Chalmers Publication Library

Copyright Notice

(Article begins on next page)

A Machine Learning Approach to Ranging Error Mitigation for UWB Localization

Henk Wymeersch, *Member, IEEE*, Stefano Maranò, *Student Member, IEEE*,
Wesley M. Gifford, *Student Member, IEEE*, Moe Z. Win, *Fellow, IEEE*

Abstract—Location-awareness is becoming increasingly important in wireless networks. Indoor localization can be enabled through wideband or ultra-wide bandwidth (UWB) transmission, due to its fine delay resolution and obstacle-penetration capabilities. A major hurdle is the presence of obstacles that block the line-of-sight (LOS) path between devices, affecting ranging performance and, in turn, localization accuracy. Many techniques have been proposed to address this issue, most of which make modifications to the localization algorithm. Since many localization algorithms work with distance or angle estimates, rather than received waveforms, information inherent in the wideband waveform is lost, leading to sub-optimal ranging error mitigation. To avoid this information loss, we present a novel approach to mitigate ranging errors directly in the physical layer. In contrast to existing techniques, which *detect* the non-line-of-sight (NLOS) condition, our approach directly *mitigates* the bias incurred in both LOS and non-LOS conditions. In particular, we apply two classes of non-parametric regressors to form an estimate of the ranging error. Our work is based on, and validated by, an extensive indoor measurement campaign with FCC-compliant UWB radios. The results show that the proposed regressors provide significant performance improvements in various practical localization scenarios, compared to conventional approaches.

Index Terms—Localization, UWB, Ranging Error Mitigation, Support Vector Machine, Gaussian Processes, Bayesian Learning.

I. INTRODUCTION

THE ability to locate people and assets, to navigate beyond GPS coverage, and to tag sensor data with geographical information will enable a myriad of applications, in both the commercial and the military sectors [1]–[4]. Ultra-wide bandwidth (UWB) transmission [5]–[8] represents a promising technology for localization in harsh environments and accuracy-critical applications [9]–[15], due to its robust signaling [16], [17], as well as through-wall propagation [18], [19], and high-resolution ranging capabilities [20], [21]. However, practical deployment of UWB systems has been impeded by

a number of technical challenges, including signal acquisition [22], multi-user interference [23], [24], multipath effects [25]–[27], and non-line-of-sight (NLOS) propagation [27]–[29].This latter issue is critical for high-resolution localization systems [11], [12], [15], [20], [21], since NLOS propagation results in positively biased range estimates [29], which in turn degrade localization performance. NLOS conditions occur frequently in many practical harsh environments, including indoors, in urban canyons or under tree canopies. Therefore, it is imperative to understand the impact of NLOS conditions on localization systems, and to develop techniques that mitigate their effects.

Different approaches to address the NLOS problem have been proposed, which we classify coarsely as *NLOS identification* [30]–[34] and *NLOS mitigation* [34]–[42]. In NLOS identification, the goal is to detect when a range estimate corresponds to a NLOS condition. This can be achieved by analyzing received waveforms [30], [34], or a collection of range estimates from a single source [31]–[33]. In NLOS mitigation, the goal is to reduce the effect of the ranging error in NLOS conditions. NLOS mitigation can be combined with explicit NLOS identification by assigning different weights to LOS and NLOS signals [34], or by only using NLOS estimates to constrain the set of possible location solutions [35]. Alternatively, NLOS identification can be omitted by performing an exhaustive search over subsets of range measurements, to find a set of consistent LOS ranges [36]–[38], or by considering the LOS/NLOS condition to be a random parameter to be averaged over [39], or by explicitly accounting for the geometry of the environment [40]–[42]. An overview of NLOS identification and mitigation techniques can be found in [43], [44], and references therein. In our recent contribution [45], we have evaluated a *non-parametric* approach to NLOS identification, followed by NLOS mitigation, based directly on measured UWB waveforms. This approach performs identification and mitigation under a common framework, without requiring a statistical characterization of waveforms under LOS and NLOS conditions. We found that first classifying waveforms as LOS or NLOS is a crude way to deal with ranging errors, since the ranging bias introduced by obstacles depends on the materials and the physical environment. Our goal is to develop a more general approach, without relying on the distinction between LOS and NLOS conditions.

In this paper, we consider the general problem of ranging error mitigation without explicit NLOS identification. Building on tools from machine learning, we propose two non-parametric regression techniques to estimate the ranging error, based solely on the received waveform and the estimated

Figure 1. Histogram of the ranging error for the LOS and NLOS condition.

Figure 2. In some situations there is a clear difference between LOS (upper waveform) and NLOS (lower waveform) signals.

distance. The first technique employs support vector machine (SVM) regression to find a hyperplane that approximates the ranging error as a function of the training data. The second technique employs a Gaussian process (GP) to determine the a posteriori distribution of the ranging error, based on training data. The estimated ranging error, in combination with a measure of certainty, can be passed to a localization algorithm. Our regression techniques have the added benefit that they can be applied even when training data is not labeled with LOS or NLOS information. To the best of our knowledge, no other technique exists that performs ranging error mitigation based on features extracted directly from received waveforms, without relying on multiple range estimates or side-information regarding the environment. Our findings are validated using a database of UWB waveforms, obtained from an extensive measurement campaign in a typical office environment using FCC-compliant UWB radios.

The remainder of the paper is organized as follows. Section II describes the problem statement, and Section III provides background information on the regression techniques used later in this paper. These regression techniques are employed in Section IV to perform ranging error mitigation. The impact of ranging error mitigation on localization performance is evaluated in Section V. Finally, conclusions are given in Section VI.

Notation: $\|\mathbf{x}\|_{\alpha}$ denotes the ℓ_{α}-norm of the vector $\mathbf{x}$, defined as

$$\|\mathbf{x}\|_{\alpha} = \left[\sum_i |x_i|^{\alpha} \right]^{1/\alpha};$$

$\mathbf{x}^T$ is the transpose of the vector $\mathbf{x}$; $\mathbf{x} \succeq \mathbf{y}$ means $x_i \geq y_i, \forall i$; $\mathcal{N}(\mathbf{m}, \mathbf{K})$ represents a real multi-variate Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$.

