Robust Estimators for Variance-Based Device-Free Localization and Tracking

Yang Zhao and Neal Patwari

Abstract—Device-free localization systems, such as variance-based radio tomographic imaging (VRTI), use received signal strength (RSS) variations caused by human motion in a static wireless network to locate and track people in the area of the network, even through walls. However, intrinsic motion, such as branches moving in the wind or rotating or vibrating machinery, also causes RSS variations which degrade the performance of a localization system. In this paper, we propose a new estimator, least squares variance-based radio tomography (LSVRT), which reduces the impact of the variations caused by intrinsic motion. We compare the novel method to subspace variance-based radio tomography (SubVRT) and VRTI. SubVRT also reduces intrinsic noise compared to VRTI, but LSVRT achieves better localization accuracy and does not require manually tuning additional parameters compared to VRTI. We also propose and test an online calibration method so that LSVRT and SubVRT do not require "empty-area" calibration and thus can be used in emergency situations. Experimental results from five data sets collected during three experimental deployments show that both estimators, using online calibration, can reduce localization root mean squared error by more than 40 percent compared to VRTI. In addition, the Kalman filter tracking results from both estimators have 97th percentile error of 1.3 m, a 60 percent reduction compared to VRTI.

Index Terms—Wireless sensor networks, sensing, statistical signal processing

1 Introduction

Device-free localization (DFL) using radio frequency (RF) sensor networks is the use of channel measurements between static radio transceivers deployed around an area of interest to locate people and objects moving within it. DFL has particular application in security and emergency applications, e.g., detecting intruders in industrial facilities, and helping police and firefighters track people inside a building during an emergency [1]. In these scenarios, the people being located cannot be expected to participate in the localization system by carrying radio devices, thus standard radio localization techniques [2] are not useful.

Various RF measurements including ultra-wideband (UWB) and received signal strength (RSS) have been proposed and applied to detect, locate and track objects and people who do not carry radio devices in an environment [3], [4], [5], [6], [7], [8]. Compared to cameras and infrared sensing methods, RF sensors have the advantage of penetrating non-metal walls and smoke [1]. While UWB measurements are expensive, RSS measurements are inexpensive and available in standard wireless devices and have been used in different device-free localization studies with surprising accuracy [6], [8], [9]. These RSS-based localization methods essentially use a windowed variance of RSS measured on static links. For example, [8] deploys an RF sensor network around a residential house and uses sample variance during a short window to track people walking inside the house; [9] places RF sensors on the ceiling of a room, and track people using the received signal strength indicator (RSSI) dynamic, which is essentially the variance of RSS measurements, with and without people moving inside the room. In this paper we focus on using RSS measurements to locate and track human motion. We use windowed variance to describe the various functions of RSS measurements recently used in different localization and tracking studies [6], [8], [9], and we call them methods for variance-based device-free localization.

This paper introduces and studies methods to make variance-based DFL more robust to noise. For variance-based localization and tracking, variance can be caused by two types of motion: *extrinsic motion* and *intrinsic motion*. Extrinsic motion is defined as the motion of people and other objects that enter and leave the environment. Intrinsic motion is defined as the motion of objects that are intrinsic parts of the environment, objects which cannot be removed without fundamentally altering the environment. If a significant amount of windowed variance is caused by intrinsic motion, then it may be difficult to detect extrinsic motion. For example, rotating fans, leaves and branches swaying in wind, and moving or rotating machines in a factory all may impact the RSS measured on static links. Also, if RF sensors are vibrating or swaying in the wind, their RSS measurements change as a result. Even if the receiver moves by only a fraction of its wavelength, the RSS may vary by several orders of magnitude as a result of small-scale fading [10], [11]. We call variance caused by intrinsic motion and extrinsic motion, the *intrinsic signal* and *extrinsic signal*, respectively. We consider the intrinsic signal to be "noise" because it does not relate to extrinsic motion which we wish

- *Y. Zhao is with Distributed Intelligent Systems Lab, General Electric Global Research Center, Niskayuna, NY 12309. E-mail: yang.zhao@utah.edu.*
- *N. Patwari is with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84103. E-mail: npatwari@ece.utah.edu.*

Manuscript received 7 Oct. 2013; revised 26 July 2014; accepted 17 Dec. 2014. Date of publication 22 Dec. 2014; date of current version 31 Aug. 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2014.2385710

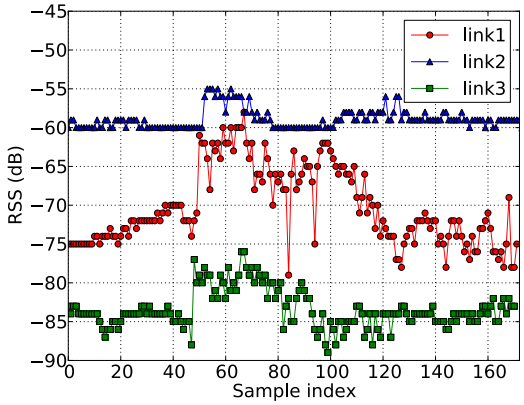


Fig. 1. Intrinsic signal measurements: RSS measurements from three links during an offline calibration period (when no people are present in the environment) of one experiment, in which we observe significant wind-induced intrinsic motion.

to detect and track. One example of intrinsic signal is shown in Fig. 1, in which the RSS measurements are recorded during an "empty-area" offline calibration period, when tree branches and leaves sway strongly in wind [12]. Considering there is no human motion in the network, that is, no extrinsic motion during the calibration, the high variations of RSS measurements are caused by intrinsic motion, in this case, wind-induced motion. To be practical, variance-based localization methods must identify and reduce the intrinsic signal.

In this paper, we propose and compare two estimators that reduce the effect of intrinsic motion and improve the robustness of VRTI [8] under noisy conditions. The first estimator uses the subspace decomposition method, which has been used in spectral estimation, sensor array processing, and network anomaly detection [13], [14], [15], [16]. We apply this method to VRTI, which leads to subspace variance-based radio tomography (SubVRT). SubVRT was introduced in [12], and is explored and tested in greater depth in this paper, and compared to a second estimator which is first presented here.

The major contribution of this paper is to propose and test a new robust estimator – least squares variance-based radio tomography (LSVRT). While both SubVRT and LSVRT are significantly more robust to intrinsic motion than VRTI, one advantage of LSVRT over SubVRT is that it does not require manual tuning of any additional parameters compared to VRTI. In contrast, SubVRT has a parameter k that must be manually set. Further, experimental results show that LSVRT has consistently better performance than SubVRT across all experiments conducted for this paper.

Another contribution of this paper is the testing of an online calibration procedure for SubVRT and LSVRT. An "empty-area" calibration (offline calibration) using measurements recorded during a period without people in the network is effective for statistical characterization of the intrinsic signal [12]. However, knowing when the area is empty may not be possible in emergency situations. To enable SubVRT and LSVRT in those situations, we propose an online calibration method in which RSS measurements are collected for statistical characterization of the intrinsic signal when the maximum pixel intensity of a radio tomographic image is low.

This paper presents new quantitative comparisons of the new methods with VRTI. Localization results from five data sets, collected during three experiments, show that both estimators using online calibration can reduce the root mean squared error (RMSE) of the location estimate by more than 40 percent compared to VRTI. The LSVRT method proposed in this paper consistently outperforms SubVRT in localization accuracy across all experiments. A new experiment presented here shows that as the level of intrinsic noise increases, the ability of both SubVRT and LSVRT to reduce the level of noise increases as well. Further, we use the Kalman filter to track people using localization estimates from SubVRT and LSVRT. The results show that tracking errors from SubVRT have a 97th percentile error of 1.4 m, a 65 percent improvement compared to VRTI, while the 97th percentile error from LSVRT is less than 1.2 m, a 70 percent improvement.

To summarize, in this paper we 1) introduce a new estimator, least-squares variance-based radio tomography; 2) introduce a new online calibration method which does require the user to know when the area of interest is empty, 3) compare and evaluate SubVRT and LSVRT estimators by using data sets collected in new and past experiments in different environmental conditions; and 4) quantify tracking improvements using a Kalman filter in combination with SubVRT and LSVRT.

The rest of this paper is organized as follows: Section 2 formulates the subspace decomposition and least squares (LS) methods, and proposes an online calibration method. Section 3 describes five data sets collected in three experiments including new experiments first reported in this paper, Section 4 shows the experimental results, and Section 5 investigates the Kalman filter tracking. Related work is presented in Section 6, and the conclusion is given in Section 7.

2 Methods

In this section, we describe variance-based radio tomographic imaging (VRTI), a through-wall imaging method introduced in [8], which serves as a baseline for comparison for our new methods. Then we extend the VRTI method by introducing the SubVRT and LSVRT estimators, and compare the two.

2.1 Problem Statement

We assume an RF sensor network with N sensors (radio transceivers) is deployed in an area of interest. We use $\mathbf{z}_{s,j}$ to denote the coordinate of sensor j for $j \in \{1, \ldots, N\}$. The deployment area has both extrinsic motion and intrinsic motion, as defined in the Introduction. The goal of all of the methods in this paper is to estimate an image of the presence of extrinsic motion. We discretize space into P pixels of a physical space and define $\mathbf{x} = [x_1, \ldots, x_P]^T$, where $x_i = 1$ if extrinsic motion occurs in pixel i, and $x_i = 0$ otherwise. We denote the center coordinate of pixel i as $\mathbf{z}_i$.

To enable motion image estimation, each sensor measures the RSS of packets received from many other sensors. We use $s_{l,t}$ to denote the RSS measured at node i_l sent by node j_l at time $t \in \mathbb{Z}$, where i_l and j_l are the receiver and transmitter number for link l, respectively. We assume L directional

links measure RSS, where in general, $L \leq N(N-1)$ since not all pairs of sensors may be connected.

