Modified MM Algorithm and Bayesian Expectation Maximization-Based Robust Localization Under NLOS Contaminated Environments

CHEE-HYUN PARK[ID] AND JOON-HYUK CHANG[ID], (Senior Member, IEEE)

School of Electronic Engineering, Hanyang University, Seoul 133-791, South Korea

Corresponding author: Joon-Hyuk Chang (jchang@hanyang.ac.kr)

ABSTRACT Robust localization methods that employ distance measurements to predict the position of an emitter are proposed in this paper. The occurrence of outliers due to the non-line-of sight (NLOS) propagation of signals can drastically degrade the localization performance in crowded urban areas and indoor situations. Hence, robust positioning methods are considered to mitigate the effects of outliers. Specifically, localization methods based on robust statistics are considered. Modified multi-stage ML-type method (MM) based weighted least squares (WLS), maximum *a posteriori* (MAP) expectation maximization (EM) WLS and variational Bayes (VB) EM WLS algorithms are developed under various outlier-contaminated environments. Simulation results show that the position estimation accuracy of the proposed modified MM WLS method, which uses the novel weight, is higher than that of the other methods under most outlier-contaminated conditions. Furthermore, the MAP-EM WLS and VB-EM WLS methods are the most accurate among algorithms that do not require statistical testing. Additionally, the mean square error (MSE) and asymptotic unbiasedness of the proposed algorithms are analyzed.

INDEX TERMS Expectation maximization, localization, maximum a posteriori, multi-stage maximum likelihood-type (MM) estimator, robust, variational Bayes, weighted least squares.

I. INTRODUCTION

Emitter positioning involves predicting the coordinates of a point emitter using the observations from each receiver. It is of considerable interest in diverse areas of study, such as telecommunication, signal processing and remote sensing. Emitter positioning algorithms are classified into direct and indirect localization methods. The direct localization method estimates the location information directly from a signal model based on the grid search method. In contrast, the indirect localization method estimates the source position using information such as the time difference of arrival (TDOA), time of arrival (TOA), received signal strength indicator (RSSI) and angle of arrival (AOA). The direct localization method has better accuracy in the low signal-to-noise ratio (SNR) condition, but incurs a higher computational

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan[ID].

burden compared with that of the indirect algorithm. The indirect positioning algorithm is examined in this study. The localization methods based on TOA, TDOA and AOA require expensive and energy-consuming hardware. The TDOA and TOA-based positioning methods also require synchronization between sensors or between sensors and the emitter. However, the TOA-based positioning method shows the best location performance among the aforementioned algorithms. The localization systems using RSSI can be implemented at a reasonable price, but their localization accuracy using the RSSI is lower than that of the techniques that use the TOA, TDOA and AOA. Moreover, this technique suffers from problems such as model-mismatch, multipath effects and background interferences, which render the distance observation inaccurate. Therefore, the TOA is utilized as the measurement in this paper. Previous studies have extensively investigated the localization problems in the line-of-sight (LOS) context [1]–[3]. However, certain problems remain to be solved.

A critical theme among position prediction problems is the localization in mixed LOS/non-line-of-sight (NLOS) environments. For example, the LOS path between a source and sensors may be blocked in urban or indoor situations. Robust statistics are required because the conventional localization methods based on a sample mean are quite vulnerable to outliers in mixed LOS/NLOS environments. The outlier is an observation that is distant from most gathered observations.

In general, studies on position prediction for the mixed LOS/NLOS problems are classified into three main areas: 1) robust statistics, 2) LOS and LOS/NLOS mixture sensor identification and 3) mathematical programming. Some examples of localization methods based on robust statistics are the M and least median squares (LMedS) algorithms [4]–[9]. The M estimator exploits a robust cost function to minimize the effects of outliers [4], [5] and the LMedS estimator yields the smallest median of squared residuals for the candidate estimates [6], [7]. The identification and discard method identifies the outlier-contaminated sensors and discards them prior to localization [10], [11]. Quadratic programming, linear programming and interior point optimization are mathematical programming methods [12]–[14], wherein, some constraints associated with NLOS measurements are employed. Some papers deal with the TOA localization under asynchronous receivers formulation in wireless networks using the expectation maximization (EM) and factor graph. In this study, we focus on robust-statistics-based localization.

The MM estimator has been a popular method because it attenuates the adverse influences of NLOS observation [4], [5]. The existing MM estimation-based localization algorithm uses a transformed range observation to obtain the position estimate and the weight to reflect the different accuracy of each sensor [15]. The transformed observation and transformed distance estimate based on the MM algorithm are assumed statistically independent in the existing MM-WLS algorithm, but this is not the case. Therefore, novel weights are derived using the fact that the transformed range measurements and transformed distance estimate based on MM estimate are correlated. From now on, the modified MM weighted least squares (WLS) estimation is abbreviated as the MM WLS-M algorithm. Moreover, the MM WLS-M method requires statistical testing to discriminate the inliers and outliers. M denotes the modified. This necessitates the determination of an optimal threshold for discriminating inliers and outliers. However, this is a non-trivial task under rapidly varying environments or channels, as the threshold needs to be determined considering the changing conditions. Thus, maximum *a posteriori* (MAP) EM and variational Bayesian (VB) EM WLS localization without statistical testing are proposed to avoid this limitation. Statistical testing is not required as the parameters, such as mean and variance, are estimated from the outlier-contaminated observations. In the remainder of this paper, the MAP EM and VB EM combined with the WLS method are denoted as the MAP-EM WLS and VB-EM WLS algorithms, respectively. We utilize the EM

algorithm because it has been widely used in the estimation of parameters of mixture model. The MAP EM method is the Bayesian version of EM algorithm in which the posterior is maximized and can outperform the standard EM algorithm in the low SNR or small sample size condition. The VB learning method has the advantages compared to the MAP algorithm. The posterior to be maximized is difficult to be known in certain case. In this situation, the posterior can be estimated using the mean-field theorem of the VB EM algorithm. Also, the VB learning is less prone to the overfitting than MAP estimation [16]. The essential of the VB EM algorithm is that the posterior can be represented as the multiplication of the variational distributions. Therefore, the performance of the VB EM algorithm may be degraded if the variational distribution is far from the true posterior. Although the Monte-Carlo-based algorithms such as Gibbs sampling and Metropolis-Hastings algorithms can be used for estimating parameters, the proposed MAP-EM and VB-EM algorithms are computationally efficient than the Monte-Carlo methods. Also, the convergence of the algorithm can be confirmed more easily than the Monte-Carlo methods. The novelty and main contributions of this study are summarized as follows:

- We develop a robust MM WLS-M localization method in which the weight calculation is different from that of existing MM WLS localization algorithm. Namely, the statistical dependence between transformed range measurements and the MM estimate are considered in the proposed method for derivation of the weight. In general, the weight is determined as the variance of response variables. As the weight is more accurate, the accuracy of the estimation algorithm is better.

- The existing statistical testing method requires an optimal threshold for discriminating the inliers and outliers. However, this is not an easy task under rapidly changing environments or channels, as the threshold should be predetermined according to the changing conditions. To address these drawbacks, we propose a localization technique that does not require hypothesis testing and an empirical threshold.

- To the best of the authors' knowledge, the MM estimation-based localization using the novel weight proposed in this paper has not been investigated yet. Moreover, the combination of Bayesian EM and WLS methods has not been reported yet.

The remainder of this paper is organized as follows. Section II describes the mixed LOS/NLOS position determination problem covered in this paper. Section III describes the existing estimation algorithms. Section IV describes the proposed robust MM WLS-M, MAP-EM WLS and VB-EM WLS positioning algorithms. Section V analyses the mean square error (MSE) and asymptotic unbiasedness of the proposed localization algorithms. Section VI demonstrates the superiority of the proposed techniques over the conventional algorithms via simulation results. Finally, Section VII presents our conclusions and future work.

II. PROBLEM FORMULATION

The localization method based on distance measurements aims to determine the location of an emitter accurately, where the error risk function, e.g., the MSE, should be minimized. In the problem of mixed LOS/NLOS emitter localization, the observation equation is expressed as

$$
\begin{aligned}
r_{i,j} &= d_i + n_{i,j} \\
&= \sqrt{(x - x_i)^2 + (y - y_i)^2} + n_{i,j},
\end{aligned} \tag{1}
$$

where $n_{i,j}$ is a Gaussian mixture random variable defined as $(1 - v_{i,2})\mathcal{N}(0, \sigma_{i,1}^2) + v_{i,2}\mathcal{N}(\mu_{i,2}, \sigma_{i,2}^2)$, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, P$ with M and P denoting the number of receivers and observations in each receiver, respectively. The inliers follow $\mathcal{N}(0, \sigma_{i,1}^2)$ and the outliers follow $\mathcal{N}(\mu_{i,2}, \sigma_{i,2}^2)$ ($\mathcal{N}(\mu_{i,l}, \sigma_{i,l}^2)$ denotes a Gaussian probability density function (PDF) with mean $\mu_{i,l}$ and variance $\sigma_{i,l}^2$ ($l=1, 2$)). The inliers have a probability of $1 - v_{i,2}$ and the outliers have a probability of $v_{i,2}$. The Gaussian mixture noise distribution with two-mode has been utilized extensively for modeling mixed LOS/NLOS situations [17]–[21]. More specifically, the authors of [17], [19] obtained the experimental result that the mixed LOS/NLOS noise distribution is consistent with two-mode mixed Gaussian distribution. It is assumed that the statistics such as the mean and variance of the LOS and NLOS noise distributions cannot be obtained. Here, $v_{i,2} \in [0, 1]$ represents a contamination ratio, which is commonly lower than 0.1 [8]. Some papers also consider the exponential distribution for modeling the NLOS bias. However, the applicability of that model in the proposed algorithm is out of scope because it requires the thorough investigation. Moreover, $[x \ y]^T$ denotes the target location to be estimated and $[x_i \ y_i]^T$ represents the known coordinates of the ith receiver. Additionally, $r_{i,j}$ is the range observation from the emitter to the ith receiver at the jth time instance and d_i is the true distance from the source to the ith receiver. Squaring (1) and rearranging the expressions yield the following equation:

$$
\begin{aligned}
x_i x + y_i y &- 0.5R + m_{i,j} \\
&= 0.5(x_i^2 + y_i^2 - r_{i,j}^2), \\
&\quad i = 1, 2, \ldots, M, \quad j = 1, 2, \ldots, P
\end{aligned} \tag{2}
$$

where $R = x^2 + y^2$, $m_{i,j} = -d_i n_{i,j} - \frac{1}{2}n_{i,j}^2$. We can obtain the following equation by expressing (2) in a matrix form

$$
\mathbf{A}\mathbf{x} + \mathbf{m}_j = \mathbf{b}_j, \qquad j = 1, \cdots, P \tag{3}
$$

where $\mathbf{m}_j = [m_{1,j}, \cdots, m_{M,j}]^T$, $\mathbf{x} = [x \ y \ R]^T$,

$$
\mathbf{A} = \begin{pmatrix} x_1 & y_1 & -0.5 \\ \vdots & \vdots & \vdots \\ x_M & y_M & -0.5 \end{pmatrix} \tag{4}
$$

and $\mathbf{b}_j = [b_{1,j} \cdots b_{M,j}]^T = \frac{1}{2}\begin{pmatrix} x_1^2 + y_1^2 - r_{1,j}^2 \\ \vdots \\ x_M^2 + y_M^2 - r_{M,j}^2 \end{pmatrix}$. (5)

