

Improving Ultra Wideband Indoor UWB Indoor Localization Via Per-Anchor Bias Calibration and Reliability-Weighted Multilateration

Temidayo Oluwafunke Otunniyi

Department of Electrical Engineering, Faculty of Engineering and Technology, Vaal University of Technology, South Africa.

Abstract

Ultra-wideband (UWB) ranging can provide accurate indoor localization, yet practical deployments exhibit substantial anchor-to-anchor heterogeneity caused by hardware-dependent timing offsets and environment-dependent ranging noise. This paper investigates a calibration strategy that explicitly separates these two effects: per-anchor systematic bias is estimated and removed, while anchor reliability is represented through inverse-variance weighting in the multilateration objective. The study uses 1,423 ranging packets collected from 17 anchors observing a stationary tag at a surveyed reference position. To prevent information leakage, packets are independently split 50/50 per anchor into calibration and evaluation subsets. Performance is assessed using 300 bootstrap trials. Relative to equal-weight multilateration, calibrated reliability-weighted multilateration reduces mean position error from 0.511 m to 0.244 m (52.3%) and median error from 0.469 m to 0.119 m (74.6%). However, the 90th-percentile error increases from 0.832 m to 1.031 m (23.9%). The key finding is therefore a distributional trade-off rather than a simple accuracy gain: calibration strongly improves typical fixes but can increase tail risk when calibration-period variance estimates are uncertain. The figures demonstrate this mechanism directly by linking anchor-level bias/noise heterogeneity to the observed localization-error distribution.

Keywords: UWB; indoor localization; multilateration; calibration; reliability weighting; bootstrap; tail risk; RTLS.

1. INTRODUCTION

Ultra-wideband (UWB) time-of-flight ranging, standardized through amendments such as IEEE 802.15.4z [30], is increasingly used for real-time locating systems (RTLS) [21], warehouses, hospitals, industrial environments, and other indoor applications requiring accurate asset localization [1]-[3], complementing a broader landscape of indoor positioning technologies that has been surveyed extensively elsewhere [10], [11]. Nevertheless, high nominal ranging resolution does not imply homogeneous ranging performance across all anchors. Practical deployments exhibit device- and environment-dependent ranging offsets and noise [4].

Two mechanisms are particularly important. First, antenna-delay and hardware-dependent timing offsets can produce systematic ranging bias that differs between devices [5], [6]. Second, propagation conditions, clock effects, and multipath or non-line-of-sight (NLOS) conditions can introduce additional anchor-dependent uncertainty [7]-[9]; classical NLOS mitigation strategies such as residual weighting [15] and constrained weighted least squares [16], [26] address this uncertainty at the algorithmic level. Range-based multilateration itself has a long history rooted in closed-form hyperbolic [17] and Taylor-series [18] position estimators, and the achievable accuracy of any such estimator is further shaped by anchor geometry through the geometric dilution of precision (GDOP) [19], [20]. Consequently, treating every anchor as equally reliable may be statistically inefficient.

Prior studies have addressed antenna-delay calibration, uncertainty characterization, NLOS mitigation, and anchor selection or weighting. Shalaby et al. reported a 46% localization improvement using antenna-delay calibration and uncertainty-aware filtering [6]. Shah et al. investigated simultaneous node calibration and ranging [5], while other studies addressed NLOS and channel-level error mitigation [8], [9]. However, many evaluations emphasize mean error, RMSE, or a single accuracy threshold.

This paper asks a more stringent question: does improving average localization accuracy necessarily improve the worst part of the error distribution? A mean-only evaluation can hide a deployment risk in which calibration produces more accurate central estimates but occasionally causes an unreliable anchor to receive excessive influence. To answer this question without assuming a parametric error model, performance is assessed via non-parametric bootstrap resampling [27], and the mitigation strategies proposed later in the paper draw on classical robust-statistics concepts, namely robust M-estimation [28] and shrinkage estimation of variance [29], to prevent a single unreliable anchor's calibration-period variance estimate from dominating the reliability-weighted objective.

The contributions of this paper are as follows:

- 1) Empirical characterization of anchor-dependent ranging bias and noise.
- 2) Leakage-controlled per-anchor calibration.
- 3) Inverse-variance reliability weighting for multilateration.
- 4) Bootstrap evaluation of mean, median, and upper-tail errors.
- 5) Identification and explanation of a calibration-induced tail-error trade-off.
- 6) Practical mitigation strategies based on variance shrinkage and weight capping.

