INDOOR ULTRA-WIDE BAND NETWORK ADJUSTMENT USING MAXIMUM LIKELIHOOD ESTIMATION

KEY WORDS: ultra-wide band, maximum likelihood estimation, network adjustment, error assessment, indoor positioning

ABSTRACT:

This study is the part of our ongoing research on using ultra-wide band (UWB) technology for navigation at the Ohio State University. Our tests have indicated that the UWB two-way time-of-flight ranges under indoor circumstances follow a Gaussian mixture distribution that may be caused by the incompleteness of the functional model. In this case, to adjust the UWB network from the observed ranges, the maximum likelihood estimation (MLE) may provide a better solution for the node coordinates than the widely-used least squares approach. The prerequisite of the maximum likelihood method is to know the probability density functions. The 30 Hz sampling rate of the UWB sensors enables to estimate these functions between each node from the samples in static positioning mode. In order to prove the MLE hypothesis, an UWB network has been established in a multi-path density environment for test data acquisition. The least squares and maximum likelihood coordinate solutions are determined and compared, and the results indicate that better accuracy can be achieved with maximum likelihood estimation.

1. INTRODUCTION

The most commonly used and widely-accepted approach to determine the coordinates of a surveying network is based on the least squares (LSE) or weighted least squares estimation, which is not robust enough, when bad measurements, outliers are present in the data or the functional model is not complete. For this reason, in these situations, more robust estimation methods are recommended. Maximum likelihood estimation (MLE) is a robust estimation method, which requires that the probability density function (PDF) is known. In most cases the PDF's are derived from prior assumptions, experts' knowledge, etc. Ultimately, the pdf can be determined from samples, if they are available. The ultra-wide band (UWB) ranging systems are capable for producing sufficient number of measurements to estimate these PDF's due to their relatively high sampling frequency that can be more than 30 Hz, if the dynamics of the application allows for multiple observations.

This study is the part of our continuing UWB research at the SPIN Lab at The Ohio State University. Earlier, the calibration possibilities and the accuracy of the dynamic trajectory points derived from UWB ranges were investigate (Koppanyi et al., 2014). It was found that cm level accuracy can be achieved outdoor, while indoors, under line-of-sight circumstances, the accuracy was above half a meter. These tests also showed that the distributions of the ranges do not follow Gaussian distribution, rather a mixture of Gaussian distributions, These assumptions motivated the use of the MLE approach instead of LSE.

The MLE approach for wireless geolocation has been investigated in non-line-of-sight (Qi, 2003; Qi et al., 2006a) and multi-path density environment (Qi et al., 2006b). The UWB ranging under similar conditions was studied in (Lee et al., 2002). These papers are rather theoretically and focus on

positioning. The goal of our paper is to adjust the static network instead of positioning and to assess the solutions. An indoor UWB network with accurate reference was established and the single ranges obtained by UWB were used for estimating the node coordinates with MLE. Based on experimental data, the LSE and MLE solutions were compared, and confirmed the hypothesis.

2. METHODS

2.1 The functional model

The UWB networks consist of nodes, where the nodes measure ranges based on time-of-arrival (ToA) between each other. One of the unknown positions can be determined with circular lateration. First, assume that the coordinates of three stations are known, and an unknown 2D position has to be calculated from the ranges between the unknown position and the stations. Note that each measurement defines a circle around each node. If at least three distances are known, the circles have to intersect each other at one point, which is the unknown position, see Fig. 1.

The conditions can be described by the following system of equations:

$$(x - x_i)^2 + (y - y_i)^2 - d_i^2 = 0, \qquad (1)$$

where
x, y = the unknown coordinates,
x_i, y_i = the ith station coordinates,
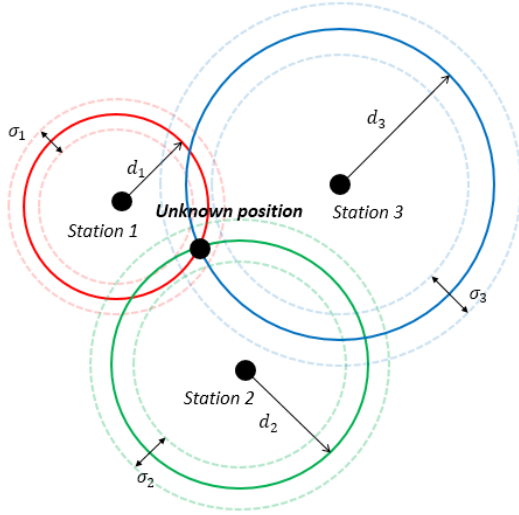d_i = the measured range between the unknown position and the ith station.

