

Characterization of Non-Line-of-Sight (NLOS) Bias via Analysis of Clutter Topology

Muzammil Hussain
Department of Computer Science
University of Oxford
Oxford, United Kingdom
Email: muzh@cs.ox.ac.uk

Yusuf Aytar
Department of Engineering Science
University of Oxford
Oxford, United Kingdom
Email: yusuf@robots.ox.ac.uk

Niki Trigoni, Andrew Markham
Department of Computer Science
University of Oxford
Oxford, United Kingdom
Email: {ntrigoni, amarkham}@cs.ox.ac.uk

Abstract—Clutter-prone environments are challenging for range-based localization, where distances between anchors and the unlocalised node are estimated using wireless technologies like radio, ultrasound, etc. This is so due to the incidence of Non-Line-Of-Sight (NLOS) distance measurements as the direct path between the two is occluded by the presence of clutter. Thus NLOS distances, having large positive biases, can severely degrade localization accuracy. Till date, NLOS error has been modelled as various distributions including uniform, Gaussian, Poisson and exponential. In this paper, we show that clutter topology itself plays a vital role in the characterization of NLOS bias. We enumerate a feature-set for clutter topologies, including features that can be practically deduced without complete knowledge of the clutter topology. We then analyze the significance of these features, both individually and in combination with each other, in the estimation of the NLOS rate as well as the NLOS bias distribution for arbitrary clutter topologies. We show that we can obtain the NLOS rate with an error of only 0.03 for a given clutter topology using only those clutter topology features that can be practically realized in a real deployment. We show that estimating the NLOS bias distribution is more challenging which give a small number of poor estimations.

I. INTRODUCTION

Location awareness has become an integral part of the modern lifestyle. For example, GPS enabled smartphone and GPS receivers enable us to navigate to our destination or provide us information pertinent to our current location. Fleet managers are able to locate and monitor cargo moving in transit. Office and home areas are instrumented to furnish information relating to the time and area we are currently at. Most location services function with the help of *anchors*, nodes aware of their own positions, and distances measured to them. When there are obstacles in the environment, the accuracy of these measured distances, and thus the estimated position, suffers significantly. This degradation is caused by the existence of Non-Line-of-Sight (NLOS) distances which have large positive biases.

Until now, a large body of research has been conducted to address this issue. Some techniques [1], [2] try to detect and separate these erroneous distances from the position estimation process, while others [3], [4], [5], [6] incorporate these distances in a way that mitigates their negative influence on the location accuracy.

NLOS bias is hard to model accurately as it is shown to depend on the (arbitrary) underlying topology/arrangement of

the obstacles in the environment [7], [8], [9], [10], [11], [12]. For this reason, it is typically assumed to be uniform [6], [5], [13], [1], Gaussian [14] or exponential [1], [15], [4] for the sake of convenience. Jourdan et al. [12] mention the use of empirically derived NLOS bias distributions in calculating the PEB error bound for localization in a cluttered environment.

This paper addresses the problem by explicitly considering NLOS characteristics, namely NLOS rate and the NLOS bias distribution, as being dependent on the clutter topology. In other words, we illustrate the variation of NLOS rates and NLOS bias distributions with different types of clutter arrangements. To this end, we identify various features for a given clutter topology and analyse the influence of each feature, individually as well as in combination with other features. We then use machine learning tools to estimate both the NLOS rate as well as the bias distribution for an arbitrary clutter topology, using its features as input.

The remaining part of the paper is organized as follows: we state the problem in Section (II). In Section (III) we define various features of the clutter topology as well as other factors affecting NLOS bias. In Section (IV), we discuss the methodology of how we generate and analyze data in the paper. In Sections (V) and (VI) we describe and discuss our techniques to learn and predict NLOS incidence probability and NLOS bias distribution from a given clutter topology. We discuss related work and the need for clutter-based NLOS characterization in Section (VII). Finally we discuss future work and summarize our conclusions in Section (VIII).

II. PROBLEM STATEMENT

In this section we define the problem of NLOS bias characterization. We first discuss the causes for NLOS distance measurements. Later we state the parameters for characterizing NLOS error that we use in the paper.

A. Preliminaries

Distance measurements are obtained between a transmitter and receiver using a *ranging* signal. The ranging signal can either be Ultrawideband (UWB) radio or even a combination of radio and ultrasound such as the MIT Cricket nodes [16]. The time duration of the signal flight from the transmitter to the receiver, in other words, the difference between the time of

transmission and the time of reception is calculated and thus the distance is estimated. It is assumed that a suitable method is used to synchronize the receiver to the transmitter, in order to accurately calculate the signal flight duration. In case the direct path between the two is blocked by clutter, the ranging signal still reaches the receiver after reflecting off surfaces in the environment. Thus, the inflated flight time duration leads to the *positive bias* in the estimated distance measurement, which can be significant when compared to the actual distance. Thus, we can have a Line of Sight (LOS) distance measurement when there is a clear direct path between the two, and a Non-Line-Of-Sight (NLOS) distance measurement arising otherwise.

If d denotes the true distance between the transmitter and receiver, a distance measurement \hat{d} can be represented by

$$\hat{d} = d + \hat{n} + \hat{b} \quad (1)$$

where \hat{n} represents a Gaussian distributed random error and \hat{b} represents the NLOS bias. In the case of a LOS distance measurement, $\hat{b} = 0$. In case \hat{d} is NLOS in nature, we assume that the positive bias \hat{b} render the Gaussian error \hat{n} insignificant.

We assume that the ranging signal will be reflected off obstacles in its path of propagation, such as ultrasound signal [16]. In other words, the ranging signal *does not propagate through the clutter*.

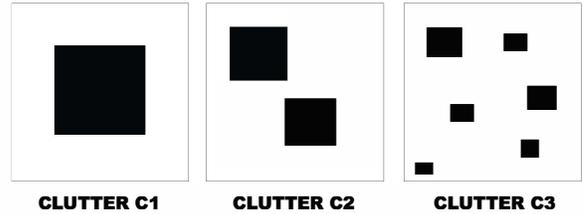
B. NLOS Characterization

We can characterize the NLOS error for a given clutter environment using two components : 1) *NLOS Incidence Probability* (f) : This is the probability of a distance measurement being NLOS, as opposed to LOS, for an arbitrary pair of points in the clutter topology. 2) *NLOS Bias Distribution* (\mathcal{P}) : This denotes the distribution of NLOS biases for an arbitrary pair of points in the clutter topology.