The objective is to determine the parameter $\mathbf{x}$ by fusing the transformed distance measurements optimally,

i.e., $[\mathbf{b}_1, \cdots, \mathbf{b}_P]$. The first-step solution is obtained using the aforementioned equations. Furthermore, the second-step procedure is employed to enhance the accuracy of the first-step solution. In this paper, a vector is represented by a lowercase boldface letter and a matrix by an uppercase boldface letter. The operator $[\cdot]^T$ denotes a transpose.

III. REVIEW OF THE CONVENTIONAL APPROACHES UNDER GAUSSIAN MIXTURE MODELS

The MM estimator has been employed in the presence of outliers. The rationale of this algorithm is to mitigate the adverse effects of outliers by utilizing a weight function, that is inversely proportional to the squared residual. The mixture EM method has been used to estimate the parameters (mixing coefficient, mean and variance) of the mixture model. Moreover, these parameters have been used for robust positioning in the presence of outliers [17], [21]. The lower bound of the marginal likelihood function is maximized when the marginal likelihood function is difficult to be directly maximized. The mixture EM method can be extended to the MAP-EM and VB-EM algorithms using a prior distribution. The principle of the MAP-EM and VB-EM algorithms is to estimate the responsibility of the observation via conditional expectation.

A. MM ESTIMATOR [8], [9]

The MM estimation involves the calculation of the scale parameter based on the S estimation, followed by location estimation using the M algorithm. The details of MM estimator can be found in [8], [9], [15].

B. MAP-EM ESTIMATOR [22]

The EM method is a recursive method to find the local maximum of a complete likelihood function, where the model depends on unobserved latent variables [23], [24]. Also, EM method can be employed for finding the parameters maximizing the log-posterior function and it is termed as MAP-EM technique in this study. The MAP-EM method is divided into an expectation (E) step and a maximization (M) step. The expectation of the log-posterior evaluated using the current estimate is determined in the E-step and the parameters that maximize the log-posterior function are determined in the M-step. The objective of the MAP-EM algorithm is to find the MAP estimate of the posterior function using the following two-step.

E-step:

$$
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E_{\mathbf{z}|\mathbf{r}, \boldsymbol{\theta}^{(k)}}[\ln p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r})], \tag{6}
$$

where $\ln(\cdot)$ denotes the natural logarithm. In the M-step, the expected log-posterior at the current estimate is maximized as follows:

M-step:

$$
\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}). \tag{7}
$$

C. VB-EM ESTIMATOR [16], [22], [25], [26]

The VB-EM algorithm estimates the random parameter $\boldsymbol{\theta}$ and the latent variable $\mathbf{z}$ of a model, given the observation

r. The Bayes' rule can be applied for inferring the posterior as follows:

$$p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r}) = \frac{p(\mathbf{r}|\boldsymbol{\theta}, \mathbf{z})p(\boldsymbol{\theta}, \mathbf{z})}{p(\mathbf{r})}. \tag{8}$$

The evidence $p(\mathbf{r})$ is determined to be

$$p(\mathbf{r}) = \int_{\boldsymbol{\theta}, \mathbf{z}} p(\mathbf{r}, \boldsymbol{\theta}, \mathbf{z}) d\boldsymbol{\theta} d\mathbf{z}. \tag{9}$$

The main limitation of the Bayesian estimation is that the above integral is intractable, except for simple models. The variational algorithm is designed to detour infeasible integrals, approximating true posterior distribution $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r})$ using the variational posterior distribution $q(\boldsymbol{\theta}, \mathbf{z})$. The Kullback-Leibler (KL) divergence is used for measuring the discrepancy between the true and variational posterior distributions, and is defined as

$$\begin{aligned} D_{\mathrm{KL}}(q||p) &= \int_{\boldsymbol{\theta}, \mathbf{z}} q(\boldsymbol{\theta}, \mathbf{z}) \ln \frac{q(\boldsymbol{\theta}, \mathbf{z})}{p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r})} d\boldsymbol{\theta} d\mathbf{z} \\ &= E_q \left\{ \ln \frac{q(\boldsymbol{\theta}, \mathbf{z})}{p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r})} \right\} \end{aligned} \tag{10}$$

where $E_q\{\cdot\}$ denotes the expectation over distribution q. The marginal distribution $\ln p(\mathbf{r})$ can be rewritten from (10) as follows:

$$\ln p(\mathbf{r}) = B(q) + D_{KL}(q||p) \tag{11}$$

where $B(q) = \int q(\boldsymbol{\theta}, \mathbf{z}) \ln \{\frac{p(\mathbf{r}, \boldsymbol{\theta}, \mathbf{z})}{q(\boldsymbol{\theta}, \mathbf{z})}\} d\boldsymbol{\theta} d\mathbf{z}$ and $D_{KL}(q||p) = -\int q(\boldsymbol{\theta}, \mathbf{z}) \ln \{\frac{p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{r})}{q(\boldsymbol{\theta}, \mathbf{z})}\} d\boldsymbol{\theta} d\mathbf{z}$. $B(q)$ becomes the lower bound of the marginal distribution because $D_{KL}(q||p) \geq 0$ and $B(q)$ is maximized for estimating $\boldsymbol{\theta}$, which is equivalent to minimizing the KL divergence. The variational posterior $q(\boldsymbol{\theta}, \mathbf{z})$ can be factorized as $q(\boldsymbol{\theta})q(\mathbf{z})$ in the variational Bayesian inference. The variational Bayesian EM method alternates between E-step and M-step, i.e., the optimal $q(\mathbf{z})$ is determined by assuming that $q(\boldsymbol{\theta})$ is fixed and the optimal $q(\boldsymbol{\theta})$ is determined with $q(\mathbf{z})$ fixed. The solutions are represented as follows:

$$q(\mathbf{z}) \propto \exp(E_{q(\boldsymbol{\theta})}\{\ln p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{r})\}) \tag{12}$$

$$q(\boldsymbol{\theta}) \propto \exp(E_{q(\mathbf{z})}\{\ln p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{r})\}). \tag{13}$$

IV. PROPOSED MM WLS-M, MAP-EM, VB-EM WLS LOCALIZATION METHODS

In this section, the MM WLS-M method is described in detail. Moreover, the MAP-EM and VB-EM algorithms are combined with the WLS method.

A. MM WLS-M ALGORITHM

A detailed explanation of the modified MM estimation-based WLS localization algorithm is presented in this subsection. The $b_{i,1:P} = [b_{i,1} \cdots b_{i,P}]^T$ is the transformed distance set and merged employing the MM estimation as follows:

$$\hat{b}_i^{MM} = \frac{\sum_{j=1}^{P} w_{i,j}^{MM} b_{i,j}}{\sum_{j=1}^{P} w_{i,j}^{MM}} \tag{14}$$

where $w_{i,j}^{MM} = \begin{cases} [1 - (\frac{b_{i,j}-\hat{b}_i^{MM}}{4.685\,\hat{\sigma}_i})^2]^2 & \text{if} |\frac{b_{i,j}-\hat{b}_i^{MM}}{\hat{\sigma}_i}| \leq 4.685; \\ 0, & \text{if} |\frac{b_{i,j}-\hat{b}_i^{MM}}{\hat{\sigma}_i}| > 4.685. \end{cases}$

$\hat{b}_i^{MM}$ is the estimate of true transformed distance that combines $b_{i,1:P}$ and $\hat{\sigma}_i$ is the S-estimate for the ith sensor. The initial point of $\hat{b}_i^{MM}$ is selected as the sample median of $b_{i,1:P}$ [8], [9]. Mostly, it can not be known whether the measurement is an inlier or outlier; thus, the inliers and outliers should be discerned utilizing the statistical testing given in the below of (14). The estimate of true transformed distance $\hat{b}_i^{MM}$ is then found recursively since both sides of (14) contain unknown $\hat{b}_i^{MM}$. The variance of the statistic $\hat{b}_i^{MM}$ is then calculated based on algebraic methods as follows:

$$\mathrm{Var}[\hat{b}_i^{MM}] = \frac{\sum_{q'_i s} (w_{i,q_i}^{MM})^2 \mathrm{Var}[b_{i,q_i}]}{(\sum_{q'_i s} w_{i,q_i}^{MM})^2}, \tag{15}$$

where $\mathrm{Var}[b_{i,q_i}] = \sum_{q'_i s} (b_{i,q_i} - me_{i1})^2 / Q_i$, q'_is are sample indices, which are determined as inliers in the ith sensor based on the statistical testing appeared in the below of (14), and Q_i is the total number of measurements decided to be inliers. In addition, $me_{i1} = \frac{1}{Q_i} \sum_{q_i=1}^{Q_i} b_{i,q_i}$ in the ith sensor. The output of this algorithm is given by:

$$y_{i,j} = \begin{cases} b_{i,j}, & |\frac{b_{i,j}-\hat{b}_i^{MM}}{\hat{\sigma}_i}| \leq 4.685; \\ \hat{b}_i^{MM}, & |\frac{b_{i,j}-\hat{b}_i^{MM}}{\hat{\sigma}_i}| > 4.685. \end{cases} \tag{16}$$

