UWB Indoor Localization Based on XGBoost NLOS Identification and DS-TWR Ranging

Xiaoli YAO [1], Zhijiang XU [2], Guojun LIU [3]

[1] Inst. of Intelligent Transportation, Zhejiang Polytechnic University of Mechanical and Electrical Engineering, Binwen Rd. 528, 310053 Hangzhou, China
[2] School of Automation, Zhejiang Polytechnic University of Mechanical and Electrical Engineering, Binwen Rd. 528, 310053 Hangzhou, China
[3] Zhejiang Zhenhang Industrial Group Company Ltd., No. 2, Shanha Road, Nanmingshan Sub-district, 323000 Lishui, China

yaoxiaoli@zime.edu.cn, xuzhijiang@zime.edu.cn

Submitted January 27, 2026 / Accepted April 19, 2026 / Online first April 27, 2026

Abstract. *Indoor environments present significant challenges for ultra-wideband (UWB) localization due to ranging errors and non-line-of-sight (NLOS) propagation. This paper proposes a robust UWB indoor localization framework that integrates double-sided two-way ranging (DS-TWR), XGBoost-based NLOS identification, residual-weighted localization, and Kalman filter (KF). The main contribution of this work is the unified use of NLOS identification in both ranging correction and localization fusion, significantly improving localization accuracy in complex environments. Experimental results demonstrate improvements in ranging accuracy of up to 53.7% and 47.22% under human-body and wooden-board occlusions. In dynamic experiments, the proposed method outperforms conventional UWB localization, KF, and weighted least squares methods with positioning accuracy improvements of 38.64%, 28.95%, and 12.9%, respectively. These results confirm the framework's effectiveness in mitigating NLOS impact and enhancing UWB localization robustness.*

Keywords

UWB, indoor localization, NLOS identification, DS-TWR, XGBoost algorithm

1. Introduction

Accurate indoor localization is a key enabling technology for a wide range of applications, including intelligent transportation systems, smart buildings, industrial automation, and asset tracking. However, achieving high positioning accuracy in indoor environments remains challenging due to signal blockage, attenuation, and severe multipath propagation caused by walls, furniture, and human activities. Although global navigation satellite systems (GNSS), such as the Global Positioning System (GPS), provide reliable positioning performance in outdoor environments, their accuracy degrades significantly indoors, which severely limits their applicability for indoor localization.

To overcome these limitations, various radio-frequency-based indoor positioning technologies have been extensively studied. Among them, UWB technology has attracted considerable attention owing to its large bandwidth and high temporal resolution, which enable centimeter-level positioning accuracy under line-of-sight (LOS) conditions [1]. Nevertheless, in practical indoor environments, UWB localization performance is still affected by hardware imperfections, clock drift, measurement noise, and, most critically, NLOS propagation. NLOS conditions introduce positive ranging bias and significantly degrade localization accuracy and system robustness, especially in complex and dynamic indoor scenarios.

Conventional UWB localization approaches mainly rely on signal processing and filtering techniques, such as least squares estimation and KF, to suppress measurement noise. While these methods perform well under LOS conditions, they do not explicitly model or compensate for NLOS-induced ranging bias. Consequently, localization accuracy deteriorates markedly when NLOS propagation dominates the ranging measurements.

In recent years, NLOS identification has emerged as an effective strategy to enhance UWB localization robustness. By distinguishing between LOS and NLOS signals prior to localization, unreliable ranging measurements can be selectively corrected or down-weighted. With the advancement of machine learning, data-driven NLOS identification methods based on channel impulse response (CIR) features have demonstrated promising performance compared with conventional threshold-based techniques. Muqaibel et al. proposed an NLOS ranging-error mitigation strategy using CIR parameters and a neural network classifier to improve localization accuracy [2]. Santamaria-Pedrón et al. developed a deep learning-based method that predicts ranging errors from CIR features and compensates them in UWB systems [3]. Additionally, Majeed et al. introduced an attention-based deep learning architecture to classify LOS/NLOS conditions more effectively, which enhances the robustness of UWB positioning under challenging propagation conditions [4].

DOI: 10.13164/re.2026.0232

However, many existing studies focus on NLOS identification or localization optimization independently, without fully integrating NLOS recognition, ranging correction, and positioning fusion into a unified framework.

Motivated by these challenges, this paper proposes a robust UWB indoor localization framework that integrates DS-TWR, machine learning-based NLOS identification, residual-weighted localization, and KF-based fusion. By explicitly incorporating NLOS identification results into both the ranging correction and positioning stages, the proposed framework effectively mitigates NLOS-induced errors and significantly improves localization accuracy and robustness in complex indoor environments.

1.1 Related Works

Extensive research has been conducted to improve the accuracy and robustness of UWB-based indoor localization systems, particularly under NLOS conditions. Existing studies mainly focus on ranging error mitigation, NLOS identification, and reliability-aware localization.

KF has been widely applied to suppress random noise in UWB ranging measurements. Ren et al. [5] and Chen et al. [6] demonstrated that KF and clock drift compensation can improve ranging stability under LOS conditions. Tien Le Minh et al. [7] further applied KF to DS-TWR-based positioning in mixed LOS/NLOS scenarios. However, these filtering-based approaches do not explicitly model NLOS-induced positive bias, and their performance degrades significantly in severe NLOS environments.

Several studies addressed TOA/TOF ranging errors through optimization-based methods. Wang et al. [8] proposed adaptive error mitigation for TOF-based indoor tracking, while Svecova and Kocur [1] and Guo et al. [9] investigated TOA/TOF-based localization techniques in constrained environments. Although these methods improve ranging accuracy, they are sensitive to strong NLOS interference.

To reduce the influence of unreliable measurements, weighted least squares (WLS)-based localization methods have been introduced. Guvenc et al. [10] and Marano et al. [11] proposed NLOS identification and weighted localization frameworks using multipath channel statistics, achieving improved robustness in NLOS environments. Djosic et al. [12] further developed multi-algorithm UWB localization methods for mixed LOS/NLOS scenarios. Nevertheless, these approaches often rely on heuristic or statistical weighting strategies and do not fully exploit data-driven NLOS identification.

With the advancement of machine learning, data-driven NLOS identification based on CIR features has gained increasing attention. Xiao et al. [13] and Zeng et al. [14] demonstrated the effectiveness of CIR-based NLOS identification for UWB localization. More recently, Park et al. [15] employed deep learning models to achieve high LOS/NLOS classification accuracy, while Cortesi, Dreher et Magno [16] explored learning-based localization approaches. Despite their promising performance, many learning-based methods require large labeled datasets or high computational complexity, which may limit real-time deployment. Beyond CIR-based identification and conventional learning-based localization, Yao et al. [17] investigated a learning-assisted indoor localization framework by integrating dual-modal AOA/TOA fusion, data-driven error correction, and particle-filter-based estimation. Their work showed that learning-based correction can also be incorporated into multimodal localization frameworks to improve positioning performance in complex indoor environments.

In summary, existing studies on UWB-based indoor localization under NLOS conditions mainly concentrate on either NLOS identification or ranging and localization optimization as separate problems. Conventional signal processing and filtering-based methods are effective in suppressing random measurement noise under LOS conditions; however, their performance deteriorates significantly in complex NLOS environments, as NLOS-induced bias is not explicitly identified or compensated. Meanwhile, machine learning-based NLOS identification methods using CIR features have shown promising classification performance. Nevertheless, in many existing approaches, the identified NLOS information is not sufficiently exploited in subsequent ranging correction and localization fusion processes, which limits the overall performance improvement.

To overcome these limitations, this paper develops an integrated UWB indoor localization framework that combines DS-TWR ranging, machine learning-based NLOS identification, residual-weighted localization, and KF-based fusion. Unlike existing methods that address these components independently, the proposed framework explicitly incorporates NLOS identification results into both the ranging correction and positioning stages, thereby effectively mitigating NLOS-induced errors and enhancing localization accuracy and robustness in complex indoor environments.

1.2 Contributions

The main contributions of this work are summarized as follows:

XGBoost-based NLOS identification using CIR features: An NLOS identification model is developed using the XGBoost algorithm. Key features extracted from the CIR, including total impulse response power, amplitudes of the first three propagation paths, and noise standard deviation, are employed to accurately distinguish between LOS and NLOS propagation conditions.