We denote the windowed RSS variance as

$$y_{l,t} = \frac{1}{m-1}\sum_{i=0}^{m-1}(\bar{s}_{l,t} - s_{l,t-i})^2, \tag{1}$$

where m is the length of the window, and $\bar{s}_{l,t} = \frac{1}{m}\sum_{i=0}^{m-1} s_{l,t-i}$ is the sample average in this window period. We let $\mathbf{y}^{(t)} = [y_{1,t}, \ldots, y_{L,t}]^T$ be the vector of windowed RSS variance from all L links at time t. If we do not need to represent time, we simplify the notation to $\mathbf{y} = [y_1, \ldots, y_L]^T$.

We further use $\mathbf{y}_c$ to denote the measurements collected during the calibration period, when no people are present in the environment; and we use $\mathbf{y}_r$ to denote the measurements from the real-time operation period.

We note two things. Transmit power does not affect $y_{l,t}$ because $\bar{s}_{l,t}$ and $s_{l,t-i}$ in (1) are equally affected by any change in transmit power. However, transmit power must be high enough so that a receiver can reliably estimate RSS, that is, packets are received even when a person obstructs the link. Next, we assume any changes in transmit power are slow such that it can be considered constant within m samples so that the short-term changes in $s_{l,t}$ are due to the channel, not to the transmitter.

2.2 Baseline Method: VRTI

Work in [8] has shown the efficacy of a linear model that relates the motion image $\mathbf{x}$ to the RSS variance $\mathbf{y}_r$:

$$\mathbf{y}_r = W\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is an L-length noise vector including intrinsic motion and measurement noise, and W is an $L \times P$ matrix quantifying how much of motion in each pixel impacts each link measurement. The weighting of pixel p on link l is formulated as [8]

$$W_{l,p} = \frac{1}{\sqrt{d_{i_l,j_l}}}\begin{cases}\phi & \text{if } d_{i_l,p} + d_{j_l,p} < d_{i_l,j_l} + d_w \\ 0 & \text{otherwise,}\end{cases} \tag{3}$$

where d_{i_l,j_l} is the Euclidean distance between two sensors i_l, j_l on link l located at $\mathbf{z}_{s,i_l}$ and $\mathbf{z}_{s,j_l}$; $d_{j_l,p}$ is the Euclidean distance between sensor j_l and $\mathbf{z}_p$, the center coordinate of pixel p; $d_{i_l,p}$ is the Euclidean distance between sensor i_l and pixel p; d_w is a tunable parameter controlling the ellipse width, and ϕ is a constant scaling factor. Essentially, column p of the W matrix describes a model for how link variances are affected by a moving person's presence in pixel p.

VRTI is the method proposed in [8] to estimate image $\mathbf{x}$ from the link measurements $\mathbf{y}_r$. Regularization is necessary for this ill-posed inverse problem, and Tikhonov regularization is used [8], which results in the image estimator,

$$\begin{aligned}\hat{\mathbf{x}} &= \Pi_1 \mathbf{y}_r \\ \Pi_1 &= (W^T W + \alpha Q^T Q)^{-1} W^T,\end{aligned} \tag{4}$$

where α is a regularization parameter, and Q is the Tikhonov matrix, which is calculated by using the difference operations in both the vertical and horizontal directions of an image, as discussed in [8]. One benefit of (4) is that Π_1 does not need to be recomputed as long as the sensors are stationary.

The vector $\hat{\mathbf{x}}$ has in element i the estimated motion in pixel i. Pixel i is centered at coordinate $\mathbf{z}_i$, and we display an arbitrary motion image by plotting x_i at coordinate $\mathbf{z}_i$ for each pixel $i = 1, \ldots, P$.

The calculation of the projection matrix in (4) requires the inverse of a $P \times P$ matrix, which has complexity order P^3. However, this is calculated once, and the real-time estimation of $\hat{\mathbf{x}}$ requires only one matrix multiply, which uses at most LP operations.

2.3 Subspace Decomposition Method

The subspace decomposition method has been widely used in spectral estimation, sensor array processing, etc. [13], [16] to improve estimation performance in noise. It is closely related to principal component analysis (PCA), which is widely used in finding patterns in high dimensional data [17]. Essentially, it divides the space of the measurements into two orthogonal subspaces, one we believe is very noisy and one that is not, and only use the portion of the measurement which is contained in the latter subspace.

To find this subspace decomposition, we first estimate the covariance matrix $C_{\mathbf{y}_c}$ as

$$C^*_{\mathbf{y}_c} = \frac{1}{M-1}\sum_{t=0}^{M-1}\left(\mathbf{y}_c^{(t)} - \boldsymbol{\mu}_c\right)\left(\mathbf{y}_c^{(t)} - \boldsymbol{\mu}_c\right)^T, \tag{5}$$

where M is the number of sample measurements, $\mathbf{y}_c^{(t)}$ is the calibration measurement vector $\mathbf{y}_c$ at time t, $\boldsymbol{\mu}_c = \frac{1}{M}\sum_{t=0}^{M-1}\mathbf{y}_c^{(t)}$ is the sample average. Then, we perform singular value decomposition (SVD) on $C^*_{\mathbf{y}_c}$

$$C^*_{\mathbf{y}_c} = U\Lambda U^T, \tag{6}$$

where the unitary matrix $U = [\mathbf{u}_1, \ldots, \mathbf{u}_L]$ and the diagonal matrix $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_L\}$. Right multiplying U on both sides of (6), we have

$$C^*_{\mathbf{y}_c}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \tag{7}$$

where $\mathbf{u}_i$ is the eigenvector corresponding to the ith eigenvalue λ_i. If the eigenvalues are in descending order, the first principal component $\mathbf{u}_1$ points in the direction of the maximum variance in the calibration measurements, the second principal component $\mathbf{u}_2$ points in the direction of the maximum variance remaining in the measurements, and so on. Examples of eigenvalues from calibration measurements are shown in Fig. 6 in Section 4.1. The first few eigenvalues are much larger than the others, thus most of the variance in the calibration measurements is in the subspace spanned by these few principal components. After obtaining eigenvalues from calibration measurements, we decide how many principal components, $k \ll L$, are necessary to capture the majority of the variations (we discuss selection of k in more detail in Section 4.3). In subspace decomposition, we will remove any part of the measurement vector $\mathbf{y}$ which falls in the subspace spanned by these k principal components, and by doing so, we eliminate a large proportion of the intrinsic noise. Note that these k components also contain portions of the extrinsic signal and so k should be kept as low as possible. Regardless, we describe the space spanned by the first k principal components, $\hat{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k]$, as the intrinsic signal subspace; and we describe the space spanned by the

next $L-k$ principal components, $\tilde{U}=[\mathbf{u}_{k+1},\mathbf{u}_{k+2},\ldots,\mathbf{u}_L]$, as the extrinsic signal subspace.

Once the two subspaces are constructed, we decompose the measurement vector $\mathbf{y}$ into two components—intrinsic signal component $\hat{\mathbf{y}}$ and extrinsic signal component $\tilde{\mathbf{y}}$:

$$\mathbf{y}=\hat{\mathbf{y}}+\tilde{\mathbf{y}}. \tag{8}$$

The linearity of this approximation is by definition here, as the two subspaces are orthogonal by the properties of the SVD. However, linearity in RSS has been justified in prior literature [8], [18]. We also note that the "spatial characteristics" of intrinsic vs. extrinsic signal are fairly different. For example, intrinsic signal caused by wind appears at the same time everywhere in the area that has branches or bushes, for example, whereas people will not tend to be simultaneously moving in exactly those same positions.

Since the principal components are orthogonal, the intrinsic signal component $\hat{\mathbf{y}}$ and the extrinsic signal component $\tilde{\mathbf{y}}$ can be formed by projecting $\mathbf{y}$ onto the intrinsic subspace and the extrinsic subspace, respectively:

$$\hat{\mathbf{y}}=\Pi_I\mathbf{y}=\hat{U}\hat{U}^T\mathbf{y} \tag{9}$$

$$\tilde{\mathbf{y}}=\Pi_E\mathbf{y}=(I-\hat{U}\hat{U}^T)\mathbf{y}, \tag{10}$$

where $\Pi_I=\hat{U}\hat{U}^T$ is the projection matrix for the intrinsic subspace, and $\Pi_E=I-\Pi_I$ is the projection matrix for the extrinsic subspace.

The key idea of SubVRT is to use the decomposed extrinsic signal component of the measurements in VRTI. We project the real-time measurement vector $\mathbf{y}_r$ onto the extrinsic subspace to obtain the extrinsic signal component $\tilde{\mathbf{y}}_r=(I-\hat{U}\hat{U}^T)\mathbf{y}_r$. Then, we replace $\mathbf{y}_r$ in (4) with $\tilde{\mathbf{y}}_r$ and obtain the solution of SubVRT:

$$\begin{aligned}\hat{\mathbf{x}}_{Sub}&=\Pi_2\mathbf{y}_r\\ \Pi_2&=(W^TW+\alpha Q^TQ)^{-1}W^T(I-\hat{U}\hat{U}^T).\end{aligned} \tag{11}$$

From (11), we see that the solution is a linear transformation of the measurement vector. The transformation matrix Π_2 is the product of the transformation matrix Π_1 in (4) with the projection matrix for the extrinsic subspace Π_E: $\Pi_2=\Pi_1\Pi_E$. Since the transformation matrix Π_2 does not depend on instantaneous real-time measurements, it can be precalculated, and it is easy to implement SubVRT for real-time applications. Calculation of $\hat{\mathbf{x}}$ from $\mathbf{y}_r$ requires LP multiplications and additions.

2.4 Least Squares Method

As shown above, SubVRT performs SVD on the covariance matrix of the calibration measurements. Here, we introduce our LSVRT estimator formulated as a least squares solution, which uses the inverse of the covariance matrix.