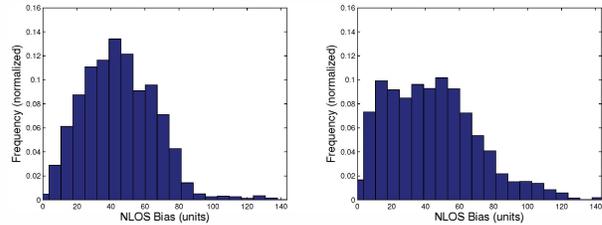
Given a set of sample distances, some being LOS and others being NLOS, for arbitrary pairs of points in the clutter topology, f can be obtained as the fraction of NLOS distances to the total number of distances. We can calculate the NLOS bias distribution using the non-parametric histogram model method. Fig. (1) shows the variation of f and \mathcal{P} for various clutter topologies. Next we will discuss various features and parameters of the clutter topology which cause this variation in NLOS characteristics.

III. FACTORS INFLUENCING NLOS ERROR

In this section we will enumerate various factors that determine the non-line-of-sight (NLOS) bias obtained during ranging measurements in an obstacle-prone environment. *Communication range*, the maximum distance an ultrasound or radio ranging signal can reach even after colliding with an unbounded number of surfaces in the clutter, is entirely independent of the clutter. However, the remaining features are related to the clutter topology and they are in fact various ways to characterize the clutter topology.

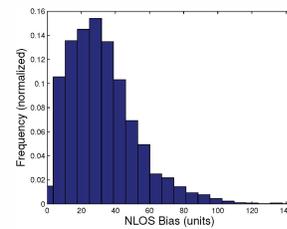


(a) Sample Clutter Topologies



(b) \mathcal{P}_{c1} ($f_{c1} = 0.4264$)

(c) \mathcal{P}_{c2} ($f_{c2} = 0.4498$)



(d) \mathcal{P}_{c3} ($f_{c3} = 0.5146$)

Fig. 1. Dependence of NLOS incidence probability and bias distribution on the underlying clutter topology. f and \mathcal{P} obtained with 10000 samples.

1) *Clutter Area Fraction*: The ratio of the clutter area to the total area, ca , is an important indicator of the level of clutter in the environment.

$$ca = \frac{\text{Total Area of Clutter}}{\text{Total Area of Enclosed Environment}}$$

In case of semi-open areas, a convex hull can be constructed across the openings in the environments in order to calculate the total area of the environment. However, it cannot independently help estimate NLOS incidence or bias error as NLOS distance measurements are dependent both on the spacing in the environment between the clutter and the enclosure boundaries, and the total amount of clutter in the environment.

2) *Communication Range*: Communication range \mathcal{C} , or the signal transmission energy, of the ranging device is vital in determining the NLOS bias. NLOS distance measurements are obtained only when the ranging signals have enough energy to travel indirect paths, reflecting off obstacles on the way. In other words, the communication range should be at least greater than the Euclidean distance between the two ranging devices involved in the distance estimation. Additionally, the loss of energy in the reflected signal as it bounces off obstacles

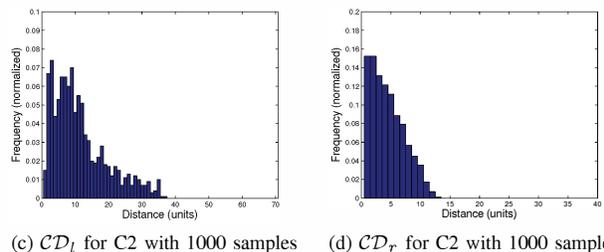
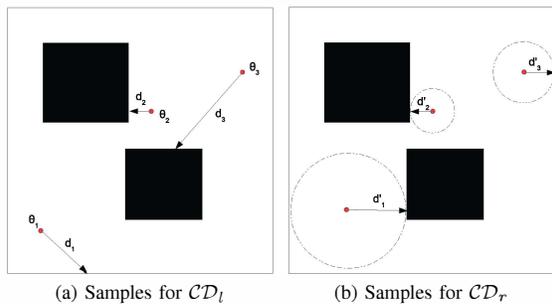


Fig. 2. Building the two clutter space distributions, CD_l and CD_r . Here, for the case of CD_l , three random points are chosen along with three random direction θ_1 , θ_2 and θ_3 , that give the corresponding spacing distances d_1 , d_2 and d_3 . In case of CD_r , if we take the same three positions, the corresponding spacing distances d'_1 , d'_2 and d'_3 represent the radii of the smallest circles that can be fitted before touching, in the *first* instance, a clutter piece or the enclosure wall. In practice, one can build these distributions with 500 or 1000 samples points.

is also vital to determine the effective communication range through the clutter.

3) *Clutter Spacing Distribution*: The space in midst of the obstacles and bounding enclosure plays a vital role in determining the NLOS distance biases. We define two types of clutter spacing distributions:

- *Linear Clutter Spacing Distribution CD_l* : This is the distribution obtained by measuring the space from a random position in the clutter topology (outside any obstacle) in a random direction till it strikes an obstacle or the enclosure walls in its path, as seen in Fig. (2a).
- *Radial Clutter Spacing Distribution CD_r* : This is the distribution obtained by measuring the maximum circle that can be drawn centered at a (random) point such that it does not intersect an obstacle or the enclosure walls, as seen in Fig. (2b).

Fig. (2) illustrates the construction of the two spacing distributions. A desirable feature of CD_l and CD_r is that they can be obtained through practical means without the need for explicit knowledge of the clutter topology. For example, a laser rangefinder can be used to obtain CD_l while a laser range scanner can be used to obtain CD_r .

4) *Occupancy Grid*: The occupancy grid of the clutter topology is a literal representation of the actual map of the clutter topology scaled by a factor s . In other words, if occ_1 represents the matrix representation of the clutter topology, with 1s denoting the enclosure boundaries and clutter and 0s denoting free space, occ_s is the corresponding matrix with

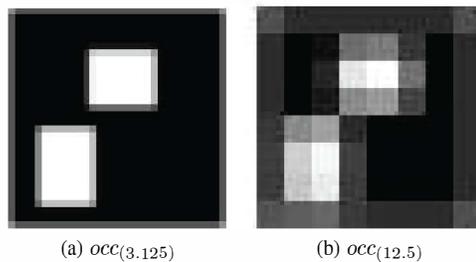


Fig. 3. Occupancy grid representation of clutter topology C2

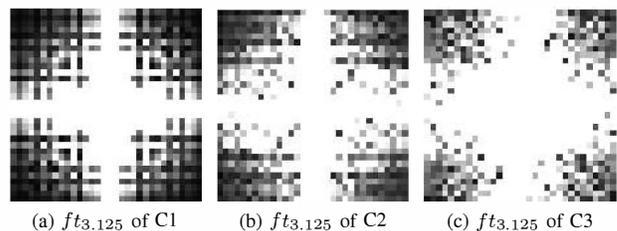


Fig. 4. Fourier transform characterization of clutter topologies C1, C2, C3

dimensions scaled by a factor of S , where $S = \frac{norm((X_L^O, Y_L^O))}{norm((X_L, Y_L))}$ and (X_L^O, Y_L^O) are the original dimensions of the clutter map image and (X_L, Y_L) are the dimensions of the scaled image. For example, if occ_1 is a 100x100 matrix, then $occ_{12.5}$ will be represented by a 8x8 matrix and $occ_{3.125}$ by a 32x32 matrix. We select bilinear interpolation scheme for matrix compression. Fig. (3) shows the graphical representations of $occ_{3.125}$ and $occ_{12.5}$ for clutter topology C2. The idea behind this representation is to capture the low-level structure of the clutter topologies while abstracting the higher level representation.