Residual-weighted localization for NLOS error mitigation: A residual-weighted least squares localization method is designed to suppress NLOS-induced ranging bias. By dynamically adjusting the weights of base stations according to residual errors, the influence of unreliable NLOS measurements on positioning results is significantly reduced.

Fusion localization based on DS-TWR and KF: Corrected DS-TWR ranging measurements are fused with a KF to enhance localization robustness. This fusion strategy

effectively smooths positioning trajectories and improves both static and dynamic localization accuracy in complex indoor environments.

Comprehensive experimental validation: Extensive static and dynamic experiments are conducted under human-body and obstacle occlusion scenarios. Experimental results demonstrate that the proposed method consistently outperforms conventional UWB localization, weighted least squares, and KF-based approaches in terms of ranging and positioning accuracy.

1.3 Organization

The remainder of this paper is organized as follows. Section 2 introduces the UWB ranging principles, including TOA and DS-TWR, and presents the proposed XGBoost-based NLOS identification method together with CIR feature extraction. Section 3 describes the residual-weighted localization algorithm and the fusion localization framework based on KF. Section 4 presents the experimental setup and performance evaluation in both static and dynamic indoor environments. Finally, Section 5 concludes the paper and discusses future research directions.

2. UWB Ranging Method

2.1 TOA Ranging

In positioning, TOA usually refers to the time difference of the wireless signal from the transmitter to the receiver, which is the time for the signal to travel. By measuring the TOA, the distance of the signal propagation can be calculated, so as to realize the localization. Figure 1 shows the principle of 2-D TOA ranging. In Fig. 1, there are three base stations *A*, *B* and *C*, and the coordinates of each base station are *A* (x_1, y_1), *B* (x_2, y_2), *C* (x_3, y_3), and the real coordinate position of tag *TS* is (x, y). In the positioning process, the time synchronization of the base station and the tag is first synchronized, and then the time t_1, t_2, t_3 when the subscript signs in to each base station to send the UWB signal is recorded. Finally, by multiplying the propagation time by the speed of light, the distance relationship between the tag and the base station can be obtained according to the conventional TOA ranging principle [18], as shown in (1):

$$d_i = \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2} \tag{1}$$

where (x, y) denotes the unknown two-dimensional coordinates of the tag, and (x_i, y_i) represents the known coordinates of the i-th base station. d_i is the Euclidean distance between the tag and the i-th base station.

In the TOA positioning technique, if the clock between the UWB base station and the tag is not synchronized, then the recorded arrival time will be biased [19]. In this way, the circle formed by the difference of the arrival time of the UWB pulse signal received by the various base stations in Fig. 1 cannot intersect at a point, but forms a region. This means that the location of the tag cannot be accurately determined, but only the approximate range in which it is located.

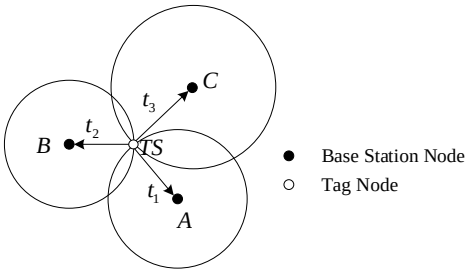

Fig. 1. Schematic diagram of TOA ranging.

2.2 DS-TWR Ranging Algorithm

Due to the complexity of the indoor environment, conventional wireless positioning techniques such as RSSI and fingerprinting are usually difficult to locate accurately. However, the DS-TWR ranging technology is not affected by the multipath effect and signal attenuation, so it has higher positioning accuracy and reliability [20] in indoor positioning and navigation. At the same time, DS-TWR is an improved method based on TOA ranging, which takes into account the propagation time difference between the signal in the process of sending and receiving. Therefore, DS-TWR algorithm is used for ranging. As shown in Fig. 2, the ranging principle of DS-TWR algorithm is to obtain the final TOA time by calculating the average of three times of TOA time. In this way, the DS-TWR algorithm reduces the four communications required by SDS-TWR to three, thereby improving ranging efficiency and reducing ranging errors.

Through the three-message exchange process, the TOF estimation in the DS-TWR scheme can be derived according to the standard DS-TWR formulation [21], as expressed in (2)–(4):

$$T_{\text{tof}} = \frac{T_{\text{round1}} \times T_{\text{round2}} - T_{\text{reply1}} \times T_{\text{reply2}}}{T_{\text{round1}} + T_{\text{round2}} + T_{\text{reply1}} + T_{\text{reply2}}} \tag{2}$$

where

$$T_{\text{round1}} = T_{\text{a2}} - T_{\text{a1}},$$
$$T_{\text{round2}} = T_{\text{b3}} - T_{\text{b2}},$$
$$T_{\text{reply1}} = T_{\text{b2}} - T_{\text{b1}},$$
$$T_{\text{reply2}} = T_{\text{a3}} - T_{\text{a2}}.$$

since

$$T_{\text{round1}} = T_{\text{reply1}} + 2T_{\text{tof}},$$
$$T_{\text{round2}} = T_{\text{reply2}} + 2T_{\text{tof}}.$$

therefore,

$$\begin{aligned} &\frac{T_{\text{round1}} \times T_{\text{round2}} - T_{\text{reply1}} \times T_{\text{reply2}}}{T_{\text{round1}} + T_{\text{round2}} + T_{\text{reply1}} + T_{\text{reply2}}} \\ &= \frac{\left(T_{\text{replyl}} + 2T_{\text{tof}}\right)\left(T_{\text{reply2}} + 2T_{\text{tof}}\right) - T_{\text{reply1}}T_{\text{reply2}}}{2T_{\text{reply1}} + 4T_{\text{tof}} + 2T_{\text{reply2}}} = T_{\text{tof}}. \end{aligned} \tag{3}$$

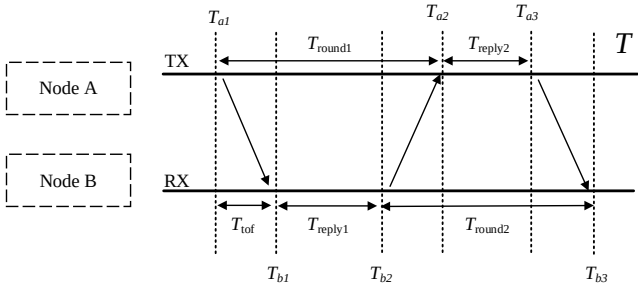

Fig. 2. Schematic diagram of DS-TWR ranging.

There may be a difference between the internal clock frequency of node A and node B, then there will be a clock deviation problem, when this clock deviation will lead to ranging error in practical applications. Considering further, if the current clock skew is k_a, k_b times of the original clock, the skew can be expressed as follows:

$$\text{err} = T_{\text{tof}} \times \left(1 - \frac{k_a + k_b}{2}\right). \tag{4}$$

Actually, in a real test system, the skew is small, so k_a, k_b will be close to 1. This makes the ranging error in the calculation of clock synchronization greatly reduced. In addition, by using the DS-TWR ranging method, we can estimate the time offset of each base station and tag, and use these estimates to eliminate the ranging error. In this way, the accuracy of ranging can be greatly improved and the error caused by clock drift can be suppressed.

3. CIR-Based NLOS Identification Using XGBoost

TOA positioning method plays an important role in indoor positioning because of its high accuracy and wide application. However, its accuracy is easily affected by NLOS signal propagation, which leads to large positioning errors. In order to solve this problem, this paper proposes a NLOS recognition model based on XGBoost algorithm and CIR signal. The model can effectively identify NLOS propagation signals by extracting features of CIR signals and using XGBoost for classification and recognition. Then, the residual weighted localization method is used to correct the identified NLOS signal to further improve the localization accuracy.

The objective function of XGBoost consists of two parts: a loss function and a regularization term, which are of the following form:

$$\partial_\theta(\theta) = \sum_i \mathcal{L}(y_i, \hat{y}_i) + \Omega(f). \tag{5}$$

where $\mathcal{L}(y_i, \hat{y}_i)$ is the loss function, which represents the difference between the predicted value $\hat{y}_i$ and the true value y_i; $\Omega(f)$ is the regularization term, which controls the complexity of the model.

