinting is well recognized as an *active localization* method as it requires the prior information on the transmit power of the source. Therefore, in *passive localization* conditions where the source emitting power is unknown, the fingerprinting method cannot be directly applied.

Furthermore, due to the abnormal circumstances, errors may occur in the sensing data for various forms. Hence, how to formulate the errors and take measures to eliminate their impairments is becoming an important research point in recent years. In [28]–[30], the non-line-of-sight (NLOS) condition is considered in the TOA based localization problems, several convex relaxation methods are proposed to mitigate the effect of NLOS errors on the localization performance. Considering the environment changes case, the authors in [31] performed experiments to quantify how changes in an environment affect the localization accuracy. Then, a correlation method for selecting channels is presented to decrease the localization error rate. The authors in [32] and [33] considered the heterogeneous device influence and tried to solve it using a fingerprint that is related to the signal strength differences. Specifically, [32] explored the use of normalized logarithmic signal strength ratios. Furthermore, [33] provided a theoretical analysis of the signal strength difference method and designed a series of substantial tests under different conditions. The results showed that both proposed methods in [32] and [33] can mitigate the effects of the hardware variations and raise the localization performance effectively. Notably, these methods were designed for cooperative or active source localization and can not be applied in passive localization cases directly. Also, considering the mismatch problem between training and runtime fingerprints, the authors in [34] proposed a multi-channel fingerprint-based indoor localization system that employs modern mathematical concepts based on the sparse representations and matrix completion theories. In [35], we developed a robust RSS-based localization algorithm to filter out the abnormal data and eliminate the errors' influence effectively. Nevertheless, to the best knowledge of the authors, the impact of abnormal data in fingerprinting-based passive source localization has not been reported.

C. Contributions

In this paper, we develop a robust passive source localization method to address the issues mentioned above. Aimed at the complicated environments in which traditional passive localization methods may suffer from great precision degradation, this paper figures out the fingerprint correlations property and then develops a new relative fingerprinting-based passive localization framework. The key idea is that by exploring the correlations between the source fingerprint and the fingerprint database, the new corrected factors can be achieved to facilitate the design of fingerprint method in passive localization. Meanwhile, considering the existence of either accidental equipment failures or random malicious behaviors, every sensing node could sporadically and randomly produce abnormal data, which results in great degradation of the estimation accuracy. To overcome the problem, this paper utilizes a data cleansing scheme to effectively filter out the abnormal data, which will eliminate the impairment it brings. Furthermore, considering the sparse property of the source distribution, we propose a sparse Bayesian learning approach. This method enables the estimated source position to be obtained by finding out the best sparse vector from a large dictionary of potential candidates. Specifically, the contributions of this paper are summarized as follows:

- A framework of relative fingerprinting-based passive localization is proposed. By analyzing the correlations between the source fingerprint vector and the collected reference fingerprint database, the application of the fingerprinting method into the passive localization case is firstly introduced. Different from the traditional fingerprint-based localization which directly uses the fingerprint database, the proposed algorithm introduces corrected factors which are calculated in the relative fingerprint constructing phase and further transforms the original one into a new relative fingerprint database. The transformation will apply the fingerprint in the passive localization for better performance.
- A generalized modeling of abnormal data in localization problem is formulated, which can involve the combined effects of unexpected equipment failures and malicious data falsifications. To solve that, a data cleansing scheme is proposed which exploits the sparsity of the abnormal data. After the process, the new source data matrix that cleansed out abnormal data component is achieved.
- A sparse Bayesian learning method is developed in the matching phase to achieve accurate location estimation. The method can precisely estimate the source position by solving an under-determined system of equations. The in-depth simulations are presented to demonstrate the effectiveness of the proposed robust algorithm, which show a higher accuracy performance in passive localization of the proposed algorithm over the state-of-the-art schemes.

TABLE I
KEY NOTATIONS AND SYMBOLS USED IN THIS PAPER.

Symbol	Definition
M	Number of the sensing nodes
N	Number of the grid points
H	Number of the detected channels
P_t	Transmit power of the source
$P_{SN,i}^{r,S}$	Received power of $i-\mathrm{th}$ sensing node from the unknown source
$P_{i,j,k}^{r}$	Received power of $i-\mathrm{th}$ sensing node at $j-\mathrm{th}$ point in $k-\mathrm{th}$ channel from a source
$n_{i,j,k}^{r}$	Noise of $i-\mathrm{th}$ sensing node at $j-\mathrm{th}$ point in $k-\mathrm{th}$ channel from a source
$a_{i,j,k}^{r}$	Abnormal data of $i-\mathrm{th}$ sensing node at $j-\mathrm{th}$ point in $k-\mathrm{th}$ channel from a source
$P_{p,i}^{r,R}$	Received power of $i-\mathrm{th}$ sensing node from the reference node
$\vec{F}_S$	Fingerprint vector of the unknown source
$\vec{F}_{R,p}^{k}$	Fingerprint vector of the reference node at $p-\mathrm{th}$ point in $k-\mathrm{th}$ channel
$\vec{F}_{CH,k}$	Fingerprint vector of received the signal in $k-\mathrm{th}$ channel
$\vec{F}_{o,k}$	Fingerprint vector of the received pure signal in $k-\mathrm{th}$ channel
$\vec{W}_k$	Fingerprint vector of the received noise in $k-\mathrm{th}$ channel
$\vec{A}_k$	Fingerprint vector of the abnormal data in $k-\mathrm{th}$ channel
$\mathbf{Y}$	Coarse data matrix of the received signal of all channels
$\mathbf{X}$	Pure data matrix of the received signal of all channels
$\mathbf{W}$	Noise data matrix of the received signal of all channels
$\mathbf{A}$	Abnormal data matrix of the received signal of all channels
$\mathbf{D_k}$	Reference fingerprint database of the $k-\mathrm{th}$ channel
SN	Number of the occupied channel by the unknown source
$\mathbf{D_{new}}$	New relative fingerprint database
$\vec{C}_p$	Corrected vector at $p-\mathrm{th}$ grid point

The rest of this paper is structured as follows. In Section II, we describe the system model of our work, which contains the tradition signal model and the general abnormal data model. In Section III, we figure out the fingerprint correlations property, which will play a pivotal role in the proposed algorithm. Then, the robust relative fingerprinting-based passive localization algorithm via a data cleansing approach is proposed in Section IV. The simulation results are provided to show the effectiveness of the new proposed algorithm in Section V. Finally, we summarize the main conclusions in Section VI. In addition, the corresponding Cramer-Rao lower bound of the abnormal data model is derived In Appendix A. Furthermore, to facilitate the reading, the key notations and symbols used in this paper are given in Table 1.

II. System Model

A. Traditional Signal Model

We consider a scenario that there are M sensing nodes with known positions and a target source with unknown position. The sensing nodes can be gridding or randomly distributed while the unknown source obeys random distribution. The purpose of the passive localization is to estimate the source position by the sensing nodes through the received power of the source, without the prior information of the source transmit power. Under the radio propagation path loss model, the average received power P_i of the $i\,\mathrm{th}$ sensing node can be expressed as [13]

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

where P_t is the source transmit power; K is a constant value for the reference distance d_0; γ is the pass loss exponent; d_i is the true distance between the source and the $i-\mathrm{th}$ sensing node; n_i is the measurement noise which is assumed to be Gaussian distribution $n_i \sim N\left(0, \sigma_i^2\right)$.