5) *Fourier Transformation*: The Fourier Transform is popular technique in the image processing research community [17], [18], [19]. It highlights the dominant spatial frequencies as well as the dominant orientations of the structures contained in the image. The Fourier transformation of an image provides its representation in the frequency domain. We use the two dimensional Fourier transformation (ft_s) of the occupancy grid occ_s , where S is the scale factor of the occupancy grid, as a characterization of the spacing and structure of the clutter. We use the magnitude of the 2D Fourier transform as the feature in our analysis. Fig. (4) shows the pictorial representation of $ft_{3.125}$ for clutter topologies C1, C2 and C3. Here the centers of each image represent the intensity of lower frequency sinusoids while the higher frequency sinusoids are represented towards the boundaries. Fig. (4c) shows the increased intensity of high frequency sinusoidal waves towards the boundaries representing the numerous, dispersed clutter in Fig. (1d).

6) *GIST characterization*: GIST [20] (\mathcal{G}) characterization of an image, widely used in the area of image classification, defines a set of 'perceptual' dimensions, such as naturalness, openness, roughness, expansion, ruggedness, that represent the dominant spatial structure of the image. A set of 384 Gabor

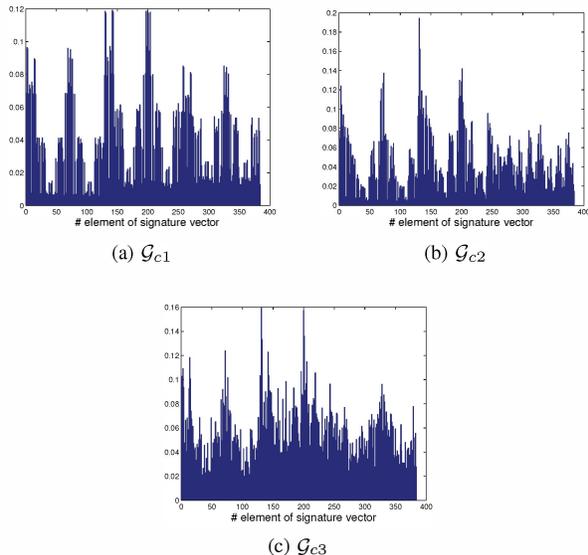


Fig. 5. GIST characterization of clutter topology C1, C2 and C3

filters each with a distinct set of parameters like frequency, orientation, etc. are individually convolved with the image to give a 384 length signature vector. The technique can be used to capture low level details of the image, for example in our case, the spacing and shapes of the clutter topology, while abstracting away the high-dimensional detail. Fig. (5) shows the variation in the GIST characterization for various clutter topologies.

IV. EXPERIMENTAL SETUP

In this section we will discuss the setup and fidelity of the experimental data. We assume, for the results discussed in this paper, that the ranging signal follows laws of reflection when it strikes obstacles or enclosure boundaries and do not *pass through clutter itself*, for example, ultrasound ranging signals. We first briefly describe the ray tracer tool we use to generate NLOS distance measurements in a given clutter topology. We then analyze the consistency of the data, which is vital for comparing results later on in the paper.

A. 2D Ray Tracer

In order to simulate NLOS distances between two points in a two dimensional enclosed environment that are occluded from each other by clutter, we use a ray tracer tool as shown in Fig. (6). The ray tracer works by shooting rays in various directions around the source point and traces the reflections across the clutter topology. While a number of such rays may eventually be incident on the destination point, we choose the *shortest indirect ray* as the NLOS distance. Note that in case there is no clutter between a pair of points, the distance between them is considered to be line-of-sight (LOS) and is simply the shortest *direct* Euclidean distance between the two.

The ray tracer has a number of parameters :

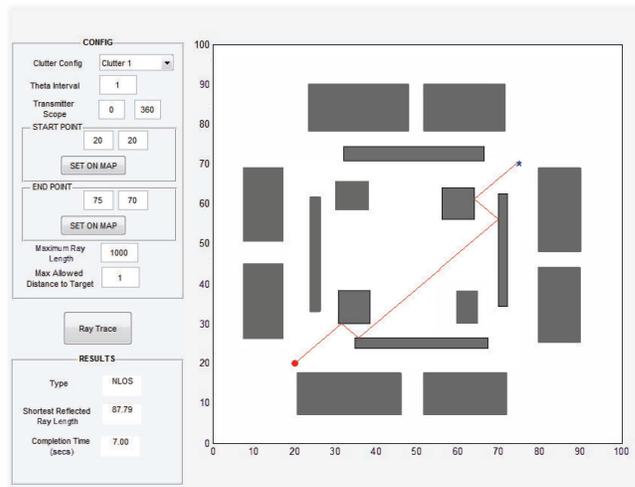


Fig. 6. 2D raytracer simulator outputs the shortest indirect path between two points in the clutter topology when clutter obscures the direct path between the two.

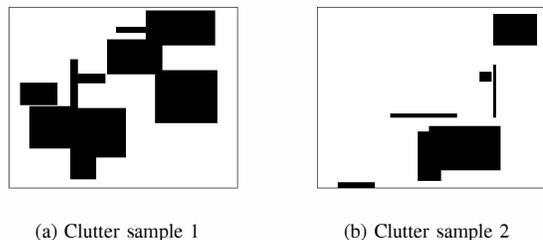


Fig. 7. Arbitrary clutter samples generated by overlaying simple rectangular boxes in the enclosure area.

- 1) *Direction Granularity* (RT_{gran}) : It is the measure of the angle interval between each consecutive ray shot out from the source point. For example, if RT_{gran} is 0.1° , then rays will be shot out at angles $0^\circ, 0.1^\circ, 0.2^\circ$, etc.
- 2) *Maximum Ray Distance* ($RT_{maxdist}$) : It denotes the maximum distance a ray can travel reflecting off edges in the clutter topology before it is deemed inviable to continue. It can be used to represent the communication range of a transceiver device.
- 3) *Maximum Distance from Destination* ($RT_{maxdest}$) : It denotes the maximum distance allowed for a ray to pass through in proximity of the destination point and still be considered incident on the destination itself. It is an approximation of the ray tracer device. For example, if we set $RT_{maxdest}$ to 2 units, then during the ray tracing process if a ray passes within 2 units from the destination point, it is considered incident on the actual destination point.