The loss function reflects the prediction error of the model. For the regression task, the loss function is MSE, which is formulated as follows:

$$\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2. \tag{6}$$

For binary classification, the loss function is the log loss, as shown in the following equation:

$$\mathcal{L}(y_i, \hat{y}_i) = -[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]. \tag{7}$$

The regularization term $\Omega(f)$ is used to control the model complexity and prevent overfitting. The form of the regularization term $\Omega(f)$ is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2. \tag{8}$$

where T is the number of leaf nodes of the tree, w_j is the weight of each leaf node of the tree, γ and λ of are hyper parameters that regulate the model complexity.

To minimize the objective function via gradient optimization, XGBoost performs a Taylor expansion of the loss function. Assuming the previous model has $(t-1)$ already been trained, the current goal is to improve the prediction by adding a new model. Then the Taylor expansion of the loss function at the t-th iteration is as follows:

$$\mathcal{L}(\hat{y}_i) \approx \mathcal{L}\left(\hat{y}_i^{(t-1)}\right) + \left(\frac{\partial \mathcal{L}}{\partial \hat{y}_i^{(t-1)}}\right) f_t(x) + \frac{1}{2}\left(\frac{\partial^2 \mathcal{L}}{\partial (\hat{y}_i^{(t-1)})^2}\right) f_t^2(x) \tag{9}$$

where $\frac{\partial \mathcal{L}}{\partial \hat{y}_i^{(t-1)}}$ is the first degree of order and $\frac{\partial^2 \mathcal{L}}{\partial (\hat{y}_i^{(t-1)})^2}$ is the second derivative.

By computing the gradient and second derivative for each sample, XGBoost is able to optimize the objective function. At each iteration, the weights of the leaf nodes of the tree are updated by the first-order and second-order gradients. The gradient and the second derivative are shown in (10) and (11):

$$g_i = \frac{\partial \mathcal{L}}{\partial \hat{y}_i}, \tag{10}$$

$$h_i = \frac{\partial^2 \mathcal{L}}{\partial \hat{y}_i^2}. \tag{11}$$

These gradients and second derivatives are used to assess the quality of each split point and to select the best split point by a gain calculation. XGBoost selects the best split point by calculating the gain. The gain represents the amount of error reduction by splitting at the current node. The gain is calculated as follows:

$$\text{Gain} = \frac{1}{2}\left(\frac{(\sum_{i\in L} g_i)^2}{\sum_{i\in L} h_i + \lambda} + \frac{(\sum_{i\in R} g_i)^2}{\sum_{i\in R} h_i + \lambda} - \frac{(\sum_{i\in S} g_i)^2}{\sum_{i\in S} h_i + \lambda}\right). \tag{12}$$

where g_i and h_i are the gradient and second derivative of the g_i and h_i are the gradient and second derivative of the sample, respectively; L and R are the left and right child nodes, and S is the parent node.

The leaf node weights w_j of each tree are updated using the following formula:

$$w_j = -\frac{\sum_{i\in j} g_i}{\sum_{i\in j} h_i + \lambda}. \tag{13}$$

This update formula makes the output of each leaf node a weighted sum of the gradients of the samples within that leaf node.

XGBoost controls the complexity of the tree by introducing a regularization term to avoid overfitting. The resulting objective function is of the form:

$$\mathrm{Obj}(\theta) = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i) + \Omega(f). \tag{14}$$

XGBoost controls the maximum depth of the tree to limit the model complexity. The maximum depth of the tree affects the number of leaf nodes T in the tree. When the maximum depth is set to max_depth, the number of leaf nodes is:

$$T = 2^{\text{max\_depth}} - 1. \tag{15}$$

By adjusting the depth of the tree, XGBoost can control the complexity and computational efficiency of the model.

The objective function of XGBoost combines a loss function and a regularization term, and the optimization process is performed based on first-order and second-order gradients. The final objective function is:

$$\mathrm{Obj}(\theta) = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i) + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2. \tag{16}$$

3.1 Feature Extraction Based on CIR

In wireless communication systems, the CIR is an important representation of signal propagation characteristics. It describes the transmission process of a signal from the transmitter to the receiver, including all propagation paths as well as their corresponding time delays and path gains. By analyzing the CIR, the propagation condition can be effectively characterized, enabling the discrimination between LOS and NLOS signal propagation.

Under LOS conditions, the signal propagates mainly along a direct path with minimal obstruction. As a result, the CIR typically exhibits a dominant pulse with a short propagation delay and a large amplitude. The CIR response of an LOS signal can therefore be approximated by a single dominant path, which can be expressed as:

$$h(t) = \delta(t - \tau_0) \tag{17}$$

where δ is the unit pulse function and τ_0 is the propagation delay of the LOS signal.

The propagation path of LOS signal is direct and obstacle-free, the signal attenuation is small, and its impulse response power is usually large.

The propagation of NLOS signals is more complex, as the signal not only travels along the direct path during propagation, but also experiences multipath effects, such as reflection, refraction, and scattering from obstacles such as buildings. For NLOS signals, the CIR response is represented by multiple pulses, which represent different propagation paths and can be expressed as follows:

$$h(t) = \sum_{i=1}^{n} \alpha_i \delta(t - \tau_i) \tag{18}$$

where α_i is the signal gain of the i-th path, τ_i is the propagation delay of the i-th path, and n is the number of paths.

The CIR response of an NLOS signal usually exhibits multiple pulses with longer delays and relatively smaller amplitudes due to multipath propagation, whereas LOS signals are generally characterized by a dominant first-arrival path and a more concentrated energy distribution. For each measurement sample, the raw CIR data output by the UWB receiver are first collected, and the propagation paths are indexed according to their arrival order. As a preprocessing step, the CIR samples are ordered according to the path arrival sequence, and the same feature extraction procedure is applied to all LOS and NLOS samples. Based on the processed CIR data, the total CIR power, the amplitudes of the first three resolvable propagation paths, and the noise standard deviation are extracted as representative features for LOS/NLOS classification, which are described in detail as follows:

1. Total CIR power:

This feature reflects the total energy of the signal in the propagation process, indicating the overall strength of the signal through multipath propagation. In the case of LOS, the signal energy is concentrated in the first path, and the power value is large. However, in the case of NLOS, due to the multi-path propagation, the total power of the signal may be dispersed by the reflected signals of multiple paths, so the total power may be dispersed. The calculation formula is given as follows:

$$P_{\text{total}} = \sum_{m=1}^{M} |h_m|^2 \tag{19}$$

where h_m represents the signal amplitude of the m-th path, and M is the total number of paths.

2. Amplitude of the first path:

For LOS signals [22], the amplitude of the first path is usually the largest, because the signal propagation path is the shortest and the energy loss is the least. However, in NLOS signals, the amplitude of the first path may be relatively small due to the multipath effect, and there may be multiple reflected paths with similar amplitudes to the first path. The calculation formula is as follows:

$$A_1 = |h_1| \tag{20}$$

where h_1 is the amplitude of the first path in CIR.

3. Amplitude of the second path:

In NLOS propagation, the signal transmission in the

process of reflection or block, the amplitude of the second path is larger, usually indicates that signal reflection path compared with the direct path also played an important role. For LOS signals, the amplitude of the second path is usually small or even zero. The calculation formula is as follows:

$$A_2 = |h_2| \tag{21}$$

where h_2 is the amplitude of the second path in CIR.

4. Amplitude of the third path:

In complex NLOS environments, the amplitudes of the third path and higher paths are larger, especially in scenes with many reflecting surfaces. These paths may be larger than the amplitudes of the first path, reflecting the multipath characteristics of the environment. Calculation formula as shown in (22):

$$A_3 = |h_3| \tag{22}$$

where h_3 is the amplitude of the third path in CIR.

5. Standard deviation of noise:

NLOS signals are usually accompanied by higher noise because the multipath effect introduces more signal interference, resulting in a lower SNR. By calculating the standard deviation of noise, it can be further determined whether the signal is affected by shade or reflection, and indirectly determine whether it is the NLOS signal. The calculation formula is shown in (23):

$$\sigma_n = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_i - \bar{h})^2} \tag{23}$$

where $\bar{h}$ is the mean of the amplitudes of all paths in CIR.