II. PROBLEM STATEMENT

A. Localization Setup

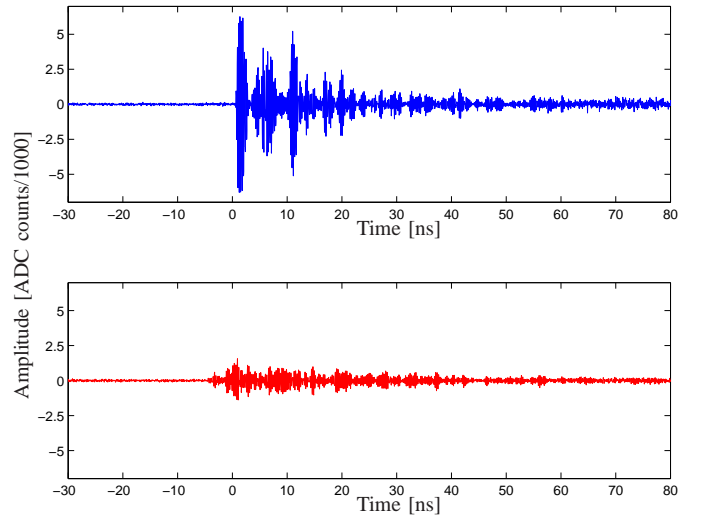A location-aware network consists of two types of nodes: *anchors* (or beacons) are nodes with known positions, while *agents* are nodes with unknown positions. We focus on the setting where a single agent with unknown position $\mathbf{p}$ is surrounded by N_{b} anchors with known positions, $\mathbf{p}_i$, $i = 1, \ldots, N_{\mathrm{b}}$. We denote the distance between the agent and anchor i by $d_i(\mathbf{p}, \mathbf{p}_i) = \|\mathbf{p} - \mathbf{p}_i\|_2$ and the agent's estimate of this distance by $\hat{d}_i$. We further introduce

$$\mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_{\mathrm{b}}}) = [d_1(\mathbf{p}, \mathbf{p}_1), d_2(\mathbf{p}, \mathbf{p}_2), \ldots, d_{N_{\mathrm{b}}}(\mathbf{p}, \mathbf{p}_{N_{\mathrm{b}}})]^T$$

and $\hat{\mathbf{d}} = [\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{N_{\mathrm{b}}}]^T$, as well as the ranging error $\Delta_i = \hat{d}_i - d_i(\mathbf{p}, \mathbf{p}_i)$.

In the absence of side-information regarding LOS or NLOS conditions and any statistical information regarding the distance estimates, a robust estimator of $\mathbf{p}$ is obtained by minimizing an appropriate norm:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left\| \mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_{\mathrm{b}}}) - \hat{\mathbf{d}} \right\|_{\alpha}. \tag{1}$$

The ℓ_1-norm is known to be more robust against outliers than the ℓ_2-norm, as those outliers incur only a linear cost in ℓ_1, whereas their cost is quadratic in ℓ_2 [46, Sec. 6.1.2]. When statistical information regarding the distance estimates is available, a common estimator is the maximum likelihood (ML) estimator:

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} p(\hat{\mathbf{d}}|\mathbf{p}). \tag{2}$$

Note that if the ranging errors are independent and identically distributed with a zero-mean Gaussian distribution (resp. Laplacian distribution), the ML estimator (2) reverts to ℓ_2-norm (resp. ℓ_1-norm) minimization.

B. Ranging Errors

In practice, range estimates are subject to different error sources, due to the environment, signal blockage, thermal noise, or algorithm artifacts. While there are many different models with varying complexity, it is difficult to capture all of these effects with a simple model. Rather than working with a complex theoretical model of these ranging errors, we

have performed an extensive ranging measurement campaign on the MIT campus, using FCC-compliant UWB radios [45]. From this campaign, we created a database, including 1024 measurements: 512 in LOS and 512 in NLOS. Here, the term LOS is used to denote the existence of a *visual* LOS. Specifically, a measurement is labeled as LOS when the straight line between the transmitting and receiving antenna is unobstructed. Each waveform $r(t)$, which is affected by thermal noise, is sampled every $T_{\text{sample}} = 41.3$ ps over an observation window of 190 ns. The range estimate was obtained by a round-trip time-of-arrival (RTOA) protocol,[1] embedded on the radio. The actual position of the radio during each measurement was manually recorded, and the ranging error was calculated with the help of computer-aided design (CAD) software. The collected waveforms were then aligned in the delay domain using a simple threshold-based method for leading edge detection.

From the measured data, we can gain more insight into the effects of LOS and NLOS conditions on a received waveform. Fig. 1 shows histograms of the ensemble of range measurements under LOS and NLOS conditions. Two typical waveforms under LOS and NLOS conditions are depicted in Fig. 2. Based on the measurement data and Figs. 1–2, we can make a number of observations:

1) The ranging error, considered over the entire ensemble of measurements, does not exhibit a Gaussian distribution. The ranging errors we observed were all non-negative (i.e., $\hat{d} \geq d$). This is due to the leading edge detection (LED) algorithm, which determines the time of arrival of the first path. LED is based on a simple threshold that is set so as to avoid false alarms (i.e., detecting noise spikes as a signal path). Hence, the ranging errors are due to missed detection of the first path, thus leading to a positive bias.

2) The ranging errors in LOS and NLOS conditions have different properties. We observed that, for LOS conditions, 98% of the measurements have a ranging error less than 1 meter, while for NLOS conditions, only 28% have a ranging error less than 1 meter.

3) The received waveforms in LOS and NLOS conditions tend to have different characteristics (as is apparent from Fig. 2). These characteristics can be exploited to identify NLOS waveforms and to compensate for the positive ranging bias.

4) The ranging error not only depends on the LOS or NLOS condition, but also on material properties, as well as the presence and positions of scatterers. This implies that the distinction between LOS and NLOS conditions provided NLOS identification techniques is rather coarse.

Based on these observations, we propose to not classify a waveform as LOS or NLOS, but rather to quantify the ranging error based on features extracted directly from the received waveform. This represents a departure from conventional

approaches and leads to (i) performance improvements; and (ii) reduction in complexity.

III. REGRESSION: MATHEMATICAL FRAMEWORK

A. Introduction

In regression, the goal is to infer an unobserved scalar ($y \in \mathbb{R}$), which depends on set of observed variables ($\mathbf{x} \in \mathbb{R}^n$). In particular, we assume a linear relationship of the form

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) \tag{3}$$

where $\varphi(\cdot)$ is a predetermined function,[2] and $\mathbf{w}$ represents the unknown parameter of the regressor. The parameter $\mathbf{w}$ can be considered a deterministic unknown which is to be estimated from a training set $\{\mathbf{x}_k, y_k\}_{k=1}^N$. Alternatively, the parameter $\mathbf{w}$ can be considered a random variable with a certain a priori distribution, for which we can then determine the a posteriori distribution from the training set. These two different viewpoints are taken by support vector machines and Gaussian processes, respectively.