B. Clutter Topology Generation

We build complex clutter topologies by sequentially overlaying simple boxes over each other. We use a 2D rectangular enclosure area with dimensions L_x and L_y . We then generate ncl boxes with random start points (x_i, y_i) and

dimensions lx and ly , where $lx = \mathcal{U}(x_i, x_i + \frac{L_x}{dimfac})$ and $ly = \mathcal{U}(y_i, y_i + \frac{L_y}{dimfac})$. ncl is a randomly chosen natural number between $[1,10]$ and $dimfac$ is set to 3 in order to allow boxes to have dimensions at most a third of the enclosure dimensions. Each box is validated to fit inside the rectangular enclosure. Fig. (7) shows a couple of random clutter samples used in our analysis.

C. NLOS Characterization Data

Given a clutter topology, we generate the NLOS incidence probability f and the NLOS bias distribution \mathcal{P} by taking \mathcal{N} samples of uniform random generated points in the clutter topology. f is calculated as the ratio of the NLOS distances observed in the \mathcal{N} samples taken. The histogram \mathcal{P} is built using the bias of the NLOS distances in the \mathcal{N} samples. One can use techniques like LOESS [21], [12] in order to obtain a smoothed probability distribution from \mathcal{P} .

In order to establish the consistency of the NLOS characterization data, we vary the number of samples \mathcal{N} and measure the consistency of the f and \mathcal{P} values. In other words, we want to determine, if we take \mathcal{N} random samples, multiple times, how close will the values of f (or of \mathcal{P}) be to each other. We use the three clutter topologies shown in Fig. (1a) for our analysis. For each clutter topology, we repeatedly (10 times) take \mathcal{N} samples and calculate f and \mathcal{P} . For each clutter topology, we compute the dissimilarity between the 10 instances of f (and similarly between the 10 instances of \mathcal{P}).

For f , we use average pair-wise distances between the 10 instances of f , for 3 clutter topologies each, to measure consistency. err_f represents the average of the pair-wise distances over three clutter topologies. In case of the bias distribution \mathcal{P} , we use the Kullback Leibler Divergence (KLD) distance [22] to measure similarities (or dissimilarities) between instances of \mathcal{P} . The KLD distance is an inherently non-symmetric metric and for two distributions P and Q , is given by

$$D_{KL}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \quad (2)$$

$$(3)$$

where both P and Q sum up to 1, $Q(i) > 0$ every time $P(i) > 0$ and \ln represents the natural logarithm. We use the *symmetric* version of the KLD distance given by Eqn. (4) to evaluate the dissimilarity between two histograms/distributions throughout the paper.

$$D_{KS}(P, Q) = D_{KL}(P||Q) + D_{KL}(Q||P) \quad (4)$$

The consistency of \mathcal{P} is measured by taking the average of D_{KS} between the first instance and each of the remaining nine instances of \mathcal{P} for a given clutter topology. $err_{\mathcal{P}}$ represents the average over three clutter topologies and similarly $\sigma(err_{\mathcal{P}})$ denotes the standard deviation.

Tables (I) and (II) represent the consistency of the NLOS characterization data we generate for our analysis. We choose

Sample Size (\mathcal{N})	err_f	$\sigma(err_f)$
100	0.0565	0.0536
500	0.0244	0.0171
1000	0.0199	0.0141
5000	0.0083	0.0055
10000	0.0070	0.0056
50000	0.0023	0.0017

TABLE I
MEASURE OF CONSISTENCY LEVEL OF NLOS INCIDENCE PROBABILITY (f) IN GENERATED DATA

Sample Size (\mathcal{N})	$err_{\mathcal{P}}$	$\sigma(err_{\mathcal{P}})$
100	1.7861	0.6555
500	0.3253	0.1061
1000	0.1364	0.0517
5000	0.0287	0.0165
10000	0.0072	0.0024
50000	0.0016	0.0004

TABLE II
MEASURE OF CONSISTENCY LEVEL OF NLOS BIAS DISTRIBUTION (\mathcal{P}) IN GENERATED DATA

\mathcal{N} as 50000 for our data generation, which yields a consistency error of approximately 0.002 for both f and \mathcal{P} (D_{KS} value).

V. NLOS INCIDENCE PROBABILITY ESTIMATION

In this section, we will look at the estimation of the NLOS probability of an arbitrary clutter topology when we are given only the features that characterize the clutter topology, as enumerated in Section (III). We use a Support Vector Regressor (SVR) which learns a given training set and is able to predict the NLOS incidence probability f .

A. Support Vector Regressor

An ϵ -SVR [23], [24] solves the following optimization problem

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \quad (5) \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \zeta_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \end{aligned}$$

where ζ_i and ζ_i^* are slack variables and ϵ is the precision. The unknown parameters, w and b , are determined based on the training set $\{x_k, y_k\}_{k=1}^N$, where $x_k \in \mathbb{R}^n$ is the input and $y_k \in \mathbb{R}$ the respective output. In our case, the input x_i represents the feature set \mathcal{F}_f^n of the i^{th} clutter topology sample, where n denotes the number of features we are considering at a time. The output is the NLOS incidence probability f_i . We found that non-linear ϵ -SVR together with the Radial Basis (RBF) kernel gives the best results during our analysis. Fig. (8) gives a graphical overview of the process of estimation of f using the SVR.

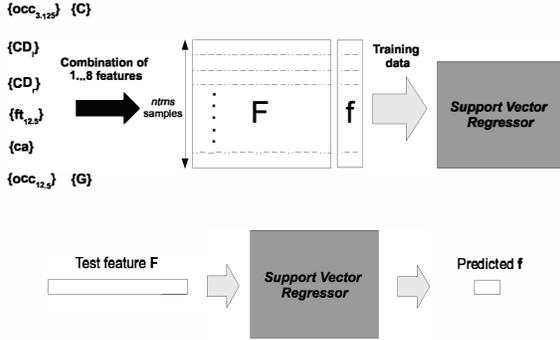


Fig. 8. Estimation of NLOS incidence probability f using Support Vector Regressor (SVR)

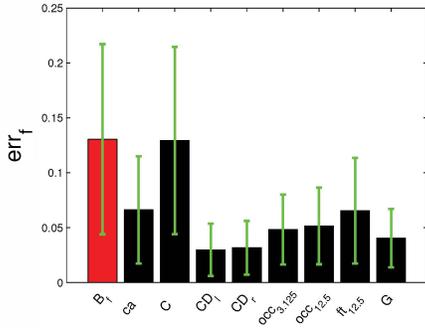


Fig. 9. Estimation error of f , err_f , for individual features

B. Results

We use a training set of n_{trn} (=10000) and test set of n_{ts} (=5000) distinct randomly generated clutter topologies. Each sample is split into features (or input) and labels (output). Three values have been used for \mathcal{C} : 100, 200 and 500. While the labels are set to f , the features, drawn from those defined in Section (III), are varied in number and composition. The SVR is first trained with the training set and we then supply it with a feature set \mathcal{F}_f from the test set and ask it to predict the value of f .