B. General Abnormal Data Model

We know that for the passive source localization problem, all the signal property information of the source is unknown. Besides the unknown source transmit power, we should firstly track the source signal among the monitored channels and find out which channel is occupied by the source. Benefiting from the multi-channel detecting techniques such as wideband spectrum sensing process [37]–[39], we are able to record the sensing data of every detected channel for each sensing node. In this paper, we suppose that totally H channels are detected and the unknown source is occupying any one of them. Hence that the M sensing nodes have to detected all the H channels and recorded the sensing data. The sensing data matrix can be modeled as that in Fig. 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

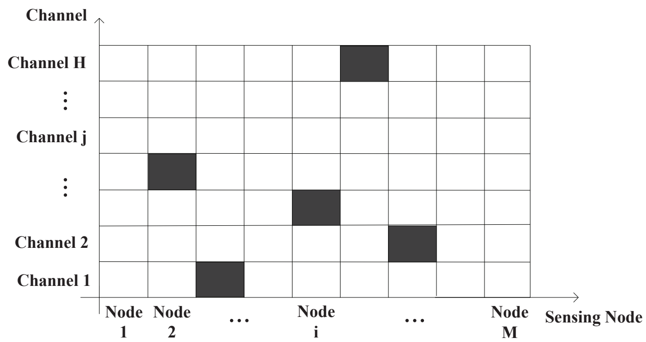


Fig. 1. Abnormal data model of the passive source localization

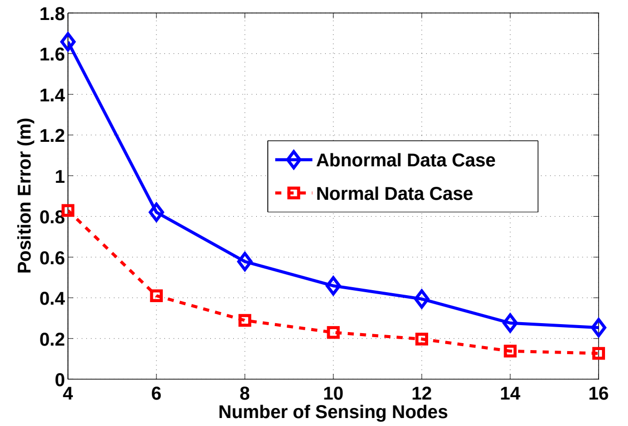


Fig. 2. CRLB comparison of the normal and abnormal data cases

However, due to the unexpected equipment failures and malicious behaviors, every sensing node could randomly and sporadically contribute abnormal data. Meanwhile, when coming up rugged environments, the transmit signal may suffer from severe loss, which could also cause the sensing data error.

Considering the abnormal data, the signal power $P^r_{i,k}$ which based on Eq. (1) can be given as

$$\begin{aligned} P^r_{i,k} &= P_t + K - 10\gamma\log_{10}\left[\frac{d}{d_0}\right] + n_{i,k} + a_{i,k} \\ &= P^o_{i,k} + n_{i,k} + a_{i,k} \end{aligned}, \tag{2}$$

where $P^r_{i,k}$ is the data that recorded by the $i-$th sensing node of the signal in $k-$th channel. $P^o_{i,k} = P_t + K - 10\gamma\log_{10}\left[\frac{d}{d_0}\right]$ is the original signal data which based on the signal propagation model, without considering the influence of noise and abnormal data. $n_{i,k}$ is the noise component which follows zero-mean Gaussian distribution. $a_{i,k}$ denotes the abnormal data component. When $a_{i,k} = 0$, the sensing data returns to the normal data.

Due to the various unexpected events (e.g., node equipment failures, malicious data falsification), the abnormal data component can be also in various forms, such as a constant or a random value. The abnormal data models are often seen in the papers which solve security problems, especially in the spectrum sensing field (see, e.g. , "Always Right/Wrong", "Always Adverse" and some more comprehensive circumstances [40]–[43]). In this paper, we do not distinguish concrete data state, but just formulate a general abnormal data model which only depicting the data variation. As can be seen, our abnormal data model is still general enough to cover most of the other data models and is able to descript other unexpected events. For detailed, the abnormal data is modeled as an i.i.d. random process with mean μ_a and variance σ_a^2 in this paper [7].

The existence of abnormal data will make the final localization process inaccurate. In Fig. 2, we give a simple comparison of the Cramer-Rao lower bound (CRLB) based on two data models. It is obviously that the existence of the abnormal data brings much impairment to the source localization, if no defense method is used for the abnormal data. Hence, the design of a robust passive localization to eliminate the abnormal data influence is very necessary. The detailed CRLB analysis of the abnormal data model is given in Appendix A.

III. Fingerprint Correlation Property

As that the source transmit power is unknown in passive localization, it is a big probability case that the transmit power of the source and the reference node are not the same. After the fingerprint database is established, even though the source and the specific point are perfectly matched, the calculation results may not be the best due to the reason that the received signal data are not the same. Hence, almost all the searching algorithms which used in traditional fingerprinting based localization are ineffective for finding the correct position in passive localization problem. In order to tailor the fingerprint method into the passive localization case, we will exploit the inner correlations of the fingerprint to develop a new fingerprinting based localization framework in the following , which will be proved to be effective for precise passive source location estimation.

Suppose that M sensing nodes are deployed for localization and the test area is divided into N grids. Based on Eq. (2), the received signal power P_r of the $i-$th sensing node at $j-$th grid point in $k-$th channel can be written as

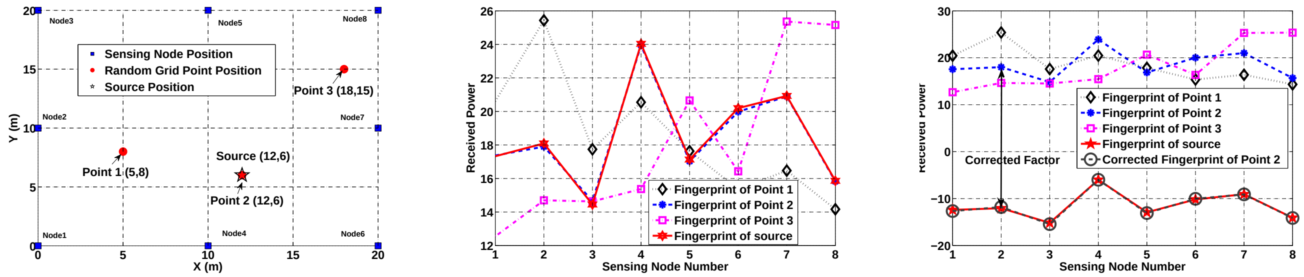
$$P^r_{i,j,k} = P_t + K - 10\gamma\log_{10}\left[\frac{d}{d_0}\right] + n^r_{i,j,k} + a^r_{i,j,k}. \tag{3}$$

To facilitate the analysis, we do not specify the channel information in this part. At any given point p, the fingerprint of any two sources are measured as $\vec{F}_{S1} = \left[P^{r,S1}_{p,1}, P^{r,S1}_{p,2}, P^{r,S1}_{p,3}, \cdots, P^{r,S1}_{p,M}\right]$ and $\vec{F}_{S2} = \left[P^{r,S2}_{p,1}, P^{r,S2}_{p,2}, P^{r,S2}_{p,3}, \cdots, P^{r,S2}_{p,M}\right]$. In fingerprinting based localization problem, these two sources can be figured as the unknown source and the reference node, respectively. $P^{r,S1}_{p,i}$ and $P^{r,S2}_{p,i}$ are the received powers of $i-$th sensing node from the two sources, respectively. We have the following Theorem 1.

Theorem 1: For a given position p in fingerprinting-based localization, the fingerprint vectors of any two sources have an inner correlation whatever there transmit powers are, which can be written as

$$\vec{F}_{S1} - \vec{F}_{S2} = \vec{V} + \vec{n} + \vec{a}, \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access



(a) configuration of the network (b) traditional fingerprint case (c) relative fingerprint case

Fig. 3. Illustration of the relative fingerprint concept. Random choose three grid points and deploy reference node with transmit power $P_{t,R} = 30\text{dBm}$ to achieve each reference fingerprint. In tradition fingerprint case, the source transmit power is the same with reference node $P_{t,S} = P_{t,R} = 30\text{dBm}$; In relative fingerprint case, the source transmit power is $P_{t,S} = 0\text{dBm}$.

where $\vec{F}_{S1}$ and $\vec{F}_{S2}$ are the fingerprints vectors of two different sources located at the same position p, $\vec{V}$ is a constant vector, $\vec{n}$ is the noise vector and $\vec{a}$ is the abnormal data vector.

Proof: By subtracting the two vectors and we obtain

$$\vec{F}_{S1} - \vec{F}_{S2} = \left[P_{p,1}^{r,S1} - P_{p,1}^{r,S2}, \cdots, P_{p,M}^{r,S1} - P_{p,M}^{r,S2}\right]. \quad (5)$$

Since both two sources are located in the same point, the distance of them to all the sensing nodes are also precisely the same. By substituting Eq. (3) into Eq. (5), the formula can be further written as follows

$$\begin{aligned}
&\vec{F}_{S1} - \vec{F}_{S2} = \left[P_{p,1}^{r,S1} - P_{p,1}^{r,S2}, \cdots, P_{p,M}^{r,S1} - P_{p,M}^{r,S2}\right] \\
&= \Big[P_{t,S1} + n_{p,1}^{r,S1} + a_{p,1}^{r,S1} - P_{t,S2} - n_{p,1}^{r,S2} - a_{p,1}^{r,S2}, \cdots, \\
&\qquad P_{t,S1} + n_{p,M}^{r,S1} + a_{p,M}^{r,S1} - P_{t,S2} - n_{p,M}^{r,S2} - a_{p,M}^{r,S2}\Big], \\
&= \left[V + n_{p,1} + a_{p,1}, \cdots, V + n_{p,M} + a_{p,M}\right] \\
&= \vec{V} + \vec{n} + \vec{a}
\end{aligned} \quad (6)$$

where V is a constant that is only associated with the transmit power of the two unknown sources, i.e., $P_{t,S1} - P_{t,S2}$. The RSS measurement noise still performs a Gaussian distribution of with zero mean, which is $n_{p,i} = n_{p,i}^{r,S1} - n_{p,i}^{r,S2} \sim N\left(0, \sigma_{S1,i}^2 + \sigma_{S2,i}^2\right)$. While the abnormal data component also performs a Gaussian distribution, which is $a_{p,i} = a_{p,i}^{r,S1} - a_{p,i}^{r,S2} \sim N\left(\mu_{a,1} - \mu_{a,2}, \sigma_{a1,i}^2 + \sigma_{a2,i}^2\right)$. □