II. DATASET AND EXPERIMENTAL ENVIRONMENT

The dataset contains 1,423 UWB ranging packets collected from 17 physically distinct anchors observing a single stationary tag. The tag reference position is (4.071, 3.664, 1.846) m. The indoor testbed spans approximately 24.7 m by 10.8 m, with anchor heights between 0.46 m and 2.90 m. Because the true tag position is known, each anchor's systematic ranging bias and noise can be estimated directly from the observed ranging errors.

III. METHODOLOGY

A. Calibration-Evaluation Split

Packets from each anchor are independently divided into calibration and evaluation subsets. Fifty percent are used only for calibration and the remaining 50% only for evaluation.

B. Per-Anchor Bias and Variance

For anchor i , let $m_{i,k}$ denote the measured range for packet k and $d_{i,k}$ the corresponding true (geometric) range, so that the ranging error is defined as

$$e_{i,k} = m_{i,k} - d_{i,k}, \quad (1)$$

The estimated bias is

$$b_i = (1/N_i) \sum_{k=1..N_i} e_{i,k}, \quad (2)$$

and the estimated variance is

$$s_i^2 = (1/(N_i-1)) \sum_{k=1..N_i} (e_{i,k} - b_i)^2. \quad (3)$$

The bias-corrected (calibrated) range is

$$c_i = m_i - b_i. \quad (4)$$

C. Reliability-Weighted Multilateration

The estimated position p minimizes

$$J(p) = \sum_{i=1..M} w_i (\|p - a_i\| - c_i)^2, \quad (5)$$

where a_i is the anchor position. Equal weighting uses $w_i = 1$, while reliability weighting uses $w_i = 1 / (s_i^2 + \epsilon)$. (6)

D. Bootstrap Evaluation

Three hundred bootstrap trials are performed. One held-out evaluation packet is sampled independently from each anchor per trial. Writing p_{est} for the estimated position, position error is

$$E = \|p_{est} - p_{true}\|^2. \quad (7)$$

Mean, median, and 90th-percentile errors are reported.

IV. RESULTS

A. Anchor-Level Bias and Noise

Anchor	Bias (mm)	Std. Dev. (mm)
3	20.78	16.91
4	449.38	212.76
5	-93.83	28.80
6	227.82	41.65
7	-200.09	22.56
8	-83.07	15.53
10	-99.19	26.49
11	-43.65	25.67
14	471.13	23.23
15	-28.57	14.59
16	67.18	29.84
18	-23.82	73.23
20	227.92	161.28
21	-8.77	20.46
24	235.70	42.78
26	373.91	114.20
31	-21.89	249.28

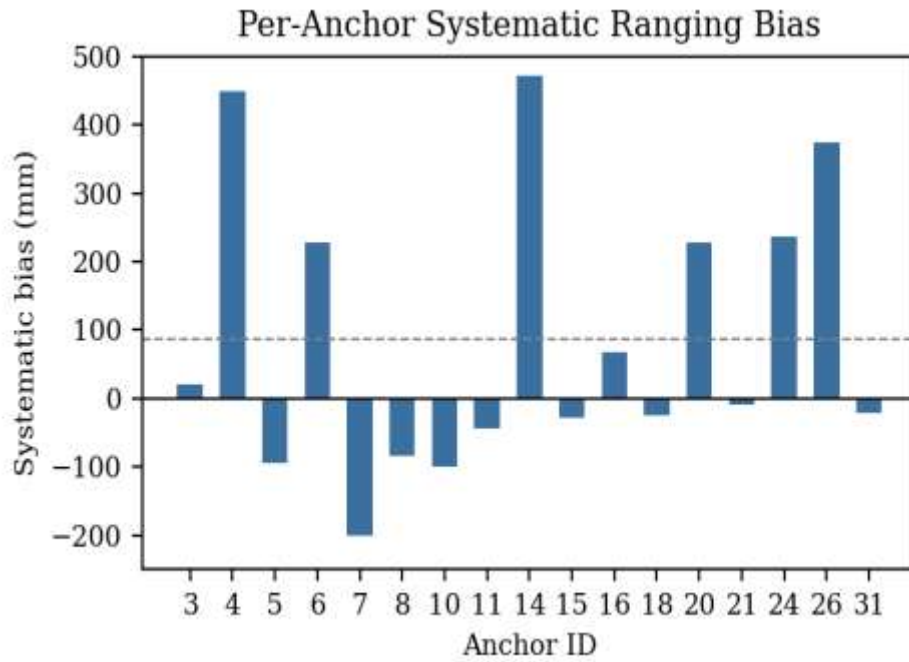
TABLE I. PER-ANCHOR SYSTEMATIC BIAS AND NOISE

Fig. 1. Per-anchor systematic ranging bias. Large positive and negative offsets demonstrate substantial anchor heterogeneity and motivate per-anchor bias correction.