2.4.1 Formulation

To derive the least squares solution to the linear model expressed in (2), the cost function can be written as [19]:

$$\begin{aligned}J(\mathbf{x})&=\|W\mathbf{x}-\mathbf{y}_r\|^2_{C_\mathbf{n}}+\|\mathbf{x}-\mathbf{x}_a\|^2_{C_\mathbf{x}}\\ &=(\mathbf{y}_r-W\mathbf{x})^TC_\mathbf{n}^{-1}(\mathbf{y}_r-W\mathbf{x})+(\mathbf{x}-\mathbf{x}_a)^TC_\mathbf{x}^{-1}(\mathbf{x}-\mathbf{x}_a),\end{aligned} \tag{12}$$

where $\|\mathbf{n}\|^2_{C_\mathbf{n}}$ indicates weighted quadratic distance $\mathbf{n}^TC_\mathbf{n}^{-1}\mathbf{n}$, $C_\mathbf{n}$ is the covariance matrix of the noise term $\mathbf{n}$, $\mathbf{x}_a$ is the prior mean of $\mathbf{x}$, and $C_\mathbf{x}$ is the covariance matrix of $\mathbf{x}$.

Taking the derivative of (12) and setting it to zero results in the LSVRT solution:

$$\hat{\mathbf{x}}_{LS}=(W^TC_\mathbf{n}^{-1}W+C_\mathbf{x}^{-1})^{-1}(W^TC_\mathbf{n}^{-1}\mathbf{y}_r+C_\mathbf{x}^{-1}\mathbf{x}_a). \tag{13}$$

Since the prior information $\mathbf{x}_a$ can be included in the tracking period, here we assume $\mathbf{x}_a$ is zero, then (13) becomes:

$$\begin{aligned}\hat{\mathbf{x}}_{LS}&=\Pi_3\mathbf{y}_r\\ \Pi_3&=(W^TC_\mathbf{n}^{-1}W+C_\mathbf{x}^{-1})^{-1}W^TC_\mathbf{n}^{-1}.\end{aligned} \tag{14}$$

The LSVRT formulation can be also justified from a Bayesian perspective. If we assume $\mathbf{y}_r$ conditioned on $\mathbf{x}$ is Gaussian distributed with mean $W\mathbf{x}$ and covariance matrix $C_\mathbf{n}$, and $\mathbf{x}$ is Gaussian distributed with mean $\mathbf{x}_a$ and covariance matrix $C_\mathbf{x}$, then the maximum a posteriori (MAP) estimator, which maximizes $p(\mathbf{x}|\mathbf{y}_r)$, is equivalent to minimizing the cost function in (12). Under the same Gaussian assumption, the minimum mean-squared error (MMSE) estimator is equivalent to the MAP estimator. Thus the LS solution (13) can also be seen as the MAP or MMSE estimator under these Gaussian assumptions.

2.4.2 Covariance of Noise

For the LSVRT solution (13), the inverse of the covariance matrix $C_\mathbf{n}$ is needed. During the calibration period, we assume that there is no extrinsic motion in the signal, so $\mathbf{x}=0$ and $\mathbf{y}_c=\mathbf{n}$. Thus $C_\mathbf{n}=C_{\mathbf{y}_c}$.

Consider the inverse of the sample covariance matrix $C^*_{\mathbf{y}_c}$. From (6), The inverse of $C^*_{\mathbf{y}_c}$ is given as,

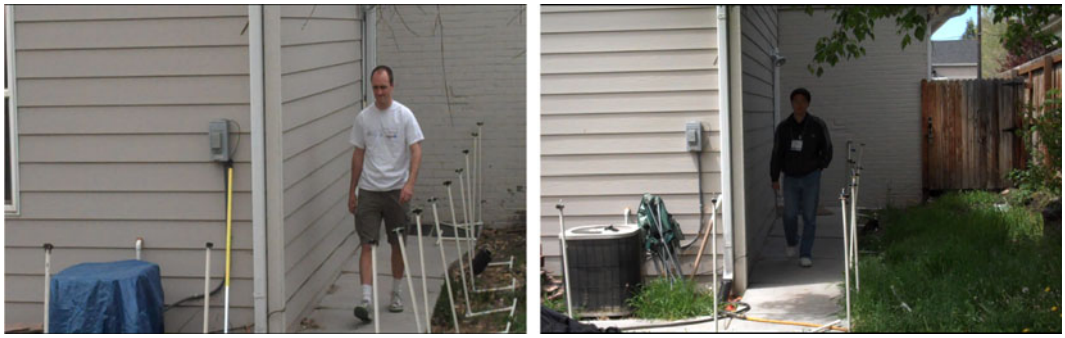
$$[C^*_{\mathbf{y}_c}]^{-1}=U\Lambda^{-1}U^T, \tag{15}$$

where Λ^{-1} is a diagonal matrix with ith element $1/\lambda_i$ [13]. The problem is that $C^*_{\mathbf{y}_c}$ is estimated using a sample of size M, and M is typically less than L, the dimension of the vector. Thus there will be at least $L-M$ zero eigenvalues λ_i in the SVD of the sample covariance matrix (equivalently, the rank of $C^*_{\mathbf{y}_c}$ is at most M). Thus its inverse is not defined.

For high dimensional covariance matrix estimation problems, many regularized covariance matrix estimators have been proposed [20], [21]. Here, we use the Ledoit-Wolf estimator, which is a linear combination of the sample covariance matrix and a scaled identity matrix, and is asymptotically optimal for any distribution [20]:

$$C_\mathbf{n}=\nu\mu I+(1-\nu)C^*_{\mathbf{y}_c}, \tag{16}$$

where $C^*_{\mathbf{y}_c}$ is the sample covariance matrix of noisy calibration measurements $\mathbf{y}_c$, μ is the scaling parameter for the identity matrix I, and ν is the shrinkage parameter that shrinks the sample covariance towards the scaled identity matrix. Again, since there is no extrinsic motion during calibration period, that is, $\mathbf{x}=0$, thus $\mathbf{y}_c=\mathbf{n}$, and we approximate $C^*_{\mathbf{y}_c}$ that is also the sample covariance of $\mathbf{n}$. We follow [20] to automatically calculate parameters ν and μ from the calibration measurements. From the Bayesian perspective,



(a) Experiment 1 (b) Experiment 2

Fig. 2. Pictures of two experiments.

this covariance matrix estimator can be seen as the combination of the prior information and sample information of the covariance matrix.

We can rewrite (16) using (6) and the fact that $UU^T = I$ since U is an orthogonal matrix

$$C_{\mathbf{n}} = \nu\mu UU^T I + (1-\nu)U\Lambda U^T = UDU^T, \tag{17}$$

where D is a diagonal matrix with ith element $\nu\mu + (1-\nu)\lambda_i$. This form allows the inverse to be written as

$$C_{\mathbf{n}}^{-1} = U\tilde{S}^{-1}U^T, \tag{18}$$

where $\tilde{S}^{-1}$ is a diagonal matrix with ith element $\frac{1}{\nu\mu+(1-\nu)\lambda_i}$. In this form, it is clear that with $\nu, \mu > 0$ there will be no divide-by-zero issues in the inverse of the estimated noise covariance matrix.

2.4.3 Covariance of Image

The LSVRT solution also requires the covariance matrix $C_{\mathbf{x}}$. As a means to generate a general statistical model for $C_{\mathbf{x}}$, we assume the positions of people in the environment can be modeled as a Poisson process. Poisson processes are commonly used for modeling the distribution of randomly arranged points in space.

Analysis of Poisson point processes leads to a covariance function that is approximately exponentially decaying [22], and the exponential spatial covariance model is shown to be effective to locate people in an RF sensor network [23]. Thus, in this paper, we use an exponentially-decaying function as the covariance matrix of the human motion

$$[C_{\mathbf{x}}]_{i,j} = \frac{\sigma_x^2}{\delta}\exp\left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|_{l_2}}{\delta}\right), \tag{19}$$

where σ_x^2 is the variance of any element of the image vector $\mathbf{x}$, δ is a space constant, and $\|\mathbf{x}_j - \mathbf{x}_i\|_{l_2}$ is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$.

2.5 Comparison of Two Methods

The SubVRT estimator and the LSVRT estimator are closely related. Both build image estimators using the covariance of the link variance measurements $\mathbf{y}_c$ estimated from calibration measurements. In this section, we show connections between these two estimators.

Recall that SubVRT uses the SVD as given in (6). We can rewrite the extrinsic subspace projector $(I - \hat{U}\hat{U}^T)$ as USU^T, where S is a diagonal matrix with its first k entries set to 0 and remaining entries set to 1. As such, the projection matrix for SubVRT can be written as

$$\Pi_2 = (W^TW + \alpha Q^TQ)^{-1}W^TUSU^T. \tag{20}$$

Next consider the LSVRT estimator from (14). Using (18), we can write the projection matrix for LSVRT as:

$$\Pi_3 = (W^TC_{\mathbf{n}}^{-1}W + C_{\mathbf{x}}^{-1})^{-1}W^TU\tilde{S}^{-1}U^T. \tag{21}$$

Comparing the two, we see that S in (20) is replaced by $\tilde{S}$ in (21). In LSVRT, the ith "principal component" u_i (the ith column of U) is weighted approximately by $1/\lambda_i$, that is, one over the variance of noise in that component. When the variance λ_i is high, the component is downweighted. In comparison in SubVRT, the ith principal component is weighted by 0 if the variance λ_i is one of the k highest; or 1 if not. Essentially, SubVRT approximates $1/\lambda_i$ as 0 or 1, completely zeroing out signals in the k noisiest dimensions and treating signals in other dimensions equally. An advantage with LSVRT is that no parameter (like k in SubVRT) is required to be tuned—recall that ν and μ are automatically determined by the Ledoit-Wolf method.

From the former part of (20) and (21), we see that the LSVRT estimator includes $C_{\mathbf{n}}^{-1}$ as a weight matrix in $W^TC_{\mathbf{n}}^{-1}W$, while the SubVRT estimator just uses W^TW. We also see that the inverse of the covariance matrix $C_{\mathbf{x}}^{-1}$ in the LSVRT solution plays the role of the regularization term αQ^TQ in the SubVRT solution.

In terms of computational complexity, both SubVRT and LSVRT require calculation of the SVD of an $L \times L$ covariance matrix, requiring on the order of L^3 operations. The Ledoit-Wolf estimation of covariance matrix adds additional complexity compared to the sample covariance matrix calculation (5), which is used in SubVRT. However, these calculations are done once when the calibration data has been collected. Real-time estimation of $\hat{\mathbf{x}}$ requires only one matrix multiply which uses at most LP operations, the same as VRTI.