B. Regression with Support Vector Machines

A SVM is a supervised machine learning technique used for classification and regression [47]–[50]. The regressor is a function $y : \mathbb{R}^n \rightarrow \mathbb{R}$, written as in (3), which can be interpreted as a hyperplane. Suppose that there exists a hyperplane such that $|y_k - y(\mathbf{x}_k)| \leq \varepsilon$ for some $\varepsilon > 0$, for all elements in the training set. Then the distance between the two bounding hyperplanes $y(\mathbf{x}) - \varepsilon = 0$ and $y(\mathbf{x}) + \varepsilon = 0$ is given by $d = 2\varepsilon / \sqrt{\|\mathbf{w}\|_2^2 + 1}$. Hence, the hyperplane that maximizes the distance between the bounding hyperplanes can be found as

$$\text{minimize } \|\mathbf{w}\|_2^2 \tag{4}$$
$$\text{s.t.} \quad y_k - \mathbf{w}^T \varphi(\mathbf{x}_k) \leq \varepsilon$$
$$y_k - \mathbf{w}^T \varphi(\mathbf{x}_k) \geq -\varepsilon.$$

In general, when ε is too small, the optimization problem becomes infeasible. To make the problem feasible, we penalize errors away from the hyperplane described in (3). The way in which errors are penalized impacts the computational complexity of determining $\mathbf{w}$, as well as the sparseness of the solution (see further). The optimization problem can be written as

$$\text{minimize } \|\mathbf{w}\|_2^2 + \gamma \sum_{k=1}^N L(y_k - \mathbf{w}^T \varphi(\mathbf{x}_k)), \tag{5}$$

where γ controls the trade-off between minimizing training errors and model complexity. The loss-function $L(\cdot)$ can take a number of forms. Popular examples include:

$$L_{\text{squared}}(e) = e^2 \tag{6}$$

$$L_{\varepsilon\text{tube}}(e) = \begin{cases} 0 & |e| - \varepsilon < 0 \\ |e| - \varepsilon & \text{otherwise} \end{cases} \tag{7}$$

---

[1]In an RTOA protocol, one radio (A) sends a request to a second radio (B). Radio B responds to the request by sending back a packet to radio A, which contains the processing time of radio B. Radio A then estimates the arrival time in its own time reference and determines the distance between A and B, assuming a known signal propagation speed.

[2]E.g., $\varphi(\mathbf{x}) = [\mathbf{x}\ 1]^T$ for a linear regressor.

In either case, the solution (called the SVM regressor) can be expressed as

$$y(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k \Phi(\mathbf{x}, \mathbf{x}_k), \qquad (8)$$

where

$$\Phi(\mathbf{x}, \mathbf{x}') = \varphi(\mathbf{x})^T \varphi(\mathbf{x}')$$

is the so-called *kernel* function. The values of α_k, $k = 1 \ldots N$, can be found using well-developed toolboxes for convex optimization. Generally, for the $L_{\varepsilon\text{tube}}$ loss, couples $(\mathbf{x}_k, y_k)$ within the tube incur no cost, leading to the corresponding $\alpha_k = 0$ and thus to a sparse solution. Given a test point $\mathbf{x}_{\text{test}}$, we can now predict the corresponding value for y as $y(\mathbf{x}_{\text{test}})$.

C. Regression with Gaussian Processes

Gaussian processes have recently gained interest from the machine learning community, as they form an elegant framework to perform regression [51]. For our situation, let y be a random variable such that, for a fixed input $\mathbf{x}$, the output is given by $y = \mathbf{w}^T \varphi(\mathbf{x}) + n$, where $n \sim \mathcal{N}(0, \sigma_n^2)$ represents measurement noise and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{w}})$. Rather than estimating $\mathbf{w}$, as in the previous section, here we average over all possible $\mathbf{w}$. Given N training points $\{\mathbf{x}_k, y_k\}_{k=1}^N$, we find that

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K} + \sigma_n^2 \mathbf{I}_N) \qquad (9)$$

where $[\mathbf{K}]_{k,l} = \varphi(\mathbf{x}_k)^T \Sigma_{\mathbf{w}} \varphi(\mathbf{x}_l) \doteq \Phi(\mathbf{x}_k, \mathbf{x}_l)$. The function $\Phi(\mathbf{x}, \mathbf{x}')$ is, similar to SVM, known as the kernel. Now, suppose we have a test point $\mathbf{x}_{\text{test}}$, and would like to determine the a posteriori distribution of the corresponding noise-free y_{test}. Under the stated assumptions, $\mathbf{y}$ and y_{test} are jointly Gaussian, with [51]

$$\begin{bmatrix} \mathbf{y} \\ y_{\text{test}} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I}_N & \mathbf{k} \\ \mathbf{k}^T & \Phi(\mathbf{x}_{\text{test}}, \mathbf{x}_{\text{test}}) + \sigma_n^2 \end{bmatrix}\right)$$

where $[\mathbf{k}]_k = \Phi(\mathbf{x}_{\text{test}}, \mathbf{x}_k)$. The a posteriori distribution $p(y_{\text{test}}|\mathbf{y})$ of y_{test} is Gaussian with mean

$$\mathbb{E}\{y_{\text{test}}|\mathbf{y}\} = \mathbf{k}^T \left(\mathbf{K} + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{y} \qquad (10)$$

and variance

$$\mathbb{E}\left\{(y_{\text{test}} - \mathbb{E}\{y_{\text{test}}|\mathbf{y}\})^2 |\mathbf{y}\right\} \qquad (11)$$
$$= \Phi(\mathbf{x}_{\text{test}}, \mathbf{x}_{\text{test}}) + \sigma_n^2 - \mathbf{k}^T \left(\mathbf{K} + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{k}.$$

We make the following comments:
- The a posteriori variance in (11) is smaller than the a priori variance $\Phi(\mathbf{x}_{\text{test}}, \mathbf{x}_{\text{test}})$, because of the training data. Also, note that neither variance depends on the training outputs. The a posteriori mean can be expressed as

$$\mathbb{E}\{y_{\text{test}}|\mathbf{y}\} = \sum_{k=1}^{N} \alpha_k \Phi(\mathbf{x}_{\text{test}}, \mathbf{x}_k), \qquad (12)$$

where α_k is the kth entry in the vector $\left(\mathbf{K} + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{y}$. Note that (12) bears close resemblance to (8). However, in the case of GP, the solution is generally not sparse, as the cost function is not insensitive to small errors.