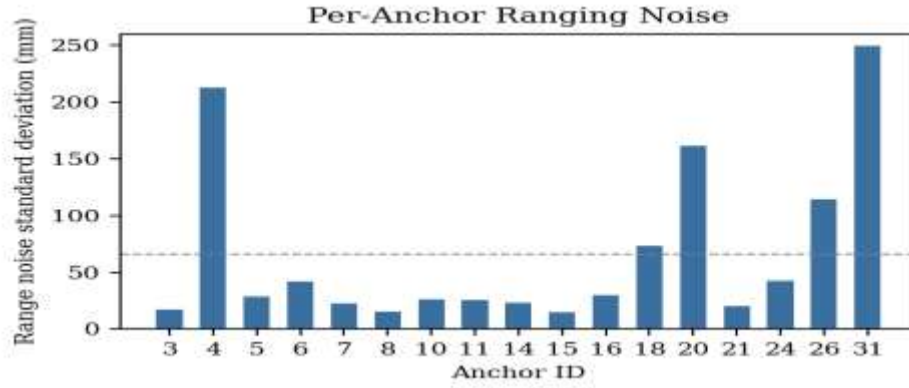


Fig. 2. Per-anchor ranging-noise standard deviation. The large variability of anchors 31, 4, and 20 illustrates why finite-sample reliability estimates can affect inverse-variance weighting.

Figure 1 demonstrates that the anchors do not share a common ranging-error characteristic. Anchors 4 and 14 show the largest positive systematic biases, approximately +449.4 mm and +471.1 mm. Other anchors also exhibit substantial positive or negative offsets. This result shows why a single global calibration constant is inadequate and supports per-anchor bias correction. Figure 2 shows similarly strong heterogeneity in stochastic error. Anchor 31 has the largest standard deviation, approximately 249.3 mm, followed by anchors 4 and 20. In contrast, anchor 15 has a standard deviation of only approximately 14.6 mm. Thus, the network contains both highly stable and highly variable ranging links.

Together, these figures justify separating calibration into two components: bias correction removes systematic error, while reliability weighting attempts to reduce the influence of noisy anchors. They also anticipate the tail-risk result because an uncertain variance estimate can lead to excessive inverse-variance weight.

B. Central Accuracy Improvement

TABLE II. NAIVE VERSUS CALIBRATED LOCALIZATION PERFORMANCE

Metric	Naive (m)	Calibrated (m)	Change
Mean	0.511	0.244	-52.3%
Median	0.469	0.119	-74.6%
90th percentile	0.832	1.031	+23.9%

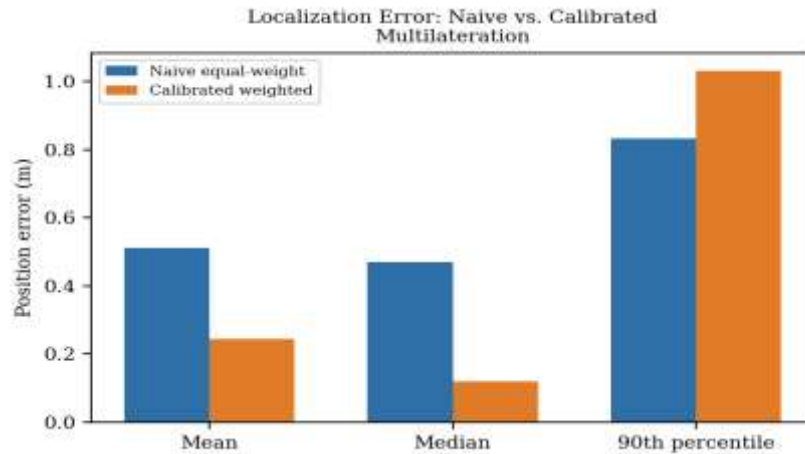


Fig. 3. Summary of localization error before and after per-anchor calibration and reliability weighting. The figure highlights the strong improvement in mean and median error together with the increase in 90th-percentile error.