2.6 Online Calibration

Note that the above SubVRT and LSVRT formulations both require calibration measurements to capture the intrinsic

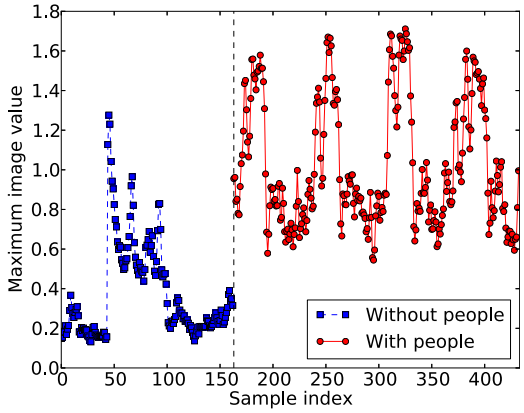


Fig. 3. Highest pixel intensities from VRTI with and without people moving in the network (using data from Experiment 2).

motion. In emergency applications, we cannot expect to be able to perform calibration when the network area is empty.

We propose an online calibration method inspired by background subtraction techniques in the field of computer vision. Many background subtraction techniques use pixel-based features, such as pixel intensity, edges, for background modeling [24], [25]. For VRTI, we find that the highest pixel intensities, when there are people moving in the network area, are typically much higher than those when no people are moving or present in the network, i.e., pixel intensities of background noise. We propose to use pixel intensity to coarsely distinguish between times when a person is or is not present.

We denote the highest pixel intensities of VRTI images when there are people in the network (foreground) and there are no people (background) as $\hat{x}_q^f$ and $\hat{x}_q^b$, respectively. An example of $\hat{x}_q^b$ and $\hat{x}_q^f$ time series is shown in Fig. 3. We see that during the online period when a person is walking inside the network (from sample index 160 to 440), all $\hat{x}_q^f$ values are above 0.5, while most background pixel intensities $\hat{x}_q^b$ are below 0.5. We also notice that the $\hat{x}_q^b$ values from sample index 50 to 100 are higher than those at other sample indexes. This is due to the effect of high wind, that is, intrinsic motion from tree leaves and branches in a high wind period causes significant RSS variations. Thus the highest pixel intensities in this period are much higher than those when the wind is not very strong.

We assume that there will be some periods of time when there are no people moving and propose the following procedure for online calibration. First, we run our SubVRT or LSVRT estimator for incoming link measurement to find $\hat{x}_q$. Note that at the start when no calibration measurements are available, both SubVRT and LSVRT estimators are replaced with VRTI. Whenever the highest pixel intensity $\hat{x}_q$ is below a threshold, we record $\mathbf{y}$ as a calibration measurement. When there are more than a threshold number of calibration measurements,[1] we use them to update the estimator's projection matrix according to SubVRT formulation (11) or LSVRT formulation (14). After we run the estimator, we calculate $\hat{x}_q$, and repeat the procedure above. A flowchart of this online calibration method is shown in Fig. 4.

1. We discuss the effect of all threshold values associated with this online calibration method in Section 4.2.

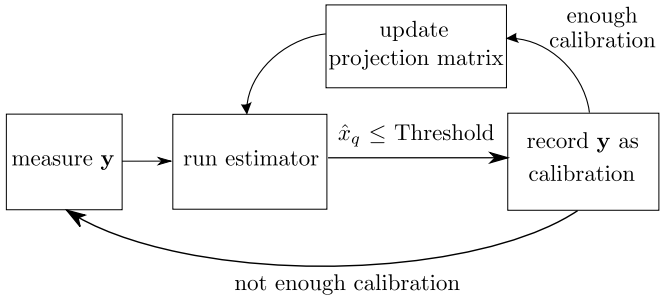


Fig. 4. Flowchart of the online calibration method.

We note that online calibration suffers from a chicken-and-egg problem of estimating the extrinsic motion image and knowing when the data is due solely to intrinsic motion. Sometimes intrinsic motion causes the max image intensity to be above the threshold, and thus that measurement is not included in the calibration. Rarely, extrinsic motion is such that image maximum is low, and it is included in the calibration. However, we show experimentally in Section 4.2 that as long as the threshold is set reasonably, sufficient calibration measurements are recorded during intrinsic motion to enable SubVRT and LSVRT to be more robust to noise.

Further, online calibration requires that, at some point during the online calibration period, no people are moving in the covered area. If people are always moving, then SubVRT and LSVRT would never be able to build a covariance model, and would revert to VRTI. In future work, one might divide a large coverage area into sub-areas, and separately perform online calibration for each sub-area. Alternatively, one may estimate where people are moving, draw large circles around their estimated locations, and then assume that links that do not intersect any circle are affected only by noise [26], and then estimate the covariance matrix using known techniques for estimation with censored data [27]. However, these ideas are not explored in this paper.

2.7 Performance Quantification

To quantify the performance of an image estimator, we use the localization error in situations when one person is present. Multiple people can certainly be tracked using radio tomography [18], [28], but localizing a single person is simple and is sufficient to quantitatively compare the performance of different imaging estimators. When a single person is present, her position $\hat{\mathbf{z}}$ is estimated as the center coordinate of the pixel with maximum value:

$$\hat{\mathbf{z}} = \mathbf{z}_q \quad \text{where} \quad q = \arg\max_p \hat{x}_p, \tag{22}$$

where $\mathbf{z}_q$ is the center coordinate of pixel q and $\hat{x}_p$ is the pth element of image estimate $\hat{\mathbf{x}}$. Then, the localization error is defined as: $e_{loc} = \|\hat{\mathbf{z}} - \mathbf{z}\|_{l_2}$, where $\mathbf{z}$ is the actual position of the person, and $\|\cdot\|_{l_2}$ indicates the Euclidean norm.

3 Experiments

3.1 RF Sensor Network Testbed

We deploy a sensor network of ZigBee TelosB nodes as our testbed. All nodes are programmed with a TinyOS token passing protocol called Spin [29]. At any particular time,

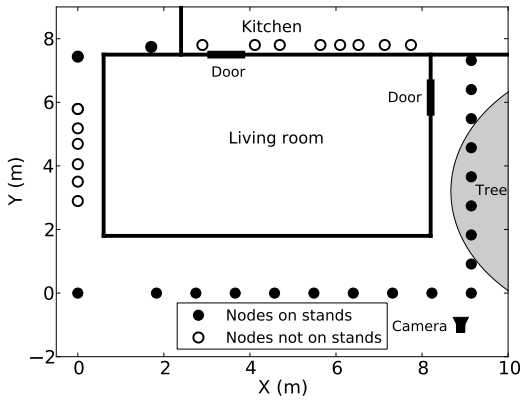


Fig. 5. Map of tested area with node locations. A tree (shaded) is at the right side of the exterior. Video was recorded from camera location indicated.

one node (the node with the token) broadcasts a packet, while all the other nodes measure the RSS of the received packet. A base station connected to a laptop also overhears the packet. The packet contains the transmitter id and a list of all of the RSS values that node has recorded in the last round of Spin. The next node (one higher in transmit id) takes the next turn broadcasting a packet. Nodes transmit in turn until all nodes have transmitted, which completes one round of Spin, and then the process repeats with the first node starting the next round of Spin.

When round t of Spin completes, all pairwise measurements from the network have been overheard by the base station and are recorded to $\{s_{l,t}\}_l$, and variance $\mathbf{y}^{(t)}$ is computed. At this point, a real-time system would immediately calculate an image estimate $\hat{x}$ that would be calculated from (4), (11) or (14). In our experiments, we record $\{s_{l,t}\}_l$ so that all methods can be tested across a range of parameters to show quantitative performance comparisons in Section 4.

Note that the transmission interval between two nodes is set by the Spin protocol so that three link measurements are recorded each second to match the speed of human motion. For faster human motion, we can increase the transmission frequency at the cost of more power consumption. When a packet is not received, or if it fails the packet CRC check, we do not use the RSS measurement for that packet. It is rare in our experimental network to see more than two packet drops in a row, so such data loss causes less than one second of delay on a link.

3.2 Experiment Location

In our experiments, thirty-four TelosB nodes are deployed outside the living room of a residential house. The living room is an addition, so the wall between the kitchen and living room was originally an external wall made of brick. The other walls are Hardy board (a concrete and plastic composite) and wood. As shown in Fig. 5, eight nodes are placed in the kitchen (on the counter), six nodes are placed outside the windows of the living room. The other 20 nodes are all placed on polyvinyl chloride (PVC) stands outside the house.

3.3 Experiment Details

We use measurements from several data sets in this paper. In all data sets, we first record data from the sensors during an empty-room calibration period. Next, a person walks around a marked path (A-B-C-D as shown in Fig. 15) in the living room at a constant speed of about 0.5 m/s, using a metronome and a metered path so that the position of the person at any particular time is known. A variety of intrinsic noise conditions are tested in three experiments:

- *Experiment 1*: We use the data set from the measurements conducted in March, 2009 reported by [8], which was performed in the same house with identical setup, as our "Experiment 1". Experiment 1 is performed on a clear winter day with no wind. As shown in our video [30] (a snapshot is shown in Fig. 2a), there are no leaves on branches, and no wind is observed during Experiment 1. The lack of leaves or wind make this data set contain the least intrinsic noise of any of those that we collected. For this experiment, have 47 seconds ($M = 140$) of measurements collected during a calibration period.
- *Experiment 2*: We conduct another experiment at the same house with identical setup on a windy day in May, with leaves on the trees and long grass in the yard next to the house. From video recorded during Experiment 2 (one snapshot is shown in Fig. 2b), one can see that wind causes grass, leaves and branches to sway [30]. The wind also causes the PVC stands supporting the nodes to move. For this experiment, we collect one minute ($M = 170$) of measurements during a calibration period.
- *Experiment 3*: The first two experiments contained intrinsic motion that we did not control. In our final experiment, we attempt to *control* intrinsic motion using four electric fans inside of the living room. The experiment is conducted on a windy summer day. The first data set (3a) is recorded with no fans on. The second data set (3b) is recorded with one rotating fan[2] turned on. The third data set (3c) is recorded with all four fans on. There are plants and objects in the room which move somewhat when the fan blows on them. For each data set 3a, 3b, and 3c, we collect one minute ($M = 170$) of measurements during a calibration period.