Without considering the influence of abnormal data $\vec{a}$, an insight can be obtained that whenever any two sources are located in the same point, the fluctuation of their fingerprints are consistent, which is made out by the received data of all sensing nodes. The only difference lies in the mean value difference of the two fingerprints written as $\vec{V}$, which is depend on the transmit power of two sources. The noise $\vec{n}$ is a zero mean random variable, which can cause slight difference of the two fingerprints besides the $\vec{V}$. However, the influence can be ignored from a statistical average view point since the mean $\mu_{n,new}$ is zero. Although the abnormal data $\vec{a}$ is also a random variable, the mean $\mu_{a,new}$ is an unpredictable nonzero variable, which will affect the difference ruleless besides $\vec{V}$. Hence, the correlation of the fingerprints cannot be accurately figured out.

Through the analysis, we observe that the mean value difference of the two fingerprints can be utilized as a corrected factor, which may help applying the fingerprint to the passive localization for better performance. By introducing a relative fingerprint framework, the fingerprint correlations property will play a pivotal role in our proposed algorithm. However, since that the influence of abnormal data cannot be omitted, a data cleansing approach should be processed first, which is also a vital step of the proposed robust localization algorithm.

IV. Robust Relative Fingerprinting-Based Passive Localization via Data Cleansing

A. Design Rationale

In order to eliminate the influence that the abnormal data brings, which is due to the accidental equipment failures or malicious operations, we exploit the sparse feature of the abnormal data and develop a data cleansing approach. By handling the sensing data to remove the negative influence from abnormal data, the position estimation accuracy can be largely improved. Meanwhile, since the fingerprint idea cannot be directly applied to the passive source localization case, we exploit the inner correlation of the fingerprint and develop a new relative fingerprinting based localization framework. Based on Theorem 1, any two fingerprints of two sources have inner correlation once they are in the same place. The expression lies in that these two fingerprints have relative consistency while the main difference is a constant vector. If the constant vector can be calculated and removed, then the two fingerprints will coincide together (the slight difference of the coincident two fingerprints in Fig. 3 is caused by the unavoidable noise), which is the core idea of the *relative fingerprint* concept. Fig. 3 gives a visualized description of the relative fingerprint. Hence, the core work of the new framework is trying to figure out the constant vectors of the source fingerprint with the fingerprint database and eliminating them, for which the relative fingerprint database is established. For accurate expression, the constant vectors to be calculated are called the *corrected factors* in the following analysis. Benefit from the new framework, the fingerprint can be effectively used in passive localization for precise source location estimation. The advantages of the proposed algorithm also include that the sparse Bayesian learning is used, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

can utilizing the source sparsity effectively and further raising the estimated precision.

Considering all the above, we divide the proposed robust passive localization method into three phases: i) the data cleansing phase, which receive the source data and perform cleansing work to get a new source fingerprint; ii) the relative fingerprint constructing phase, which first provide the reference fingerprint database and then use the cleansed source fingerprint to transform it into a new relative fingerprint database; iii) the matching phase, which use the sparse Bayesian learning to find the best point for the final estimated position. For more clear and visualized description of the proposed algorithm, the fundamental procedure of each phases and their relations is shown in Fig. 4. In the following, we will describe the three phases sequentially in detail.

B. Data Cleansing Phase

After scanning all the monitored channels, the sensing nodes recorded all sensing data and form a source data matrix $\mathbf{Y}$, which can be seen in Fig. 1. As introduced, every node could randomly and sporadically make errors in sensing data due to the various unexpected events. In this paper, we consider the case that the abnormal sensing data are randomly and sparsely distributed, which is one practical and common assumption in the literature [7]. Hence, the recorded data is composed of three types: the original data, the noise and the abnormal data. Based on Eq. (2), the fingerprint vector of the $k-th$ channel can be further rewritten as

$$\begin{aligned} \vec{F}_{CH,k} &= \left[P^r_{k,1}, P^r_{k,2}, P^r_{k,3}, \cdots, P^r_{k,M}\right] \\ &= \left[P^o_{k,1}, P^o_{k,2}, P^o_{k,3}, \cdots, P^o_{k,M}\right] + \\ &[n_{k,1}, n_{k,2}, n_{k,3}, \cdots, n_{k,M}] + [a_{k,1}, a_{k,2}, a_{k,3}, \cdots, a_{k,M}] \\ &= \vec{F}_{o,k} + \vec{W}_k + \vec{A}_k \end{aligned}. \tag{7}$$

Hence that, the recorded sensing data matrix $\mathbf{Y}$ can be showed as

$$\mathbf{Y} = \left[\vec{F}_{CH,1}, \vec{F}_{CH,2}, \vec{F}_{CH,3}, \cdots, \vec{F}_{CH,H}\right]^T \in \Re^{H\times M}. \tag{8}$$

To facilitate the following analysis, we translate the data components into matrix form.

$$\mathbf{X} = \left[\vec{F}_{o,1}, \vec{F}_{o,2}, \vec{F}_{o,3}, \cdots, \vec{F}_{o,H}\right]^T \in \Re^{H\times M}, \tag{9}$$

$$\mathbf{W} = \left[\vec{W}_1, \vec{W}_2, \vec{W}_3, \cdots, \vec{W}_H\right]^T \in \Re^{H\times M}, \tag{10}$$

$$\mathbf{A} = \left[\vec{A}_1, \vec{A}_2, \vec{A}_3, \cdots, \vec{A}_H\right]^T \in \Re^{H\times M}. \tag{11}$$

Then, we can express the sensing data in the new matrix form as

$$\mathbf{Y} = \mathbf{X} + \mathbf{W} + \mathbf{A}, \tag{12}$$

where $\mathbf{Y}$ is the recorded coarse sensing data of all detected channels by the sensing nodes. $\mathbf{X}+\mathbf{W}$ is the normal sensing data component which contains the original data matrix $\mathbf{X}$ and the noise matrix $\mathbf{W}$. $\mathbf{A}$ is the abnormal data component.

In this paper, we consider the case where only a source is occupying one of the H channels, which indicates that the matrix $\mathbf{X}$ is low-rank. Also, considering that the accidental equipment failures or random malicious behaviors, nonzero entries in the matrix $\mathbf{A}$ are randomly and sparsely distributed. Therefore, the core work of the data cleansing process is to recover the sensing data from the influence of abnormal data, by exploiting the low-rank property of $\mathbf{X}$ and the sparsity property of $\mathbf{A}$.

Based on the observations, a principal component pursuit problem is formulated as follows

$$\begin{aligned} \min_{\mathbf{X},\mathbf{A}} \quad & \mathrm{rank}\left(\mathbf{X}\right) + \lambda \left\langle \mathbf{A} \right\rangle \\ s.t. \quad & \mathbf{X}+\mathbf{W}+\mathbf{A} = \mathbf{Y} \end{aligned}, \tag{13}$$

where $\mathrm{rank}\left(\cdot\right)$ and $\left\langle \cdot \right\rangle$ stand for the rank of a matrix and the number of nonzero entries in a matrix respectively. λ is denoted as a controlling parameter. However, the optimization goal is known as intractable which needs further process. In order to tackle this problem, we introduce the nuclear norm $\|\mathbf{X}\|_* = \sum_i \upsilon_i\left(\mathbf{X}\right)$ as the convex surrogate of $\mathrm{rank}\left(\mathbf{X}\right)$ by calculating the sum of the singular values, together with the l_1- norm $\|\mathbf{A}\|_1 = \sum_{h,m} |a_{h,m}|$ as the convex surrogate of $\left\langle \mathbf{A} \right\rangle$ [36]. Hence, Eq. (13) can be rewritten to a tractable convex optimization problem

$$\begin{aligned} \min_{\mathbf{X},\mathbf{A}} \quad & \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \\ \text{s.t.} \; & \|\mathbf{Y}-\mathbf{X}-\mathbf{A}\| \le \varepsilon \end{aligned}, \tag{14}$$

where ε is a noise related parameter. By denoting μ as a tuning value, Eq. (14) can be further written into following form

$$\min_{\mathbf{X},\mathbf{A}} \; \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \mu\|\mathbf{Y}-\mathbf{X}-\mathbf{A}\|^2. \tag{15}$$

Since that Eq. (15) is a convex optimization problem, an alternating direction method of multipliers (ADMM) algorithm is used for solving due to its higher accuracy with fewer iterations. We operate the ADMM method on the augmented Lagrangian form as follows

$$L\left(\mathbf{X},\mathbf{A},\mu\right) = \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \mu\|\mathbf{Y}-\mathbf{X}-\mathbf{A}\|^2. \tag{16}$$

By operating an iterative procedure until reaches a stop condition, we will finally obtain the optimization variables $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{A}}$.