The $y_{i,1:P}$ are averaged using the sample mean to increase the accuracy, i.e., $\hat{b}_i = \frac{\sum_{j=1}^{P} y_{i,j}}{P}$. The variance of the statistic $\hat{b}_i$ was determined using algebraic methods in existing method as follows [15]:

$$\mathrm{Var}[\hat{b}_i] = \frac{\frac{S_i}{Q_i} \times Q_i + \mathrm{Var}[\hat{b}_i^{MM}] \times R_i}{P^2}, \tag{17}$$

where $S_i = \sum_{q'_i s} (y_{i,q_i} - me_{i2})^2$, R_i is the number of samples determined as outliers in the ith sensor, $P = (Q_i + R_i)$ is the total number of samples in the ith sensor, $me_{i2} = \frac{1}{Q_i} \sum_{q_i=1}^{Q_i} y_{i,q_i}$ and $\mathrm{Var}[\hat{b}_i^{MM}]$ is defined in (15). The statistical independence between b_{i,q_i} and $\hat{b}_i^{MM}$ is assumed in existing MM WLS algorithm. However, this is not true because $\hat{b}_i^{MM}$ includes b_{i,q_i}. Therefore, the variance calculation should be modified using the property that b_{i,q_i} and $\hat{b}_i^{MM}$ are correlated as follows:

$$\begin{aligned} \mathrm{Var}[\hat{b}_i] &= \mathrm{Var}\left[\frac{\sum_{q'_i s} b_{i,q_i} + (\frac{\sum_{k=1}^{P} w_{i,k} b_{i,k}}{\sum_{k=1}^{P} w_{i,k}})R_i}{P}\right] \\ &= \mathrm{Var}\left[\frac{\sum_{q'_i s}(1 + \frac{R_i \cdot w_{i,q_i}}{\sum_{k=1}^{P} w_{i,k}})b_{i,q_i} + \frac{R_i \sum_{r'_i s} w_{i,r_i} b_{i,r_i}}{\sum_{k=1}^{P} w_{i,k}}}{P}\right] \\ &= \mathrm{Var}\left[\frac{\sum_{q'_i s}(1 + \frac{R_i \cdot w_{i,q_i}}{\sum_{k=1}^{P} w_{i,k}})b_{i,q_i}}{P}\right] \\ &= \frac{[\sum_{q'_i s}(1 + \frac{R_i \cdot w_{i,q_i}}{\sum_{k=1}^{P} w_{i,k}})]^2 \mathrm{Var}[b_{i,q_i}]}{P^2} \end{aligned} \tag{18}$$

where $q_i's$ and $r_i's$ are the LOS sensor index set and LOS/NLOS sensor index set, respectively. In (18), the third line is derived using the property that the weights that belong to the outlier set are zeros. Then, the transformed measurement equation is determined as follows:

$$\mathbf{Ax} + \boldsymbol{\varepsilon} = \hat{\mathbf{b}}^{\mathbf{f}}, \quad (19)$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \cdots, \varepsilon_M]^T$, $\hat{\mathbf{b}}^{\mathbf{f}} = [\hat{b}_1, \cdots, \hat{b}_M]^T$, the superscript f is the acronym of the first-step estimate and $\boldsymbol{\varepsilon}$ is the additive noise of $\hat{\mathbf{b}}^{\mathbf{f}}$. The first-step WLS estimate based on the M estimation is determined as follows:

$$\hat{\mathbf{x}}_{\mathbf{f}} = (\mathbf{A}^T \mathbf{W}^{\mathbf{f}} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{\mathbf{f}} \hat{\mathbf{b}}^{\mathbf{f}} \quad (20)$$

where $\mathbf{W}^{\mathbf{f}} = \mathbf{C}_{\hat{\mathbf{b}}^{\mathbf{f}}}^{-1} = (\text{diag}[\text{Var}\{\hat{b}_1\} \cdots \text{Var}\{\hat{b}_M\}])^{-1}$. Furthermore, the estimation performance of the first-step estimate is enhanced utilizing the second-step approach [2], [3]. Because the two-step localization method is well known, its explanation is omitted. For more details of the two-step localization method, it can be referred to [2], [3].

B. MAP-EM WLS ALGORITHM
An EM estimate of the parameters of mixture distribution can be obtained by maximizing the complete data log-likelihood, i.e., the lower bound of the marginal likelihood function. The objective is the maximization of the marginal likelihood function with respect to the parameters of the mixture model, which is obtained as follows:

$$\mathcal{L}(\boldsymbol{\theta}_i; \mathbf{r}_i) = \ln p(\mathbf{r}_i; \boldsymbol{\theta}_i)$$
$$= \sum_{j=1}^{P} \ln \left(\sum_{l=1}^{2} v_{i,l} \, \mathcal{N}(r_{i,j} - d_i; \boldsymbol{\delta}_i) \right), \quad i = 1, \cdots, M,$$
$$(21)$$

where $\mathbf{r}_i = [r_{i,1}, \cdots, r_{i,P}]^T$, $\boldsymbol{\theta}_i = [\boldsymbol{v}_i, \boldsymbol{\mu}_i, d_i, \boldsymbol{\lambda}_i]^T$, $\boldsymbol{\delta}_i = [\boldsymbol{\mu}_i, d_i, \boldsymbol{\lambda}_i]^T$, $\boldsymbol{v}_i = [v_{i,1}, v_{i,2}]^T$, $\boldsymbol{\mu}_i = [\mu_{i,1}, \mu_{i,2}]^T$ and $\boldsymbol{\lambda}_i = [\lambda_{i,1}, \lambda_{i,2}]^T$. However, the optimization is difficult because the summation exists in the inner part of the logarithm. Instead of solving (21) directly, we maximize the complete log-likelihood function in which the latent variable, $z_{j,l}$ ($j = 1, \cdots, P, l = 1, 2$), is appended to the incomplete range observation. $z_{j,l}$ indicates the mixture component responsible for the corresponding distance observation. Moreover, the EM method can be used for optimizing the complete log-posterior and the conditional expectation for the complete log-posterior of the ith sensor is expressed as

$$l(\boldsymbol{\theta}_i; \boldsymbol{\theta}_i^{(k)}) = E_{\mathbf{z}_i|\mathbf{r}_i, \theta_i^{(k)}} \left[\ln p(\boldsymbol{\theta}_i, \mathbf{z}_i|\mathbf{r}_i) \right]$$
$$= E_{\mathbf{z}_i|\mathbf{r}_i, \theta_i^{(k)}} \left[\ln \left\{ \prod_{j=1}^{P} \sum_{l=1}^{2} p(r_{i,j}|z_{j,l}, \boldsymbol{\theta}_i) \right. \right.$$
$$\left. \left. \times p(z_{j,l}|\boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i) \right\} \right], \quad i = 1, \cdots, M, \quad (22)$$

where $\mathbf{z}_i$ is a vector enclosing latent variables $z_{j,l}$ and superscript (k) denotes the kth iteration-step. We assume that the

prior of the mixing coefficient, $\boldsymbol{v}_i$, is modeled as the Dirichlet distribution with the parameter $\boldsymbol{\alpha}_i$, where $\boldsymbol{\alpha}_i$ consists of $\alpha_{i,l} = \alpha_{i,0} + N_{i,l}$ ($i = 1, \cdots, M$, $l = 1, 2$), $N_{i,l} = \sum_{j=1}^{P} \gamma_{j,l}$, $\gamma_{j,l} = p(z_{j,l}|r_{i,j}, \boldsymbol{\theta}_i^{(k)})$ and $\alpha_{i,0}$ is the hyper-parameter. The prior mean (μ_p) is modeled as a Gaussian distribution with zero mean and the precision prior (λ_p) is modeled as a gamma distribution with the parameter (a_0, b_0). Then, (22) can be further derived as follows:

$$l(\boldsymbol{\theta}_i; \boldsymbol{\theta}_i^{(k)})$$
$$= E_{\mathbf{z}_i|\mathbf{r}_i, \theta_i^{(k)}} \left[\ln p(\boldsymbol{\theta}_i, \mathbf{z}_i|\mathbf{r}_i) \right]$$
$$= E_{\mathbf{z}_i|\mathbf{r}_i, \theta_i^{(k)}} \left[\sum_{j=1}^{P} \left\{ \ln v_{i,1} - \frac{1}{2}\ln 2\pi + \ln \lambda_{i,1} \right.\right.$$
$$- \frac{\lambda_{i,1}}{2}(r_{i,j} - d_i - \mu_{i,1})^2 + \ln v_{i,2} - \frac{1}{2}\ln 2\pi$$
$$\left.\left. + \ln \lambda_{i,2} - \frac{\lambda_{i,2}}{2}(r_{i,j} - d_i - \mu_{i,2})^2 \right\} \right]$$
$$+ \ln p(v_{i,1}) + \ln p(v_{i,2}) + \ln p(\mu_{i,1}) + \ln p(\mu_{i,2})$$
$$+ \ln p(\lambda_{i,1}) + \ln p(\lambda_{i,2})$$
$$= \sum_{j=1}^{P} (\ln v_{i,1} - \frac{1}{2}\ln 2\pi - \ln \lambda_{i,1}$$
$$- \frac{\lambda_{i,1}}{2}(r_{i,j} - d_i - \mu_{i,1})^2) p(z_{j,1}|\mathbf{r}_i, \boldsymbol{\theta}_i^{(k)})$$
$$+ (\ln v_{i,2} - \frac{1}{2}\ln 2\pi + \ln \lambda_{i,2}$$
$$- \frac{\lambda_{i,2}}{2}(r_{i,j} - d_i - \mu_{i,2})^2) p(z_{j,2}|\mathbf{r}_i, \boldsymbol{\theta}_i^{(k)})$$
$$- \ln B(\alpha_{i,1}) + (\alpha_{i,0} - 1)\ln v_{i,1} - \ln B(\alpha_{i,2})$$
$$+ (\alpha_{i,0} - 1)\ln v_{i,2}$$
$$- \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln \lambda_p - \frac{\lambda_p \mu_p^2}{2} + a_0 \ln b_0$$
$$+ (a_0 - 1)\ln b_0 + (a_0 - 1)\ln \lambda_p - b_0 \lambda_p - \ln \Gamma(a_0) \quad (23)$$

where $B(\alpha_{i,l}) = \frac{\Gamma(\alpha_{i,l})^2}{\Gamma(2\alpha_{i,l})}$ and $\Gamma(\cdot)$ denotes the gamma function. Rearranging (23) with respect to $\boldsymbol{v}_i$ yields (24):

$$E_{\mathbf{z}_i|\mathbf{r}_i, \theta^{(k)}}[\ln p(\boldsymbol{\theta}_i, \mathbf{z}_i|\mathbf{r}_i)]$$
$$= \text{const} + \sum_{j=1}^{P} \{(\ln v_{i,1})\gamma_{j,1} + (\ln v_{i,2})\gamma_{j,2}\} + (\alpha_{i,1} - 1)\ln v_{i,1}$$
$$+ (\alpha_{i,2} - 1)\ln v_{i,2} \quad (24)$$

where const represents the irrelevant terms of $\boldsymbol{v}_i = \{v_{i,1}, v_{i,2}\}$. Differentiating (24) with respect to $v_{i,l}$ and setting it to zero yield (25):

$$\frac{\partial l(\boldsymbol{\theta}_i; \boldsymbol{\theta}_i^{(k)})}{\partial v_{i,l}} = \sum_{j=1}^{P} \frac{\gamma_{j,l}}{v_{i,l}} + \frac{(\alpha_0 - 1)}{v_{i,l}} + \beta = 0, \quad (25)$$

where β represents the Lagrange multiplier and the constraint $\sum_{l=1}^{2} v_{i,l} = 1$ is used. Then, $v_{i,l}$ is obtained by solving (25)

as follows:

$$v_{i,l}^{(k+1)} = \frac{\sum_{j=1}^{P} \gamma_{j,l}^{(k)} + \alpha_0 - 1}{P + 2\alpha_0 - 2}. \tag{26}$$