First, we analyze the estimation accuracy of \mathcal{P} when the SVM is used with only one feature at a time. Fig. (9) shows the performances of the features when taken individually to predict f . The clutter spacing distributions, CD_l and CD_r , perform the best, followed by the GIST representation of the clutter topology, \mathcal{G} . Communication range, \mathcal{C} , performs the worst since it alone has least significance in determining whether a range measurement will be NLOS or not.

Next, we use combinations of various features for the estimation of \mathcal{P} . The results for feature sets that give the best results (lowest err_f in Table (III)). Here \mathcal{F}_f^n represents the feature set that produces the least error when any combination of n features is used. The estimation accuracy is denoted by the mean (absolute) estimation error, denoted by err_f and its standard deviation. We see that the clutter spacing

Estimation Technique using SVR	err_f (mean)	err_f (std)
\mathcal{B}_f	0.1305	0.0866
$\mathcal{F}_f^1 = \{CD_l\}$	0.0298	0.0238
$\mathcal{F}_f^2 = \{CD_l, CD_r\}$	0.0272	0.0230
$\mathcal{F}_f^3 = \{ca, CD_l, CD_r\}$	0.0294	0.0236
$\mathcal{F}_f^4 = \{ca, \mathcal{C}, CD_l, CD_r\}$	0.0272	0.0230

TABLE III
ESTIMATION OF NLOS INCIDENCE PROBABILITY (f) USING SVR. HERE, THE SET \mathcal{F}_f^n DENOTES THE SET OF n FEATURES WHICH ACHIEVE THE MINIMUM err_f

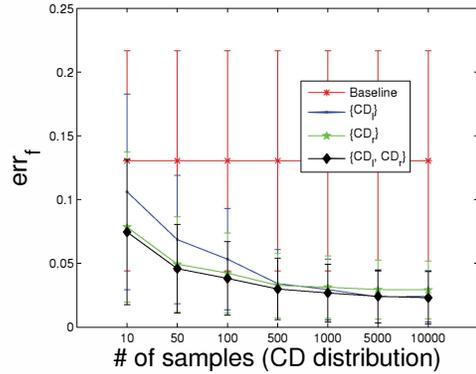


Fig. 10. Effect of the sample size for the clutter spacing distributions, CD_l and CD_r , on the accuracy of the NLOS probability (f) estimation using SVM regression.

distributions play a significant role in determining the value of f and are aided by other features like communication range (\mathcal{C}) and ca . We note that combinations of features can give non-intuitive results as well. For example, in Fig. (9), the GIST representation \mathcal{G} performs well *on its own*, better than ca and \mathcal{C} . However when taken in combination with other features, we found that the feature set $\mathcal{F}_f = \{CD_l, CD_r, \mathcal{G}\}$ performs worse than $\mathcal{F}_f = \{ca, CD_l, CD_r\}$ and $\mathcal{F}_f = \{ca, \mathcal{C}, CD_l, CD_r\}$. For comparison purposes we use a baseline \mathcal{B}_f , which is simply the average of $\{f_i\}_{i=1}^{n_{trn}}$. We see that there is a substantial improvement of the prediction accuracy with the use of the SVR over the (naive) average of all the training data results. Besides, as seen in Table (III), we can get a average estimation error of only 0.0272 for as little as two features, $\mathcal{F}_f = \{CD_l, CD_r\}$.

Finally, we investigate the effect of consistency of the spacing distributions on err_f . This is important because we find that CD_l and CD_r are the most effective features in the accurate prediction of f . Fig. (10) shows the effect on the number of samples used to build the CD_l and CD_r distributions on the actual estimation error of the NLOS probability (f). Having a sample size as low as 100 can yield an estimation error of around 5 percent. This shows the viability of these clutter features in a practical scenario where f has to be estimated in arbitrary cluttered environments.

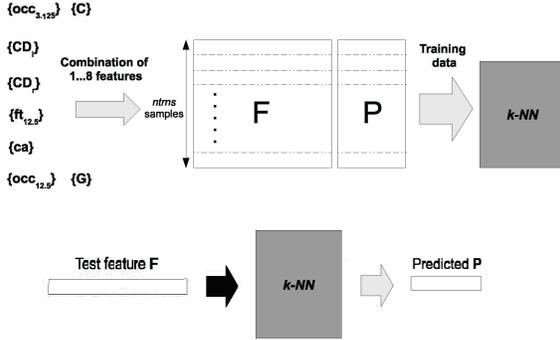


Fig. 11. Estimation of NLOS bias distribution/histogram \mathcal{P} using k-NN algorithm

VI. NLOS BIAS DISTRIBUTION ESTIMATION

In this section, we investigate the estimation of the NLOS bias distribution and analyze how various factors of the clutter topology affect the estimation process. We first briefly discuss the basics of the k-nearest neighbors (k-NN) approach we use for estimating the bias distribution. We then discuss the results and the dependence of the accuracy of the bias distribution estimation on various parameters of the k-NN estimation technique.

A. k-Nearest-Neighbor Estimation

The k-Nearest-Neighbor estimation algorithm is a method for classifying objects based on the closest *training* examples in the feature space. It is an instance-based learning technique where an object is classified by the majority vote of its neighbors, with the assumption that the object is probably the class most common amongst its k neighbors. In our case, instead of classification, we estimate the bias distribution as a function of the individual bias distributions of the k-nearest neighbors.

Suppose our feature set, $\mathcal{F}_{\mathcal{P}}$ is an arbitrary set of the features discussed in Section (III). Treating the feature set as input, we have $\{x_i\}_{i=1}^{ntrn}$ where *ntrn* is the number of training samples and $x_i \in \mathcal{R}^n$. The *ntrn* samples of the training set can be viewed to represent a set of points in *n*-dimensions. Given a new feature set input $\mathbf{x} \in \mathcal{R}^n$, we try to find the k-nearest neighbors of the point \mathbf{x} in the *n*-dimensional space. The bias distribution $\mathcal{P}(\mathbf{x})$ is obtained as a linear interpolation of the bias distributions [25] of its k-nearest neighbors using the Euclidean distances to neighbors as interpolation weights. Fig. (11) gives a graphical overview of the process of estimation of \mathcal{P} using the k-NN algorithm.