Since that we do not analyze the source detection problem in this paper, an assumption is considered that the unknown source is accurately detected among all the channels and the serial number of the occupied channel is SN. Based on the cleansed sensing data matrix $\tilde{\mathbf{X}}$ and the abnormal data matrix $\tilde{\mathbf{A}}$, the cleansed source fingerprint vector is written as

$$\vec{F}_S = \left[P^{r,S}_{SN,1}, P^{r,S}_{SN,2}, P^{r,S}_{SN,3}, \cdots, P^{r,S}_{SN,M}\right]. \tag{17}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

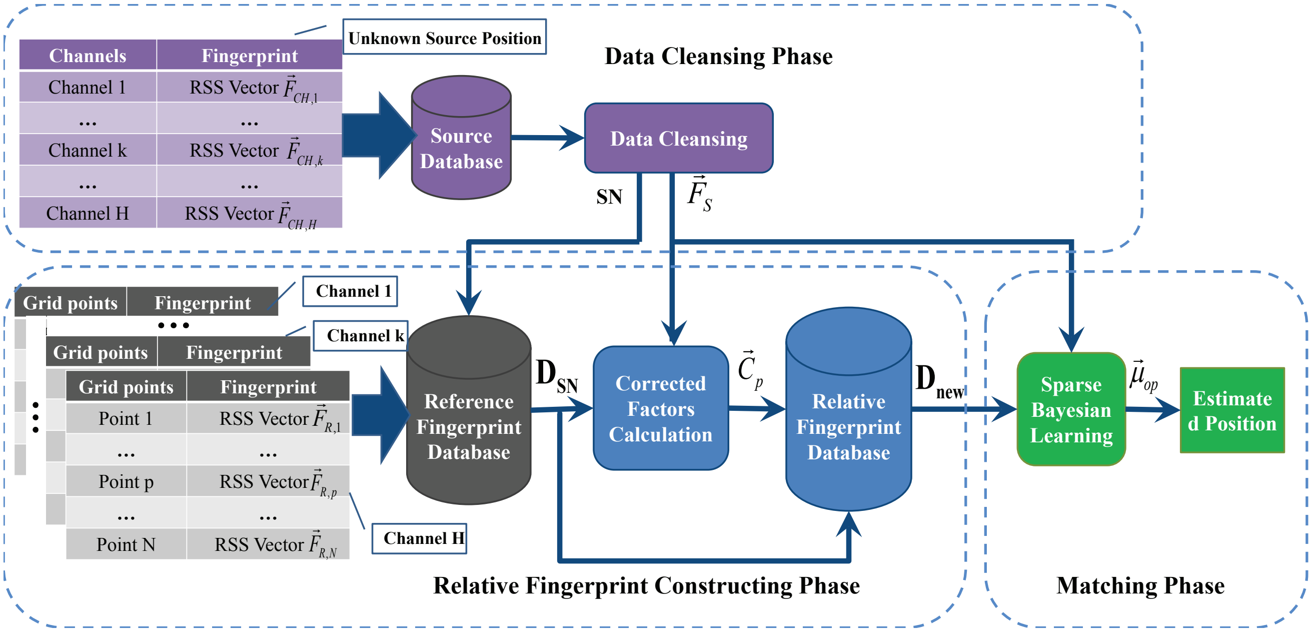


Fig. 4. Fundamental procedure of the proposed robust localization algorithm

C. Relative Fingerprint Constructing Phase

Under the fingerprinting based localization paradigm, one core task is to build the fingerprint database. However, the reference fingerprint database constituted by collected data cannot be directly used for passive localization, something change must be done in order to apply the fingerprint to the passive localization case. During this phase, we use the reference fingerprint database and the cleansed source fingerprint to calculate the corrected factors, by which the fingerprint database can be transformed into a new form, which is proved to be an effective way for the application of fingerprint in passive localization.

First, we randomly choose a known source as the reference node, without considering the specific transmit power. For channel k, an RSS vector data received by all the sensing nodes from the reference node is recorded in each grid point among the measurement area. Specifically, the fingerprint vector of a fixed grid point p can be written as

$$\vec{F}^k_{R,p} = \left[P^{r,R}_{p,1}, P^{r,R}_{p,2}, P^{r,R}_{p,3}, \cdots, P^{r,R}_{p,M}\right]. \tag{18}$$

After all the grids finished RSS vector data collection, the fingerprint database of channel k is

$$\mathbf{D}_k = \left[\vec{F}^k_{R,1}, \vec{F}^k_{R,2}, \vec{F}^k_{R,3}, \cdots, \vec{F}^k_{R,N}\right]^T. \tag{19}$$

Thus, the reference fingerprint database is comprised of all the channels' fingerprint database, which can be showed as $\{\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_H\}$.

When the channel that the source occupied is detected, the corresponding channel's fingerprint database is picked up for further process. For simplification, we write the specific fingerprint database as

$$\mathbf{D_{SN}} = \left[\vec{F}_{R,1}, \vec{F}_{R,2}, \vec{F}_{R,3}, \cdots, \vec{F}_{R,N}\right]^T. \tag{20}$$

After achieving the cleansed source fingerprint vector $\vec{F}_S$ as shown in Eq. (19), the mean value of the $\vec{F}_S$ is calculated

$$E\left(\vec{F}_S\right) = \left(\sum_{i=1}^{M} P_i^{r,S}\right) \Big/ M. \tag{21}$$

Similarly, for each grid point p, the fingerprint vector $\vec{F}_{R,p}$ of the fingerprint database $\mathbf{D_{SN}}$ can be find in Eq. (7). Hence, the mean value of specific reference fingerprint vector $\vec{F}_{R,p}$ can be calculated as

$$E\left(\vec{F}_{R,p}\right) = \left(\sum_{i=1}^{M} P_{p,i}^{r,R}\right) \Big/ M. \tag{22}$$

Here, we define C_p as the correction factor of the grid point p.

$$C_p = E\left(\vec{F}_S\right) - E\left(\vec{F}_{R,p}\right). \tag{23}$$

As is known that, for grid point p, we have $C_p^1 = C_p^2 = \cdots = C_p^M$. Hence, the corrected value vector of point p is

$$\vec{C}_p = \left[C_p^1, C_p^2, \cdots C_p^M\right]. \tag{24}$$

After calculations of the correction value in all grid points, we get a new relative fingerprint database when compared to the original one, which is written as

$$\begin{aligned} \mathbf{D}_{new} &= \left[\vec{F}_{R,1} - \vec{C}_1, \vec{F}_{R,2} - \vec{C}_2, \vec{F}_{R,3} - \vec{C}_3, \cdots, \vec{F}_{R,N} - \vec{C}_N\right]^T \\ &= \left[\vec{F}'_{R,1}, \vec{F}'_{R,2}, \vec{F}'_{R,3}, \cdots, \vec{F}'_{R,N}\right]^T \end{aligned}. \tag{25}$$

D. Matching Phase

When the relative fingerprint database and the fingerprint vector of the unknown source are achieved, then it can be a keyword to search the fingerprint database for position

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

estimation. The best matched grid point is viewed as the estimated position of the source. To find the best match, many matching metrics can be used in matching phase. In this paper, we introduce a new robust passive localization scheme that consider the sparse property of the unknown source and utilizes this information during the estimation of the source position. A critical insight regarding the position of the unknown source is that the associated location vector is inherently sparse when considering a discrete physical space.

Here we consider the sparse vector $\vec{u}$ as the unknown source position information, in which a non-zero component at p position indicates the position of unknown source at $p-th$ grid point. For instance, the vector of Eq. (26) means that the unknown source is located at the first grid point.

$$\vec{u} = [1, 0, 0, \cdots, 0]^T. \tag{26}$$

Since the new fingerprint database $\mathbf{D}_{new}$ is calculated in processing phase, the final data matching can be expressed as follows

$$\vec{F}_S = \mathbf{D}_{new}{}^T \vec{u} + \vec{\Upsilon}, \tag{27}$$

where $\vec{F}_S$ is the recorded fingerprint vector of the unknown source and the $\vec{\Upsilon}$ is the noise vector.