Figure 3 makes the central result visually explicit. Mean error decreases from 0.511 m to 0.244 m, representing a 52.3% reduction. Median error decreases from 0.469 m to 0.119 m, representing a 74.6% reduction.

The median improvement is especially important because it indicates that the gain is not confined to a few unusually good fixes. A large fraction of the evaluated positions benefits from calibration. However, the 90th-percentile error increases from 0.832 m to 1.031 m. The figure therefore shows that calibration reshapes the distribution rather than uniformly translating all errors toward zero.

C. Bootstrap CDF and Tail Behavior

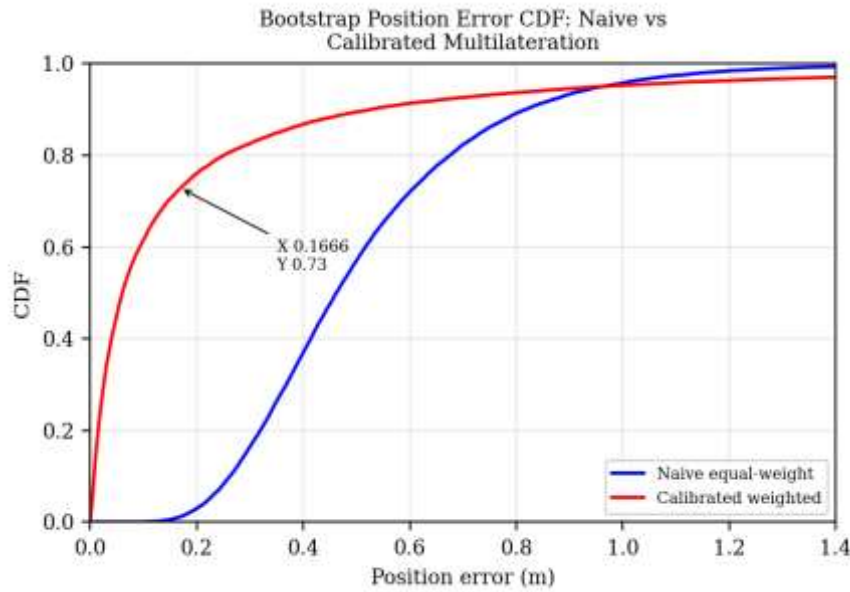


Fig. 4. Empirical bootstrap CDF of localization error for naive equal-weight and calibrated reliability-weighted multilateration. The calibrated distribution is substantially better in the central region but becomes less favorable in the upper tail.

Figure 4 provides distribution-level evidence for the trade-off. In the central region, the calibrated method produces smaller localization errors, consistent with the 74.6% median reduction. In the high-error region, the advantage diminishes and reverses, consistent with the increase in the 90th percentile.

The CDF is therefore more informative than mean error alone. It demonstrates that the proposed method improves typical-case accuracy but does not uniformly improve tail robustness.

TABLE III. INTERPRETATION OF THE MAIN EXPERIMENTAL FIGURES

Figure	Observation	Scientific implication
Fig. 1	Large positive and negative anchor biases	A global correction is insufficient; per-anchor calibration is required.
Fig. 2	Large differences in anchor noise	Anchor reliability is heterogeneous and should influence weighting.
Fig. 3	Mean and median improve while P90 worsens	Calibration improves typical accuracy but introduces tail risk.

Fig. 4	CDF improves centrally but loses advantage in the tail	Mean accuracy alone is insufficient; the full error distribution must be evaluated.
--------	--	---

V. DISCUSSION

The anchor-level results explain the central improvement. Large systematic offsets, particularly those of anchors 4 and 14, are well suited to bias correction. Once these offsets are removed, the multilateration solution is no longer systematically pulled toward incorrect positions.

The tail behavior has a different explanation. Inverse-variance weighting assumes that the calibration-period variance accurately represents future uncertainty. When an anchor is noisy or non-stationary, a finite calibration sample can underestimate its variance. The resulting weight may be too large:

$$s_i^2 \downarrow \Rightarrow w_i \uparrow \Rightarrow \text{greater influence of an unreliable measurement.} \quad (8)$$

This explains the observed pattern: calibration produces substantially better typical fixes but can occasionally produce larger errors.

Define the tail-risk ratio

$$R_{\text{tail}} = 1.031 / 0.832 \approx 1.239. \quad (9)$$

Thus, the calibrated method has approximately 23.9% greater 90th-percentile error, while reducing median error by 74.6%.