Let $\mathcal{N}_{\mathbf{x}}^k$ ($\exists p, 1 \leq p \leq ntrn$ and $\mathcal{N}_{\mathbf{x}}^k = x_p$) denote the k^{th} nearest neighbors of \mathbf{x} , separated by the Euclidean distance $d_{\mathbf{x}}^k$. If $\mathcal{P}(\mathcal{N}_{\mathbf{x}}^k)$ is the NLOS bias distribution of the k^{th} nearest neighbors of \mathbf{x} , the NLOS bias distribution $\mathcal{P}(\mathbf{x})$ is estimated as

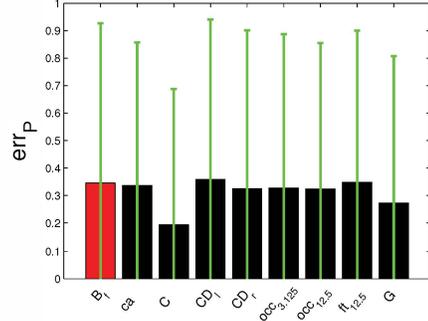


Fig. 12. Estimation error of \mathcal{P} , e_p , for individual features

$$\mathcal{P}(\mathbf{x}) = \sum_{j=1}^k \gamma_j \mathcal{P}(\mathcal{N}_{\mathbf{x}}^j) \quad (6)$$

$$\gamma_j = \frac{e^{-(d_{\mathbf{x}}^j)^2}}{\sum_{i=1}^k e^{-(d_{\mathbf{x}}^i)^2}} \quad (7)$$

where γ_j denotes the j^{th} weight in the weighted sum calculation. Since the interpolation weights are normalized, the resulting $\mathcal{P}(\mathbf{x})$ is guaranteed to be a probability distribution as well.

B. Results

We use a training set of n_{trn} ($= 10000$) and test set of n_{ts} ($= 5000$) distinct clutter topologies. Here, the search space is built with features which can vary in composition. $\mathcal{F}_{\mathcal{P}}^i$ denotes the feature set with *i* features mentioned in Section (III). We exhaustively try the various features, individually as well as in combination with other features. The k-NN algorithm is initially trained with the training set. We set $k=10$ for our analysis.

We begin by analyzing the impact of individual features in estimating \mathcal{P} using the k-NN algorithm. The accuracy of the estimation of \mathcal{P} by taking D_{KS} (given by Eqn. (4)) between the estimated and actual bias distributions. Fig. (12) shows the performance of individual features in predicting the \mathcal{P} for the test set. We find that none of the features perform well *on average* in estimating \mathcal{P} . The reason for this is that the NLOS bias distribution is too complex for a single feature to be able to predict it accurately.

Next, we analyze the efficacy of various combinations of features in predicting \mathcal{P} for the test set. We use exhaustive combinations of features, using two, three and four features together. We list the results for feature sets that give the best results (lowest $err_{\mathcal{P}}$ in Table (IV)). $err_{\mathcal{P}}$ is calculated as the symmetric KLD metric (D_{KS}) between the estimated bias distribution and the actual bias distribution of the test data, which measures the dissimilarity between the two distributions. We use the mean and standard deviation of $err_{\mathcal{P}}$ for the entire test set to measure the accuracy of NLOS bias distribution estimation. Again, here \mathcal{F}_f^n represents the feature set that

Estimation Technique using k-NN	$err_{\mathcal{P}}$ (mean)	$err_{\mathcal{P}}$ (std)
$\mathcal{B}_{\mathcal{P}}$	0.3452	0.5827
$\mathcal{F}_{\mathcal{P}}^1 = \{\mathcal{C}\}$	0.1934	0.4947
$\mathcal{F}_{\mathcal{P}}^2 = \{\mathcal{C}, CD_r\}$	0.1620	0.3812
$\mathcal{F}_{\mathcal{P}}^3 = \{ca, \mathcal{C}, CD_r\}$	0.1551	0.3783
$\mathcal{F}_{\mathcal{P}}^4 = \{ca, \mathcal{C}, CD_l, CD_r\}$	0.2426	0.4272

TABLE IV
ESTIMATION OF NLOS BIAS DISTRIBUTION (\mathcal{P}) USING K-NN. HERE, THE SET $\mathcal{F}_{\mathcal{P}}^i$ DENOTES THE SET OF i FEATURES WHICH ACHIEVE THE MINIMUM $err_{\mathcal{P}}$

produces the least error when any combination of n features is used. $\mathcal{B}_{\mathcal{P}}$ represents a baseline solution with the given training data, where \mathcal{P} is estimated by taking the average of all NLOS bias distributions in the training data.

On studying Table (IV), we see that communication range \mathcal{C} plays a significant role in improving the estimation error $err_{\mathcal{P}}$. However we do note the unusually large variances in $err_{\mathcal{P}}$ when compared to the mean $err_{\mathcal{P}}$ values, that suggests that the k-NN algorithm is not able to find accurate estimates for a *small number* of test samples. We also note that increasing the number of features does not necessarily improve the estimation accuracy of \mathcal{P} . For instances, $err_{\mathcal{P}}$ is lower for $\mathcal{F}_{\mathcal{P}}^3$ than $\mathcal{F}_{\mathcal{P}}^4$ as seen in the Table (IV).

While we find that the k-NN algorithm is not able to predict \mathcal{P} for good accuracy for every element of the test set, we nevertheless find that in a large number of cases the $err_{\mathcal{P}}$ is low. We can deduce that large outliers seen in the boxplots Fig. (13a) and Fig. (13b) are the reason for the high standard deviation seen in Fig. (12) and Table (IV) respectively. Fig. (13c) shows the empirical cumulative distribution of $err_{\mathcal{P}}$, for feature sets $\mathcal{F}_{\mathcal{P}}^2$ and $\mathcal{F}_{\mathcal{P}}^3$ where we see that in case of $\mathcal{F}_{\mathcal{P}}^3$, more than 80% of the test set yields an $err_{\mathcal{P}} < 0.2$. Furthermore, we see that there is a substantial improvement in the accuracy of predicting \mathcal{P} when compared to using the naive average of all training set \mathcal{P} ($\mathcal{B}_{\mathcal{P}}$).

Next, we provide a visual representation of the estimated \mathcal{P} for various features. Fig. (14) shows a comparison of the estimation of \mathcal{P} by various features sets for a single test clutter topology. Fig. (14d) - (14k) show the results for individual features, while Fig. (14a) and (14b) show actual bias distribution and the *baseline* bias distribution (obtained by averaging all \mathcal{P} in the training set) respectively. Fig. (14c) shows the estimation of \mathcal{P} using the feature set $\mathcal{F}_{\mathcal{P}} = \{ca, \mathcal{C}, CD_r\}$, gives a low dissimilarity measure D_{KS} of 0.0095.

Finally, in order to analyse the effect of neighbor size k , we perform the k-NN estimation for varying values of k . Fig. (15) demonstrates the effect of varying the maximum nearest neighbor set size on $err_{\mathcal{P}}$. We conclude that a neighbor set of size $k=10$ is the optimal size for $\mathcal{F}_{\mathcal{P}}^2$ and $\mathcal{F}_{\mathcal{P}}^3$ which also yields the lowest $err_{\mathcal{P}}$.