As can be seen in Eq. (27), the passive source localization problem then turns to be an accurate detection issue of the non-zero coefficient of the sparse vector $\vec{u}$.

Here we employ a sparse Bayesian learning with relevance vector machine (RVM) approach for the final location estimation. The Bayesian framework associated with RVM, given a dictionary-dependent sparsity penalty, presents invariance properties leading to accurate sparse signal estimation, especially for structured dictionaries [34]. Given the cleansed source fingerprint vector $\vec{F}_S$ and the new relative fingerprint database $\mathbf{D}_{new}$, the main goal is to formulate a posterior probability distribution for $\vec{u}$. The adopted probabilistic framework introduces a prior over the sparse vector by N independent, hyperparameters $\vec{b} = [b_1, b_2, b_3, \cdots, b_N]^T$. Each one of the hyperparameters is associated with corresponding position in the area of interest that mitigate the prior and individually controlling the strength of the prior over its associated weight.

Sparse Bayesian learning defines a zero-mean Gaussian prior with precision b_i on each element of the sparse vector $\vec{u}$:

$$\begin{aligned} p\left(\vec{u}\left|\vec{b}\right.\right) &= \prod_{i=1}^{N} \mathrm{N}\left(u_i \left|0, b_i^{-1}\right.\right) \\ &= (2\pi)^{-\frac{N}{2}} \prod_{i=1}^{N} b_i^{\frac{1}{2}} \exp\left(-\frac{b_i u_i^2}{2}\right) \end{aligned}. \tag{28}$$

Meanwhile, the noise vector of the Sparse Bayesian framework is modelled probabilistically as independent zero-mean Gaussian with variance σ^2.

$$p\left(\vec{\Upsilon}\right) = \prod_{i=1}^{N} \mathrm{N}\left(\Upsilon_i \left|0, \sigma^2\right.\right). \tag{29}$$

Based on Eq. (27), the source fingerprint vector $\vec{F}_S$ is modeled as

$$p\left(\vec{F}_S \left|\vec{u}, \sigma^2\right.\right) = \left(2\pi\sigma^2\right)^{-\frac{M}{2}} \exp\left(-\frac{\left\|\vec{F}_S - \mathbf{D}_{new}\vec{u}\right\|^2}{2\sigma^2}\right). \tag{30}$$

By combining the likelihood and prior within Bayes' rule, the posterior probability for the sparse vector $\vec{u}$ is defined as:

$$\begin{aligned} p\left(\vec{u}\left|\vec{F}_S, \vec{b}, \sigma^2\right.\right) &= \frac{p\left(\vec{F}_S|\vec{u},\sigma^2\right)p\left(\vec{u}|\vec{b}\right)}{p\left(\vec{F}_S|\vec{b},\sigma^2\right)} \\ &= (2\pi)^{-\frac{N}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\vec{u}-\vec{\mu})^T \mathbf{\Sigma}^{-1}(\vec{u}-\vec{\mu})\right) \end{aligned}, \tag{31}$$

where $|\cdot|$ is the determinant of a matrix. The mean $\vec{\mu}$ and the covariance matrix $\mathbf{\Sigma}$ are given as

$$\mathbf{\Sigma} = \left(\mathbf{V} + \sigma^{-2}\mathbf{D}_{new}{}^{\mathbf{T}}\mathbf{D}_{new}\right)^{-1}, \tag{32}$$

$$\vec{\mu} = \sigma^{-2}\mathbf{\Sigma}\mathbf{D}_{new}{}^{\mathbf{T}}\vec{F}_S, \tag{33}$$

and $\mathbf{V} = diag\,(b_1, b_2, b_3, \cdots, b_N)$. Hence, finding the sparse vector $\vec{u}$ then translates in estimating the unknown variables $\vec{\mu}$ and $\mathbf{\Sigma}$, which is final fall over the estimation of the hyperparameters $\vec{b}$ of the sparse vector $\vec{u}$.

The sparse Bayesian learning is formulated as the local maximisation with respect to the hyperparameters $\vec{b}$ of the marginal likelihood. It is an iterative process where each iteration estimates $\vec{b}$ and σ^2 that maximize the marginal likelihood.

$$\begin{aligned} \ell\left(\vec{b}\right) &= \log p\left(\vec{F}_S\left|\vec{b}, \sigma^2\right.\right) \\ &= -\tfrac{1}{2}\left[M\log 2\pi + \log|\mathbf{C}| + \vec{F}_S^T\mathbf{C}^{-1}\vec{F}_S\right] \end{aligned}, \tag{34}$$

with $\mathbf{C} = \sigma^2\mathbf{I} + \mathbf{D}_{new}\mathbf{V}^{-1}\mathbf{D}_{new}{}^{\mathbf{T}}$. After a number of iterations, a hyperparameter b_{op} remains relatively small indicating the non-zero components of the sparse vector $\vec{u}_{op}$. Consequently, the estimated location of the unknown source is the grid point op that corresponds to the maximum component of $\vec{u}_{op}$.

$$\eta = \underset{(\vec{b},\sigma^2)}{\arg\max}\,\vec{u} = \underset{(\vec{b},\sigma^2)}{\arg\max}\, p\left(\vec{F}_S\left|\vec{u}, \vec{b}, \sigma^2\right.\right). \tag{35}$$

E. Summary of the Robust Passive Localization Algorithm

Based on the above analysis, the overall procedure of the robust passive localization algorithm is summarized in Algorithm 1. Focusing on the abnormal data problem, the proposed algorithm performs a data cleansing approach on the received source sensing data, which can effectively eliminate the influence that the abnormal data brings. Focusing on the complicated environment problem, the proposed algorithm introduces a new relative fingerprinting based framework, which can transform the fingerprint database into a new form and realize effective application of the fingerprint to the passive localization. At last, the sparse Bayesian learning method is used in matching phase, which can use the sparse property of the source and raise the precision of the final position estimation effectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

Algorithm 1: Robust Relative Fingerprinting-Based Passive Localization via Data Cleansing

1 **Parameter setting:** Grid points M; Sensing nodes N; Detected channels H; Iterations NUM.
2 **Input:** Reference fingerprint database $\{\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_H\}$; Sensing data matrix of the channels $\mathbf{Y}$.
3 **Phase I: Data Cleansing Phase**
4 **for** $i = 1, 2, \cdots, NUM$ **do**
5 Do the singular value decomposition (svd) process $(\mathbf{P}, \mathbf{S}, \mathbf{Q}) = \mathrm{svd}\left(\mathbf{Y} - \tilde{\mathbf{A}}^{(i)}\right)$;
6 Update the cleansed data matrix $\tilde{\mathbf{X}}^{(i+1)} = \mathbf{P}\Psi_{\mu}(\mathbf{S})\,\mathbf{Q}$;
7 Update the abnormal data matrix $\tilde{\mathbf{A}}^{(i+1)} = \Psi_{\lambda\mu}\left(\mathbf{Y} - \tilde{\mathbf{X}}^{(i+1)}\right)$;
8 **return** $\tilde{\mathbf{X}}, \tilde{\mathbf{A}}$; //Cleansed data matrix and abnormal data matrix
9 **return** $\vec{F_S}$, SN ; //Cleansed source fingerprint vector and corresponding serial number
10 **Phase II: Relative Fingerprint Constructing Phase**
11 **for** $p = 1, 2, \cdots, M$ **do**
12 Calculate the mean value of $\vec{F_S}$ as $E\left(\vec{F_S}\right)$;
13 Calculate the mean value of the $p\mathrm{th}$ grid fingerprint $\vec{F}_{R,p}$ of $\mathbf{D}_{SN}$ as $E\left(\vec{F}_{R,p}\right)$;
14 Calculate the corrected factor $C_p = E\left(\vec{F_S}\right) - E\left(\vec{F}_{R,p}\right)$ and expand to the vector $\vec{C_p}$;
15 Calculate the new fingerprint database $\mathbf{D}_{new}$;
16 **return** $\mathbf{D}_{new}$; //Corrected reference fingerprint database
17 **Phase III: Matching Phase**
18 Determine prior distribution on each point $p\left(\vec{u}\middle|\vec{b}\right)$;
19 Estimate the hyperparameters $\vec{b}$ via maximizing the marginal likelihood $\ell\left(\vec{b}\right)$;
20 Estimate the posterior probability of the sparse vector $\vec{u}$ as $p\left(\vec{u}\middle|\vec{F_S}, \vec{b}, \sigma^2\right)$;
21 Estimate the final source location $\vec{s}$ via $\arg\max \vec{u}$;
22 **Output:** $\vec{s}$.