Table I
EXTRACTED FEATURES

Name	Equation
Energy	$\mathcal{E}_r = \int_T |r(t)|^2 \, dt$
Maximum amplitude	$r_{\max} = \max_t |r(t)|$
Rise time[a]	$t_{\text{rise}} = t_{\text{H}} - t_{\text{L}}$
Mean excess delay[b]	$\tau_{\text{MED}} = \int_T t\psi(t)dt$
RMS delay spread[c]	$\tau_{\text{RMS}} = \int_T (t - \tau_m)^2 \psi(t)dt$
Kurtosis[d]	$\kappa = \frac{1}{\sigma_{|r|}^4 T} \int_T \left(|r(t)| - \mu_{|r|}\right)^4 dt$
Estimated distance	$\hat{d}$

[a] $t_{\text{L}} = \min\{t : |r(t)| \geq \alpha\sigma_n\}$ and $t_{\text{H}} = \min\{t : |r(t)| \geq \beta r_{\max}\}$, where σ_n is the standard deviation of the thermal noise. The values of $\alpha > 0$ and $0 < \beta \leq 1$ are chosen empirically; in our case, we used $\alpha = 6$ and $\beta = 0.6$ so as to minimize the false alarm probability.
[b, c] $\psi(t) = |r(t)|^2 / \mathcal{E}_r$.
[d] $\mu_{|r|} = \frac{1}{T} \int_T |r(t)|dt$ and $\sigma_{|r|}^2 = \frac{1}{T} \int_T (|r(t)| - \mu_{|r|})^2 dt$.

- A popular choice for the kernel is

$$\Phi(\mathbf{x}, \mathbf{x}') = \theta_0 \exp\left(-\frac{\theta_1}{2} \|\mathbf{x} - \mathbf{x}'\|_2^2\right) + \theta_2 \mathbf{x}^T \mathbf{x}', \quad (13)$$

where the hyperparameters $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]$ are usually estimated from the training data. Note that the choice $\boldsymbol{\theta} = [1, 0, 1]$ corresponds to conventional linear regression, with $\varphi(\mathbf{x}) = [\mathbf{x}\ 1]^T$.
- The SVM with the squared loss function (6) can be shown to be equivalent to the solution of a GP [52]. For that reason, we will only consider SVM with loss function (7) in Sections IV–V.

IV. RANGING ERROR MITIGATION

In this section, we will describe how SVM and GP can be applied to perform ranging error mitigation, based on features extracted from the received waveform, without requiring knowledge of the ranging error distribution. The features will serve as the observed input $\mathbf{x}$, while the ranging error will be the unobserved output y. We first explain the features we consider, and then provide implementation details of the SVM and GP regression techniques.

A. Feature Selection

As in our related work on obstruction detection [45], we have selected features based on the following observations. Due to reflections or obstructions, NLOS signals are considerably more attenuated and present smaller energy than LOS signals. In the LOS case, the strongest path corresponds to the first path and the received signal exhibits a short rise time. In the NLOS case, some weak multipath components precede the strongest path, as a result the rise time is longer. The root-mean-square (RMS) delay spread, which captures the temporal dispersion of the signal energy due to the multipath channel, is larger in NLOS signals. We also include features that have been considered in the literature. Taking these considerations into account, the features we extract from a received signal

Table II
SUMMARY OF THE MITIGATION PROCEDURE

Name	Features	Output	Parameters	Software
SVM	$\mathbf{x} = \log\left[\mathcal{E}_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, \tau_{\text{RMS}}, \kappa, \hat{d}\right]^T$	$y = \Delta$	$\varepsilon = 0.5$, $\gamma = 10^{-7}$, $\theta_0 = 1$, $\theta_1 = 1$	[53]
GP	$\mathbf{x} = \log\left[\mathcal{E}_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, \tau_{\text{RMS}}, \kappa, \hat{d}\right]^T$	$y = \Delta$	maximum likelihood	[51]
SVM–log	$\mathbf{x} = \log\left[\mathcal{E}_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, \tau_{\text{RMS}}, \kappa, \hat{d}\right]^T$	$y = \log \Delta$	$\varepsilon = 0.1$, $\gamma = 10^{-7}$, $\theta_0 = 1$, $\theta_1 = 1$	[53]
GP–log	$\mathbf{x} = \log\left[\mathcal{E}_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, \tau_{\text{RMS}}, \kappa, \hat{d}\right]^T$	$y = \log \Delta$	maximum likelihood	[51]

$r(t)$, observed for a duration T, are as follows: (i) the energy $\mathcal{E}_r$; (ii) the maximum amplitude $r_{\max}$; (iii) the rise time t_{rise}; (iv) the mean excess delay τ_{MED}; (v) the RMS delay spread τ_{RMS}; (vi) the kurtosis κ; and (vii) the estimated distance $\hat{d}$. We provide the analytical expression of each feature in Table I.

B. Mitigation Procedure

The database S consists of 1024 training samples. Every training sample is a vector consisting of 7 elements (the features), as described above in Section IV-A, along with the corresponding ranging error (the unobserved output). Our goal is to learn a function of the form (8), that maps the features to a ranging error. When determining the function-value for a specific input $\mathbf{x}_k$, care must be taken to avoid training the SVM or the GP with that same input. For this reason, we use 10-fold cross-validation [51], and divide up the database into ten disjoint parts: $S = S_1 \cup \ldots \cup S_{10}$, with $S_i \cap S_j = \emptyset$, for $i \neq j$. In the nth fold, we determine the functions (8) for SVM or (10)–(11) for GP, based on the training set $S \setminus S_n$. Then, the resulting function is applied to the test set S_n, giving the predicted outputs for S_n. For numerical reasons, the inputs $\mathbf{x}_k$ are converted to the logarithmic domain prior to training. We will consider four cases, two for SVM and two for GP. The details are listed in Table II. In all cases we use the kernel described in (13), with $\theta_2 = 0$. The output y of the mitigation procedure is either the ranging error Δ or its logarithm $\log \Delta$. In the latter case, the mitigation procedure will be denoted by GP–log or SVM–log. Note that $\log \Delta$ is well-defined, since all ranging errors are non-negative (see Fig. 1). Moreover, this approach will ensure that estimates of the ranging errors will also be non-negative.