By analyzing the above features, LOS and NLOS signals can be distinguished according to the following criteria:

(1) When the amplitude of the first arriving path is greater than the second and subsequent amplitude, and total power is focusing on the first path signal and shorter delay time of arrival, you can say the signals for LOS.

(2) When the amplitude of the path of the second, third, and so on multiple paths increases obviously, and the signal arrival time delay increased significantly, it can be thought of as the signal for the NLOS signal. At this time, the noise level is also high, and the interference of the signal is large.

3.2 XGBoost Model Training

Since the CIR descriptors used in this study are structured and low-dimensional statistical features rather than high-dimensional raw signals, a tree-boosting framework is adopted to model their nonlinear interactions. XGBoost is well suited to this setting because it can effectively capture the coupled effects among path amplitudes, total power, and noise statistics, while maintaining good generalization ability on a moderate-sized dataset through regularization and additive boosting. In comparison with kernel-based classifiers, bagging-based tree models, and neural-network architectures, this framework provides a more suitable balance between discrimination capability, robustness, and computational efficiency for the CIR-based LOS/NLOS identification task.

The XGBoost classifier is implemented in a supervised binary classification setting, the model is configured with a maximum tree depth of 12, a learning rate of 0.1, and 500 estimators. The classifier is trained using the CIR-based feature vectors extracted from the measured signals, and the boosting procedure is performed iteratively to construct the final decision model.

For each ranging observation, the UWB receiver outputs the corresponding CIR measurement together with the measured signal information. In this study, the CIR data used for model development are obtained directly from the measured UWB signals. During the offline processing stage, the CIR measurements are organized and used to calculate the total CIR power, extract the amplitudes of the first three propagation paths, and compute the noise standard deviation, thereby forming the CIR-based feature vectors for classifier training and testing. During online deployment, the same feature extraction procedure is applied to the incoming CIR measurements, and the extracted features are then fed into the trained XGBoost classifier for LOS/NLOS identification.

Figure 3 illustrates the structure of the XGBoost-based NLOS recognition model. The workflow begins with data acquisition, where the required CIR signals are collected. Relevant features are then extracted from the CIR, including the total CIR power, the amplitudes of the first, second, and third paths, and the noise standard deviation. These features are subsequently used as input to the XGBoost classifier.

During the training process, XGBoost incrementally constructs an ensemble of classification trees by successively adding new trees to optimize the loss function and enhance the classification performance. Each tree is split according to the information gain of its nodes, and samples reaching the leaf nodes are assigned corresponding weights [23]. The contribution of each feature to the classification result is reflected by the learned tree structure and leaf weights. After multiple training iterations and tree expansions, the model learns decision rules that effectively distinguish between LOS and NLOS propagation conditions, thereby completing the NLOS recognition task.

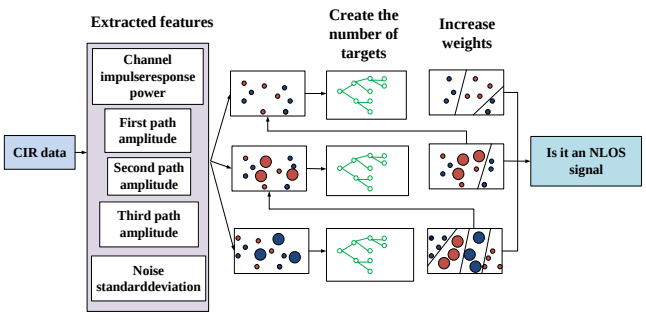

Fig. 3. NLOS recognition process based on XGBoost algorithm.

Step by step in the process of model training, XGBoost generates more classification tree, each generated a tree, the model will learn a new objective function, makes the final prediction results more close to the real value.

3.3 Positioning Algorithm Based on Residual Weighting

After the classification of LOS/NLOS signals based on XGBoost algorithm, the corresponding localization algorithm was selected for different signal types. For LOS signal, using the traditional method of TOA calculation; for NLOS signals, the residual weighted localization algorithm is used. The algorithm reduces the influence of NLOS signal on positioning results by weighting. Firstly, the residual error of each base station was calculated to evaluate its influence on the positioning results. Secondly according to the size of residual distribution of weight. Finally, the weighted least square method is used to optimize the positioning results. By giving LOS base station bigger weight and NLOS base stations smaller weight, the algorithm effectively reduces the influence of NLOS error to the final positioning result. Weighted residuals positioning algorithm based on NLOS identification process is shown in Fig. 4.

For a target location (x, y) and multiple base stations (x_i, y_i), the difference between the distance measured by each base station d_i and the theoretically calculated distance d_i' is called the residual r_i, which represents the measurement error. Namely:

$$r_i = d_i - d_i' \tag{24}$$

where $d_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$.

In order to reduce the influence of NLOS signal on the positioning results, the residual error r_i of the base station is used to dynamically adjust the weight w_i of each base station. The weighting function is as follows:

$$w_i = \frac{1}{1+\lambda r_i^2} \tag{25}$$

where λ is to control the error of the influence of weight factor. The larger the residual, the greater the base station's measurement error, and the smaller the corresponding weight.

After the weights are calculated, the weighted least squares method is used to optimize the target position (x, y). Optimization goal is to minimize the sum of squares of the weighted residuals, formula is as follows:

$$\min_{x,y} \sum_{i=1}^{n} w_i \left(d_i' - \sqrt{(x-x_i)^2 + (y-y_i)^2} \right)^2 . \tag{26}$$

Among them, n is the number of base stations. This process is optimized by an iterative method. In each iteration, the new target position estimate is used to update the residual r_i, thereby updating the weight w_i of each base station. In each iteration, the sum of squares of the weighted residuals was gradually reduced, and finally converged to the optimal solution, which was the final estimate of the target position. In order to avoid over-reliance on the weights of some base stations, it is usually necessary to normalize the weights of all base stations to ensure that the sum of the weights is 1. The normalized weights are calculated using the following formula:

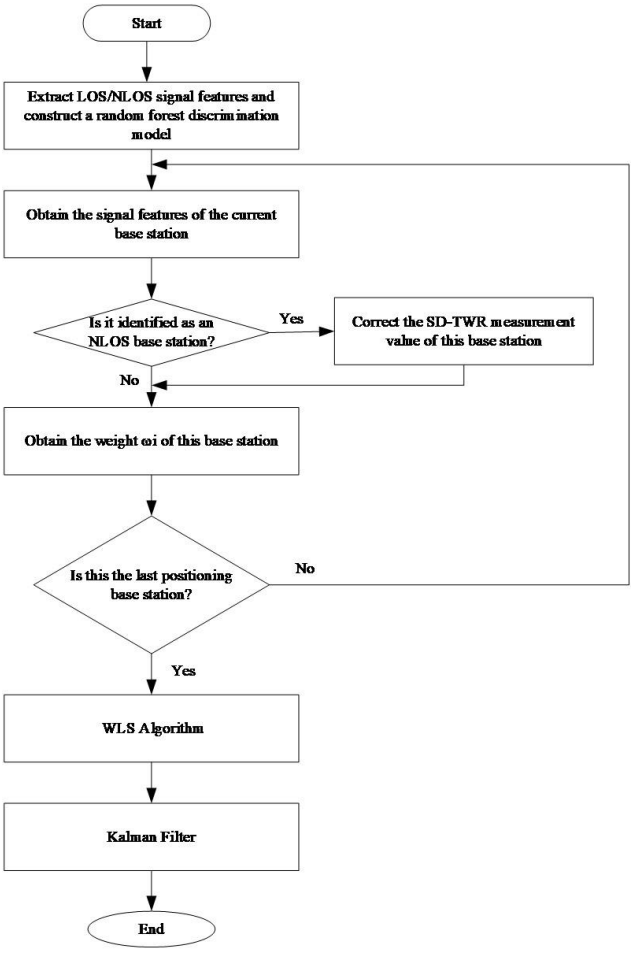

Fig. 4. Weighted residuals positioning algorithm based on NLOS identification process.

$$w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j} . \tag{27}$$

This normalization enables the weights of all base stations to be compared on the same scale, thus avoiding the influence of some base stations on the positioning results. The final optimization objective function becomes:

$$\min_{x,y} \sum_{i=1}^{n} w_i' \left(d_i' - \sqrt{(x-x_i)^2 + (y-y_i)^2} \right)^2 . \tag{28}$$

Through normalization, the residual weighted localization algorithm can use the measurement results of each base station more fairly, so as to improve the accuracy of the final localization results.