Figure 1. Circular lateration

If more than one range is available between the nodes and the unknown position, the equalities may not be satisfied, when errors are present. In this case, the error model is the following:

$$v_{i,k} = \sqrt{(x - x_i)^2 + (y - y_i)^2} - d_{i,k}, \qquad (2)$$

where $v_{i,k}$ = the error in the kth range of ith station,
       $d_{i,k}$ = the kth distance between the unknown position, and the ith station.

When the goal is to determine the coordinates of the network nodes, this formula could be extended to consider that the station coordinates are also unknowns (in which case, the unknown position is also a node):

$$v_{i,j,k} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} - d_{i,j,k}, \qquad (3)$$

where x_i, y_i = the ith station coordinates,
       $d_{i,j,k}$ = the kth measured range between the ith and jth station,
       $v_{i,j,k}$ = the residual between them.

Note that the problem becomes ill-determined, when not enough ranges (connections) are available, and at least, one node coordinates and the orientation of the network are not known. These networks are called anchored-free network, while if these parameters are known, the network is anchored.

2.2 Least squares and maximum likelihood estimation

When coordinates have to be estimated from multiple measurements, the most popular approach is the least squares estimation, which will find those x_i, y_i coordinates where the function of the sum of the squared value of $v_{i,j}$ is minimal:

$$\min_{x,y} \sum_{i,j,k} v_{i,j,k}^2 (x_i, y_i, x_j, y_j, d_{i,j,k}). \qquad (4)$$

In most applications, the LSE can provide proper results, but its outlier sensitivity is a well-known issue, thus if the numbers of the outliers is significant or the functional model is incomplete, one of the more robust estimation methods is preferred, such as contrast the maximum likelihood estimation (MLE) method:

$$\max_{x,y} \prod_{i,j} f_{i,j}(x_i, y_i, x_j, y_j), \qquad (5)$$

where $f_{i,j}$ = the probability density function of the distances between the ith and jth station.

Note that the MLE requires the PDF's of the range measurements. The PDF can be derived from priori assumptions or it can be estimated from the samples. In the latter case, the determined function $(\hat{f}_{i,j})$ is the empirical probability density function (EPDF).

Note that the well-known M-estimators are the generalization of the maximum likelihood estimation (MLE). The idea behind using the MLE concept is its robustness. M-estimator uses a pre-defined function instead of using the likelihood function (Huber, 2009).

2.3 Motivation for using MLE

In an UWB ranging system, the samples are acquired at about 30-100 Hz, which means over 100 ranges are available within 2-3 seconds. So in static applications, the EPDF can be derived from these samples, and it can be used in maximum likelihood estimation directly instead of defining it from prior assumptions.

In a narrow corridor indoor environment, tests indicate that UWB ranges do not typically follow Gaussian distribution. For an example, Fig. 2 shows the relative histogram of the ranges between the same nodes. It is clearly demonstrating the presence of more than one peak. Also, this distribution rather consists of a set of Gaussian distributions than a single one; this type of distribution is called Gaussian mixture distribution.
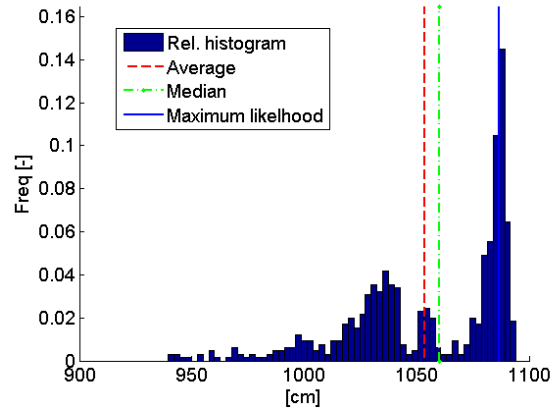

Figure 2. Sample indoor range

The multi-path signal propagation may cause this distribution. The conventional UWB systems use correlation-based time-of-flight method for estimating ranges. Theoretically, the large bandwidth decrease the impact of the multi-path mismatches, but multiple correlation peaks may still remain, and filtering these peaks is important to be able to provide the accurate distances. The experienced multiple peaks in the histogram can be suggested by the multi-path mismatches (Gezici, 2005). We emphasize that these statements are just our hypotheses, it is not clear for us that these peaks are caused by the multi-path environment, but the fact is that similar distributions did not occur in outdoor circumstances.