VII. RELATED WORK

NLOS identification and mitigation is a mature area of research, with a substantial amount being done in the area of cellular networks, where it is desired that a mobile terminal

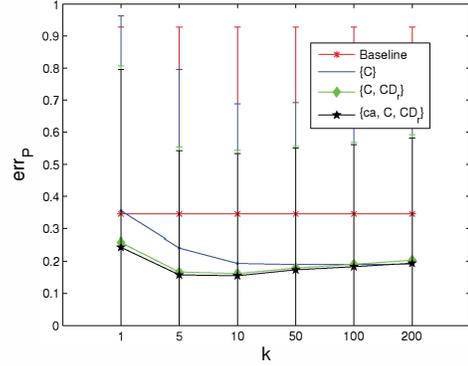
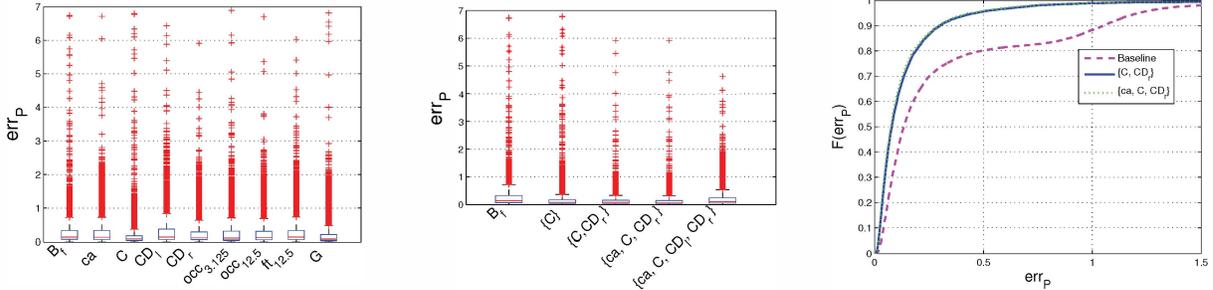


Fig. 15. Effect of the maximum number of nearest neighbor k considered, on the accuracy of \mathcal{P}

(MT) can be located using timing signals from the cellular basestations [26], [1], [2]. Current techniques can be classified into NLOS-elimination and NLOS-mitigation. In the first category, the NLOS distance measurement is identified and removed completely from the position estimation [13], [1], [2]. The second category, instead of discarding the information contained in NLOS distances, uses them in position estimation, albeit mitigating their effect [4], [5].

A number of papers use the characterization of NLOS biases itself as an input for their NLOS mitigation techniques. Wylie et al. [2] attempt to identify NLOS distances by comparing the variance of the distances over time to a pre-defined threshold. Similarly, Gezici et al. [7] compare the distribution of ranges to a known measurement error distribution in order to classify the signal as NLOS or not. Guvenc et al. [4] use multipath channel statistics of UWB signals in the weighted least squares position estimation. The authors use the kurtosis and the mean excess delay of the multipath channel, to determine the weights to assign for each reference. Marano et al. [3] use various characteristics of the impulse response of the received UWB signal in order to detect and even estimate the approximate NLOS bias. The paper builds a database from an UWB distance measurement campaign and then uses a Least-Squares Support Vector Machine (LS-SVM) to classify and estimate the NLOS bias.

A number of papers require knowledge of the NLOS bias distribution and make assumptions accordingly. Jourdan et al. [12] derive the Position Error Bound (PEB) using empirically derived NLOS bias distributions. A number of papers [15], [4] assume exponential error distribution for NLOS bias for UWB signals, while [5], [13], [6] assume the NLOS bias to be uniformly distributed within a predefined range. Borrás et al. [14] assume that the NLOS bias follows a Gaussian distribution while defining a decision theoretic framework for NLOS identification. Huang et al. [8] derive the Cramer Rao Bound for localization in NLOS environments where the NLOS bias distribution is obtained through a non-parametric Gaussian kernel density method.



(a) Distribution of err_P for test data when individual clutter features are used

(b) Distribution of err_P for test data when combination of clutter features are used

(c) Empirical distribution of err_P

Fig. 13. Distribution of estimation error err_P

The effect of clutter topology on the NLOS bias has been shown in a number of papers [9], [10]. Wang et al. [9] conclude that the NLOS bias for broadband channel sounder is strongly dependent on the clutter geometry and is frequency-dependent for severe clutter. Alsindi et al. [10] conclude that the NLOS bias follows a log-normal distribution with large bias forming the long tail, when there is an obstruction between the two ranging nodes. Furthermore, the authors show that the parameters of this lognormal distribution is dependent on the clutter environment and system bandwidth [11], [27].

In this paper, we did not present a NLOS identification or mitigation technique. Instead, we show that the NLOS bias distribution has a strong relationship with the clutter topology and that it is possible to characterize the NLOS error from the clutter topology itself. However it is important to note that the ranging signal technology and the interaction of the ranging signals with the clutter are also key in determining the NLOS bias [9], [4] and should ideally be taken into account in conjunction with the clutter topology. In this paper, our focus is only on the impact of the clutter topology.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we investigated the influence of clutter topology on the NLOS error, namely the NLOS incidence probability and the NLOS bias distribution. We also showcased an approach whereby we can estimate NLOS error characteristics as a function of the clutter topology. We enumerate a number of features that help characterize clutter topologies and show that there is a significant influence of various features of the clutter topology on NLOS bias error. We estimated the NLOS error characteristics using various machine learning tools. We conclude that while NLOS rate can be accurately predicted with an error of only 2.7%, NLOS bias distribution can be difficult to predict with good accuracy. We also see that clutter features like the spacing distributions we defined in Section (III) are much better in estimating NLOS characteristics than features like GIST and occupancy grid, which unlike the former, do require concrete information about the clutter topology beforehand. For future work, we are looking into the characterization of NLOS error taking into account both clutter topology and the interaction of the ranging signal with

actual clutter material. We are currently working towards the verification of these results in an actual test-bed using MIT-Cricket [16] motes for distance measurements.

IX. ACKNOWLEDGEMENTS

This work is supported by the EPSRC grants EP/F064209/1 and EP/F064578/1 on Actuated Acoustic Sensor Networks for Industrial Processes (AASN4IP). We would also like to thank the Oxford Supercomputing Center (OSC) for allowing the use of their facilities in the generation of ray-trace samples for clutter topologies. We would also like to thank Andrew Symington and Mikael Wallman of University of Oxford for their valuable feedback.