F. Complexity Analysis

In this section, the computational complexity of the proposed algorithm is analyzed. As can be seen in Algorithm 1, there are mainly three phases. Phase II only contains the basic mathematical operations, which holds the lowest computational cost. The main work of Phase I is computing SVD, which requires to compute those singular vectors of $\mathbf{Y} - \tilde{\mathbf{A}}^{(i)}$. However, since that the matrix $\mathbf{X}$ is low-rank, an approximate SVD can be used to avoid a significant computation task, by which computing several largest nonzero singular values in front. The dominant computational cost lies in Phase III, which is proportional to the number of partitioned grids. The reason is twofold. On one hand, more grid numbers will increase the iterations for finding the best $\vec{u}$ in line 20. On the other hand, large grid numbers also increase the computational cost of the marginal likelihood $\ell\left(\vec{b}\right)$ in line 19 for each iteration, which is directly related with the fingerprint database.

V. Performance Evaluation

A. Basic Simulation Setup

In the following simulations, a fundamental scene of a $20\mathrm{m} \times 20\mathrm{m}$ square measured area is considered, and eight sensing nodes are deployed at the known coordinates (0,0), (0,10), (0,20), (10,0), (10,20), (20,0), (20,10), (20,20) in meters. Also, the measured area is divided into 20×20 grids, hence that each grid is a $1\mathrm{m} \times 1\mathrm{m}$ square. The unknown source is distributed randomly in the measured area. Meanwhile, since that the occupied channel of the source is unknown, we have to scan much more channels to find which channel is used. We assume that there are 10 channels to be detected and the source is occupying any one of them. For signal propagation modeling, the source transmit power is $P_{t,S} = 0\mathrm{dBm}$ and the reference node transmit power is set to be $P_{t,R} = 30\mathrm{dBm}$. The pass loss exponent is $\gamma = 2$. The AWGN performs a Gaussian distribution with zero mean and variance $\sigma = 1$. For the abnormal state modeling, since the abnormal data is an i.i.d. random variable, we set the mean value of the abnormal data to be 10 and called abnormal data strength in the following analysis. The number of abnormal data referred to the specific source vector is 1, and is randomly distributed among all the eight data of the sensing nodes. We will detailed illustration of the abnormal data strength and the abnormal data number in the corresponding analysis.

B. Effectiveness Analysis of The Relative Fingerprint Framework

This subsection compares the effectiveness of the relative fingerprint framework among several circumstances, which is shown in Fig. 5. In this part, the cumulative density function (CDF) is used to show the performance analysis of each method. We randomly choose 100 points in the measured area as the source positions to be located. For each point, the distance between the real position and the estimated position is calculated, which called error distance. The CDF figure can give a direct outlook of the satisfied point rate under the specific error distance conditions. Also, several localization schemes under different configurations are used to demonstrate the effectiveness of the relative fingerprint framework.

- A baseline scheme called "No Framework + KNN" (N-F-K scheme), which means that the localization method only uses the K-nearest neighbor (KNN) matching method for the final position estimation.
- A baseline scheme called "No Framework + Sparse Learning" (N-F-SL scheme), which means that the localization method only uses the sparse Bayesian learning for the final position estimation.
- A baseline scheme called "Framework + KNN" (F-K scheme), which means that the localization method uses

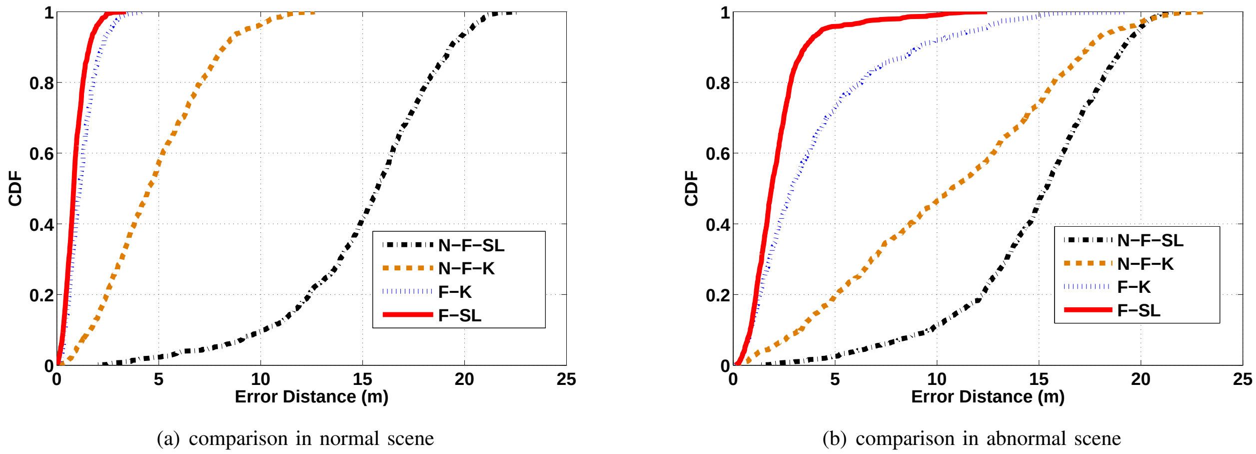


Fig. 5. Effect of the relative fingerprint framework in two scenes.

the relative fingerprint framework and the KNN matching method for the final position estimation.

- A baseline scheme called “Framework + Sparse Learning” (F-SL scheme), which means that the localization method uses the relative fingerprint framework and the sparse Bayesian learning for the final position estimation.

Figure 5 gives two sub-graphs which show the performance comparison of the four schemes in both the normal scene and the abnormal scene. The normal scene means that the received sensing data from the source are without any "polluted", while the abnormal scene means that the sensing data contains abnormal data. As can be seen from the two figures, the F-SL scheme and the F-K scheme which use the relative fingerprint framework can reach great localization performance when compared to the N-F-SL scheme and the N-F-K scheme. For the reason, the relative fingerprint framework can eliminate the average difference between the source fingerprint and the reference fingerprint and then transform the passive localization problem to an ordinary active case. Hence, the relative fingerprint framework can help improving the passive localization performance greatly. On the other hand, when comparing the F-SL scheme with the F-K scheme, it is clearly that the proposed sparse Bayesian learning is more effective than the widely used KNN method in matching phase. In the following simulations, more detailed analysis will be given to show the effectiveness of our proposed algorithm which uses both the relative fingerprint framework and the sparse Bayesian learning method.

C. *Effectiveness Analysis of The Proposed Passive Localization Method*

In this part, we use the root mean square error (RMSE) criterion to measure the position estimation performance, which can clearly present the performance variation under different parameters constraint. The RMSE is written as

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\|\tilde{s}_i - s_i\|^2}. \tag{36}$$

where $\tilde{\vec{s}}$ is the estimated position and the $\vec{s}$ is the real position of the unknown source. K is the simulation times.

Also, in order to show the effectiveness of the proposed robust localization algorithm, several localization schemes are also performed for comparison. Since that the above subsection has analyzed the importance of the relative fingerprint framework, in the following we configure all the schemes with the specific framework and mainly analyze the effectiveness of the data cleansing work and the sparse Bayesian learning process. The schemes are listed as follows.

- A baseline scheme called “Abnormal" (AB scheme), which means that the localization scheme directly use the sensing data without any pretreatment and then perform the position estimation by the sparse Bayesian learning method.
- A baseline scheme called “Abnormal + KNN Cleansing” (AB-K-C scheme), which means that the localization scheme uses the original abnormal data without any pretreatment and then uses the data cleansing method together with the KNN matching method for the final position estimation.
- A baseline scheme called “Abnormal + Proposed Cleansing” (AB-P-C scheme), which means that the localization scheme uses the original abnormal data without any pretreatment and then uses the data cleansing method together with the sparse Bayesian learning method for the final position estimation.
- A baseline scheme called “Abnormal + Perfect Cleansing” (AB-PER scheme), which means that the localization scheme uses the sensing data which are assumed to be perfected cleansed and then uses the sparse Bayesian learning method for the matching process.
- A baseline scheme called “Normal” (NL scheme), which means that the localization scheme use the sensing data that do not have any abnormal data and then perform the position estimation by the sparse Bayesian learning method.