To clearly illustrate the workflow of the proposed method, the pseudocode is given in Algorithm 1.

Algorithm 1. Pseudocode of the proposed UWB localization framework

```
Input:
UWB raw signal measurements (CIR data)
Pre-trained XGBoost model for NLOS classification
Base station locations
Dynamic movement data (coordinates, velocities)
Output:
Estimated position of the mobile tag
1: Initialize position estimate (x, y) = (0, 0)
2: Initialize Kalman filter parameters (state, covariance matrix)
3: for each UWB signal measurement do
4:     Extract CIR features (total impulse response power, path amplitudes, noise standard deviation)
5:     Use the pre-trained XGBoost model to query NLOS conditions
6:     if (NLOS detected) then
7:         Apply residual- WLS to correct the ranging bias
8:     else
9:         Use raw UWB measurements for localization
10:    end if
11:    Input corrected or raw measurements into Kalman filter
12:    Update position estimate based on Kalman filter
13:    if (Position estimate stable) then
14:        Exit loop // When position stabilizes, exit
15:    end if
16: end for
17: Output final position estimate from Kalman filter
End Algorithm
```

4. Fusion Localization Algorithm

Firstly, the weight of DS-TWR correction value was determined. For each location BS, the known LOS/NLOS data set can be used to calculate the LOS and NLOS probability density function of the BS, and then the weight of the BS can be determined according to the ratio of the probability density function. For the base station identified as a NLOS base station, the corrected DS-TWR value can be used as the ranging value of the base station. Finally, the estimated position of the mobile tag is obtained by using the weighted least square method.

Then, the estimated value is optimized to obtain a more accurate positioning result. Specifically, in the KF algorithm, the DS-TWR measurement value is taken as the observation value, and the coordinates of the moving target are taken as the state variable. The state transition matrix and the observation matrix are used to predict and update the state, so as to obtain the optimal estimate value of the moving target. In this way, the DS-TWR localization algorithm under NLOS can be realized, and more accurate localization result can be obtained [24], [25].

4.1 Weighted Least Squares Method

The estimated value of DS-TWR ranging time in LOS environment can be expressed as:

$$\hat{t} = d / c + n_\tau \tag{29}$$

where n_τ is the additive Gaussian white noise of the time measurement, d denotes the true distance between the base station and the mobile tag, c represents the speed of light in vacuum.

At a given distance d and $\hat{t}$ the corresponding probability density function is:

$$f_i\left(t \mid d, \mathrm{LOS}\right) = \left(2\pi\sigma_\tau^{\,2}\right)^{-1/2} \exp\left\{-[t-(d/c)]^2 / 2\sigma_\tau^{\,2}\right\} \tag{30}$$

where σ_τ is standard deviation of DS-TWR measurement noise.

In the NLOS environment, the estimated value of DS-TWR ranging time is positively biased with high probability, which can be expressed as:

$$\hat{t} = d / c + n_\tau + b_\tau \tag{31}$$

where b_τ is NLOS error.

Then the probability density function in the NLOS environment is:

$$\begin{aligned} &f_i\left(t \mid d, \mathrm{NLOS}\right) = \\ &\frac{1}{2\lambda_\mathrm{b}} \exp\left(\sigma_\tau^{\,2} / 2\lambda_\mathrm{b}^{\,2}\right) \exp\left[-\left(t-d/c\right)/\lambda_\mathrm{b}\right] \\ &\times\left\{1+\mathrm{erf}\left[\frac{\lambda_\mathrm{b}\left(t-d/c\right)-\sigma_\tau^2}{\sqrt{2}\sigma_\mathrm{t}\lambda_\mathrm{b}}\right]\right\}. \end{aligned} \tag{32}$$

When $J_i(t,d) = f_\mathrm{t}(t/d,\mathrm{LOS}) / f_\mathrm{t}(t/d,\mathrm{NLOS})$ then the weight w_i of the coordinate of the i-th base station is calculated by

$$w_i = \begin{cases} k_1, & J_i\left(t,d\right) \le \Delta_1 \\ k_2, \Delta_1 < J_i\left(t,d\right) \le \Delta_2 \\ k_3, & J_i\left(t,d\right) > \Delta_2 \end{cases} \tag{33}$$

where $k_1 < k_2 < k_3$, k_1 and k_3 are the weights of DS-TWR ranging time values in NLOS environment and LOS environment, respectively.

According to the distance relationship between the mobile tag and the base station, Equation (34) is established:

$$d_i = \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2}; i = 1, 2, \cdots, M. \tag{34}$$

By processing (34), the following can be obtained.

$$X_{i,1}x + Y_{i,1}y + d_{i,1}d_1 = \left(K_i - K_1 - d_{i,1}^2\right)/2 \tag{35}$$

where $X_{i,1} = X_i - X_1$, $Y_{i,1} = Y_i - Y_1$, $K_i = X_i^2 + Y_i^2$.

The position estimate of the moving tag can be obtained by using the weighted least squares method:

$$\mathbf{Z} = \left(\mathbf{G}_\mathrm{a}^{\mathrm{T}}\mathbf{W}\mathbf{G}_\mathrm{a}\right)^{-1}\mathbf{G}_\mathrm{a}^{\mathrm{T}}\mathbf{W}\mathbf{h} \tag{36}$$

where

$$\mathbf{Z} = \begin{bmatrix} x \\ y \\ d_1 \end{bmatrix}, \mathbf{G}_\mathrm{a} = \begin{bmatrix} X_{2,1} & Y_{2,1} & d_{2,1} \\ X_{3,1} & Y_{3,1} & d_{3,1} \\ \vdots & \vdots & \vdots \\ X_{M,1} & Y_{M,1} & d_{M,1} \end{bmatrix}, \mathbf{h} = \frac{1}{2}\begin{bmatrix} K_2 - K_1 - d_{2,1}^2 \\ K_3 - K_1 - d_{3,1}^2 \\ \cdots \\ K_M - K_1 - d_{M,1}^{\,2} \end{bmatrix},$$

W is obtained according to (33).

4.2 Kalman Filter Algorithm

In localization tasks, the distance between the mobile tag and the base station can be used to estimate the position. Specifically, by measuring the distance between the mobile tag and the base station, a set of equations is generated, which can be solved using the nonlinear least squares method to obtain the coordinates of the mobile tag. However, in practical applications, measurement errors and environmental interference can lead to deviations in the positioning results. These deviations are commonly modeled as noise, which is assumed to be independent and uncorrelated with zero mean.

To address the uncertainties in the localization process, we combine the KF with NLOS identification techniques. This approach updates each measurement in real-time, thereby improving positioning accuracy and enhancing system robustness. The iterative process of the KF is discussed in detail in reference [7].

5. Experimental Results and Analysis

5.1 Analysis of NLOS Identification Experiments

To evaluate the effectiveness of XGBoost in NLOS identification, UWB devices manufactured by Haoru Technology Co., Ltd. were used to collect CIR data for model training and testing. The devices are equipped with a high-resolution time-of-flight (TOF) measurement system, and support multiple frequency bands ranging from 3.5 GHz to 6.5 GHz and offers three different transmission rates: 110 kbps, 850 kbps, and 6.8 Mbps, which can be adjusted according to user needs.

To achieve accurate localization, the base station uses the TOA positioning method, which is suitable for indoor localization tasks. The localization base station equipment used in the experiment is shown in Fig. 5(left). The tag device used in the experiment is shown in Fig. 5(right). During testing, the pedestrian wears the tag on their wrist, and the tag interacts with the localization base station through wireless communication to collect positioning data, which is then processed for coordinate calculation on the backend server.