4 Results

4.1 Eigenvalues and Eigen-Networks

We propose in Section 2.3 that most of the intrinsic noise can be contained within a low-dimensional subspace. To verify this via experiment, we use PCA on the calibration measurements from Experiments 1 and 2. The eigenvalues of $C_{\mathbf{y}_c}$ from these two experiments are shown in Fig. 6. Because there is more intrinsic motion in Experiment 2, we see that the largest eigenvalue from Experiment 2 is almost twice as large as that from Experiment 1. For Experiment 1, the first four eigenvalues are much larger than the other eigenvalues, thus the subspace spanned by the four corresponding eigenvectors can capture most of the intrinsic signal. However, for Experiment 2, there are more large-valued

2. The fan is motor-controlled to change its facing direction, in addition to rotating the fan blades.

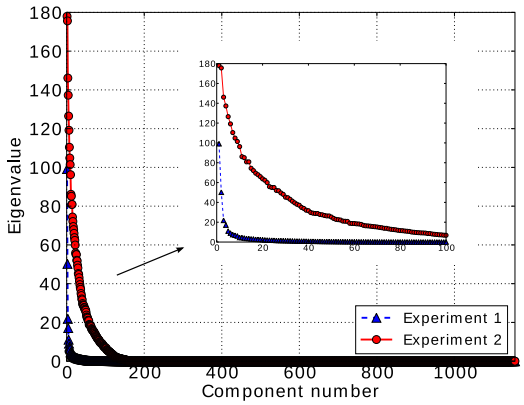


Fig. 6. Scree plot.

eigenvalues. Thus more eigenvectors are necessary to capture the majority of the variations in the measurements.

Since each of the principal components used to construct the intrinsic subspace is an eigenvector of the covariance matrix of the network measurements, and each element in an eigenvector is from an individual link, we refer these eigenvectors $\mathbf{u}_i$ as "eigen-networks". The first eigen-network $\mathbf{u}_1 = [u_{11}, u_{12}, \ldots, u_{1L}]^T$ points in the direction of the maximum variance of the calibration measurements $\mathbf{y}_c$. We show the first eigen-network $\mathbf{u}_1$ graphically in Fig. 7. We see the links with u_{1l} values higher than 30 percent of the maximum value are all in the lower right side of the house. This is consistent with our observation that the leaves and branches on the tree located to the right side of the house cause significant temporal changes in the RSS measured on links likely to propagate through the branches and leaves. Note that links with high u_{1l} values all have at least one end point near the tree. In particular, links which are likely to see significant diffraction around the bottom-right corner of the house have high u_{1l} values. The leaves and branches almost touch this corner, as seen in Fig. 2b. Not only do these links measure high RSS variance during the calibration period, they do so simultaneously. That is, the fact that these links have high positive u_{1l} values indicates that when one of these links experiences increased RSS variance, the other links also measure increased RSS variance. Thus, the first eigen-network $\mathbf{u}_1$ becomes a spatial signature for intrinsic motion-induced RSS variance. When we see this linear combination in $\mathbf{y}_r$, we should attribute it to intrinsic, rather than extrinsic motion. These observations about the source of RSS variance on links support the idea that intrinsic motion in the environment causes increased RSS variance simultaneously on multiple links.

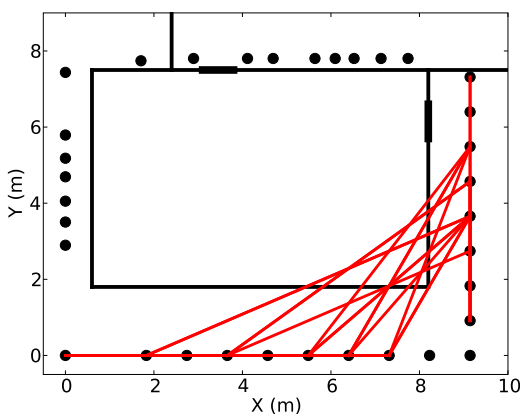


Fig. 7. First eigen-network: Links with $u_{1l} > 30$ percent of $max_l\ u_{1l}$.

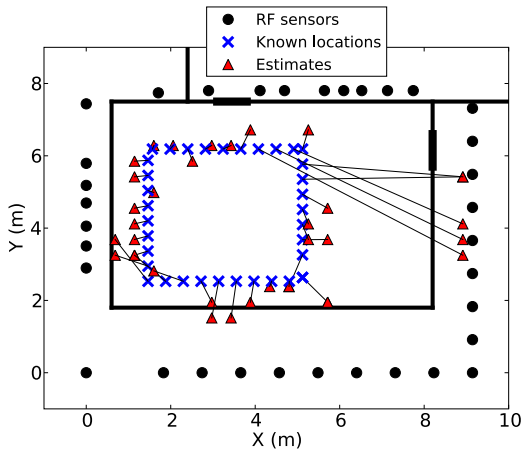


Fig. 8. Estimates from VRTI using measurements recorded when a person walks the last round of the square path in Experiment 2.

4.2 Localization Results

In this section, we compare the performance of VRTI, SubVRT and LSVRT.

First, we compare in detail how LSVRT and VRTI perform in a particular segment of Experiment 2. The VRTI estimates of Experiment 2 are shown in Fig. 8. For clarity, we only show the actual/estimated positions when the person walks one round (out of the four rounds) of the square. We show the last round because it is particularly affected by wind and thus illustrates the effects of intrinsic noise on VRTI. In Fig. 8, some VRTI estimates are greatly biased to the right side of the experimental area (i.e., five estimates > 4.0 m error). However, for LSVRT, the impact of intrinsic motion is greatly reduced. As shown in Fig. 9, the estimates from LSVRT are more accurate than VRTI. There are no estimate errors larger than 2.0 m.

Note that some estimates are outside the house. The algorithms presented do not include any prior information of the house map or physical barriers which would prevent certain trajectories. Incorporation of such prior knowledge

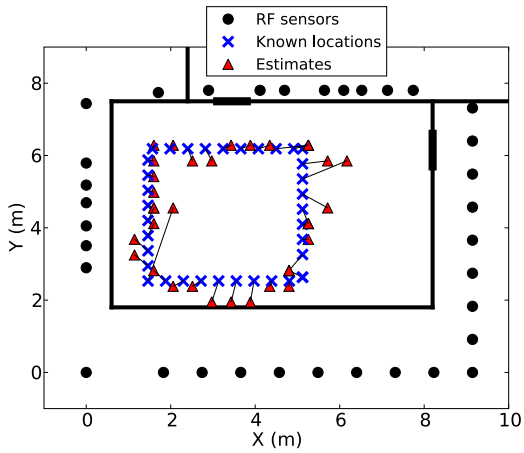


Fig. 9. Estimates from LSVRT using the same measurements as used in Fig. 8.

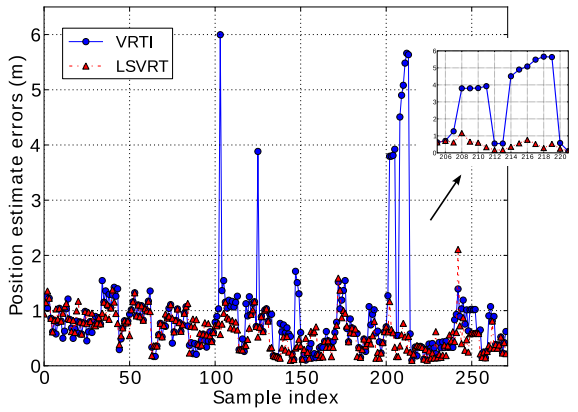


Fig. 10. Estimate errors from VRTI and LSVRT.

might be used to obtain better estimates, but at the expense of requiring more information to deploy the system.

Quantitatively, we next compare the localization errors from VRTI, SubVRT and LSVRT for the full data sets. The comparison between VRTI and SubVRT is shown in Fig. 9 of [12], and the comparison between VRTI and LSVRT is shown in Fig. 10. The localization errors from SubVRT are all below 1.8 m. For VRTI, there are several estimates with errors above 3.0 m. These large errors are due to the impact of intrinsic motion on static link measurements. Specifically, we compare the localization errors during a period with strong wind, from sample index 205 to 221, as shown in the inset of Fig. 9 of [12]. During this period, the average localization error from VRTI is 3.0 m, while the average error from SubVRT is 0.62 m, a 79 percent improvement, and for LSVRT, it is only 0.50 m, a 83 percent improvement.

We also compare the RMSE of the estimates, which is defined as the square root of the average squared localization error over the course of the entire experiment. The RMSEs from all of the experiments are summarized in Table 1. For Experiment 1, the RMSE from VRTI is 0.70 m, while the RMSE from SubVRT is 0.65 m, a 7.0 percent improvement and the RMSE from LSVRT is 0.63 m, a 9.6 percent improvement. For Experiment 2, the RMSE from VRTI is 1.26 m, while SubVRT and LSVRT are more robust to impact of intrinsic motion. The RMSE from SubVRT is 0.74 m, a 41.3 percent improvement, and the RMSE from LSVRT is 0.69 m, a 45.3 percent improvement.

For the third experiment, the RMSEs are generally higher for all methods. We know that the small-scale position of sensors in the network can significantly affect overall localization performance [31]. We suspect that this deployment had more links in deep fades.

The results from Experiment 3 show two things consistently:

1) LSVRT consistently has superior performance compared to SubVRT. The improvement in RMSEs for LSVRT are 14 to 19 percent better than SubVRT for each of 3a, 3b, and 3c, as seen in Table 1.
2) Compared to data set 3a in which no fans were on, the improvement in RMSE increases as the quantity of intrinsic motion increases (as more fans are turned on) in 3b and then in 3c. The improvement increases in 3b compared to 3a, and in 3c compared to 3b, as seen in Table 1.