C. Mitigation Performance

In Fig. 3 we show the CDF of the residual ranging error, i.e., the remaining error after mitigation. For the SVM (resp. SVM–log), these residuals have a mean of -3 cm (resp. 12 cm), and a standard deviation of 1.09 m (resp. 1.07 m). For the GP (resp. GP–log), the mean is 3 cm (resp. 17 cm), and the standard deviation 1.12 m (resp. 1.06 m). The fraction of residual errors less than one meter have increased from 63% (without mitigation) to around 90% (with mitigation). Note that the residual ranging errors can be negative, as they are defined as $\hat{d} - \hat{\Delta}$, where $\hat{\Delta}$ is the estimate of the ranging error output by the regressor. For GP–log and SVM–log, $\hat{\Delta} \geq 0$.
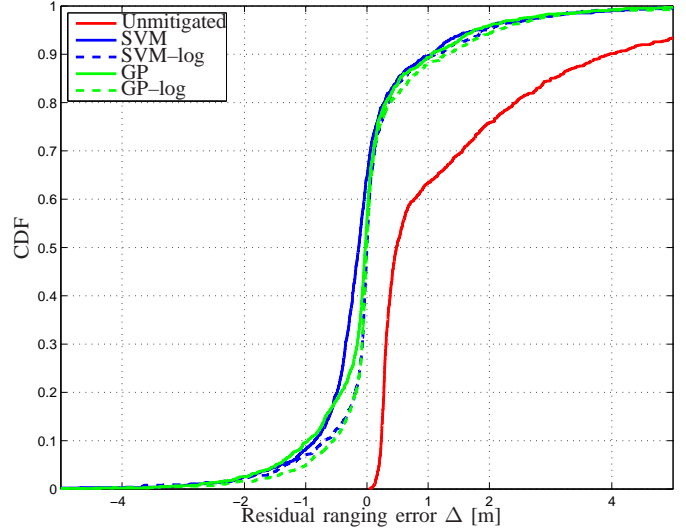

Figure 3. CDF of residual ranging error without mitigation, and using SVM and GP-based mitigation.

V. LOCALIZATION: STRATEGIES AND PERFORMANCE

In this section, we will evaluate the localization performance for a fixed number of anchors $N_b = 5$ and a varying probability of NLOS condition $0 \leq P_{\text{NLOS}} \leq 1$. We place an agent at position $\mathbf{p} = (0,0)$. For every anchor i ($1 \leq i \leq N_b$), we draw a measured waveform from the experimental database (described in Section II-B): with probability P_{NLOS} we draw from the NLOS database and with probability $1 - P_{\text{NLOS}}$ from the LOS database. The ith anchor is placed at position

$$\mathbf{p}_i = d_i(\mathbf{p}, \mathbf{p}_i)(\sin(2\pi(i-1)/N_b), \cos(2\pi(i-1)/N_b)), \quad (14)$$

where $d_i(\mathbf{p}, \mathbf{p}_i)$ is the true distance corresponding to that waveform. The estimate of the distance between the agent and the ith anchor ($\hat{d}_i$), is determined by the agent using the RTOA protocol.[3] The agent then estimates its position using one of the localization strategies to be described below, yielding a position estimate $\hat{\mathbf{p}}$.

To capture the accuracy and availability of localization, we introduce the notion of *outage probability*. For a certain scenario (say, a fixed N_b and P_{NLOS}, and a given localization strategy) and an allowable error e_{th} (say, 1 meter), the agent is said to be in outage when its position error $\|\mathbf{p} - \hat{\mathbf{p}}\|$ exceeds e_{th}. The outage probability is then given by the

---

[3]As our focus is on ranging error mitigation, rather than the placement of the anchors, we assume sufficient angular separation among anchors.
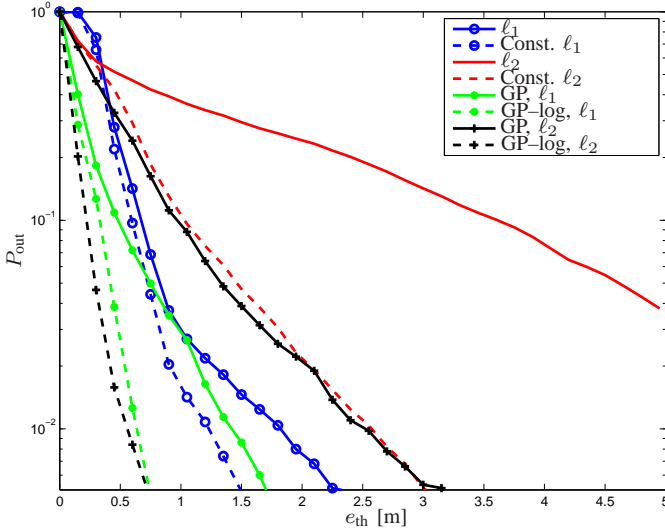
Figure 4. Outage probability for $N_b = 5$ anchors, with $P_{\mathrm{NLOS}} = 0.2$.

Figure 5. Outage probability for $N_b = 5$ anchors, with $P_{\mathrm{NLOS}} = 0.8$.

complementary CDF of the localization error:

$$P_{\mathrm{out}}(e_{\mathrm{th}}) = \mathrm{Prob}\left\{\|\mathbf{p} - \hat{\mathbf{p}}\|_2 > e_{\mathrm{th}}\right\}. \qquad (15)$$

The outage probability is determined through Monte Carlo simulation, by generating 5000 networks for every scenario.

A. Localization Strategies

We consider four different localization strategies that do not require knowledge of the statistics of the ranging error, or the LOS/NLOS condition. Given the N_b anchor's positions and a vector $\hat{\mathbf{d}}$ of N_b distance estimates, the estimate of $\mathbf{p}$ is found by solving one of the following four optimization problems.

- **Norm minimization:** A standard approach is to simply minimize the norm of the residuals:

$$\arg\min_{\mathbf{p}} \ \left\|\mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b}) - \hat{\mathbf{d}}\right\|_\alpha \qquad (16)$$

  for $\alpha \in \{1, 2\}$.

- **Constrained norm minimization:** We can exploit the knowledge that the distance estimates (see Fig. 1) are positively biased, i.e., $\hat{d}_i \geq d_i(\mathbf{p}, \mathbf{p}_i)$, through an additional constraint:

$$\arg\min_{\mathbf{p}} \ \left\|\mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b}) - \hat{\mathbf{d}}\right\|_\alpha \qquad (17)$$
$$\text{s.t. } \hat{\mathbf{d}} - \mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b}) \succeq \mathbf{0},$$

  for $\alpha \in \{1, 2\}$.