In the experiments, for XGBoost-based LOS/NLOS classification, a labeled dataset was constructed using the

Fig. 5. Experimental data collection setup: UWB localization base station (left) and TOA positioning tag (right).

Parameter	Value/Setting
Number of anchor stations	4
Anchor locations	(0, 0.75), (0, 5), (6, 0.75), (6, 5)
Distance	3 meters (static experiment), tag moves in dynamic experiment
Scenario lettings	Static environment (with occlusion), dynamic environment (no occlusion)
Sampling rate	100 Hz

Tab.1. Experimental parameters table.

Parameter	Value/Setting
Classification task	Binary classification (LOS/NLOS)
Dataset size	6,200 samples
Class distribution	3,100 LOS / 3,100 NLOS
Train/Test dplit	70% / 30%
CIR features	Total CIR power; first-path amplitude; second-path amplitude; third-path amplitude; noise standard deviation
Objective function	binary:logistic
Evaluation metric	logloss
Maximum tree depth	12
Learning rate	0.1
Number of estimators	500
Subsample	0.8
colsample_bytree	0.8
min_child_weight	1
gamma	0
reg_alpha	0
reg_lambda	1
Random seed	42
Early stopping	Not used
Training mode	Offline
Inference mode	Online

Tab. 2. XGBoost model settings.

UWB experimental platform under both static occlusion scenarios and dynamic indoor scenarios. During data acquisition, the UWB tag communicated with the base stations, and the corresponding ranging measurements were recorded to extract CIR-based features for classification. The final dataset contained 6,200 samples, including 3,100 LOS samples and 3,100 NLOS samples. All samples were randomly divided into a training set (70%) and a test set (30%). The experimental parameters are summarized in Tab. 1.

The data processing pipeline and the XGBoost training configuration are summarized as follows. In this work, both the offline training stage and the online processing stage were conducted in MATLAB. For each sample, five CIR-derived features were extracted, namely the total CIR power, the amplitudes of the first three propagation paths, and the

noise standard deviation. The XGBoost classifier was implemented with the objective function set to binary:logistic and the evaluation metric set to logloss. During the offline stage, the CIR features were extracted and used to train the XGBoost classifier, whereas during online deployment the same feature extraction procedure was applied to incoming CIR measurements and the pre-trained model was used for LOS/NLOS identification. For NLOS signals, residual-WLS was employed to reduce the ranging bias, and the corrected measurements were then fed into the KF for position estimation. The main XGBoost model settings are listed in Tab. 2.

Figure 6 shows the convergence behavior of different classification models during the training process. After 200 iterations, the final loss value of the SVM model was 0.036063, while the CNN model achieved a final loss value of 0.027063. In contrast, the XGBoost model converged more rapidly and reached a significantly lower loss value of 0.0120979 under the same number of iterations. These results indicate that the XGBoost model exhibits faster convergence and better optimization performance compared with the other models.

As shown in Tab. 3, XGBoost achieves the highest accuracy, recall, and F1-score among all evaluated methods. Specifically, the XGBoost model improves the classification accuracy by 7.1% and 4.8% compared with SVM and KNN, respectively. The experimental results demonstrate that XGBoost can effectively extract discriminative features from CIR signals and exhibits superior recognition accuracy and generalization capability in NLOS identification tasks.

5.2 Localization Experiment Analysis

The UWB bracelet tag and the UWB base station were vertically placed in a LOS environment. The base station was installed at a height of 4.5 m above the ground, and the ground-truth distance between the bracelet tag and the base station was 3 m. The average ranging error and maximum ranging error of different algorithms were calculated, as shown in Fig 7.

Figure 7 shows that, after applying the data correction described in (13), the average ranging errors of the raw UWB measurements, KF, weighted least squares, and the proposed method under LOS conditions are 0.15 m, 0.135 m, 0.117 m, and 0.092 m, respectively. The corresponding maximum ranging errors are 0.33 m, 0.25 m, 0.23 m, and 0.20 m, respectively. These results indicate that the proposed method achieves the lowest average and maximum ranging errors among the evaluated algorithms under LOS conditions.

The NLOS scenario was generated by introducing human-body and wooden-board occlusions at a distance of 0.50 m from the bracelet tag, with each occlusion lasting 30 s, as shown in Fig. 8. During the experiment, a total of 1,000 ranging samples were collected. The distance distributions obtained from the ground-truth distance, raw UWB measurements, weighted least squares, KF, and the proposed method are illustrated in Fig. 9. Figure 9 (left) shows that different types of occlusions impose varying impacts on

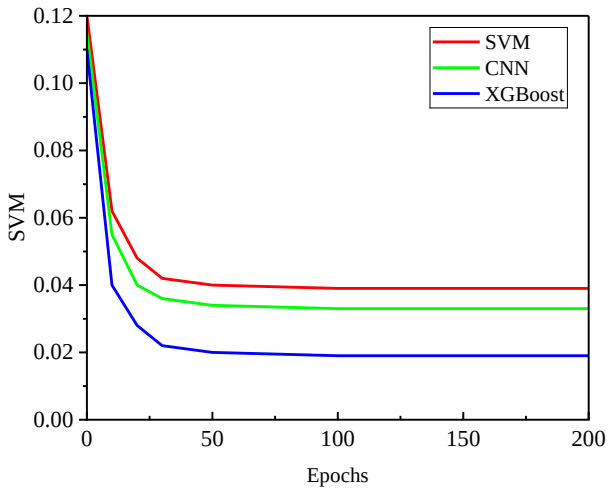

Fig. 6. Different iterative process model diagram.

Algorithm	Accuracy (%)	Recall (%)	F1-score (%)
SVM	85.2	82.1	83.6
KNN	87.5	84.3	85.9
XGBoost	92.3	90.1	91.2

Tab. 3. Comparison of experimental results for different models.

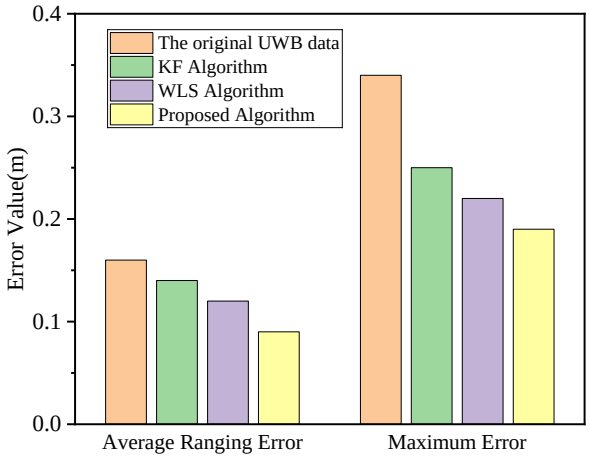

Fig. 7. Comparison of errors of various algorithms in LOS environment.

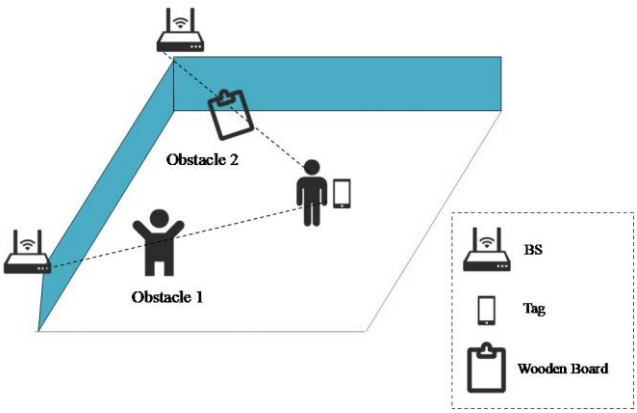

Fig. 8. Experimental scene diagram.