Now try to estimate the distance based on the histogram. Estimating the expected value is a widely-used approach. For the LSE estimation method, the estimation of the expected value is the sample mean statistic, which is shown by the red dashed line; the median statistic, which can be the L1-norm estimation of the expected value, is depicted with green dashed line. Note that these values are located at a third smallest peak, which is likely to be a bad estimation of the distance, even if it is an unbiased estimation of the expected value. If only this histogram is known and considering the multi-path mismatches, the location of the maximum peak is likely to be the valid estimation of the real distance, and this maximum is chosen by the maximum likelihood estimation (blue line in the figure).

This example demonstrates that if large number of outliers is present in the measurements or the functional model is not complete (which peak is the real distance), the MLE can provide better results due to its robustness. However, the MLE may not be unbiased or efficient in statistical sense; yet it is a very robust estimation method that fits better to our UWB ranges.

2.4 Estimating the probability density function

Estimating the PDF is necessarily to use MLE. First, relative histograms can be estimated from a set of distances obtained by the UWB sensors. Note the histogram may not be the best description of the phenomena; it is not continuous, its shape depends on the bin size, and it may fluctuate. In order to address these issues, kernel density estimation is used, where the kernel is the standard normal function:

$$\hat{f}_{i,j}^h(x) = \frac{1}{nh}\sum_{i=1}^n \Phi(\frac{x-d_{i,j,k}}{h}),\qquad(6)$$

where $\quad\hat{f}_{i,j}^h$ = the estimated empirical PDF from the distances between the ith and jth station,
Φ = the standard normal kernel function,
n = the sample number,
h = the bandwidth of the kernel.

Choosing the appropriate value for the bandwidth is important, because larger bandwidth smooths out the relevant peaks, but if it is too small, the fluctuation remains. In this paper, the performance was found to be the best at the 0.005 bandwidth.

2.5 Estimating the node coordinates with MLE

After estimating the empirical probability density function, the maximum likelihood estimation of the node positions is those coordinates of which distances maximize the likelihood functions:

$$\max_{x,y} \prod_{i,j} \hat{f}_{i,j}^h(\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}).\qquad(7)$$

Note that here the distances are the independent variables of the EPDF, while the errors are the independent variables of the score function in LSE case (see Eq. 4). Because of numerical properties, instead of solving the Eq. 7, log likelihood function is maximized:

$$\max_{x,y} \sum_{i,j} \ln \hat{f}_{i,j}^h(\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}).\qquad(8)$$

The equation can be solved by one of the numerical optimization methods. The line search algorithm was selected in

this work, which requires good initial value to avoid trapped at the local maximum instead of finding the global critical point. For that reason, the LSE solution is calculated first, and then its result is used as the initial value for the MLE. The reliability can be further improved with global optimization methods, such as different stochastic-based algorithms.

3. TEST

The UWB network is established in a multi-path challenged environment in the Bolz Hall at The Ohio State University. Seven TimeDomain's PulsOn 400 units are placed on the floor (http://www.timedomain.com/). The distances between them vary between 2 – 20 meters, see right side of Fig. 3, network arrangement is shown in the left side.

Note that the error propagation property inside the network is not favourable; especially in X direction due to the fact that the Y extent of the network is much larger than the X extent. The network is measured with surveying tape to determine reference coordinates; the estimated accuracy of these coordinates is better than 5 cm.
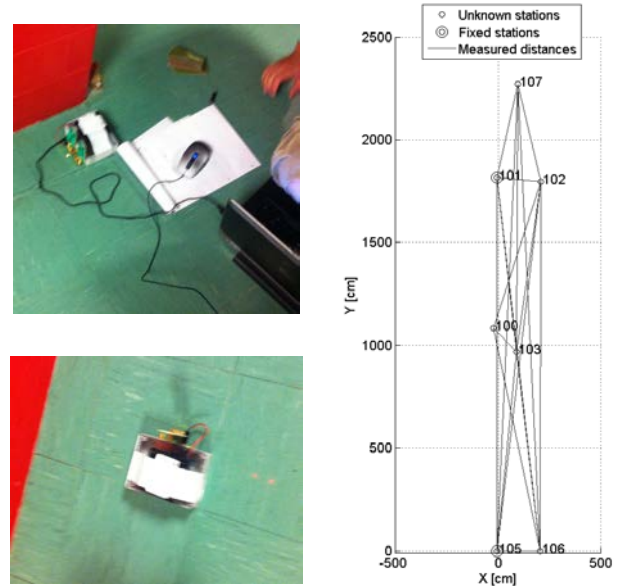