TABLE 1
Localization RMSEs (in meter) from VRTI, SubVRT, and LSVRT

Methods	VRTI	SubVRT		LSVRT	
Results	RMSE	RMSE	Improve	RMSE	Improve
Exp. 1	0.70	0.65	7.0%	0.63	9.6%
Exp. 2	1.26	0.74	41.3%	0.69	45.3%
Exp. 3a	1.31	1.28	2.3%	1.10	16.0%
Exp. 3b	1.44	1.15	20.1%	0.88	38.9%
Exp. 3c	1.89	1.35	28.6%	1.0	47.1%

TABLE 2
Localization RMSEs (in meter) from SubVRT and LSVRT Using Online and Offline Calibration

Methods	SubVRT		LSVRT	
Results	Online	Offline	Online	Offline
Exp. 1	0.66	0.65	0.64	0.63
Exp. 2	0.77	0.74	0.70	0.69
Exp. 3a	1.31	1.28	1.07	1.10
Exp. 3b	1.03	1.15	0.77	0.88
Exp. 3c	1.32	1.35	0.99	1.0

We believe that during the one-minute calibration for Experiment 3a, there was very little wind present, even though there was wind later while the person was moving in the room. As a result, for Experiment 3a, SubVRT and LSVRT show small reductions in RMSE compared to VRTI.

Finally, we compare the localization RMSEs from SubVRT and LSVRT using online and offline calibration methods in Table 2. We see that the RMSEs from both estimators using online calibration are very similar to those using offline calibration for all experiments. Remember that in our online calibration method, we perform calibration only after enough measurements are collected, as described in Section 2.6. We find as long as the number of measurements used in the calibration is above 100 (which corresponds to about 30 seconds) the performance of both estimators using online calibration is significantly better than VRTI. Thus, we choose the threshold of the number of calibration measurements to be 100 in our tests. We note that a result is that new intrinsic noise conditions, for example a new weather condition, would take at least the duration of the calibration (here, 30 s) before the system properly reduced its intrinsic noise.

Also recall that we use an intensity threshold value in online calibration. Here, we test the effect of this threshold on the localization accuracy. We choose different threshold values and show RMSEs from Experiments 1 and 2 in Fig. 11. We see that as long as the threshold value is in the range of 0.3 to 0.8, RMSE is not sensitive to its value. For values < 0.3, very little data will be available for estimation of the covariance matrix. For thresholds > 0.8, many data points from the period with a person present are included, which can lead SubVRT and LSVRT to learn to ignore data corresponding to actual people in the area, and lead to increased RMSE. When the threshold is between 0.3 and 0.8, there is sufficient intrinsic noise and little extrinsic noise, which allows the two methods to reduce the RMSE. From Fig. 3, we see that all highest pixel intensity values

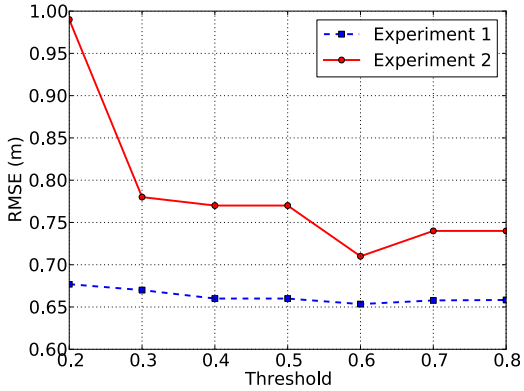


Fig. 11. RMSEs vs. pixel intensity threshold values.

with people in the network are above 0.5, thus we simply choose the threshold value to be 0.5 in our tests.

We note that adaptive thresholding for determining when an area is empty has been developed for mean-based RTI [28], but is not directly applicable to VRTI because people disappear when they stop moving. Future work could explore applying such adaptive methods in applications in which image intensities are unpredictable.

In sum, our SubVRT and LSVRT estimators using online calibration achieve approximately equivalent localization accuracy compared to using offline calibration.

4.3 Discussion

The parameters that we use in VRTI, SubVRT and LSVRT are listed in Table 3. We show the effect of the number of nodes on these three algorithms. We also discuss the effects of the number of principal components k and the image variance parameter σ_x^2 on the performances of SubVRT and LSVRT, respectively.

To see the effect of node number on the localization performance, we run VRTI, SubVRT and LSVRT algorithms using RSS measurements from only a randomly chosen subset N less than the 34 total nodes used in Experiment 1. For example, when we use $N = 20$ nodes, we randomly choose 20 of the measured nodes, and then run our algorithms using the RSS measurements collected between pairs of these 20 nodes. For each N, we repeat the above procedure 100 times, and each time calculate the RMSEs of the position estimates. The average RMSEs and the RMSE standard deviations of the three algorithms from Experiment 1 are shown in Fig. 12, for $N = 20$ to 34 (a node density of 0.27 per m^2 to 0.47 per m^2). We find if we only use 26 nodes ($L = 650$) to cover this 9 m by 8 m area, the average RMSEs from three algorithms are all above

TABLE 3
Parameters in VRTI, SubVRT, LSVRT, and Kalman Filter

Param.	Value	Description
α	100	Regularization parameter
m	4	Window length to calculate variance
k	4/40	Principal components, Exp. 1 / Exp. 2, 3
σ_x^2	0.001	Image variance parameter
σ_w^2	2	Process noise parameter
σ_v^2	5	Measurement noise parameter

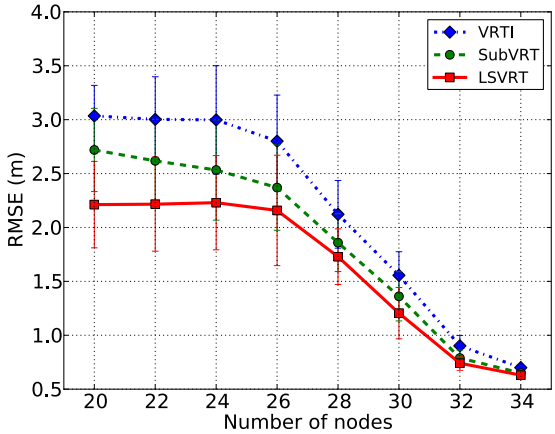


Fig. 12. Localization RMSEs from Experiment 1 vs. number of nodes.

2 m. Comparing results from $N = 26$ vs. $N = 32$, the RMSE reduces by a factor of $3 - 3.6$ for the three methods. For all methods, increasing N may lead to diminishing returns beyond $N = 32$. We also find that the performance of LSVRT is consistently better than SubVRT and VRTI independent of number of nodes.

An important parameter for SubVRT is the number of principal components used to construct the intrinsic subspace. As discussed in Section 2.3, the first k components are used to calculate the projection matrix for the intrinsic subspace Π_I. If $k = 0$, $\Pi_I = 0$, then $\Pi_1 = \Pi_2$, SubVRT is simplified to VRTI. The RMSEs of SubVRT using a range of k are shown in Fig. 13. Since the first eigen-network $\mathbf{u}_1$ captures the strongest intrinsic signal, when $k = 1$, the RMSE of Experiment 2 decreases substantially from 1.26 m to 0.82 m. Since Experiment 1 has less intrinsic motion, the RMSE decreases from 0.70 m when $k = 0$ to 0.65 m when $k = 4$, a less substantial decrease. We note that as k increases, more and more information in the measurement is removed, and the RMSE stops decreasing dramatically, and even increases, at certain k. This is because when k becomes very large, the information removed also contains a great amount of signal caused by extrinsic (human) motion. Thus, the performance of SubVRT could be degraded if k is chosen to be too large. The parameter k is a tradeoff between removing intrinsic motion impact and keeping useful information from extrinsic motion. For experiments without much intrinsic motion, such as Experiment 1, we choose a small k.

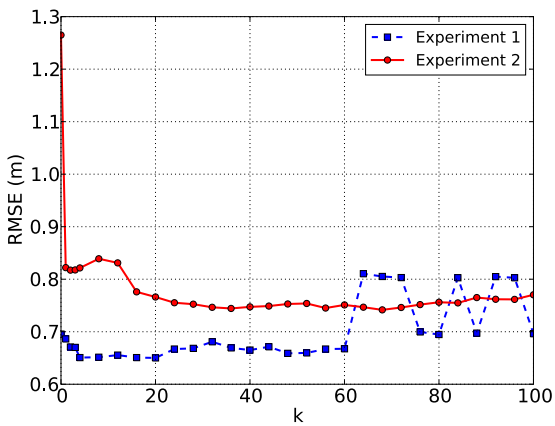


Fig. 13. Localization RMSEs vs. principal component number k.

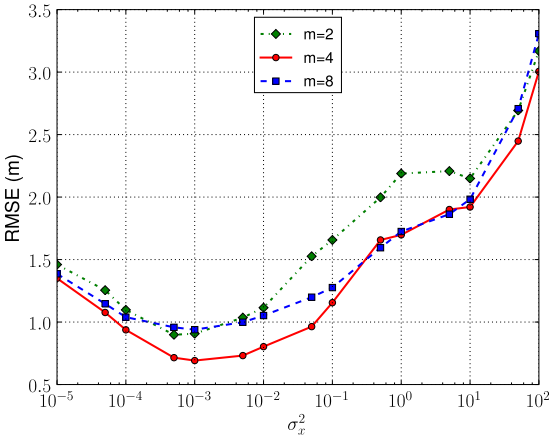


Fig. 14. Localization RMSEs vs. σ_x^2.

However, for Experiment 2, with strong impact from intrinsic motion, we use a large k. As listed in Table 3, we use $k = 4$ and $k = 40$ for all experiments.