- **Mitigation followed by norm minimization:** Using either SVM or GP, we can obtain an estimate $\hat{\Delta}_i$ of the ranging error Δ_i, which we can subtract from the estimated range, leading to a mitigated range, $\tilde{d}_i = \hat{d}_i - \hat{\Delta}_i$. Using the vector of mitigated ranges, $\tilde{\mathbf{d}}$, we can minimize the norm of the residuals:

$$\arg\min_{\mathbf{p}} \ \left\|\mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b}) - \tilde{\mathbf{d}}\right\|_\alpha. \qquad (18)$$

  Note that now we cannot perform constrained optimization, since $\tilde{\mathbf{d}} \succeq \mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b})$ cannot be guaranteed (see also Fig. 3).
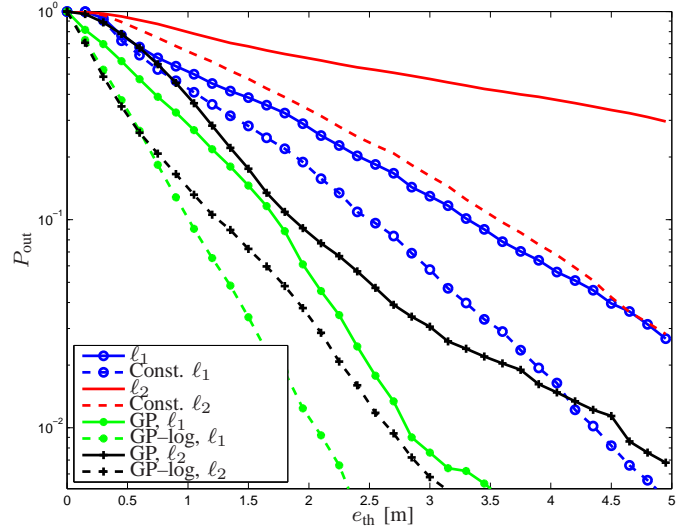
- **Log-domain mitigation followed by norm minimization:** Using either SVM–log or GP–log, we can obtain an estimate of $l_i(\mathbf{p}, \mathbf{p}_i) = \log\left(\hat{d}_i - d_i(\mathbf{p}, \mathbf{p}_i)\right)$. Norm minimization can be performed as follows

$$\arg\min_{\mathbf{p}} \ \left\|\mathbf{l}(\mathbf{p}, \mathbf{p}_{1:N_b}) - \mathbf{y}\right\|_\alpha, \qquad (19)$$

  for $\alpha \in \{1, 2\}$, where

$$\mathbf{l}(\mathbf{p}, \mathbf{p}_{1:N_b}) = [l_1(\mathbf{p}, \mathbf{p}_1), \ldots, l_{N_b}(\mathbf{p}, \mathbf{p}_{N_b})]^T \qquad (20)$$

  and

$$\mathbf{y} = [y_1(\mathbf{x}_1), \ldots, y_{N_b}(\mathbf{x}_{N_b})]^T \qquad (21)$$

  is the vector of outputs from the regressor. Note that there is an implicit constraint in (19), as the logarithm can only be applied to positive arguments.

B. Localization Performance

Overall, based on our investigations, we found that GP and SVM perform similarly, with GP performing slightly better than SVM. In the remainder of this section, we will focus on GP.

We first consider the outage performance for $P_{\mathrm{NLOS}} = 0.2$ in Fig. 4 and $P_{\mathrm{NLOS}} = 0.8$ in Fig. 5. In low P_{NLOS}, Fig. 4 indicates that, except for very small allowable errors e_{th}, ℓ_1-norm minimization outperforms ℓ_2-norm minimization. This is because the ℓ_1-norm is more robust against outliers, caused by NLOS conditions. Additionally, we observe that for any e_{th}, constrained ℓ_1- or ℓ_2-norm minimization uniformly outperforms unconstrained minimization, as we would expect. The performance difference is especially significant for ℓ_2-norm minimization, as adding the constraints can counteract the effect of outliers. For very small e_{th}, ℓ_1-norm minimization exhibits poor performance since it will attempt to find sparse solutions by driving some components of the ranging error vector $\hat{\mathbf{d}} - \mathbf{d}(\mathbf{p}, \mathbf{p}_{1:N_b})$ to zero, at the cost of larger errors in the remaining components. We see a performance
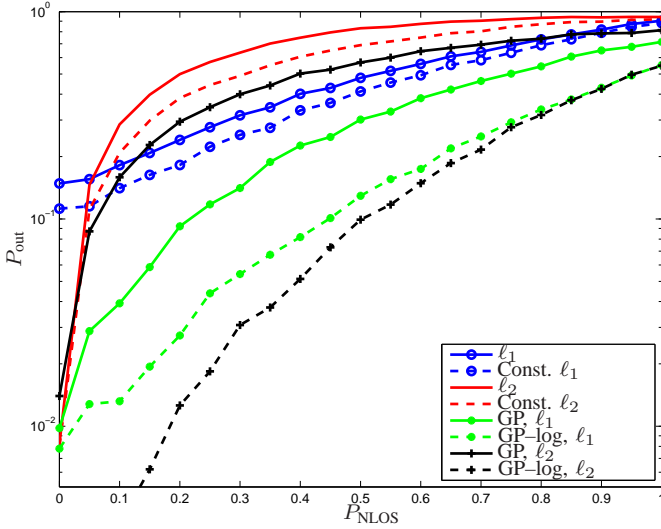
Figure 6. Outage probability for $N_b = 5$ anchors, with $e_{th} = 50$ cm.

Figure 7. Outage probability for $N_b = 5$ anchors, with $e_{th} = 2$ m.

improvement when using GP error mitigation with ℓ_1- or ℓ_2-norm minimization (18), compared to when no mitigation is applied. For GP error mitigation with the ℓ_1-norm, this gain is particularly visible for small e_{th}, while for the ℓ_2-norm, order-of-magnitude gains are achievable for $e_{th} > 50$ cm. Overall, GP error mitigation with ℓ_1-norm minimization outperforms GP error mitigation ℓ_2-norm minimization. For all considered values of e_{th}, GP–log (19) error mitigation achieves the best performance for both ℓ_1- and ℓ_2-norm minimization. In high P_{NLOS}, we see from Fig. 5 that without mitigation, the situation is similar, with ℓ_1-norm minimization outperforming ℓ_2-norm minimization, and constrained minimization reducing P_{out} compared to unconstrained minimization. When mitigation is employed, significant performance gains are visible in this high P_{NLOS} scenario. The strategy (19) again yields the best performance, with the ℓ_1-norm outperforming the ℓ_2-norm for all considered values of e_{th}.