Figure 3. A UWB unit on the floor (left lower) and another one attached to the logging laptop (left upper); the network arrangement with the measured distances (right)

Nodes 101, 102, 103, 105, and 106 are part of the network, stations, and ranges are measured from these nodes. Nodes 100 and 107 are also measured by the network points, but no observation is executed on these nodes. The nodes 101 and 105 are assumed to be fixed stations (marked by double circle in the figure) to avoid the rotation or translation of the network for comparing it with the reference coordinates; thus, the network is anchored.

Laptop based data acquisition software records the raw ranges sent by the sensors via USB cable (right upper side of Fig. 3). Approximately 1000 distances are measured with two-way time-of-flight range estimation at 30 Hz in every case, but note that there are failed observations too.

4. DISCUSSION

The LSE and MLE solutions are shown in Fig. 4. The black circles show the reference coordinates measured by surveying tape in the floor map, shown in Fig. 4a. The double circle marks the fixed stations, nodes 101 and 105. The points with red asterisks present the LSE solution, while the blue ones show the MLE solutions. It is clearly seen that the MLE coordinates are closer to the reference coordinates than the LSE coordinates.

Fig. 4b-e show the histograms between nodes 102 and 101, 102 and 103, 101 and 103, 106 and 103, respectively. The empirical probability density functions smoothed by the standard normal kernel function are plotted with green line. The red dashed line shows the LSE, the blue dashed line is the MLE estimated distances. In the Fig. 4b, the LSE and MLE solutions provide nearly same result, while in Fig 4c the estimated distances are different, however both appear to follow Gaussian distribution.
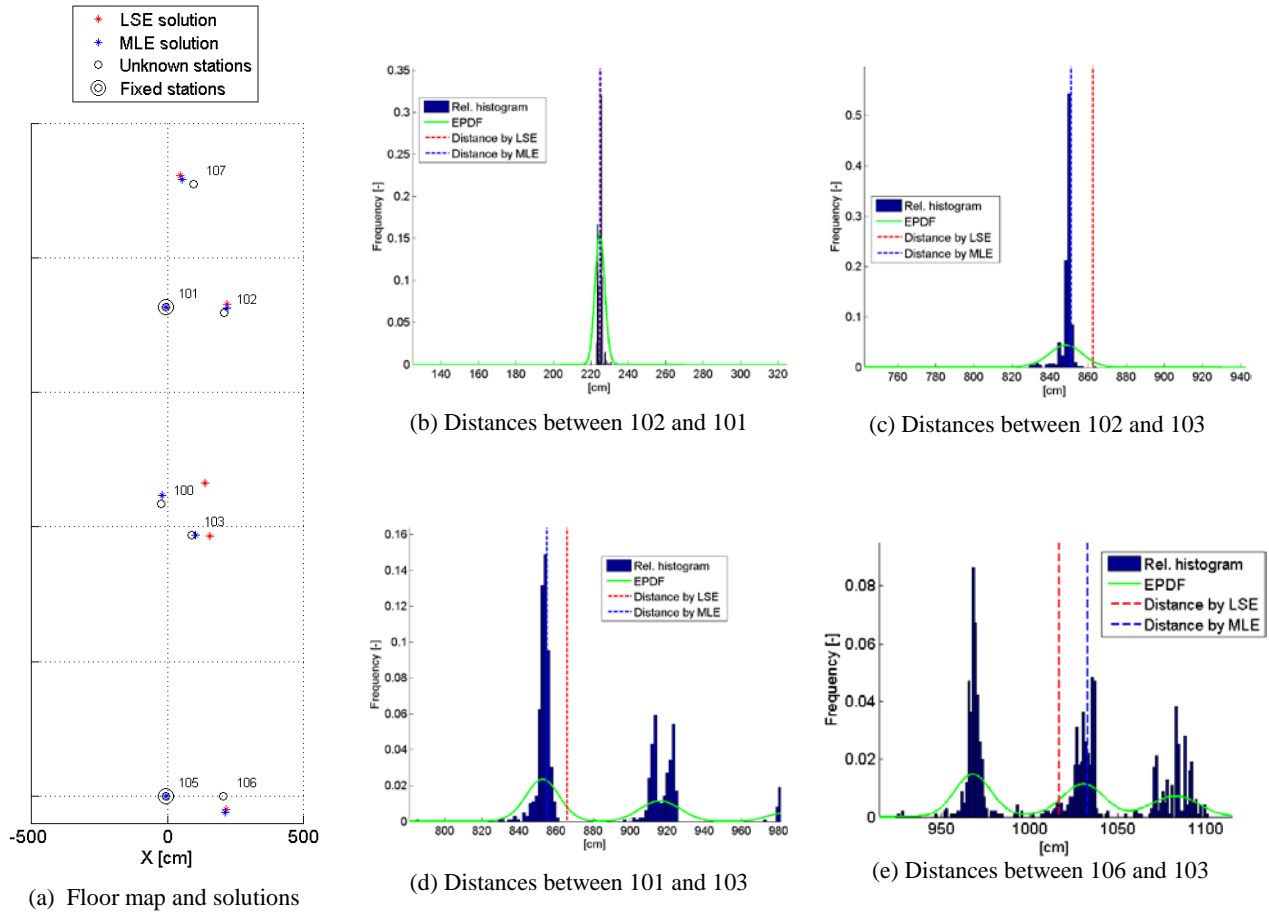