An advantage of LSVRT over SubVRT is that LSVRT can change its parameters automatically based on calibration measurements, thus we do not need to manually tune parameters like k in SubVRT. Thus, we only investigate parameter σ_x^2 in LSVRT, which plays the same role of the regularization parameter α in SubVRT. From Fig. 14, we see the RMSE from LSVRT reaches the minimum at 0.63 m, when $\sigma_x^2 = 0.001$ and $m = 4$. However, the localization RMSEs from LSVRT are shallow functions of σ_x^2 in the range from 10^{-4} to 10^{-1}. That is, LSVRT is not very sensitive to this regularization parameter in a wide range. Another advantage of LSVRT is that its localization accuracy is higher than SubVRT for all five experiments, as listed in Table 1. As discussed in Section 2.4, the inverse of the covariance matrix $C_{\mathbf{x}}$ is used as the regularization term in the LSVRT formulation. This regularization scheme provides better smoothing of the images, compared to the regularization term (difference operations of the image) in VRTI and SubVRT formulations. Thus, the motion images from LSVRT are generally cleaner than those from VRTI and SubVRT, and LSVRT has even better localization accuracy than SubVRT.

5 Tracking

In this section, we apply a Kalman filter to the localization estimates shown in Section 4.2 to better estimate moving people's positions over time. Then, we compare the tracking results from VRTI with those from SubVRT and LSVRT, and show that the Kalman filter tracking results from SubVRT and LSVRT are more robust to large localization errors.

5.1 Kalman Filter

In the state transition model of the Kalman filter, we include both position (P_x, P_y) and velocity (V_x, V_y) in the Cartesian coordinate system in the state vector $\mathbf{s} = [P_x, P_y, V_x, V_y]^T$, and the state transition model is

$$\mathbf{s}[t] = G\mathbf{s}[t-1] + \mathbf{w}[t], \tag{23}$$

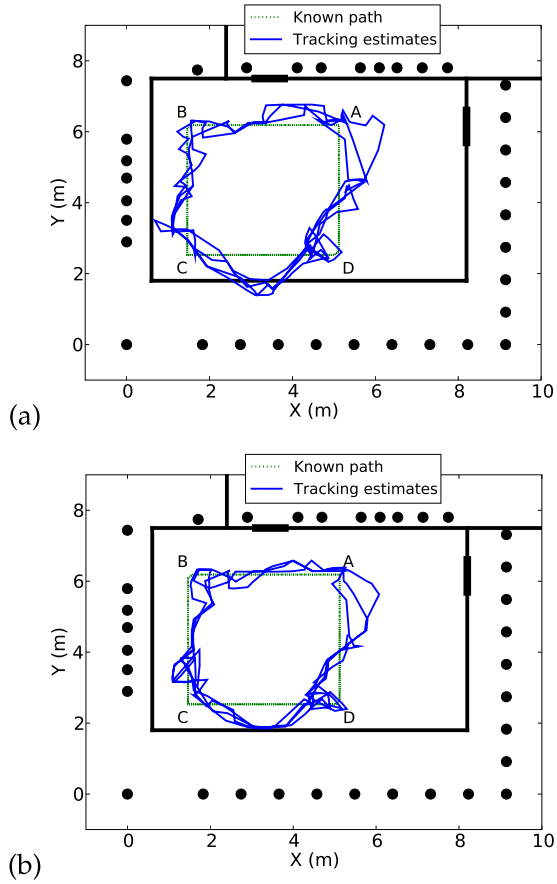


Fig. 15. Kalman filter tracking results of Experiment 2 from SubVRT (a) and LSVRT (b).

where $\mathbf{w} = [0, 0, w_x, w_y]^T$ is the process noise, and G is

$$G = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{24}$$

The observation inputs $\mathbf{r}[t]$ of the Kalman filter are the localization estimates from VRTI, SubVRT or LSVRT at time t, and the observation model is

$$\mathbf{r}[t] = H\mathbf{s}[t] + \mathbf{v}[t], \tag{25}$$

where $\mathbf{v} = [v_x, v_y]^T$ is the measurement noise, and H is

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{26}$$

In the Kalman filter, v_x and v_y are zero-mean Gaussian with variance σ_v^2, w_x and w_y are zero-mean Gaussian with variance σ_w^2 [32]. The parameters σ_v^2 and σ_w^2 of the measurement noise and process noise are listed in Table 3.

5.2 Tracking Results

We use the Kalman filter described above to track the positions of the person. The tracking results of Experiment 2 from SubVRT and LSVRT are shown in Fig. 15. We see that for Experiment 2, with significant intrinsic motion, the Kalman filter tracking results using SubVRT and LSVRT estimates generally have errors less than 1 meter. The tracking

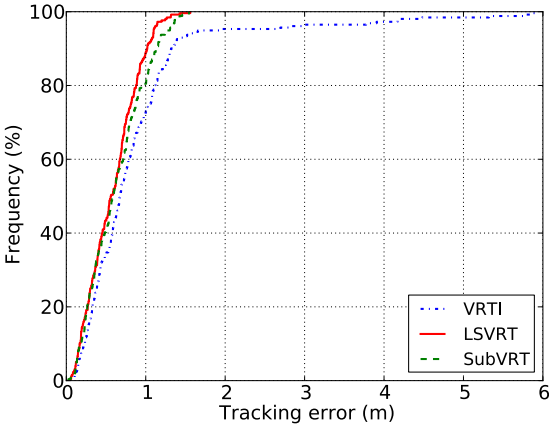


Fig. 16. CDFs of tracking errors.

results from LSVRT, proposed in this paper, are even better than those from SubVRT. The cumulative distribution functions (CDFs) of the tracking errors from Experiment 2 are shown in Fig. 16. We see that the Kalman filter tracking results from VRTI have many more large errors than SubVRT and LSVRT. For VRTI, 97 percent of the tracking errors are less than 3.91 m, while 97 percent of the tracking errors from SubVRT are less than 1.36 m, a 65.2 percent improvement. For LSVRT, 97 percent of the errors are less than 1.15 m, a 70.6 percent improvement compared to VRTI. We use the 97th percentile of errors to show the robustness of the tracking algorithm to large errors, and the CDFs show the tracking results from SubVRT and LSVRT are more robust to these large errors.

We also compare the RMSEs of the tracking results from VRTI, SubVRT and LSVRT, which are listed in Table 4. For Experiment 1, the tracking RMSEs from SubVRT and LSVRT are both 0.57 m, a 13.6 percent improvement compared to the RMSE of 0.66 m from VRTI. For Experiment 2, the tracking RMSE from SubVRT is reduced by 40.5 percent to 0.72 m compared to 1.21 m RMSE from VRTI, and the RMSE from LSVRT is reduced by 45.5 percent to 0.66 m. We note that the tracking RMSEs from VRTI, SubVRT and LSVRT of Experiment 2 are both larger than Experiment 1 due to the impact of intrinsic motion. However, for VRTI the tracking RMSE from Experiment 2 has a 83.3 percent increase compared to Experiment 1, while for SubVRT and LSVRT, they only increase 26.3 and 15.8 percent, respectively. For the data sets recorded during Experiment 3, the tracking errors from LSVRT are all below 1 meter. Overall, the tracking RMSEs from SubVRT and LSVRT are more robust to intrinsic motion than VRTI, with LSVRT performing better than SubVRT. Also, the performance gain vs. VRTI increases as the quantity of intrinsic motion increases.

6 Related Work

For device-free localization of people in wireless sensor networks, different measurements, algorithms and frameworks have been proposed [6], [8], [9], [33], [34]. For RSS-based localization, we may divide methods into *fingerprint-based* and *model-based* algorithms. Like fingerprint-based real-time location service (RTLS) systems, fingerprint-based device-free localization methods use a database of training measurements, and estimate a person's location by comparing a measurement taken during the online phase with the training measurement database [35], [36], [37], [38]. To collect one training measurement, a person stands in one position while link measurements are made, and associates the collected data with her known coordinate. Several such measurements are made throughout the area of interest. One active area of research is developing tracking methods for multiple people which do not require an exponential increase in the number of training measurements that must be made [36]. Another area is the use of channel state information (instead of RSS) for fingerprint-based DFL [37], [38]. Finally, significant research addresses fingerprint-based activity recognition rather than localization alone [39], [40]. Since fingerprint-based methods require a person to perform training at many known locations inside the area of interest, it is not suitable in emergency scenarios in which presence in the area of interest is dangerous.

TABLE 4
Tracking RMSEs (in Meter) from VRTI, SubVRT, and LSVRT

Methods	VRTI	SubVRT		LSVRT	
Results	RMSE	RMSE	Improve	RMSE	Improve
Exp. 1	0.66	0.57	13.6%	0.57	13.6%
Exp. 2	1.21	0.72	40.5%	0.66	45.5%
Exp. 3a	1.17	1.15	1.7%	0.98	16.2%
Exp. 3b	1.31	0.98	25.2%	0.79	39.7%
Exp. 3c	1.86	1.33	28.5%	0.99	46.8%

Model-based algorithms [8], [18], [23], [26], [41], [42], [43], [44], [45] require a model for the relationship between changes in RSS as a function of people's positions. With this forward model, these methods solve localization as an inverse problem. Model-based algorithms typically require 1) calibration measurements taken when the area is empty of people, and 2) known node coordinates. However, VRTI does not require calibration, and methods have been developed to learn the "empty" condition of each link [25], [45]. Regarding 2), node self-localization methods such as GPS may provide sufficient node localization for RTI [46]; additionally, a model-based algorithm may simultaneously perform device-free localization and improve the node location estimates [18].

Radio tomographic imaging (RTI) is a particular type of model-based algorithm which describes a statistic of the measured RSS as a linear combination of the effect caused by each pixel in an area. Mean-based RTI relates link attenuation to the loss due to people or objects in the environment [23], [43], but does not perform well in non-LOS multipath-rich environments. Further, static building structure (walls and open areas) can be mapped [47]. Variance-based RTI (VRTI) [8] relates link attenuation to motion in the environment, and does not require any calibration. As such, it is well-suited for emergency applications. However, VRTI cannot locate people if they stand still without any motion, and it is sensitive to other motion in the environment, as shown in this paper.