A practical safeguard is variance shrinkage:

$$s_{i,\text{shrink}}^2 = \lambda s_i^2 + (1 - \lambda) s_{\text{pool}}^2, \quad (10)$$

followed by

$$w_{\text{ireg}} = 1 / (s_{i,\text{shrink}}^2 + \epsilon). \quad (11)$$

Weight clipping can also prevent one anchor from dominating:

$$w_{\text{iclip}} = \min(w_{\text{ireg}}, w_{\text{max}}). \quad (12)$$

These methods are proposed as future safeguards rather than as additional experimental results.

VI. COMPARISON WITH PREVIOUS WORK

Shalaby et al. reported approximately 46% localization improvement through antenna-delay calibration and uncertainty-aware EKF fusion [6]. The present results show a 52.3% reduction in mean error and a 74.6% reduction in median error.

The comparison should be interpreted cautiously because the datasets, environments, algorithms, and metrics differ. The principal distinction of this study is its distribution-level evaluation. By reporting the 90th percentile and bootstrap CDF, the paper identifies a calibration trade-off that mean-only reporting could conceal.

It is also useful to situate the calibrated multilateration approach adopted here relative to two alternative families of indoor localization methods. First, fingerprinting-based approaches, which match an observed signal vector against a pre-surveyed radio map rather than inverting an explicit ranging model, range from early RSSI radio-map matching [12], [13] to more recent channel-state-information (CSI) deep-learning methods [14]. These approaches can achieve high accuracy without an explicit anchor-geometry or bias model, but they require substantial site-specific training data and must be resurveyed whenever the environment changes, whereas the calibration procedure evaluated here requires only a single surveyed reference position per anchor. Second, an orthogonal line of work addresses ranging uncertainty temporally rather than statically, fusing sequential UWB range measurements with inertial data via Kalman-family filters [22], [24], [25] or particle filters [23]. Such temporal fusion could plausibly mitigate the tail-risk increase identified in this study by smoothing occasional large errors over time; the present work instead isolates the calibration-and-weighting problem for a stationary tag at a single reference position, so that the accuracy/tail-risk trade-off can be attributed unambiguously to the weighting scheme rather than to filter dynamics, and leaves the combination of reliability weighting with temporal fusion to future work.

VII. LIMITATIONS AND FUTURE WORK

The tag was evaluated at one fixed reference position. Therefore, the results should not yet be generalized to every position in a large indoor environment. Future experiments should repeat the procedure at multiple locations and under different propagation conditions.

The calibration sample sizes also vary between anchors, making variance-estimation uncertainty important. Future work should evaluate variance shrinkage, Bayesian variance estimation, robust multilateration, Huber-type losses, and explicit weight caps.

Online recalibration is another important extension. A deployed system could continuously update anchor reliability using recent observations while constraining the rate at which weights change.

Finally, channel-impulse-response or IQ features could be combined with empirical ranging reliability to identify NLOS or degraded links before multilateration.

VIII. CONCLUSION

This study demonstrates that per-anchor bias calibration combined with inverse-variance-weighted multilateration can substantially improve typical UWB indoor localization accuracy while simultaneously creating a measurable tail-risk trade-off.

Mean error decreases by 52.3% and median error by 74.6%, whereas the 90th-percentile error increases by 23.9%. The figures provide a coherent explanation of this behavior: anchor-dependent systematic bias creates a strong opportunity for calibration, while anchor-dependent and uncertain noise characteristics limit the reliability of aggressive inverse-variance weighting. The practical conclusion is therefore precise: calibration is highly valuable when typical localization accuracy is the primary objective, but applications with strict upper-tail error requirements should combine calibration with regularized reliability weighting, weight capping, robust losses, or related safeguards.

The broader methodological lesson is that calibrated localization systems should be evaluated not only by how close their average estimate is to the truth, but also by how calibration changes the complete localization-error distribution.