Furthermore, $\mu_{i,l}^{(k+1)}$, $d_i^{(k+1)}$ and $\lambda_{i,l}^{(k+1)}$ can be obtained similarly as follows:

$$\mu_{i,l}^{(k+1)} = \frac{\sum_{j=1}^{P} \gamma_{j,l}^{(k)}(r_{i,j} - d_i^{(k)})}{\sum_{j=1}^{P} \gamma_{j,l}^{(k)} + 1} \tag{27}$$

$$d_i^{(k+1)} = \frac{\sum_{j=1}^{P} \{\gamma_{j,1}^{(k)}\lambda_{i,1}^{(k)}(r_{i,j} - \mu_{i,1}^{(k)}) + \gamma_{j,2}^{(k)}\lambda_{i,2}^{(k)}(r_{i,j} - \mu_{i,2}^{(k)})\} + \lambda_p\mu_p}{\sum_{j=1}^{P}\{\gamma_{j,1}^{(k)}\lambda_{i,1}^{(k)} + \gamma_{j,2}^{(k)}\lambda_{i,2}^{(k)}\} + \lambda_p} \tag{28}$$

$$\lambda_{i,l}^{(k+1)} = \frac{\sum_{j=1}^{P} \gamma_{j,l}^{(k)} - 2}{\sum_{j=1}^{P}(r_{i,j} - d_i^{(k)} - \mu_{i,l}^{(k)})^2 \gamma_{j,l}^{(k)}} \tag{29}$$

In the following, the superscript (k) is omitted for brevity of notation. Subsequently, the WLS method is derived using the mean and covariance of $\mathbf{r}_i = [r_{i,1} \cdots r_{i,P}]^T$ $(i = 1, \cdots, M)$. The mean of $\mathbf{r}_i$ is represented as follows:

$$\mu_{\mathbf{r}_i} = v_{i,1}(d_i + \mu_{i,1}) + v_{i,2}(d_i + \mu_{i,2}). \tag{30}$$

Also, the variance of $\mathbf{r}_i$ is expressed as follows:

$$\sigma_{\mathbf{r}_i}^2 = v_{i,1}(\sigma_{i,1}^2 + (d_i + \mu_{i,1})^2 - \mu_{\mathbf{r}_i}^2)$$
$$+ v_{i,2}(\sigma_{i,2}^2 + (d_i + \mu_{i,2})^2 - \mu_{\mathbf{r}_i}^2). \tag{31}$$

Subsequently, the mean and variance of the transformed observation are obtained as follows:

$$\mu_{\mathbf{b}_i} = \frac{x_i^2 + y_i^2 - \mu_{\mathbf{r}_i}^2}{2} \tag{32}$$

$$\sigma_{\mathbf{b}_i}^2 \simeq \mu_{\mathbf{r}_i}^2 \sigma_{\mathbf{r}_i}^2 \tag{33}$$

where $\mathbf{b}_i = [b_{i,1}, \cdots, b_{i,P}]^T$. It should be noticed that the mixture EM algorithm has not been combined with the WLS method in existing literatures. The MAP-EM WLS method does not require statistical testing because the mean and variance of the observations are calculated with the outliers included. This property is also applied to the VB-EM WLS method introduced in the following section. The transformed measurement equation for the MAP-EM WLS method is expressed as follows:

$$\mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon} = \mathbf{b}^{\mathrm{m}}, \tag{34}$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \cdots, \varepsilon_M]^T$, $\mathbf{b}^{\mathrm{m}} = [\mu_{\mathbf{b}_1}, \cdots, \mu_{\mathbf{b}_M}]^T$, the superscript m is the abbreviation of the mean and $\boldsymbol{\varepsilon}$ is the noise component of $\mathbf{b}^{\mathrm{m}}$. Then, the first-step WLS estimate is determined as follows:

$$\hat{\mathbf{x}}_{\mathrm{f}} = (\mathbf{A}^T \mathbf{W}^{\mathrm{m}} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{\mathrm{m}} \mathbf{b}^{\mathrm{m}} \tag{35}$$

where $\mathbf{W}^{\mathrm{m}} = \mathbf{C}_{\mathbf{b}^{\mathrm{m}}}^{-1} = (\mathrm{diag}[\sigma_{\mathbf{b}_1}^2 \cdots \sigma_{\mathbf{b}_M}^2])^{-1}$. Furthermore, the accuracy of the first-step estimate can be improved using the second-step approach and it is identical to that of the MM WLS-M algorithm.

C. VB-EM WLS ALGORITHM

The VB-EM-based method has been used for approximating the posterior probability using unobserved latent variables given some observed data. The approximated distribution is restricted to belong to an exponential family to make the approximated distribution similar to the true posterior. The joint distribution of $\mathbf{r}_i$, $\boldsymbol{\theta}_i = [\boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i]^T$ and $\mathbf{z}_i$ is expressed as follows:

$$p(\mathbf{r}_i, \mathbf{z}_i, \boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i) = p(\mathbf{r}_i|\mathbf{z}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i)p(\mathbf{z}_i|\boldsymbol{v}_i)p(\boldsymbol{v}_i)p(\boldsymbol{\mu}_i|\boldsymbol{\lambda}_i)p(\boldsymbol{\lambda}_i). \tag{36}$$

The variational distribution is factorized with respect to the latent variables and the parameters so that

$$q(\mathbf{z}_i, \boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i) = q(\mathbf{z}_i)q(\boldsymbol{v}_i)q(\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i). \tag{37}$$

First, the variational distribution of $\mathbf{z}_i$ is derived using the mean field theory [25], [26] as follows:

$$q(\mathbf{z}_i)$$
$$= \exp\left\{E_{\boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i}\left[\ln\left(\prod_{j=1}^{P}\sum_{l=1}^{2} p(r_{i,j}|z_{j,l}, \boldsymbol{\theta}_i)p(z_{j,l}|\boldsymbol{\theta}_i)\right)\right]\right\}$$
$$= \exp\left\{E_{\boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i}\left[\sum_{j=1}^{P}\left\{\ln\left(v_{i,1}\frac{\sqrt{\lambda_{i,1}}}{\sqrt{2\pi}}\right.\right.\right.$$
$$\times\exp\left(-\frac{\lambda_{i,1}(r_{i,j} - d_i - \mu_{i,1})^2}{2}\right)$$
$$\left.\left.\left. + v_{i,2}\frac{\sqrt{\lambda_{i,2}}}{\sqrt{2\pi}}\exp\left(-\frac{\lambda_{i,2}(r_{i,j} - d_i - \mu_{i,2})^2}{2}\right)\right)\right\}\right]\right\}$$
$$= \exp\left\{E_{\boldsymbol{v}_i, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i}\left[\sum_{j=1}^{P}\left\{\ln v_{i,1} - \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_{i,1}\right.\right.$$
$$- \frac{\lambda_{i,1}(r_{i,j} - d_i - \mu_{i,1})^2}{2} + \ln v_{i,2} - \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_{i,2}$$
$$\left.\left. - \frac{\lambda_{i,2}(r_{i,j} - d_i - \mu_{i,2})^2}{2}\right\}\right]\right\}. \tag{38}$$

Absorbing any terms independent of $\mathbf{z}_i$ into the constant yields:

$$q(\mathbf{z}_i) = \exp\left[\sum_{j=1}^{P}\sum_{l=1}^{2} z_{j,l}\ln\rho_{j,l} + \mathrm{const}\right] \tag{39}$$

where $\rho_{j,l} = \exp\left\{E[\ln v_{i,l}] + \frac{1}{2}E[\ln\lambda_{i,l}] - \frac{1}{2}\ln 2\pi - \frac{1}{2}E_{\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i}\left[\frac{\lambda_{i,l}(r_{i,j} - d_i - \mu_{i,l})^2}{2}\right]\right\}$, $E[\ln v_{i,l}] = \psi(\alpha_{i,l}) - \psi(\sum_{k=1}^{2}\alpha_{i,k})$, $\psi(\cdot)$ is the digamma function, $E[\ln\lambda_{i,l}] = \psi(\frac{v_{i,l}}{2}) + \ln 2 + \ln\lambda_{i,l}$, $E_{\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i}\left[\frac{\lambda_{i,l}(r_{i,j} - d_i - \mu_{i,l})^2}{2}\right] = \beta_{i,l}^{-1} +$

$v_{i,l}(r_{i,j} - d_i - \mu_{i,l})^2$, $\beta_{i,l} = \beta_{i,0} + \sum_{j=1}^P \rho_{j,l}$, $v_{i,l} = v_{i,0} + \sum_{j=1}^P \rho_{j,l}$, $\beta_{i,0}$ and $v_{i,0}$ are hyper-parameters and const is the term independent of $\mathbf{z}_i$. Then, $q(\mathbf{z}_i)$ can be expressed as follows:

$$q(\mathbf{z}_i) = \prod_{j=1}^P \prod_{l=1}^2 u_{j,l}^{z_{j,l}}, \quad i = 1, \cdots, M, \tag{40}$$

where $u_{j,l} = \frac{\rho_{j,l}}{\sum_{k=1}^2 \rho_{j,k}}$.