Let us now evaluate the outage probability as a function of P_{NLOS} for a fixed e_{th}. Figs. 6–7 show P_{out} for $e_{th} = 50$ cm and $e_{th} = 2$ m, respectively. For $e_{th} = 50$ cm, Fig. 6 shows how ℓ_1-norm minimization performs better than ℓ_2-norm minimization, except for very small P_{NLOS}. When $P_{NLOS} \to 0$, ℓ_2-norm minimization yields excellent performance, since all the distance estimates have almost no error (see also Fig. 1) in pure LOS conditions. On the other hand, ℓ_1-norm minimization, tries to find a sparse solution. This means ℓ_1-norm minimization will try to set some errors to zero, while the other errors remain large (i.e., a solution $\hat{p}$ that lies on the intersection of two or more circles, and far away from the remaining circles), thus leading to poorer outage performance. GP error mitigation with ℓ_1-norm minimization exhibits good performance, outperforming ℓ_1-norm minimization for all P_{NLOS}. Finally, GP–log error mitigation yields the best performance, with the ℓ_2-norm slightly outperforming ℓ_1-norm. When relaxing the value of e_{th} to 2 m, outage probabilities for all localization strategies will drop, as observed in Fig. 7. Again, ℓ_2-norm minimization has the poorest performance, except when $P_{NLOS} \to 0$, in which case no outages were observed for 5000 network
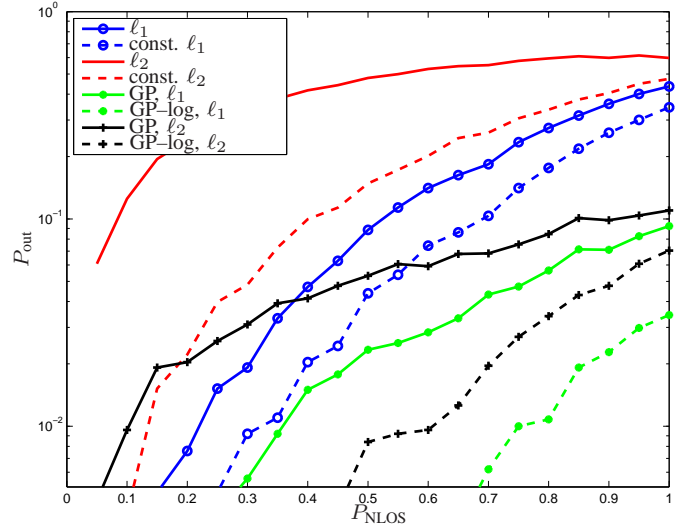
realizations. Constrained ℓ_2-norm minimization achieves better performance, but is still consistently outperformed by ℓ_1-norm minimization (both constrained and unconstrained). GP error mitigation has good performance, with outages remaining below 10% for all P_{NLOS}. Again, GP error mitigation with ℓ_1-norm minimization turns out to be better than ℓ_2-norm minimization. Finally, GP–log error mitigation again exhibits the best performance for all P_{NLOS}. In the higher e_{th} regime, GP–log error mitigation with ℓ_1 norm minimization wins out due to its robustness.

VI. CONCLUSION

Conventional approaches to deal with the challenge of localization in cluttered environments typically involve first detecting the NLOS condition, and then taking appropriate measures to account for the NLOS condition. However, the wide variety of materials and diverse operating environments can impact ranging performance in unique ways, indicating that the coarse distinction between LOS and NLOS is not always meaningful. Based on this observation, we have taken a different approach in this paper. Our approach employs non-parametric machine learning techniques (SVM and GP) to estimate the ranging error directly from the received waveform, without any a priori or a posteriori knowledge of the NLOS condition. Based on an extensive indoor measurement campaign with FCC-compliant UWB radios, we evaluated the localization performance in terms of outage probability for different localization strategies.

Our results revealed that: (i) ℓ_1-norm minimization is more robust in coping with outliers than ℓ_2-norm minimization, for localization without mitigation; (ii) constraints can provide significant gains, especially when localization requirements are not too stringent; (iii) SVM or GP regression techniques provide additional performance gains for all considered scenarios; (iv) SVM or GP regression techniques, combined with knowledge of constraints on the ranging error, provide the best performance for the scenarios under consideration.

The strategy of combining SVM or GP regression techniques with knowledge of constraints on the ranging error provides orders of magnitude performance improvements compared to traditional approaches. This highlights the fact that non-parametric ranging error mitigation has the potential to significantly improve localization performance.

VII. Acknowledgment

The authors wish to thank A. Globerson for helpful discussions. The authors also wish to thank the editor and the anonymous reviewers for helpful comments which greatly improved the manuscript.

References

[1] M. Z. Win, A. Conti, S. Mazuelas, Y. Shen, W. M. Gifford, D. Dardari, and M. Chiani, "Network localization and navigation via cooperation," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 56–62, May 2011.

[2] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.

[3] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[4] U. A. Khan, S. Kar, and J. M. F. Moura, "DILAND: An algorithm for distributed sensor localization with noisy distance measurements," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1940–1947, Mar. 2010.

[5] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[6] ——, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, Apr. 2000.

[7] A. F. Molisch, "Ultra-wide-band propagation channels," *Proc. IEEE*, vol. 97, no. 2, pp. 353–371, Feb. 2009.

[8] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[9] D. Dardari, A. Conti, J. Lien, and M. Z. Win, "The effect of cooperation on localization systems using UWB experimental data," *EURASIP J. Appl. Signal Process.*, vol. 2008, pp. Article ID 513 873, 1–11, 2008, special issue on *Cooperative Localization in Wireless Ad Hoc and Sensor Networks*.

[10] H. Wymeersch, U. Ferner, and M. Z. Win, "Cooperative Bayesian self-tracking for wireless networks," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 505–507, Jul. 2008.

[11] D. B. Jourdan, D. Dardari, and M. Z. Win, "Position error bound for UWB localization in dense cluttered environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 2, pp. 613–628, Apr. 2008.

[12] Y. Shen and M. Z. Win, "On the accuracy of localization systems using wideband antenna arrays," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 270–280, Jan. 2010.