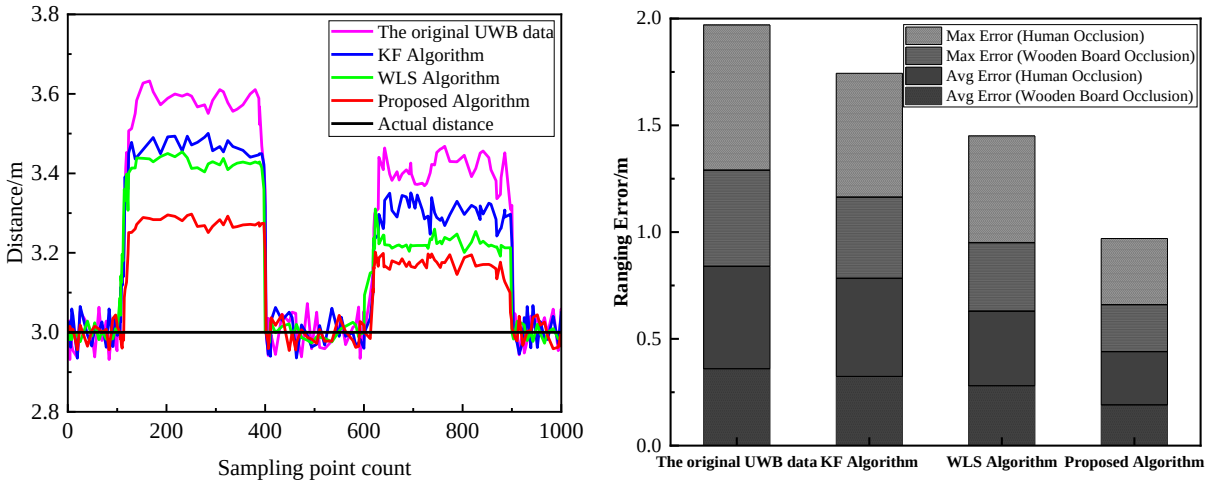

Fig. 9. Error comparison of each algorithm in NLOS static experiment.

UWB ranging performance. The corresponding error statistics after sorting and comparison are presented in Fig. 9 (right). Under human-body occlusion, the average ranging errors of the raw UWB measurements, KF, weighted least squares, and the proposed method are 0.54 m, 0.52 m, 0.42 m, and 0.25 m, respectively, while the maximum ranging errors are 0.68 m, 0.67 m, 0.50 m, and 0.31 m, respectively.

When the UWB signal is blocked by a wooden board, the average ranging errors of the raw UWB measurements, KF, weighted least squares, and the proposed method are 0.36 m, 0.324 m, 0.245 m, and 0.19 m, respectively, and the corresponding maximum errors are 0.45 m, 0.40 m, 0.37 m, and 0.22 m, respectively. Overall, the proposed method exhibits superior NLOS error suppression capability. Compared with raw UWB measurements, KF, and weighted least squares methods, the proposed approach improves ranging accuracy under human-body occlusion by 53.7%, 51.92%, and 40.48%, respectively. In the wooden-board occlusion scenario, the ranging accuracy improvements reach 47.22%, 41.36%, and 22.45%, respectively. To clearly present the results of the static experiments, the corresponding error statistics are summarized in Tab. 4.

To further analyze the reliability of the experimental results, we calculated the 95% confidence intervals for the localization errors of each method. In the static experiment, the XGBoost framework had an average localization error of 0.092 meters under human-body occlusion, with a 95% confidence interval of [0.085 m, 0.100 m], while the traditional UWB localization method had an error of 0.15 meters, with a 95% confidence interval of [0.14 m, 0.16 m]. Additionally, we performed an independent samples t-test, and the results showed that the difference in localization accuracy between the XGBoost framework and the traditional method is statistically significant (p-value < 0.05), indicating that the proposed method has an advantage in improving localization accuracy.

Algorithm	Mean error (m)	Maximum error (m)	RMSE (m)	Percentage improvement (%)
UWB	0.15	0.33	0.23	-
KF	0.135	0.25	0.19	10.0
WLS	0.117	0.23	0.18	21.3
Proposed algorithm	0.092	0.20	0.15	53.7

Tab. 4. Static experiment results.

To evaluate the robustness and stability of the proposed localization algorithm, dynamic experiments were conducted using UWB equipment in a typical indoor environment. The experimental site was a university conference room, as shown in Fig. 10 (left). The room contained a conference table, multiple chairs, bookcases, and a glass wall, which introduced severe multipath propagation and NLOS interference, thereby representing a complex indoor environment.

A rectangular area of 5 m × 4 m was defined as the experimental region, with corner coordinates at (0.5, 0.5), (0.5, 4.5), (5.5, 0.5), and (5.5, 4.5). During the experiment, the UWB tag was worn at the waist at a height of 0.8 m, and its initial position was set to (0.5, 0.5). The tag moved in the two-dimensional x–y plane with a constant horizontal and vertical velocity of 0.1 m/s. The schematic diagram of the experimental setup is shown in Fig. 10 (right). Four UWB base stations were deployed around the experimental area, with coordinates at (0, 0.75), (0, 5), (6, 0.75), and (6, 5), respectively.

Figure 11 (left) shows the trajectories obtained using the ground-truth path, raw UWB measurements, weighted least squares, KF, and the proposed method under NLOS conditions. All methods are capable of recovering the motion trajectory of the mobile target. However, compared with raw UWB measurements, weighted least squares, and KF, the proposed method produces a trajectory with reduced deviation from the ground-truth path. Moreover, the overall trajectory more closely follows the real path, and the trajectory at corner regions is smoother.

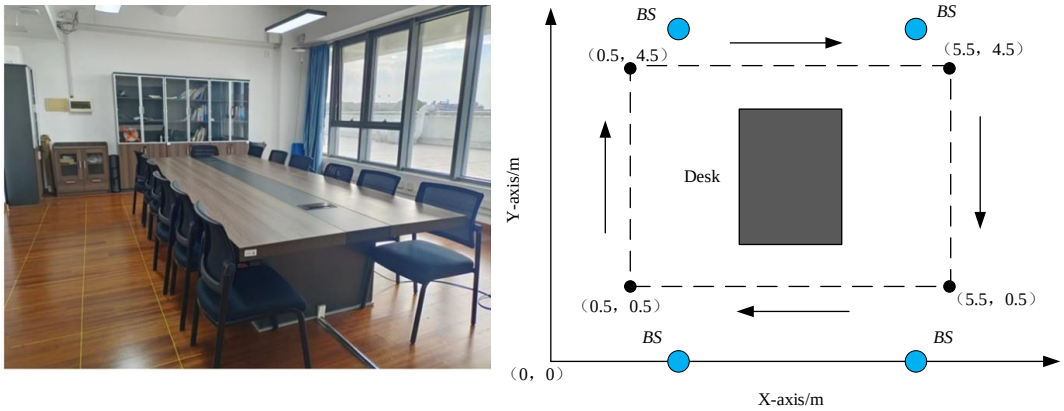

Fig. 10. Experimental scene diagram (left) and experimental diagram (right).

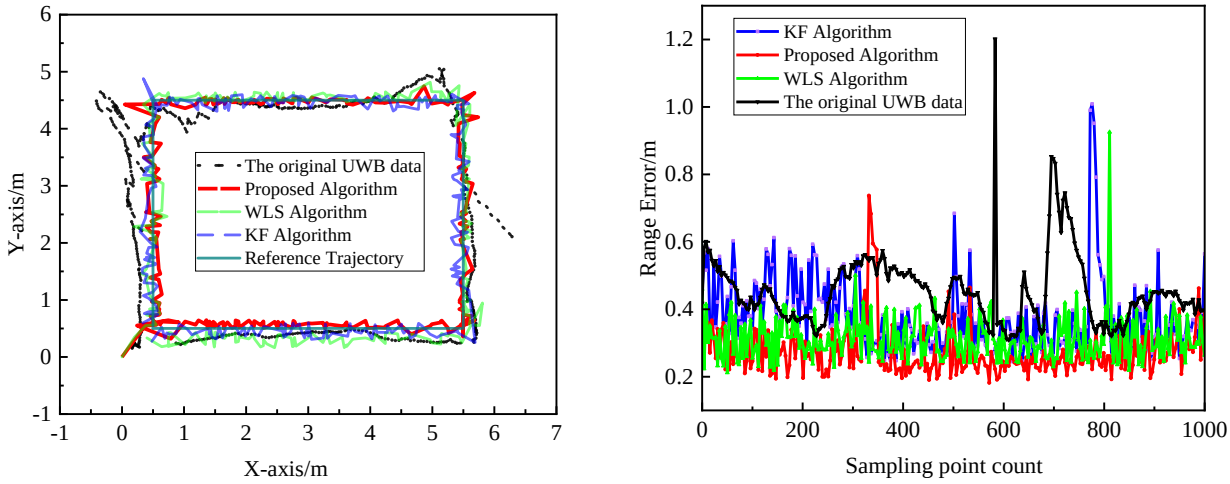

Fig. 11. Trajectories of different methods (left) and distance error distributions of different methods (right) in NLOS dynamic scenarios.