REFERENCES

- [1] A. Alarifi et al., "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.
- [2] M. F. R. Al-Okby et al., "UWB-based real-time indoor positioning systems: A comprehensive review," *Applied Sciences*, vol. 14, no. 23, p. 11005, 2024.
- [3] J. Qiao et al., "Advancements in indoor precision positioning: A comprehensive survey of UWB and Wi-Fi RTT positioning technologies," *Network*, vol. 4, no. 4, pp. 545-566, 2024.
- [4] A. R. Jimenez Ruiz and F. Seco Granja, "Comparing Ubisense, BeSpoon, and DecaWave UWB location systems: Indoor performance analysis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 2106-2117, 2017.
- [5] S. Shah, L. Kovavisaruch, K. Kaemarungsi, and T. Demeechai, "Node calibration in UWB-based RTLSs using multiple simultaneous ranging," *Sensors*, vol. 22, no. 3, p. 864, 2022.
- [6] M. A. Shalaby et al., "Calibration and uncertainty characterization for ultra-wideband two-way-ranging measurements," *arXiv:2210.05888*, 2022.
- [7] J. Sidorenko et al., "Decawave UWB clock drift correction and power self-calibration," *Sensors*, vol. 19, no. 13, p. 2942, 2019.
- [8] J. Xin et al., "A Bayesian filtering approach for error mitigation in ultra-wideband ranging," *Sensors*, vol. 19, no. 3, p. 440, 2019.
- [9] H. Yang et al., "Ultra-wideband ranging error mitigation with novel channel impulse response feature parameters and two-step non-line-of-sight identification," *Sensors*, vol. 24, no. 5, p. 1703, 2024.
- [10] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568-2599, 2019.
- [11] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C*, vol. 37, no. 6, pp. 1067-1080, 2007.

- [12] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in Proc. IEEE INFOCOM, vol. 2, 2000, pp. 775-784.
- [13] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in Proc. 3rd Int. Conf. Mobile Syst., Appl., Services (MobiSys), 2005, pp. 205-218.
- [14] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," IEEE Trans. Veh. Technol., vol. 66, no. 1, pp. 763-776, Jan. 2017.
- [15] P.-C. Chen, "A non-line-of-sight error mitigation algorithm in location estimation," in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), vol. 1, 1999, pp. 316-320.
- [16] Y. T. Chan, W. Y. Tsui, H. C. So, and P. C. Ching, "Time-of-arrival based localization under NLOS conditions," IEEE Trans. Veh. Technol., vol. 55, no. 1, pp. 17-24, Jan. 2006.
- [17] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," IEEE Trans. Signal Process., vol. 42, no. 8, pp. 1905-1915, Aug. 1994.
- [18] W. H. Foy, "Position-location solutions by Taylor-series estimation," IEEE Trans. Aerosp. Electron. Syst., vol. AES-12, no. 2, pp. 187-194, 1976.
- [19] N. Levanon, "Lowest GDOP in 2-D scenarios," IEEE Proc.-Radar, Sonar Navig., vol. 147, no. 3, pp. 149-155, June 2000.
- [20] I. Sharp, K. Yu, and Y. J. Guo, "GDOP analysis for positioning system design," IEEE Trans. Veh. Technol., vol. 58, no. 7, pp. 3371-3382, 2009.
- [21] T. Sathyan, D. Humphrey, and M. Hedley, "WASP: A system and algorithms for accurate radio localization using low-cost hardware," IEEE Trans. Syst., Man, Cybern. C, vol. 41, no. 2, pp. 211-222, 2011.
- [22] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME J. Basic Eng., vol. 82, no. Series D, pp. 35-45, 1960.
- [23] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Process., vol. 50, no. 2, pp. 174-188, 2002.
- [24] B. Li, Z. Hao, and X. Dang, "An indoor location algorithm based on Kalman filter fusion of ultra-wide band and inertial measurement unit," AIP Adv., vol. 9, no. 8, p. 085210, 2019.
- [25] L. Markovic, M. Kovac, R. Milijas, M. Car, and S. Bogdan, "Error state extended Kalman filter multi-sensor fusion for unmanned aerial vehicle localization in GPS and magnetometer denied indoor environments," arXiv:2109.04908, 2021.
- [26] R. Fan and X. Du, "NLOS error mitigation using weighted least squares and Kalman filter in UWB positioning," arXiv:2205.05939, 2022.
- [27] B. Efron, "Bootstrap methods: Another look at the jackknife," Ann. Statist., vol. 7, no. 1, pp. 1-26, 1979.
- [28] P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist., vol. 35, no. 1, pp. 73-101, 1964.
- [29] W. James and C. Stein, "Estimation with quadratic loss," in Proc. 4th Berkeley Symp. Math. Statist. Probab., vol. 1, 1961, pp. 361-379.
- [30] IEEE Standard for Low-Rate Wireless Networks—Amendment 1: Enhanced Ultra Wideband (UWB) Physical Layers (PHYs) and Associated Ranging Techniques, IEEE Std. 802.15.4z-2020, 2020.