Bistatic radar can be used to detect and locate motion from behind walls using bistatic WiFi signals inside of the building [48]. Monostatic or multi-antenna through-wall

radar systems [49], [50] operate with one or very few devices, but use specialized receivers and require GHz of bandwidth.

In this paper, we apply subspace decomposition and least squares-based formulations to reduce the impact of intrinsic noise in VRTI. We note that noise reduction would presumably benefit both fingerprint-based and model-based algorithms, and future work may apply these techniques to other device-free localization methods.

7 Conclusion

In this paper, we propose to use subspace decomposition and least squares methods to reduce noise for variance-based device-free localization and tracking. We discuss how intrinsic motion, such as leaves moving in the wind, increase measured RSS variance in a way that is "noise" to a localization system. Our new estimator LSVRT outperforms SubVRT in localization accuracy, and it requires tuning of fewer parameters compared to SubVRT. We also propose a new online calibration method so that both SubVRT and LSVRT can use real-time online measurements to perform calibration instead of using "empty-area" offline calibration measurements. We perform new experiments to better investigate the effect of intrinsic motion and the performance of SubVRT and LSVRT. Experimental results show that SubVRT and LSVRT can reduce localization RMSE by more than 40 percent, and our online calibration method achieves similar localization accuracy as offline calibration. We further apply a Kalman filter on SubVRT and LSVRT estimates for tracking and show that SubVRT and LSVRT are more robust to large errors than VRTI.

Acknowledgments

This material is based upon work supported by the US National Science Foundation under Grant Nos. #0748206 and #1035565. Yang Zhao is the corresponding author.

References

[1] N. Patwari and J. Wilson, "RF sensor networks for device-free localization: Measurements, models and algorithms," *Proc. IEEE*, vol. 98, no. 11, pp. 1961–1973, Nov. 2010.

[2] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Comput. Netw.*, vol. 51, no. 10, pp. 2529–2553, 2007.

[3] C. Chang and A. Sahai, "Object tracking in a 2D UWB sensor network," in *Proc. 38th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2004, vol. 1, pp. 1252–1256.

[4] L.-P. Song, C. Yu, and Q. H. Liu, "Through-wall imaging (TWI) by radar: 2-D tomographic results and analyses," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 12, pp. 2793–2798, Dec. 2005.

[5] M. C. Wicks, "RF tomography with application to ground penetrating radar," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 2017–2022.

[6] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. ACM Int. Conf. Mobile Comput. Netw.*, 2007, pp. 222–229.

[7] A. M. Haimovich, R. S. Blum, and L. J. Cimini, "MIMO radar with widely separated antennas," *IEEE Signal Process.*, vol. 25, no. 1, pp. 116–129, Jan. 2008.

[8] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 612–621, May 2011.

[9] D. Zhang, J. Ma, Q. Chen, and L. M. Ni, "An RF-based system for tracking transceiver-free objects," in *Proc. 5th Annu. IEEE Int. Conf. Pervasive Comput. Commun.*, 2007, pp. 135–144.

[10] B. Sklar, "Rayleigh fading channels in mobile digital communication systems .i. characterization," *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 90–100, Jul. 1997.

[11] G. D. Durgin, *Space-Time Wireless Channels*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[12] Y. Zhao and N. Patwari, "Noise reduction for variance-based device-free localization and tracking," in *Proc. 8th IEEE Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Salt Lake City, Utah, USA, Jun. 2011, pp. 179–187.

[13] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997.

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[15] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[16] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, Aug. 2004, pp. 219–230.

[17] I. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[18] S. Nannuru, Y. Li, Y. Zeng, M. Coates, and B. Yang, "Radio-frequency tomography for passive indoor multitarget tracking," *IEEE Trans. Mobile Comput.*, vol. 12, no. 12, pp. 2322–2333, Dec. 2013.

[19] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004.

[20] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, pp. 365–411, Feb. 2004.

[21] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero III, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.

[22] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4024–4036, Aug. 2009.

[23] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, May 2010.

[24] A. Elgammal, R. Duraiswami, D. Harwood, L. S. Davis, R. Duraiswami, and D. Harwood, "Background and foreground modeling using nonparametric kernel density for visual surveillance," in *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.

[25] A. Edelstein and M. Rabbat, "Background subtraction for online calibration of baseline RSS in RF sensing networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 12, pp. 2386–2398, Dec. 2013.

[26] M. Bocca, O. Kaltiokallio, and N. Patwari, "Radio tomographic imaging for ambient assisted living," in *Proc. Int. Competitions Final Workshop Eval. AAL Syst. Competitive Benchmarking*, 2013, pp. 108–130.

[27] H. Ng, P. Chan, and N. Balakrishnan, "Estimation of parameters from progressively censored data using em algorithm," *Comput. Statist. Data Anal.*, vol. 39, no. 4, pp. 371–386, 2002.

[28] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian, "Multiple target tracking with RF sensor networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 8, pp. 1787–1800, Aug. 2014.

[29] Sensing and Processing Across Networks (SPAN) Lab, Spin website [Online]. Available: http://span.ece.utah.edu/spin, 2014.

[30] Sensing and Processing Across Networks (SPAN) Lab, RTI website [Online]. Available: http://span.ece.utah.edu/radio-tomographic-imaging, 2014.

[31] M. Bocca, A. Luong, N. Patwari, and T. Schmid, "Dial it in: Rotating RF sensors to enhance radio tomography," in *Proc. IEEE Int. Conf. Sensing, Commun. Netw.*, 2014, pp. 600–608.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[33] F. Viani, P. Rocca, G. Oliveri, D. Trinchero, and A. Massa, "Localization, tracking, and imaging of targets in wireless sensor networks: An invited review," *Radio Sci.*, vol. 46, no. 5, pp. 1–12, 2010.

[34] J. Wilson and N. Patwari, "A fade level skew-laplace signal strength model for device-free localization with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 947–958, Jun. 2011.

[35] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef, "Nuzzer: A large-scale device-free passive localization system for wireless environments," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1321–1334, Jul. 2013.
[36] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *Proc. 12th Int. Conf. Inf. Process. Sensing Netw.*, Apr. 2013, pp. 79–90.
[37] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst.*, 2013, pp. 236–245.
[38] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, "Monophy: Mono-stream-based device-free WLAN localization via physical layer information," in *Proc. Wireless Commun. Netw. Conf.*, 2013, pp. 4546–4551.
[39] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Feb. 2014.
[40] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Int. Conf. Mobile Comput. Netw.*, Sept. 2013, pp. 27–38.
[41] M. A. Kanso and M. G. Rabbat, "Compressed RF tomography for wireless sensor networks: Centralized and decentralized approaches," in *Proc. 5th IEEE Int. Conf. Distrib. Comput. Sensing Syst.*, Jun. 2009, pp. 173–186.
[42] O. Kaltiokallio and M. Bocca, "Real-time intrusion detection and tracking in indoor environment through distributed RSSI processing," in *Proc. IEEE 17th Int. Conf. Embedded Real-Time Comput. Syst. Appl.*, Aug. 2011, pp. 61–70.
[43] R. K. Martin, C. Anderson, R. W. Thomas, and A. S. King, "Modelling and analysis of radio tomography," in *Proc. 4th IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, Dec. 2011, pp. 377–380.
[44] O. Kaltiokallio, M. Bocca, and N. Patwari, "Follow @grandma: Long-term device-free localization for residential monitoring," in *Proc. 7th IEEE Int. Workshop Practical Issues Building Sen. Netw. Appl.*, Oct. 2012, pp. 984–991.
[45] Y. Zheng and A. Men, "Through-wall tracking with radio tomography networks using foreground detection," in *Proc. Wireless Commun. Netw. Conf.*, Paris, France, Apr. 2012, pp. 3278–3283.
[46] D. Maas, J. Wilson, and N. Patwari, "Toward a rapidly deployable RTI system for tactical operations," in *Proc. 8th IEEE Workshop Prac. Issues Building Sensor Netw. Appl.*, Oct. 2013, pp. 203–210.
[47] Y. Mostofi, "Compressive cooperative sensing and mapping in mobile networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 12, pp. 1769–1784, Dec. 2011.
[48] K. Chetty, G. Smith, H. Guo, and K. Woodbridge, "Target detection in high clutter using passive bistatic Wi-Fi radar," in *Proc. IEEE Radar Conf.*, May 2009, pp. 1–5.
[49] F. Adib and D. Katabi, "See through walls with Wi-Fi!" in *Proc. ACM SIGCOMM Conf. SIGCOMM*, Aug. 2013, pp. 75–86.
[50] A. R. Hunt, "Image formation through walls using a distributed radar sensor array," in *Proc. 32nd Applied Imagery Pattern Recognition Workshop*, Oct. 2003, pp. 232–237.

Yang Zhao received the BS degree from Shandong University in 2003, the MS degree from the Beijing University of Aeronautics and Astronautics in 2006, and the PhD degree from the University of Utah, 2012, all in electrical engineering. He was a research assistant at the Sensing and Processing Across Networks (SPAN) lab of the University of Utah from 2009 to 2012. He was a research intern at the Human Computer Interaction (HCI) group of Microsoft Research Asia during the spring of 2012. Since 2013, he has been a research engineer at Software Sciences and Analytics (SSA) organization of General Electric Global Research Center in Niskayuna, NY. His research interests include signal processing, sensor networks, machine-to-machine.

Neal Patwari received the BS and the MS degrees from Virginia Tech in 1997 and 1999, and the PhD degree from the University of Michigan, Ann Arbor, in 2005, all in electrical engineering. He was a research engineer at Motorola Labs, FL, between 1999 and 2001. He is an associate professor in the Department of Electrical and Computer Engineering at the University of Utah. He directs the Sensing and Processing Across Networks (SPAN) Lab, which performs research at the intersection of statistical signal processing and wireless networking. He is also the director of research at Xandem Technology. He has received the Best Paper Awards from the IEEE Signal Processing Magazine in 2009, SenseApp in 2012, and IPSN in 2014, and the 2011 University of Utah Early Career Teaching Award. He is an associate editor of the *IEEE Transactions on Mobile Computing*.