REFERENCES

- [1] P. Chen, "A non-line-of-sight error mitigation algorithm in location-estimation," in *1999 IEEE Wireless Communications and Networking Conference, 1999. WCNC, 1999*, pp. 316–320.
- [2] M. Wylie and J. Holtzman, "The non-line of sight problem in mobile location estimation," in *5th IEEE International Conference on Universal Personal Communications*, vol. 2, 1996, pp. 827–831.
- [3] S. Marano, W. Gifford, H. Wymeersch, and M. Win, "Nlos identification and mitigation for localization based on uwb experimental data," *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 7, pp. 1026–1035, 2010.
- [4] I. Guvenc, C. Chong, and F. Watanabe, "Nlos identification and mitigation for uwb localization systems," in *Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE*. IEEE, 2007, pp. 1571–1576.
- [5] S. Venkatesh and R. Buehrer, "A linear programming approach to nlos error mitigation in sensor networks," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 301–308.
- [6] T. Jia and R. Buehrer, "Collaborative position location with nlos mitigation," in *Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops), 2010 IEEE 21st International Symposium on*. IEEE, 2010, pp. 267–271.
- [7] S. Gezici, H. Kobayashi, and H. Poor, "Nonparametric nonlinear-of-sight identification," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 4. IEEE, 2003, pp. 2544–2548.
- [8] J. Huang and Q. Wan, "The crlb for wsn location estimation in nlos environments," in *Communications, Circuits and Systems (ICCCAS), 2010 International Conference on*. IEEE, 2010, pp. 83–86.
- [9] W. Wang, t. Jost, and U. Fiebig, "Characteristics of the nlos bias for an outdoor-to-indoor scenario at 2.45 ghz and 52 ghz," in *IEEE Antennas and Wireless Propagation Letters*, vol. 10. IEEE, 2011.
- [10] N. Alsindi, B. Alavi, and K. Pahlavan, "Spatial characteristics of uwb toa-based ranging in indoor multipath environments," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*. IEEE, 2007, pp. 1–6.

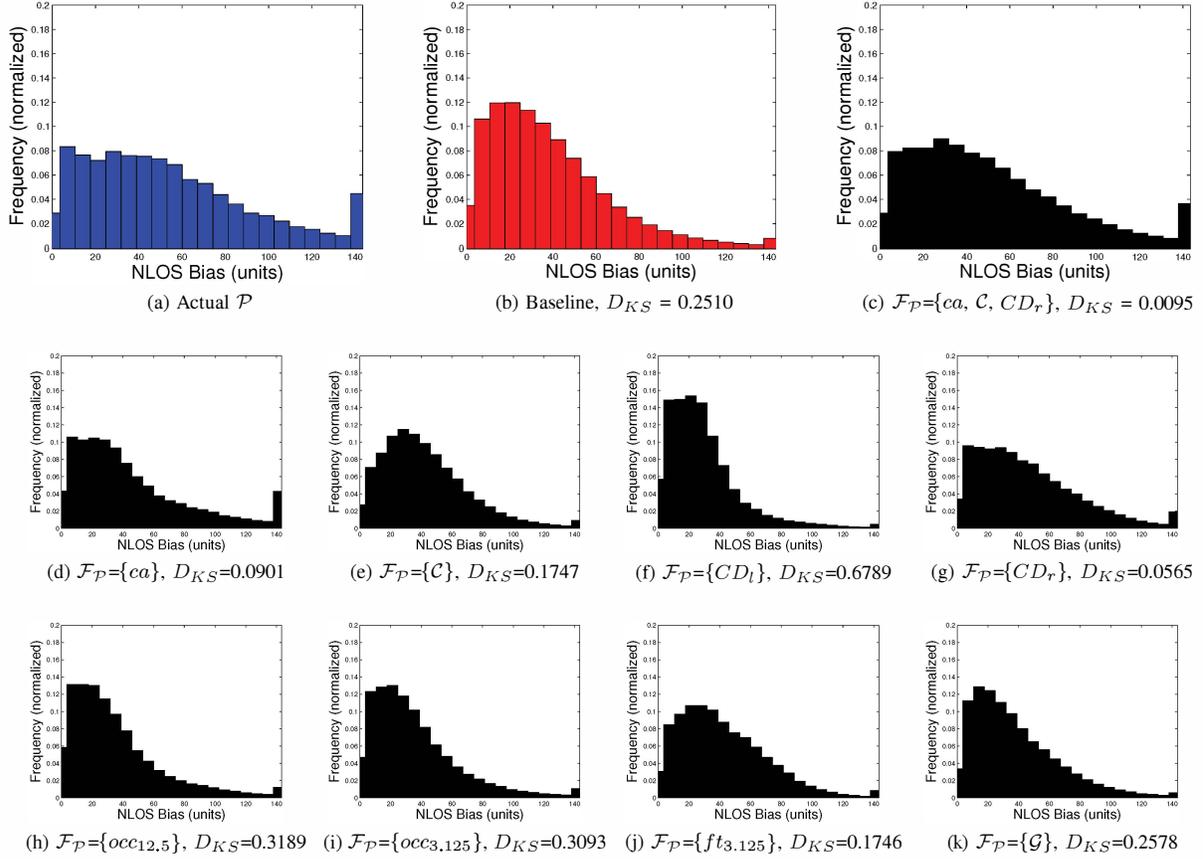


Fig. 14. Estimation of NLOS Bias distribution with k-NN

- [11] —, “Measurement and modeling of ultrawideband toa-based ranging in indoor multipath environments,” *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 3, pp. 1046–1058, 2009.
- [12] D. Jourdan, D. Dardari, and M. Win, “Position error bound for ubw localization in dense cluttered environments,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 2, pp. 613–628, 2008.
- [13] S. Nawaz and N. Trigoni, “Convex programming based robust localization in nlos prone cluttered environments,” in *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*. IEEE, 2011, pp. 318–329.
- [14] J. Borrás, P. Hatrack, and N. Mandayam, “Decision theoretic framework for nlos identification,” in *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, vol. 2. IEEE, 1998, pp. 1583–1587.
- [15] S. Gezici and Z. Sahinoglu, “Uwb geolocation techniques for ieee 802.15. 4a personal area networks,” *Mitsubishi Electric Research Laboratory Technical Report TR-2004-110*, 2004.
- [16] N. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *MOBICOM*. ACM, 2000, pp. 32–43.
- [17] R. d’Entremont, “Performance of the discrete fourier transform satellite imagery classification technique,” DTIC Document, Tech. Rep., 1980.
- [18] X. Tang and W. Stewart, “Optical and sonar image classification: wavelet packet transform vs fourier transform,” *Computer vision and image understanding*, vol. 79, no. 1, pp. 25–46, 2000.
- [19] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa, “Multiscale fourier descriptor for shape classification,” in *Image Analysis and Processing, 2003. Proceedings. 12th International Conference on*. IEEE, 2003, pp. 536–541.
- [20] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [21] W. Cleveland, “Lowess: A program for smoothing scatterplots by robust locally weighted regression,” *The American Statistician*, vol. 35, no. 1, pp. 54–54, 1981.
- [22] S. Kullback and R. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [23] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [24] A. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [25] A. Read, “Linear interpolation of histograms,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 425, no. 1, pp. 357–360, 1999.
- [26] M. Silventoinen and T. Rantalainen, “Mobile station emergency locating in GSM,” in *IEEE International Conference on Personal Wireless Communications, 1996*, 1996, pp. 232–238.
- [27] B. Alavi and K. Pahlavan, “Modeling of the distance error for indoor geolocation,” in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, vol. 1. IEEE, 2003, pp. 668–672.