The source localization performance of the aforementioned schemes is compared for different abnormal data strength in

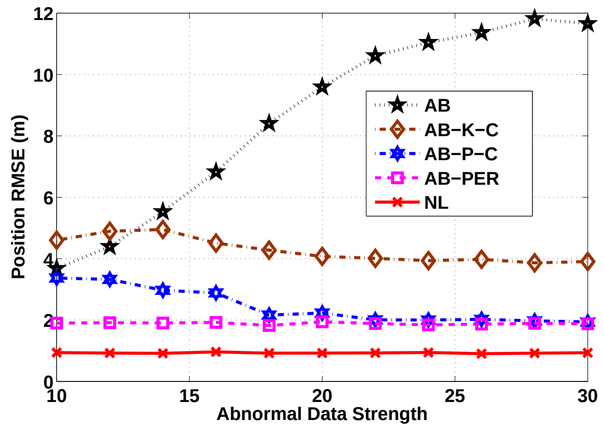


Fig. 6. RMSE of the source localization versus abnormal strength for various schemes.

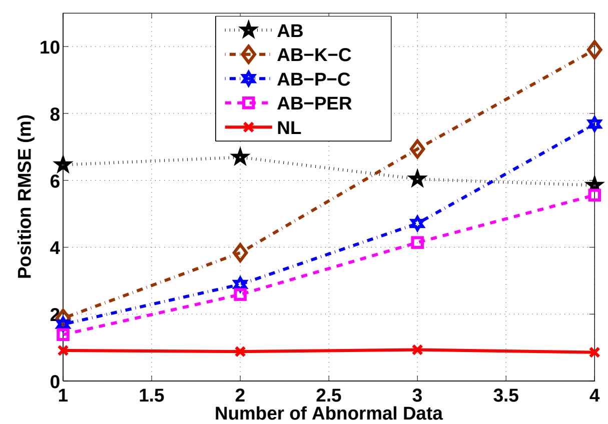


Fig. 7. RMSE of the source localization versus number of abnormal data for various schemes.

Fig. 6. Since the abnormal data is modeled as a Gaussian random variable, the real strength of the abnormal data cannot be precisely figured out. Hence, we use the mean value of the abnormal data as the abnormal data strength for the analysis. As is shown in the figure, the RMSE of the AB scheme is increasing along with the larger of the abnormal data strength. The reason is that the abnormal data with larger strength will affect the localization precision more heavily. However, by using the data cleansing method to filtering the abnormal data, the three schemes (AB-K-C scheme, AB-P-C scheme and AB-PER scheme) all follow a monotonically decreasing trend as the abnormal data strength gets larger. This is because that the larger strength can make it more easily to find out the abnormal data, hence the data cleansing work are more precise, which put a positive effect on the final source localization. On the other hand, when the strength is very weak, the influence caused by the sensing data error cleansed may be more serious than the abnormal data itself brings. This is the reason why the AB scheme can performs better than the AB-K-C scheme on the condition that the abnormal data strength is weak enough. The AB-PER scheme holds an assumption that the abnormal data are perfectly cleansed, which can get a better performance when compared to the AB-K-C scheme and AB-P-C scheme, whose sensing data after cleansing method may still exist errors. When compared the AB-K-C scheme with the AB-P-C scheme, it is clear that our proposed sparse Bayesian learning method in matching phase can reach a higher precision of the final estimation.

The effect of the abnormal data number on the RMSE of the source position estimation is given in Fig. 7. In our work, only the abnormal data in the source fingerprint $\vec{F_S}$ may affect the final position estimation. Hence, the abnormal data number in Fig. 7 means the number of abnormal data in the source fingerprint. In this simulation, we use 12 sensing nodes to perform the localization process and the abnormal data number set from 1to 4. For example, when the abnormal data number is 2, that is to say any two sensing nodes out of the whole 12 nodes are recording a wrong data, thus the final received source fingerprint contains 10 normal data and 2 abnormal data. From Fig. 7 we can see that as the number of abnormal data increases, the position RMSE of the AB-K-C scheme and the AB-P-C scheme are all goes larger, which means a decrease on the performance of localization. The reason can be seen that too many abnormal data will definitely raise the difficulties of the data cleansing process, and the precision of the cleansed data will be reduced greatly. The more the abnormal data exists, the lower precision the data cleansing work gets, and finally the worse performance of the source localization achieves. Meanwhile, although the AB-PER scheme can cleanse the abnormal data perfectly, too many abnormal data will leave less normal data to be used for localization, which can also decrease the final source localization performance. Also, we should note that when the number of abnormal data goes very large, the efficiency of data cleansing process gets very worse, together with the fewer existed data to be used for final localization, the negative influence they brings has already surpass the effect of the abnormal data (For example, when the number is 4, the RMSE of the AB scheme is smaller than that of the AB-K-C scheme and the AB-P-C scheme). An interesting phenomenon in Fig. 7 appears that when the number of abnormal data grows, the position RMSE of the AB scheme turns to be slightly smaller. A rational explanation may be that too much numbers can eliminate the sparse property of abnormal data, which will weaken the bad influence of the abnormal data.

Fig. 8 shows the The effect of sensing node number on the RMSE of the source position estimation for various schemes. From the figure we can find that as the number of sensing node increase, all the schemes performs a monotonically decreasing trend, which means that the localization performance are getting improved gradually. It is because that increasing the sensing nodes will enlarge the sensing data matrix, which can weaken the influence of the noise and further raise the localization precision. Especially, enlarging the sensing data matrix can help improving the efficiency of data cleansing method, which has stronger effect on the KNN matching method otherwise. Thus, when the number of sensing node gets smaller, the difference of the AB-K-C scheme, AB-P-C

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2817576, IEEE Access

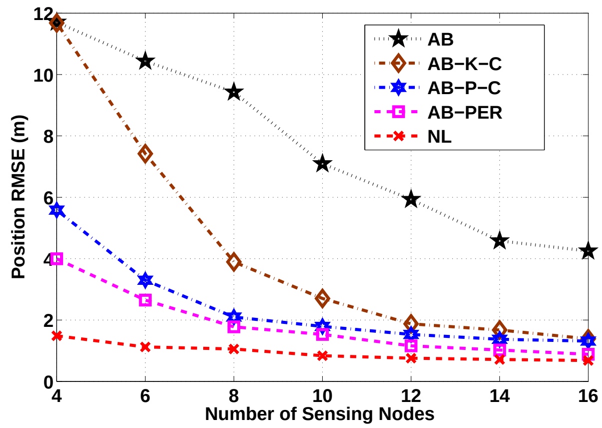


Fig. 8. RMSE of the source localization versus number of sensing nodes for various schemes.

scheme and the AB-PER scheme becomes more significantly.

VI. CONCLUSION

In this paper, we propose a novel robust relative fingerprinting-based passive localization algorithm via a data cleansing approach. Firstly, a new relative fingerprint framework is introduced. The key idea is that by exploring the correlations between the source fingerprint and the fingerprint database, the fingerprint idea can be applied into the passive localization case appropriately. Then, a generalized modeling of the abnormal data in localization problem is formulated. Based on this, a data cleansing approach is proposed to eliminate the abnormal data impairment. Finally, the sparse Bayesian learning is used for the purpose of achieving more precise estimated source position. Simulation results demonstrate that the proposed algorithm achieves a higher accuracy performance in passive localization in terms of eliminating the abnormal data impairment.

APPENDIX A
CRAMER-RAO LOWER BOUND ANALYSIS

In this part, we analyze the CRLB of the localization estimation based on the abnormal data model. The CRLB defines a lower bound on the variance of the unbiased estimator and is employed as a benchmark for evaluating the performance of the estimation. As is analyzed before, the abnormal data can be in various forms and is modeled as a Gaussian random variable in this paper. Hence, the abnormal data occurred in the $i-\mathrm{th}$ sensing node is $a_i \sim N\left(\mu_{a,i}, \sigma_{a,i}^2\right)$. The Fisher Information Matrix (FIM) of the measurement model is written as

$$\operatorname{cov}\left(\vec{\delta}\right) \geq \mathbf{J}^{-1}, \tag{37}$$

where $\vec{\delta}$ is the unknown vector to be estimated, which is written as $\vec{\delta} = [s_x, s_y]$. $\operatorname{cov}()$ is the covariance matrix of the unknown vector $\vec{\delta}$ and $\mathbf{J}$ is the FIM.

From the abnormal data model in Eq. (2), the joint probability density function of the measurement data can be written as

$$g\left(\vec{P}\,\middle|\,\vec{\delta}\right) = (2\pi)^{-\frac{M}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\vec{P}-\vec{\mu}\right)^T \mathbf{C}^{-1}\left(\vec{P}-\vec{\mu}\right)\right), \tag{38}$$

where $\vec{P} = [P_1^r, P_1^r, \cdots, P_M^r]$ is the measured signal strength, $\mathbf{C}$ is the covariance matrix and $\vec{\mu}$ is the mean of the measurement vector $\vec{P}$.