Second, the variational distribution of $\mathbf{v}_i$ is derived as follows:

$$
\begin{aligned}
&q(\mathbf{v}_i) \\
&= \exp\left\{ E_{\mathbf{z}_i, \mu_i, \lambda_i} \left[\ln \left(\prod_{j=1}^P \sum_{l=1}^2 p(r_{i,j}|z_{j,l}, \boldsymbol{\theta}_i) p(z_{j,l}|\boldsymbol{\theta}_i) \right) \right] \right\} \\
&= \exp\left\{ \sum_{j=1}^P \sum_{l=1}^2 u_{j,l} \ln v_{j,l} + (\alpha_{i,0} - 1) \sum_{l=1}^2 \ln v_{j,l} + \text{const} \right\}
\end{aligned}
\tag{41}
$$

where const is the irrelevant term of $\mathbf{v}_i$.

Finally, the variational posterior distribution $q(\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i)$ are obtained similarly using the above derivation as follows:

$$q(\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i) = \mathcal{N}(\boldsymbol{\mu}_i | \boldsymbol{\mu}_v, \text{diag}[\boldsymbol{\lambda}_v]) \, \text{Gamma}(\mathbf{a}_i, \mathbf{b}_i) \tag{42}$$

where $\boldsymbol{\mu}_v = [\frac{\sum_{j=1}^P u_{j,1}(r_{i,j}-d_i)}{\sum_{j=1}^P u_{j,1}}, \frac{\sum_{j=1}^P u_{j,2}(r_{i,j}-d_i)}{\sum_{j=1}^P u_{j,2}}]^T$, $\boldsymbol{\lambda}_v = [\frac{\sum_{j=1}^P u_{j,1}(r_{i,j}-d_i-\mu_{i,1})^2}{\sum_{j=1}^P u_{j,1}}, \frac{\sum_{j=1}^P u_{j,2}(r_{i,j}-d_i-\mu_{i,2})^2}{\sum_{j=1}^P u_{j,2}}]^T$, Gamma denotes the gamma distribution, $\mathbf{a}_i = [\frac{1}{2}, \frac{1}{2}]^T$, $\mathbf{b}_i = [\mathbf{b}_i(1) \ \mathbf{b}_i(2)]^T$, $\mathbf{b}_i(1) = \frac{1}{2} \frac{\sum_{j=1}^P u_{j,1}(r_{i,j}-d_i-\mu_{i,1})^2}{\sum_{j=1}^P u_{j,1}}$, and $\mathbf{b}_i(2) = \frac{1}{2} \frac{\sum_{j=1}^P u_{j,2}(r_{i,j}-d_i-\mu_{i,2})^2}{\sum_{j=1}^P u_{j,2}}$. The hyper-parameters of the gamma distribution were set to zero without loss of generality in the derivation of $q(\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i)$. The variational posterior means of $v_{i,l}$, $\mu_{i,l}$, $\lambda_{i,l}$ and d_i are obtained as follows:

$$v_{i,l}^{(k+1)} = \frac{\sum_{j=1}^P u_{j,l}^{(k)}}{P} \tag{43}$$

$$\mu_{i,l}^{(k+1)} = \frac{\sum_{j=1}^P u_{j,l}^{(k)}(r_{i,j} - d_i^{(k)})}{\sum_{j=1}^P u_{j,l}^{(k)}} \tag{44}$$

$$u_{i,l}^{(k+1)} = \frac{\sum_{j=1}^P u_{j,l}^{(k)}}{\sum_{j=1}^P (r_{i,j} - d_i^{(k)} - \mu_{i,l}^{(k)})^2 u_{j,l}^{(k)}}. \tag{45}$$

The range is estimated by maximizing the complete log-likelihood (22) as follows:

$$d_i^{(k+1)} = \frac{\sum_{j=1}^P \{u_{j,1}^{(k)}\lambda_{i,1}^{(k)}(r_{i,j} - \mu_{i,1}^{(k)}) + u_{j,2}^{(k)}\lambda_{i,2}^{(k)}(r_{i,j} - \mu_{i,2}^{(k)})\}}{\sum_{j=1}^P \{u_{j,1}^{(k)}\lambda_{i,1}^{(k)} + u_{j,2}^{(k)}\lambda_{i,2}^{(k)}\}} \tag{46}$$

Then, the mean and variance of the transformed observations are obtained using (30)−(33). The remaining procedures are identical to those of the MAP-EM WLS algorithm.

V. MSE AND ASYMPTOTIC UNBIASEDNESS ANALYSIS

This section analyzes the MSE and asymptotic unbiasedness of the proposed methods.

A. MSE PERFORMANCE ANALYSIS

The MSE of the estimator is the sum of the squared bias and error variance. The root MSE (RMSE) can be obtained as the square root of the MSE. The estimation error $\Delta\widehat{\mathbf{x}}_e$ can be expressed as

$$
\begin{aligned}
\Delta\widehat{\mathbf{x}}_e &= \mathbf{D}_2^{-1}\Delta\widehat{\mathbf{x}}_s \\
&= \mathbf{D}_2^{-1}(\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}(\widehat{\mathbf{h}} - \mathbf{H}\mathbf{x}_s) \\
&= \mathbf{D}_2^{-1}(\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}\mathbf{D}_1(\widehat{\mathbf{x}}_f - \mathbf{x}_f) \\
&= \mathbf{G}(\mathbf{A}^T\mathbf{C}_{\widehat{\mathbf{b}}^{MM}}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_{\widehat{\mathbf{b}}^{MM}}^{-1}(\widehat{\mathbf{b}}^{MM} - \mathbf{A}\mathbf{x}_f)
\end{aligned}
\tag{47}
$$

where $\mathbf{D}_1 = \text{diag}[2x \ 2y \ 1]$, $\mathbf{D}_2 = 2\text{diag}[x \ \ y]$, $\mathbf{G} = \mathbf{D}_2^{-1}(\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\widehat{\mathbf{h}}}^{-1}\mathbf{D}_1$ and $\mathbf{x}_f$, $\mathbf{x}_s$ are the true values of $\widehat{\mathbf{x}}_f$, $\widehat{\mathbf{x}}_s$. The error covariance matrix of $\widehat{\mathbf{x}}_e$ can be expressed as follows:

$$\text{cov}[\Delta\widehat{\mathbf{x}}_e] = \mathbf{G}(\mathbf{A}^T\mathbf{C}_{\widehat{\mathbf{b}}^{MM}}^{-1}\mathbf{A})^{-1}\mathbf{G}^T. \tag{48}$$

The MSEs of the MAP-EM and VB-EM WLS methods can be obtained similarly using the above procedure.

B. ASYMPTOTICAL UNBIASEDNESS ANALYSIS

The unbiasedness of the MM WLS-M estimator can be proven as follows. The expected value of $\widehat{\mathbf{b}}^{MM}(i)$ can be expressed as follows:

$$
\begin{aligned}
E[\widehat{\mathbf{b}}^{MM}(i)] &= E\left[\frac{x_i^2 + y_i^2 - (\hat{r}_i^{MM})^2}{2}\right] \\
&= E\left[\frac{x_i^2 + y_i^2 - (d_i + n_i^{MM})^2}{2}\right] \\
&= E\left[\frac{x_i^2 + y_i^2 - (d_i^2 + 2d_i n_i^{MM} + (n_i^{MM})^2)}{2}\right] \\
&\simeq \frac{x_i^2 + y_i^2 - d_i^2}{2}
\end{aligned}
\tag{49}
$$

where n_i^{MM} denotes the estimation error of $\hat{r}_i^{MM}$. The second-order error term, $(n_i^{MM})^2$, can be neglected under sufficiently low noise conditions. Furthermore, $E[n_i^{MM}]$ is zero because the MM estimator is an asymptotic unbiased estimator [8], [9]. Therefore, the bias of $\widehat{\mathbf{b}}^{MM}(i)$ is approximately zero and the MM WLS-M estimator ($\widehat{\mathbf{x}}_e$) is asymptotic unbiased estimator. Moreover, $\lim_{P\to\infty}\mathbf{C}_{\widehat{\mathbf{b}}^{MM}} \simeq \mathbf{0}_{M\times M}$, where $\mathbf{0}_{M\times M}$ is the zero-valued square matrix with the size of M. (Note that the MM estimator is the consistent estimator [15]). Then, MM WLS-M estimator is asymptotically consistent because $\lim_{P\to\infty}E[\widehat{\mathbf{b}}^{MM}] \simeq \mathbf{b}^o$ and $\lim_{P\to\infty}\mathbf{C}_{\widehat{\mathbf{b}}^{MM}} \simeq \mathbf{0}_{M\times M}$, where $\mathbf{b}^o$ denotes the true value of $\mathbf{b}$. [27]. Furthermore, the unbiasedness of the VB-EM method can be verified. The Gaussian mixture parameters can be estimated using the EM algorithm and they are unbiased estimates when each Gaussian component is separated sufficiently [28]. Moreover, the second-order error terms can be neglected

when the estimation is performed well. Then, $E[\mu_{\mathbf{b}_i}]$ can be expressed as follows:

$$
\begin{aligned}
E[\mu_{\mathbf{b}_i}] \\
= E\left[\frac{x_i^2 + y_i^2 - \mu_{\mathbf{r}_i}^2}{2}\right] \\
\simeq \frac{x_i^2 + y_i^2 - ((1 - v_{i,2})(d_i + \mu_{i,1}))^2 + (v_{i,2}(d_i + \mu_{i,2}))^2}{2} \\
= \frac{x_i^2 + y_i^2 - d_i^2 + (v_{i,2})^2(\mu_{i,2})^2 + 2v_{i,2}\mu_{i,2}d_i}{2} \\
\simeq \frac{x_i^2 + y_i^2 - d_i^2}{2}.
\end{aligned}
\tag{50}
$$

The following property is used in the derivation of (50): $v_{i,1} + v_{i,2} = 1$ and $\mu_{i,1} = 0$. Usually, $v_{i,2}$ is close to zero. Therefore, the estimation bias is smaller when $v_{i,2}$ is closer to zero. However, the bias term cannot be neglected as $v_{i,2}$ is closer to one.