(a) Floor map and solutions

(b) Distances between 102 and 101

(c) Distances between 102 and 103

(d) Distances between 101 and 103

(e) Distances between 106 and 103

Figure 4. Comparison of LSE and MLE solutions

Node #	Measuring tape		LSE					MLE				
	X [cm]	Y [cm]	X [cm]	Y [cm]	dX [cm]	dY [cm]	dR [cm]	X [cm]	Y [cm]	dX [cm]	dY [cm]	dR [cm]
101	-5.6	1816.3	-5.6	1816.3	0.0	0.0	**0.0**	-5.6	1816.3	0.0	0.0	**0.0**
105	-5.5	0.0	-5.5	0.0	0.0	0.0	**0.0**	-5.5	0.0	0.0	0.0	**0.0**
100	-23.5	1085.5	140.5	1161.7	-164.0	-76.2	**180.9**	-18.2	1116.8	-5.3	-31.3	**31.7**
102	207.7	1795.4	219.7	1825.6	-12.0	-30.2	**32.5**	219.3	1811.9	-11.6	-16.5	**20.1**
103	89.8	968.7	155.3	965.5	-65.5	3.2	**65.6**	104.5	968.5	-14.7	0.2	**14.7**
106	206.5	0.0	215.2	-49.4	-8.7	49.4	**50.1**	212.4	-58.9	-5.9	58.9	**59.2**
107	96.6	2272.0	49.0	2306.1	47.6	-34.1	**58.6**	54.9	2290.9	41.7	-18.9	**45.8**
AVG					59.6	38.6	**77.5**			15.8	25.2	**34.3**
ABS MAX					164.0	76.2	**180.9**			41.7	58.9	**59.2**
ABS MIN					8.7	3.2	**32.5**			5.3	0.2	**14.7**