Since that the measurement data are independently recorded by the sensing nodes, $\mathbf{C}$ is a diagonal matrix and the element of the matrix is

$$C_{ii} = \sigma_W^2 + \sigma_{a,i}^2. \quad i = 1, 2, \cdots, M \tag{39}$$

The element in the vector $\vec{\mu}$ is

$$\mu_i = P_i^o + \mu_{a,i}. \quad i = 1, 2, \cdots M \tag{40}$$

Furthermore, we transform the joint pdf into the logarithm form

$$\ln\left(g\left(\vec{P}\,\middle|\,\vec{\delta}\right)\right) = K_C - \frac{1}{2}\left(\vec{P}-\vec{\mu}\right)^T \mathbf{C}^{-1}\left(\vec{P}-\vec{\mu}\right), \tag{41}$$

where $K_C = -\frac{1}{2}\ln\left[(2\pi)^M |\mathbf{C}|\right]$ is a constant that does not depend on the unknown value $\vec{\delta}$.

Thus, the FIM is calculated as follows

$$\mathrm{J}\left(\vec{\delta}\right) = -E\left[\frac{\partial^2 \ln g\left(\vec{P}\,\middle|\,\vec{\delta}\right)}{\partial\vec{\delta}\,\partial\vec{\delta}^T}\right] = \left(\frac{\partial\vec{\mu}}{\partial\vec{\delta}}\right)^T \mathbf{C}^{-1}\left(\frac{\partial\vec{\mu}}{\partial\vec{\delta}}\right). \tag{42}$$

For detailed,

$$\frac{\partial\vec{\mu}}{\partial\vec{\delta}} = \left[\left(\frac{\partial\mu_1}{\partial s_x}\ \frac{\partial\mu_1}{\partial s_y}\right), \cdots \left(\frac{\partial\mu_i}{\partial s_x}\ \frac{\partial\mu_i}{\partial s_y}\right), \cdots \left(\frac{\partial\mu_M}{\partial s_x}\ \frac{\partial\mu_M}{\partial s_y}\right)\right]^T, \tag{43}$$

$$\frac{\partial\mu_i}{\partial s_x} = \frac{10\gamma}{\ln 10}\frac{t_{ix}-s_x}{d_i^2}, \tag{44}$$

$$\frac{\partial\mu_i}{\partial s_y} = \frac{10\gamma}{\ln 10}\frac{t_{iy}-s_y}{d_i^2}, \tag{45}$$

where $[t_{ix}\ t_{iy}]$ is the coordinate of the $i-\mathrm{th}$ sensing node and d_i is the distance of the source from $i-\mathrm{th}$ sensing node.

As can be seen in the above, CRLB is actually the inverse of the FIM. Once the FIM $\mathbf{J}$ is figured out, the CRLB of the unknown vector $\vec{\delta}$ can be calculated as $\mathbf{J}^{-1}$.

REFERENCES

[1] M. C. R. Talampas and K.-S. Low, "A Geometric Filter Algorithm for Robust Device-Free Localization in Wireless Networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp.1670-1678, Oct. 2016.

[2] F. Zhao, H. Luo, X. Zhao, Z. Pang, and H. Park, "HYFI: Hybrid Floor Identification Based on Wireless Fingerprinting and Barometric Pressure," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 1067-1080, Feb. 2017.

[3] J. Wang, Q. Gao, Y. Yu, X. Zhang, and X. Feng, "Time and Energy Efficient TOF-Based Device-Free Wireless Localization," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 158-168, Feb. 2016.

[4] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 5, no. 24, pp. 175ĺC183, Oct. 2017.

[5] T. V. Haute, B. Verbeke, E. D. Poorter, and I. Moerman, "Optimizing Time of Arrival Localization Solutions for Challenging Industrial Environments," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3 , pp. 1430-1439, Jun. 2017.

[6] W.-A. Zhang, X. Yang, L. Yu, and S. Liu, "Sequential Fusion Estimation for RSS-based Mobile Robots Localization with Event-driven WSNs," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4 , pp. 1519-1528, Aug. 2016.

[7] G. Ding, J. Wang, Q. Wu, L. Zhang, Y. Zou, Y.-D. Yao, and Y. Chen, "Robust spectrum sensing with crowd sensors," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3129-3143, 2014.

[8] R. Li, Z. Zhao, Y. Zhang, J. Palicot, and H. Zhang, "Adaptive multi-task compressive sensing for localization in wireless LANs," *IET Commun.*, vol. 8, no. 10, pp. 1736ÍC1744, Jul. 2014.

[9] G. Wang and K. Yang, "A new approach to sensor node localization using RSS measurements in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1389-1395, May 2011.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] B. P. B. Elwischger and T. Sauter, "Efficient Ambiguity Resolution in Wireless Localization Systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 888-897, Apr. 2017.

[15] H. Shen, Z. Ding, S. Dasgupta, and C. Zhao, "Multiple source localization in wireless sensor networks based on time of arrival measurement," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1938-1949, Apr. 2014.

[16] O. Jean and A. J. Weiss, "Passive localization and synchronization using arbitrary signals," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2143-2150, Apr. 2014.

[17] B. Huang, L. Xie, and Z. Yang, "TDOA-based source localization with distance-dependent noises," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 468-480, Jan. 2015.

[18] N. Wu, W. Yuan, H. Wang, and J. Kuang, "TOA-based Passive Localization of Multiple Targets with Inaccurate Receivers Based on Belief Propagation on Factor Graph," *Digital Signal Processing*, vol. 49, pp. 14-23, 2016.

[19] I. T. Haque and C. Assi, "Profiling-based indoor localization schemes," *IEEE Systems J.*, vol. 9, no. 1, pp. 76-85, Mar. 2015.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] M. Zhou, Y. Tang, Z. Tian, L. Xie, and W. Nie, "Robust Neighborhood Graphing for Semi-supervised Indoor Localization with Light-loaded Location Fingerprinting," *IEEE Internet Things J.* 2017, vol: PP, no. 99, pp. 1-1, Nov. 2017

[26] 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.

[27] 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.

[28] S. Zhang, S. Gao, G. Wang, and Y. Li, "Robust NLOS error mitigation method for TOA-based localization via second-order cone relaxation," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2210-2213, Dec. 2015.

[29] G. Wang, H. Chen, Y. Li, and N. Ansari, "NLOS error mitigation for TOA-based localization via convex relaxation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4119-4131, Aug. 2014.

[30] G. Wang, A. M.-C. So, and Y. Li, "Robust convex approximation methods for TDOA-based localization under NLOS conditions," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3281-3296, July 2016.

[31] B. Mager, P. Lundrigan, and N. Patwari, "Fingerprint-Based Device-Free Localization Performance in Changing Environments," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2429-2438, Nov. 2015.

[32] M. B. Kjrgaard and C. V. Munk, "Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength," *in Proc. IEEE PERCOM*, pp. 110-116, Nov. 2008.

[33] A. K. M. M. Hossain, Y. Jin, W. S. Soh, and H. N. Van, "SSD: A robust RF location fingerprint addressing mobile devices' heterogeneity," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 65-77, Jan. 2013.

[34] 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.

[35] C. Kan, G. Ding, Q. Wu, and T. Zhang, "Robust Localization with Crowd Sensors: A Data Cleansing Approach," *Mobile Networks and Applications*, vol. 22, no. 115, pp. 1-11, May 2017.

[36] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. MA, "Stable principal component pursuit," *in Proc. Intl. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 1518ÍC1522.

[37] Z. Tian, and G. B. Giannakis, "A Wavelet Approach to Wideband Spectrum Sensing for Cognitive Radios," *Cognitive Radio Oriented Wireless Networks and Communications, 2006. 1st International Conference on*, Jun. 2006

[38] H. Sun, A. Nallanathan, C. X. Wang , and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wilreless Commun.*, vol. 20, no. 2, pp. 74-81, Apr. 2013.

[39] C. P. Yen, Y. Tsai, and X. Wang, "Wideband Spectrum Sensing Based on Sub-Nyquist Sampling," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3028-3040, Jun. 2013.

[40] 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, future directions," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 126-136, Jul. 2013.

[41] A. W. Min, K. G. Shin, and X. Hu, "Secure cooperative sensing in IEEE 802.22 WRANs using shadow fading correlation," *IEEE Trans. Mobile Comput.*, vol. 10, no. 10, pp. 1434-1447, Oct. 2011.

[42] A. Attar, H. Tang, A. V. Vasilakos, F. R. Yu, and V. C. M. Leung, "A survey of security challenges in cognitive radio networks: Solutions and future research directions," *Proc. IEEE*, vol. 100, no. 12, pp. 3172-3186, Dec. 2012.

[43] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 428-445, 2013.