VI. NUMERICAL RESULTS

The localization performances of the proposed robust source localization algorithms are compared with those of MM WLS [15], robust EKF [29], robust extrapolated single propagation unscented Kalman filter (ESPUKF) [15], [30], [31] and semi-definite programming (SDP) [32] algorithms. The SDP method showed accurate performance under the heavily contaminated environment, thus it was included for comparing the localization performance. Also, the Kalman filter has been widely utilized for localization, the robust version of Kalman filter was compared with the proposed algorithm. We used Matlab R2019b to obtain the simulation results.

A. SIMULATION ENVIRONMENTS

The simulation settings are summarized as follows. The RMSE is defined as the estimation performance measure as follows:

$$
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{10}\sum_{k=1}^{200}[(\widehat{x}^k(i) - x(i))^2 + (\widehat{y}^k(i) - y(i))^2]}{10 \times 200}}
\tag{51}
$$

where $[\widehat{x}^k(i), \widehat{y}^k(i)]^T$ is the predicted coordinates of the point source in the ith position set and k^{th} Monte-Carlo iteration. Additionally, $x(i)$ and $y(i)$ indicate the ith true position of the emitter. Seven sensors located at [0, 0] m, [10, 0] m, [0, 10] m, [-10, 0] m, [0, −10] m, [−10, −10] m and [10, 10] m were considered. Ten locations of the emitter were generated with a uniform distribution in the 20 m×20 m. The trial number of Monte-Carlo simulation was 500 and the variance of inlier observations was postulated to be identical. Furthermore, an omni-directional emitter was assumed. The initial mean and variance components of mixture distributions in the MAP-EM and VB-EM WLS methods were set to $[0.1, 1]^T$ and $[(1.4826 \times \text{MAD}(\mathbf{r}_i))^2, (1.4826 \times \text{MAD}(\mathbf{r}_i))^2]^T$, where $\mathbf{r}_i$ denotes $[r_{i,1}, \cdots, r_{i,P}]^T$. Fig. 1 presents the deployment of the sensors and emitters, denoted by circles and asterisks, respectively.
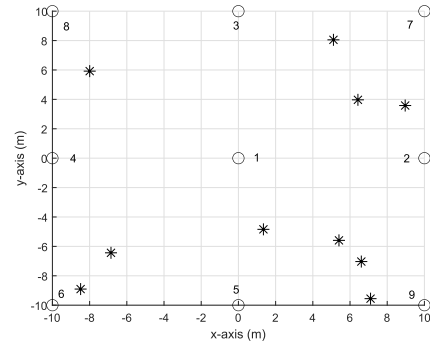

FIGURE 1. Deployment of sensors (white circles indicate sensors, whereas asterisks indicate sources).

B. RMSEs AS A FUNCTION OF STANDARD DEVIATION OF LOS AND NLOS NOISE

The positioning accuracy as a function the standard deviation of the outlier noise is illustrated in Fig. 2. The contamination ratio ($v_{i,2}$) was 20%, the sample length was 20 and the standard deviation of the inlier noise ($\sigma_{i,1}$) was 0.3 m in Fig. 2(a). The bias of the outlier observation noise ($\mu_{i,2}$) was set to 20 m, such that the outlier-contaminated distance observation was consistently larger than zero. The standard deviation of the outlier noise ($\sigma_{i,2}$) was set as a remarkably larger value than that of the inlier noise such that the observation noise is always positive. The RMSE of the proposed MM WLS-M method was lower than that of the other existing methods in Fig. 2(a). The MM WLS-M method outperformed the MM WLS algorithm because the correlation between the transformed range observations and the MM estimate was considered in the derivation of the weight. The RMSE of the VB-EM WLS method was slightly inferior to that of MM WLS method. However, the VB-EM algorithm did not require statistical testing to distinguish the inliers and outliers unlike the MM WLS method. Moreover, the Cramér-Rao lower bound (CRLB) was not displayed since its algebraic derivation for a nonlinear function and non-Gaussian distribution is non-trivial problem [17], [33]. As shown in Fig. 2(b), the contamination ratio was 40%, whereas the remaining environments were identical to those in Fig. 2(a). The RMSE of the proposed MM WLS-M method was the lowest among the localization methods. Further, the RMSEs of the VB-EM and MAP-EM WLS methods were lower that that of the MM WLS algorithm. Moreover, the proposed VB-EM and MAP-EM WLS methods outperformed the SDP and Kalman-filter-based methods. In particular, the RMSE of the VB-EM WLS method was nearly identical to that of the MM WLS-M method in Fig. 2(c). The square root of error variances of the VB-EM WLS and MM WLS-M algorithms were similar with their RMSEs in Figs. 2(a)−(c).

Fig. 3 assumes the identical situation with that appeared in Fig. 2, except that sensors 4−7 were outlier-contaminated sensors. The RMSE of the proposed MM WLS-M algorithm was lower than that of the existing and EM-based WLS algorithms. The RMSEs of the EM-based WLS algorithms were degraded. In Fig. 3(b), the contamination ratio

TABLE 1. Simulation settings.

Source location	a single source located within a 20×20 m^2 region
Number of simulated sources	10
Number of Monte-Carlo simulation	200
Number of receivers	9
Directivity of source	omni-direction
Initial value of b_i^{MM}	median of $b_{i,j}$, $i=1, \cdots, M, j=1, \cdots, P$
Initial value of $\nu_{i,l}$	0.5
Initial value of $\mu_{i,l}$	[0.1, 1]
Initial value of $u_{i,l}$	$(1.4826 \, \text{MAD}(r_{i,j}))^2$

(a) Contamination ratio (ϵ): 20 %, bias of NLOS noise (μ_2): 20 m, standard deviation of LOS noise (σ_1): 0.3 m

(b) ϵ: 40 %, σ_1: 0.3 m, μ_2: 20 m

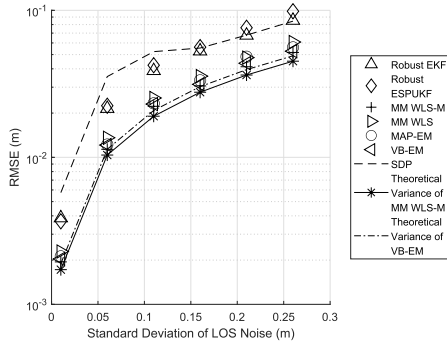(c) ϵ: 80 %, σ_1: 0.3 m, μ_2: 20 m
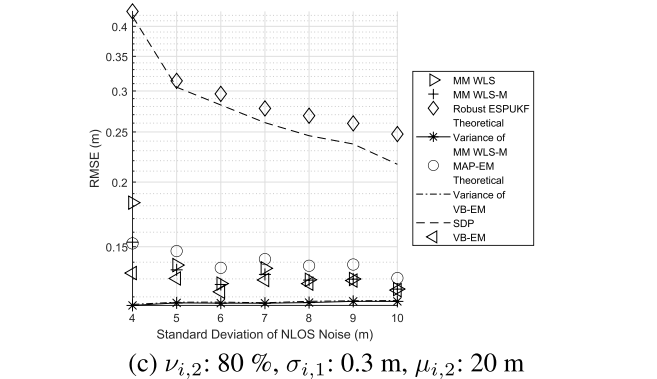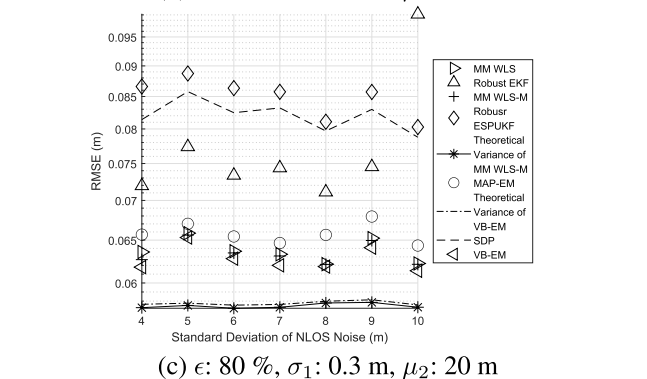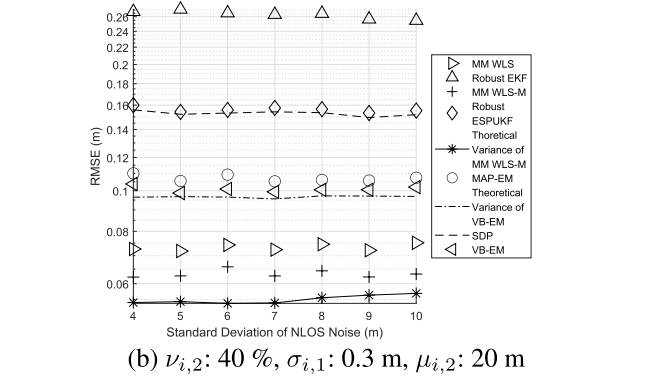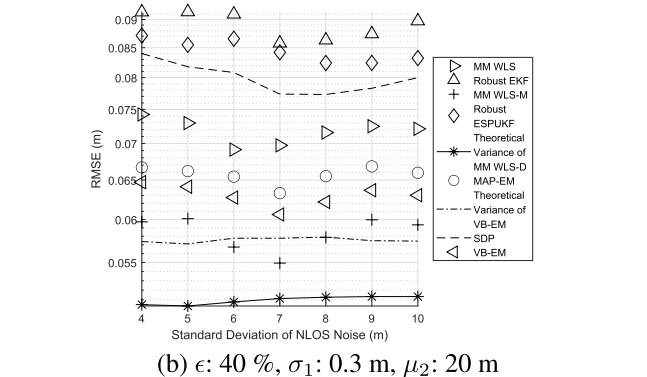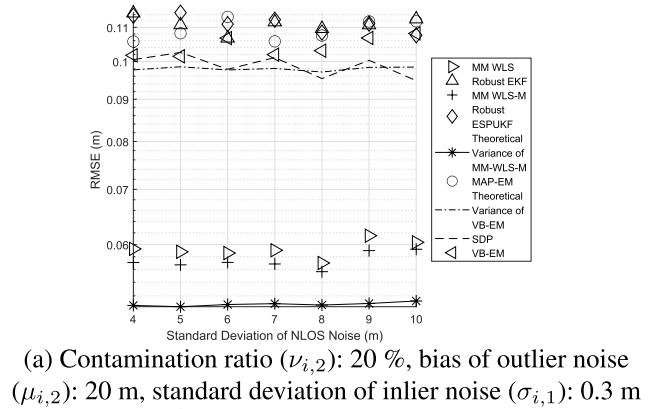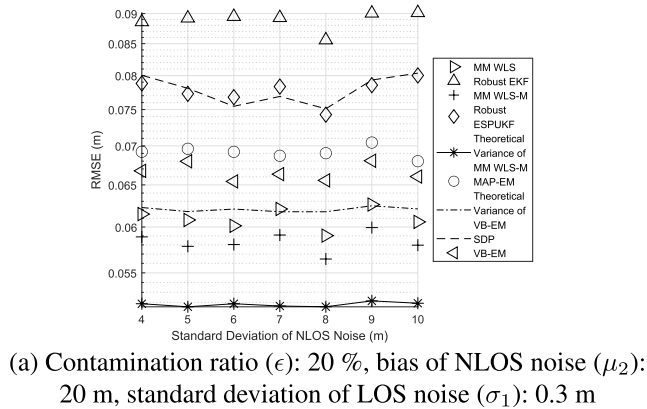
FIGURE 2. RMSEs of the positioning methods with respect to the standard deviation of NLOS noise when sensors 6−7 are outlier-contaminated sensors and the remaining sensors are LOS sensors.