Table 1. Numerical comparison of LSE and MLE solutions

Node #	Measuring tape		LSE					MLE				
	X [cm]	Y [cm]	X [cm]	Y [cm]	dX [cm]	dY [cm]	dR [cm]	X [cm]	Y [cm]	dX [cm]	dY [cm]	dR [cm]
101	-5.6	1816.3	-5.6	1816.3	0.0	0.0	**0.0**	-5.6	1816.3	**0.0**	**0.0**	**0.0**
105	-5.5	0.0	-5.5	0.0	0.0	0.0	**0.0**	-5.5	0.0	**0.0**	**0.0**	**0.0**
100	-23.5	1085.5	15.5	1106.2	-39.0	-20.7	**44.2**	-18.8	1128.0	**-4.7**	**-42.5**	**42.7**
102	207.7	1795.4	219.4	1826.3	-11.7	-30.9	**33.1**	219.2	1809.6	**-11.5**	**-14.2**	**18.3**
103	89.8	968.7	149.8	968.6	-60.0	0.1	**60.0**	89.7	969.0	**0.1**	**-0.3**	**0.4**
107	96.6	2272.0	43.2	2312.6	53.4	-40.6	**67.1**	59.7	2290.6	**36.9**	**-18.6**	**41.4**
AVG					41.0	23.1	**51.1**			**13.3**	**18.9**	**25.7**
ABS MAX					60.0	40.6	**67.1**			**36.9**	**42.5**	**42.7**
ABS MIN					11.7	0.1	**33.1**			**0.1**	**0.3**	**0.4**

Table 2. Numerical comparison of LSE and MLE solutions after removing node 106

In Fig. 4d-e, note that the LSE distances are not coinciding with any peaks; they fall between them. Obviously, it is not absolutely necessarily that the MLE chooses one of the peaks, though it tends to select one. In Fig. 4c, the estimated distance of MLE is at the maximum peak, while a lower peak is selected in Fig. 4d. This shows that it is not inevitably that the estimated distances have to be at the maximum peak, as it depends on all of the measurements, all of the EPDFs in which the node is present, such that the probabilities of the distances maximize the likelihood function.

The Table 1 shows the numerical results. In some cases, larger correction can be noticed; for example, the coordinates of nodes 100 and 103 are improved by 149.2 cm and 50.9 cm using MLE. The average of the differences between the reference points and the LSE coordinates is 77.5 cm. Using MLE, a significantly lower average differences, 34.3 cm, can be achieved. The maximum differences between the LSE and MLE solutions are 180.9 and 59.2, respectively; note the minimum discrepancies are also decreased using MLE. These numerical results confirm that the MLE estimation of this type of network provides better results than LSE.

The only node, where the LSE coordinates are closer to the reference, is the node 106. The closer analysis of the measurement and the distances show that node 106 does not work properly during the test; after removing these ranges from the dataset, the LSE and MLE solutions are improved, though the MLE solutions still show lower errors, the average difference of LSE from the reference is 51.1 cm, while the MLE average is 25.7 cm, see Table 2.

5. CONCLUSION

In this paper, ad-hoc UWB network adjustment process using maximum likelihood estimation was presented. The empirical probability density function is derived from the ranges obtained by the UWB sensors using standard normal kernel estimation. The network coordinates are calculated using LSE and MLE methods. The results indicate that the MLE estimation can provide more accurate coordinates than the LSE solution.

In further research, the use of expectation maximization to estimate the PDFs and to separate the outliers will be investigated. The cause of the Gaussian mixture distribution is also expected to be examined.

REFERENCES

Gezici, S., 2005., Zhi Tian; Giannakis, G.B.; Kobayashi, Hisashi; Molisch, AF.; Poor, H.V.; Sahinoglu, Z., ,Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks, Signal Processing Magazine, IEEE , vol.22, no. 4, pp. 70,84, July 2005 doi: 10.1109/MSP.2005.1458289

Huber, Peter J., 2009., Robust Statistics (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-12990-6.

Koppanyi, Z., 2014., Toth, C.K., Grejner-Brzezinska, D.A., Jozkow, G., Performance Analysis of UWB Technology for Indoor Positioning In *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2014, pp. 154-165.

Lee, R.A., 2002, Scholtz Ranging in a dense multipath environment using an UWB radio link, *IEEE Trans. Select. Areas Commun.*, vol. 20, no. 9, pp. 1677–1683, Dec. 2002.

Qi Y., 2003, Wireless geolocation in a non-line-of-sight environment, Ph.D. dissertation, Princeton Univ., Princeton, NJ, Nov. 2003

Qi Y., 2006a, Kobayashi, Hisashi; Suda, H., Analysis of wireless geolocation in a non-line-of-sight environment, *Wireless Communications, IEEE Transactions*, vol.5, no.3, pp.672,681, March 2006 doi: 10.1109/TWC.2006.1611097

Qi Y., 2006b Kobayashi, Hisashi; Suda, H., On Time-of-arrival Positioning in a Multipath Environment, *Vehicular Technology, IEEE Transactions*, vol.55, no.5, pp.1516,1526, Sept. 2006