[13] ——, "Fundamental limits of wideband localization – Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010. [Online]. Available: http://arxiv.org/abs/1006.0888v1

[14] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization – Part II: Cooperative networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4981–5000, Oct. 2010. [Online]. Available: http://arxiv.org/abs/1006.0890v1

[15] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, pp. 70–84, Jul. 2005.

[16] M. Z. Win and R. A. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor communications channel: A communication theoretic view," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.

[17] ——, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 51–53, Feb. 1998.

[18] A. Attiya, A. Bayram, A. Safaai-Jazi, and S. Riad, "UWB applications for through-wall detection," in *Proc. IEEE Antennas and Propagation Society International Symposium*, vol. 3, 2004, pp. 3079–3082.

[19] Y. Yang, Y. Wang, and A. Fathy, "Design of compact Vivaldi antenna arrays for UWB see through wall applications," *Progress In Electromagnetics Research*, vol. 82, pp. 401–418, 2008.

[20] J.-Y. Lee and R. A. Scholtz, "Ranging in a dense multipath environment using an UWB radio link," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1677–1683, Dec. 2002.

[21] D. Dardari, A. Conti, U. J. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proc. IEEE*, vol. 97, no. 2, pp. 404–426, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.

[22] W. Suwansantisuk and M. Z. Win, "Multipath aided rapid acquisition: Optimal search strategies," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 174–193, Jan. 2007.

[23] M. Z. Win, P. C. Pinto, and L. A. Shepp, "A mathematical theory of network interference and its applications," *Proc. IEEE*, vol. 97, no. 2, pp. 205–230, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.

[24] N. C. Beaulieu and D. J. Young, "Designing time-hopping ultrawide bandwidth receivers for multiuser interference environments," *Proc. IEEE*, vol. 97, no. 2, pp. 255–284, Feb. 2009, special issue on *Ultra-Wide Bandwidth (UWB) Technology & Emerging Applications*.

[25] M. Z. Win, G. Chrisikos, and A. F. Molisch, "Wideband diversity in multipath channels with nonuniform power dispersion profiles," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1014–1022, May 2006.

[26] M. Z. Win, G. Chrisikos, and N. R. Sollenberger, "Performance of Rake reception in dense multipath channels: Implications of spreading bandwidth and selection diversity order," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 8, pp. 1516–1525, Aug. 2000.

[27] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. F. Molisch, "A measurement-based statistical model for ultra-wideband industrial channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3028–3037, Aug. 2007.

[28] A. F. Molisch, "Ultrawideband propagation channels-theory, measurements, and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

[29] I. Güvenç, C.-C. Chong, F. Watanabe, and H. Inamura, "NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistics," *EURASIP J. Adv. in Signal Process.*, vol. 2008, pp. 1–14, 2008.

[30] F. Benedetto, G. Giunta, A. Toscano, and L. Vegni, "Dynamic LOS/NLOS statistical discrimination of wireless mobile channels," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, Los Angeles, CA, Apr. 2007, pp. 3071–3075.

[31] J. Borras, P. Hatrack, and N. B. Mandayam, "Decision theoretic framework for NLOS identification," in *Proc. 48th Annual Int. Veh. Technol. Conf.*, vol. 2, Ottawa, Canada, May 1998, pp. 1583–1587.

[32] J. Schroeder, S. Galler, K. Kyamakya, and K. Jobmann, "NLOS detection algorithms for ultra-wideband localization," in *Workshop on Positioning, Navigation and Commun. (WPNC)*, Mar. 2007, pp. 159–166.

[33] S. Gezici, H. Kobayashi, and H. V. Poor, "Non-parametric non-line-of-sight identification," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, vol. 4, Orlando, FL, Oct. 2003, pp. 2544–2548.

[34] I. Guvenc, C.-C. Chong, and F. Watanabe, "NLOS identification and mitigation for UWB localization systems," in *Proc. IEEE Wireless Commun. and Networking Conf.*, Kowloon, China, Mar. 2007, pp. 1571–1576.

[35] S. Venkatesh and R. M. Buehrer, "A linear programming approach to NLOS error mitigation in sensor networks," in *Proc. IEEE IPSN*, Nashville, Tennessee, Apr. 2006.

[36] P.-C. Chen, "A non-line-of-sight error mitigation algorithm in location estimation," in *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 1, Sep. 1999, pp. 316–320.

[37] R. Casas, A. Marco, J. J. Guerrero, and J. Falcó, "Robust estimator for non-line-of-sight error mitigation in indoor localization," *EURASIP J. Appl. Signal Process.*, no. 1, pp. 156–156, 2006.

[38] M. Tuchler and A. Huber, "An improved algorithm for UWB-based positioning in a multi-path environment," in *International Zurich Seminar on Communications*, Zurich, 2006, pp. 206–209.

[39] B. Denis and N. Daniele, "NLOS ranging error mitigation in a distributed positioning algorithm for indoor UWB ad-hoc networks," in *International Workshop on Wireless Ad-Hoc Networks*, May/Jun. 2004, pp. 356–360.

[40] S. Venkatraman, J. Caffery, Jr., and H. R. You, "Location using LOS range estimation in NLOS environments," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, vol. 2, Birmingham, AL, May 2002, pp. 856–860.

[41] A. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 24–40, 2005.

[42] S. Al-Jazzar and J. Caffery, Jr., "NLOS mitigation method for urban environments," in *Proc. IEEE Semiannual Veh. Technol. Conf.*, vol. 7, Sep. 2004, pp. 5112–5115.

[43] J. Khodjaev, Y. Park, and A. S. Malik, "Survey of NLOS identification and error mitigation problems in UWB-based positioning algorithms for dense environments," *Annals of Telecommunications*, 2009.

[44] J. Schroeder, S. Galler, K. Kyamakya, and T. Kaiser, "Three-dimensional indoor localization in non line of sight UWB channels," in *Proc. IEEE Int. Conf. on Ultra-Wideband (ICUWB)*, Singapore, Sep. 2007.

[45] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010.

[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[47] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[48] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, 2002.

[49] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[50] B. Schölkopf and A. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press, 2002.

[51] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Springer, 2006.

[52] T. Gestel, J. Suykens, G. Lanckriet, A. Lambrechts, B. Moor, and J. Vandewalle, "Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel Fisher discriminant analysis," *Neural Computation*, vol. 14, no. 5, pp. 1115–1147, 2002.

[53] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods Matlab toolbox," Perception Systemes et Information, INSA de Rouen, Rouen, France, 2005.