(a) Contamination ratio ($\nu_{i,2}$): 20 %, bias of outlier noise ($\mu_{i,2}$): 20 m, standard deviation of inlier noise ($\sigma_{i,1}$): 0.3 m

(b) $\nu_{i,2}$: 40 %, $\sigma_{i,1}$: 0.3 m, $\mu_{i,2}$: 20 m

(c) $\nu_{i,2}$: 80 %, $\sigma_{i,1}$: 0.3 m, $\mu_{i,2}$: 20 m

FIGURE 3. RMSEs of the positioning methods with respect to standard deviation of outlier noise when sensors 4−7 are outlier-contaminated sensors.

was 40%, whereas the remaining environments were same as those of Fig. 3(a). The RMSE of the proposed MM WLS-M algorithm was lower than that of the existing methods. As shown in Fig. 3(c), the contamination ratio was 80%, whereas the remaining environments were identical to those in Fig. 3(a). This situation modeled the severely

outlier-contaminated environments. The performance of the SDP method was similar to that of the EM-based WLS methods under the severely outlier-contaminated conditions.

Fig. 4(a) illustrates the RMSEs as a function of the standard deviations of inliers. Sensor 7 was postulated to be a LOS/NLOS mixture sensor, whereas the other sensors belong

(a) Contamination ratio ($\nu_{i,2}$): 40 %, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m, bias of outlier noise ($\mu_{i,2}$): 20 m, sensor 7 is outlier-contaminated sensor.
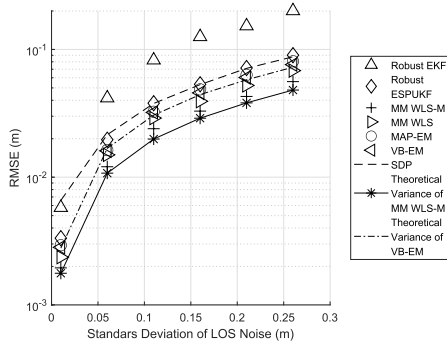
(b) $\nu_{i,2}$: 40 %, $\sigma_{i,2}$: 10 m, $\mu_{i,2}$: 20 m, sensors 4−7 are outlier-contaminated sensors.
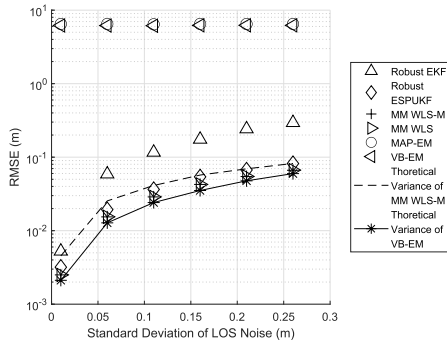
(c) $\nu_{i,2}$: 40 %, $\sigma_{i,2}$: 10 m, $\mu_{i,2}$: 20 m, all sensors are outlier-contaminated sensors.

FIGURE 4. RMSEs of the positioning methods with respect to the standard deviation of inlier noise.

to LOS sensors. The contamination ratio was 40% and the accuracy of the proposed EM-based WLS and MM WLS-M methods were slightly higher than those of the other methods in Fig. 4(a). The RMSEs of the EM-based WLS methods were nearly identical to that of the MM WLS-M method. As shown in Fig. 4(b) and (c), the MM WLS-M algorithm outperformed the existing methods and the proposed EM-based WLS algorithms. The performances of all the robust methods became inferior as the standard deviation of the LOS noise was higher. In particular, the accuracy of the EM-based WLS methods was very sensitive to the number of LOS sensors as shown in Fig. 4(c). Hence, the EM-based WLS method is not recommended unless the number of LOS sensors is more than three.
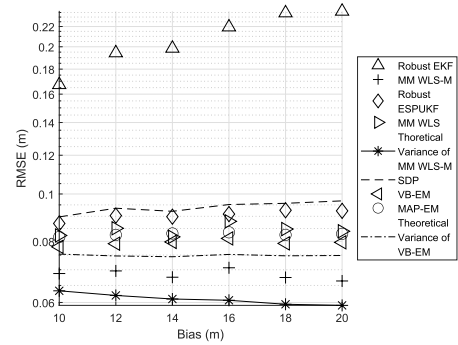

FIGURE 5. RMSEs of the positioning algorithms with respect to the bias when sensors 4−7 are outlier-contaminated sensors (contamination ratio: 40%, standard deviation of LOS noise ($\sigma_{i,1}$): 0.3 m, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m).
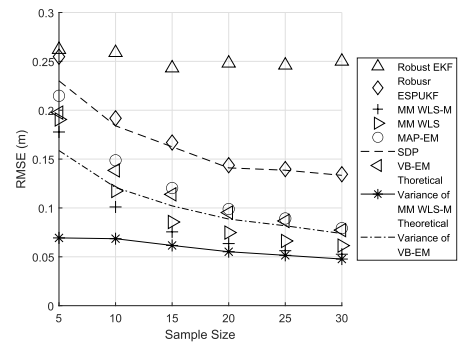

FIGURE 6. RMSEs of the positioning algorithms with respect to sample size when sensors 4−7 are outlier-contaminated sensors (bias of outlier noise ($\mu_{i,2}$): 20 m, contamination ratio: 40%, standard deviation of inlier noise ($\sigma_{i,1}$): 0.3 m, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m).

C. RMSEs AS A FUNCTION OF SEVERAL PARAMETERS

The RMSEs with respect to bias are shown in Fig. 5. The estimation accuracies of the proposed methods were nearly unchanged with the variation in the bias and the proposed MM WLS-M method was superior to the other algorithms. The RMSEs of the proposed WLS-based algorithms were not influenced by the bias because the localization was dependent on the LOS sensors. Furthermore, comparing the localization methods that do not require statistical testing, the accuracy of the proposed MAP-EM and VB-EM WLS algorithms was higher than those of the other localization algorithms.

Fig. 6 compares the accuracy as a function of the sample size. As the sample size increased, the RMSEs of all the methods decreased. The RMSE of the MM WLS-M method was lower than that of the other localization methods for all sample sizes. Moreover, comparing the localization methods that do not require statistical testing, the proposed EM-based WLS algorithms outperformed the existing SDP and Kalman-filter-based methods.

Fig. 7 displays the RMSEs as a function of the number of sensors. The number of LOS/NLOS mixture sensors was set as two and the number of LOS sensors was increased one by one. As the number of sensors increased, the RMSEs of all the methods decreased. The RMSE of the MM WLS-M method was lower than that of the MM WLS method for the entire
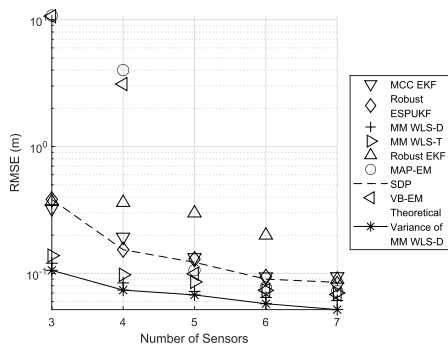
FIGURE 7. RMSEs of the positioning algorithms with respect to the number of sensors (bias of outlier noise ($\mu_{i,2}$): 20 m, contamination ratio: 40%, standard deviation of inlier noise ($\sigma_{i,1}$): 0.3 m, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m).
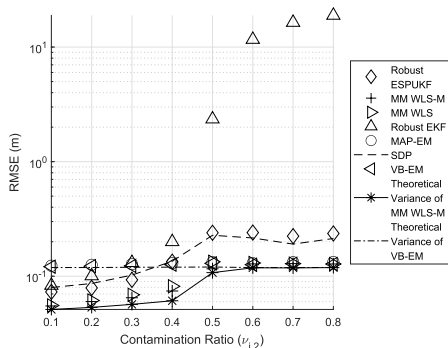

FIGURE 8. RMSEs of the localization algorithms with respect to the contamination ratio (bias of outlier noise ($\mu_{i,2}$): 20 m, standard deviation of inlier noise ($\sigma_{i,1}$): 0.3 m, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m).

number of sensors. However, the RMSEs of the MAP-EM and VB-EM WLS methods were severely degraded when the number of LOS sensors was less than three. This observation coincides with the results shown in Fig. 4(c), i.e., the EM-based WLS algorithms are not effective when the number of LOS sensors is less than three.

The RMSEs with respect to the contamination ratio ($\nu_{i,2}$) are shown in Fig. 8. The localization performances of all the methods were degraded as the contamination ratio increased. Specifically, the RMSEs of the localization algorithms were evidently degraded when the contamination ratio was higher than 0.5. The RMSE of the MM WLS-M method was lower than those of the other methods when the contamination ratio was less than 0.5, but it was similar to that of the MM WLS method when the contamination ratio exceeded 0.5.

D. EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTIONS (ECDFs) OF LOCALIZATION ERROR

Fig. 9 provides the ECDFs with respect to estimation errors. The contamination ratio ($\nu_{i,2}$) was 90% and sensor 7 was an outlier-contaminated sensor, whereas the other sensors were LOS sensors. The other conditions were same as those of Fig. 2(a). This situation represents a mildly contaminated environment. The RMSEs of the MM WLS-M, MM WLS and EM-based WLS methods were nearly identical
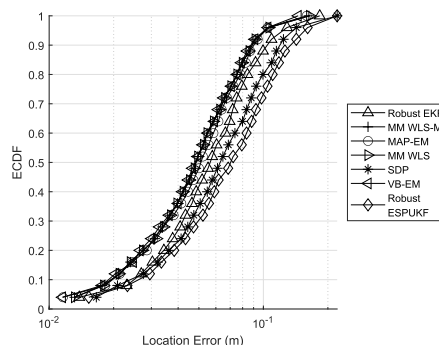
(a) Contamination ratio ($\nu_{i,2}$): 90 %, standard deviation of outlier noise ($\sigma_{i,2}$): 10 m, bias of outlier noise ($\mu_{i,2}$): 20 m, standard deviation of inlier noise ($\sigma_{i,1}$): 0.3 m, sensor 7 is outlier-contaminated sensor.
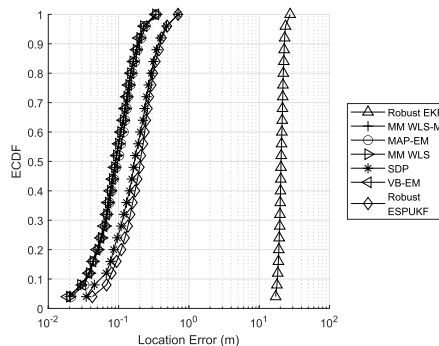
(b) $\nu_{i,2}$: 90 %, $\sigma_{i,1}$: 0.3 m, $\sigma_{i,2}$: 10 m, $\mu_{i,2}$: 20 m, sensors 4−7 are outlier-contaminated sensors.
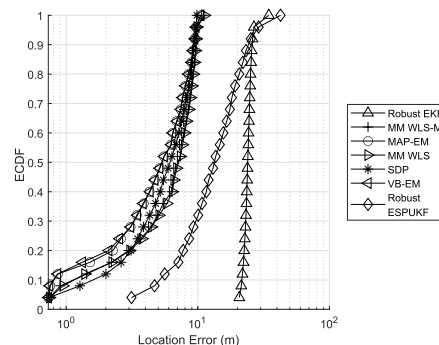
(c) $\nu_{i,2}$: 90 %, $\sigma_{i,1}$: 0.3 m, $\sigma_{i,2}$: 10 m, $\mu_{i,2}$: 20 m, sensors 2−7 are outlier-contaminated sensors.

FIGURE 9. ECDFs of the positioning methods.

as shown in Fig. 9(a). Note that the VB-EM and MAP-EM WLS methods do not require statistical testing, where the optimal threshold must be determined. This advantage can be important under rapidly varying channel conditions or environments. The conditions in Fig. 9(b) were postulated to be identical to those of Fig. 9(a), except that sensors 4−7 were outlier-contaminated sensors. These simulation settings represent moderately outlier-contaminated situations. The RMSEs of the proposed methods were lower than those of the other existing techniques. All the sensors besides sensor 1 were outlier-contaminated sensors in Fig. 9(c). This situation represents a severely outlier-contaminated condition. In Fig. 9(c), the RMSEs of the EM-based WLS

methods were lower than those of the other methods, although they were degraded significantly compared with those of the former cases. The RMSEs of all the robust methods were increased as the number of outlier-contaminated sensors was incremented. The EM-based WLS methods were significantly superior to the other methods under seriously outlier-contaminated situations. The simulation results illustrated that the MM WLS-M method was superior to the other algorithms in most environments and the EM-based WLS algorithms showed effective performances when the number of LOS sensors was more than three.

VII. CONCLUSION

Robust localization algorithms were developed using the MM WLS-M, MAP-EM and VB-EM WLS techniques. The proposed MM WLS-M method utilized a weighting matrix estimated using the property that transformed measurements are correlated with the MM estimate. The RMSE of the MM WLS-M method was lower than those of the other localization methods under all the conditions, excluding severely outlier-contaminated environments. Furthermore, the RMSEs of the MAP-EM and VB-EM WLS algorithms were the lowest among the localization algorithms that do not require statistical testing. However, the accuracies of the MAP-EM WLS and VB-EM WLS algorithms were severely degraded when the number of LOS sensors was less than three. Moreover, the MSE and asymptotic unbiasedness of the proposed methods were analyzed. The proposed localization methods are iterative methods, i.e., they may diverge under severely low-SNR or heavily outlier-contaminated noise conditions because the initial position estimate cannot be appropriately determined. Moreover, the computational complexity of the iterative method was higher than that of the closed-form technique. Therefore, algebraic robust localization algorithms shall be developed in the future work.

REFERENCES

[1] D. Torrieri, "Statistical theory of passive location systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-20, no. 2, pp. 183–198, Mar. 1984.

[2] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1905–1915, May 1994.

[3] C.-H. Park and J.-H. Chang, "Closed-form localization for distributed MIMO radar systems using time delay measurements," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1480–1490, Feb. 2016.

[4] P. Petrus, "Robust huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.

[5] G. L. Sun and W. Guo, "Bootstrapping M-estimators for reducing errors due to non-line-of-sight (NLOS) propagation," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 509–510, Aug. 2004.

[6] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Los Angeles, CA, USA, Apr. 2005, pp. 91–98.

[7] R. Casas, A. Marco, J. J. Guerrero, and J. Falcó, "Robust estimator for Non-Line-of-Sight error mitigation in indoor localization," *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 1, pp. 1–8, Dec. 2006.

[8] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory Methods*, Hoboken, NJ, USA: Wiley, 2006.

[9] V. J. Yohai, "High breakdown-point and high efficiency robust estimates for regression," *Ann. Statist.*, vol. 15, pp. 591–601, Jun. 1987.

[10] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 672–681, Mar. 2006.

[11] J. Riba and A. Urruela, "A non-line-of-sight mitigation technique based on ML-detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Quebec, QC, Canada, May 2004, pp. 153–156.

[12] 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.

[13] S. Venkatesh and R. M. Buehrer, "NLOS mitigation using linear programming in ultrawideband location-aware networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 3182–3198, Sep. 2007.

[14] S. Tomic and M. Beko, "A bisection-based approach for exact target localization in NLOS environments," *Signal Process.*, vol. 143, pp. 328–335, Feb. 2018.

[15] C. H. Park and J.-H. Chang, "Robust localization based on ML-type, multistage ML-type, and extrapolated single propagation UKF methods under mixed LOS/NLOS conditions," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5819–5832, Sep. 2020.

[16] S. Nakajima, K. Watanabe, and M. Sugiyama, *Variational Bayesian Learning Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[17] F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "EM- and JMAP-ML based joint estimation algorithms for robust wireless geolocation in mixed LOS/NLOS environments," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 168–182, Jan. 2014.

[18] Y. Feng, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "TOA-based robust wireless geolocation and Cramér-Rao lower bound analysis in harsh LOS/NLOS environments," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2243–2255, May 2013.

[19] C. Fritsche and A. Klein, "On the performance of mobile terminal tracking in urban GSM networks using particle filters," in *Proc. 17th Eur. Signal Process. Conf.*, Glasgow, U.K., Aug. 2009, pp. 1953–1957.

[20] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Robust tracking and geolocation for wireless networks in NLOS environments," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 889–901, Oct. 2009.

[21] F. Yin, C. Fritsche, D. Jin, F. Gustafsson, and A. M. Zoubir, "Cooperative localization in WSNs using Gaussian mixture modeling: Distributed ECM algorithms," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1448–1463, Mar. 2015.

[22] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the *EM* algorithm," *J. Roy. Stat. Soc.: B (Methodol.)*, vol. 39, no. 1, pp. 1–22, 1977.

[24] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[25] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[26] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, Apr. 2017.

[27] E. L. Lehmann and G. Casella, *Theory Point Estimation*, 2nd ed. Springer, 2012.

[28] D. Nityasuddhi and D. Böhning, "Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances," *Comput. Statist. Data Anal.*, vol. 41, nos. 3–4, pp. 591–601, Jan. 2003.

[29] U. Hammes and A. M. Zoubir, "Robust MT tracking based on M-Estimation and interacting multiple model algorithm," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3398–3409, Jul. 2011.

[30] S. K. Biswas, B. Southwell, and A. G. Dempster, "Performance analysis of fast unscented Kalman filters for attitude determination," *IFAC-PapersOnLine*, vol. 51, no. 1, pp. 697–701, 2018.

[31] S. K. Biswas, L. Qiao, and A. G. Dempster, "A novel a priori state computation strategy for the unscented Kalman filter to improve computational efficiency," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1852–1864, Apr. 2017.

[32] R. M. Vaghefi, J. Schloemann, and R. M. Buehrer, "NLOS mitigation in TOA-based localization using semidefinite programming," in *Proc. 10th Workshop Positioning, Navigat. Commun. (WPNC)*, Dresden, Germany, Mar. 2013, pp. 1–6.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

CHEE-HYUN PARK received the Ph.D. degree in electronics and computer engineering from Sungkyunkwan University, Suwon, South Korea, in 2011. From April 2011 to August 2012, he was with the University of Wisconsin-Madison, Madison, WI, USA, as a Postdoctoral Fellow. He is currently a Research Professor with Hanyang University, Seoul, South Korea. His research interests include detection and estimation theory, particularly in signal processing, source localization, array signal processing, and machine learning.

JOON-HYUK CHANG (Senior Member, IEEE) received the B.S. degree in electronics engineering from Kyungpook National University, Daegu, South Korea, in 1998, and the M.S. and Ph.D. degrees in electrical engineering from Seoul National University, South Korea, in 2000 and 2004, respectively. From 2000 to 2005, he was with Netdus Corporation, Seoul, South Korea, as a Chief Engineer. From 2004 to 2005, he held a post-doctoral position with the University of California at Santa Barbara, Santa Barbara, where he was involved in adaptive signal processing and audio coding. In 2005, he joined the Korea Institute of Science and Technology, Seoul, as a Research Scientist, where he was involved in speech recognition. From 2005 to 2011, he was an Assistant Professor with the School of Electronic Engineering, Inha University, Incheon, South Korea. He is currently a Full Professor with the School of Electronic Engineering, Hanyang University, Seoul. His research interests include speech coding, speech enhancement, speech recognition, audio coding, and adaptive signal processing. He was a recipient of the IEEE/IEEK IT Young Engineer of the Year 2011. He is serving as the Editor-in-Chief for the *Signal Processing Society Journal of the IEIE*.

● ● ●