As the laboratory represents a typical complex and unstructured indoor environment, Figure 11 (right) shows that raw UWB measurements, weighted least squares, and KF suffer from degraded positioning performance and large trajectory fluctuations. The maximum positioning errors of these methods reach 1.2 m, 0.9 m, and 1.0 m, respectively. The corresponding average positioning errors are 0.44 m, 0.38 m, and 0.31 m, which are insufficient to satisfy the accuracy requirements of practical indoor localization scenarios.

In contrast, the trajectory generated by the proposed method more closely follows the ground-truth path. Large positioning fluctuations mainly occur at corner regions and near the leftmost solid wall, where signal propagation is severely obstructed by surrounding furniture and obstacles. Nevertheless, the distance error distribution of the proposed method remains relatively stable, with an average positioning error of 0.27 m. Compared with raw UWB measurements, KF, and weighted least squares, the proposed method improves positioning accuracy by 38.64%, 28.95%, and 12.9%, respectively.

To evaluate the positioning performance, the cumulative distribution function (CDF) and root mean square error (RMSE) of the dynamic positioning error were calculated, and the corresponding results are shown in Fig. 12. The CDF curves indicate that 80% of the positioning errors obtained using raw UWB measurements, KF, and weighted least squares are below 0.40 m, 0.35 m, and 0.28 m, respectively, whereas the proposed method achieves an error below 0.25 m for 80% of the positioning points. For a broader comparison across conventional and learning-based methods, SVM and KNN were additionally included in the evaluation. As summarized in Tab. 5, the proposed method yields the lowest mean error, maximum error, and RMSE among all the compared approaches. These results indicate that the proposed framework not only improves dynamic localization accuracy over conventional model-based methods, but also exhibits better overall positioning performance than the additional machine-learning-based comparison methods.

Algorithm	Mean error (m)	Maximum error (m)	RMSE (m)	Percentage improvement (%)
UWB	0.44	1.2	0.68	-
KF	0.38	0.9	0.61	13.64
WLS	0.31	1.0	0.55	29.55
SVM	0.30	0.9	0.52	31.82
KNN	0.29	0.7	0.49	34.09
Proposed algorithm	0.27	0.5	0.43	38.64

Tab.5. Dynamic experiment results.

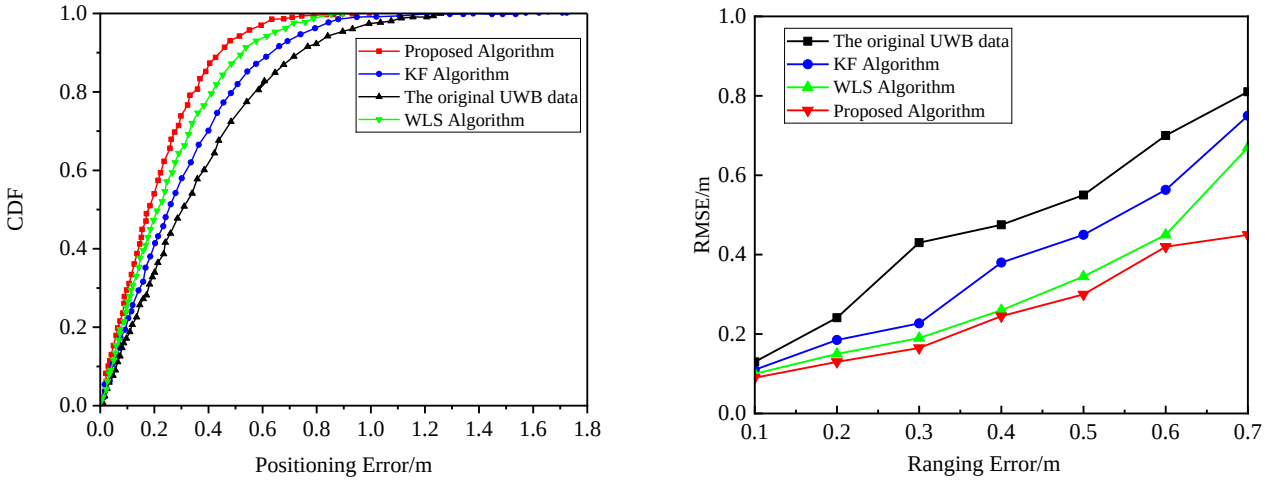

Fig. 12. CDF diagram of different algorithms (left) and RMSE diagram of different algorithms (right).

This performance advantage is mainly attributable to the joint effect of NLOS identification, residual-weighted localization, and KF-based temporal filtering. Specifically, the XGBoost-based classifier first distinguishes measurements affected by unfavorable propagation conditions, which reduces the direct influence of severely biased NLOS ranging results on position estimation. On this basis, the residual-weighted localization strategy assigns smaller contributions to measurements with larger residuals, thereby suppressing the impact of unreliable anchors in the localization solution. Furthermore, the KF-based update smooths the dynamic trajectory and mitigates instantaneous fluctuations caused by measurement noise. As a result, the proposed framework improves not only localization accuracy, but also the robustness and stability of dynamic indoor positioning.

To further analyze the reliability of the experimental results, we calculated the 95% confidence intervals for the localization errors of each method. In the dynamic experiment, where the UWB tag was moving in a typical indoor environment, the XGBoost framework had an average localization error of 0.27 meters, with a 95% confidence interval of [0.25 m, 0.29 m], while the traditional UWB localization method had an error of 0.44 meters, with a 95% confidence interval of [0.42 m, 0.46 m]. Additionally, we performed an independent samples t-test, and the results showed that the difference in localization accuracy between the XGBoost framework and the traditional method is statistically significant (p-value < 0.05), indicating that the proposed method provides a significant improvement in localization accuracy in dynamic scenarios.

In addition to localization accuracy, the practical applicability of the proposed framework also depends on its computational efficiency and real-time feasibility. For this reason, the online processing chain is further analyzed in a stage-wise manner. During online operation, the framework mainly consists of four steps: CIR feature extraction, XGBoost-based LOS/NLOS inference, residual-weighted WLS localization, and KF-based state update. The training stage of XGBoost is completed offline, whereas only the trained classifier is used during online deployment. Let T and D denote the number of trees and the maximum tree depth, respectively. Then, the inference complexity of the XGBoost classifier can be expressed as (TD) per sample, since each sample is evaluated through a sequence of tree traversals. In the feature processing stage, only five scalar CIR descriptors are used, namely the total CIR power, the amplitudes of the first three propagation paths, and the noise standard deviation, so the preprocessing burden remains low. For the localization stage, the number of anchors is fixed to four in this study, and the KF operates on a low-dimensional state vector. Therefore, both the residual-weighted WLS solver and the KF update are carried out on small matrices, resulting in limited online computational overhead. Moreover, the experimental system operates at a sampling rate of 100 Hz, corresponding to a nominal update period of 10 ms. Since coordinate calculation is executed on the backend server, the proposed framework is organized such that computationally intensive model training is completed offline, while the online stage is restricted to lightweight feature extraction, tree traversal, weighted localization, and state update. This implementation structure is therefore compatible with real-time server-side deployment.

6. Conclusion

This paper presented a robust UWB indoor localization framework for complex environments by jointly integrating DS-TWR ranging, machine learning-based NLOS identification, residual-weighted localization, and KF-based fusion. By explicitly identifying NLOS propagation using an XGBoost classifier and adaptively mitigating its impact at both the ranging and positioning stages, the proposed framework significantly improves localization accuracy and robustness under severe NLOS conditions.

Extensive static and dynamic experiments were conducted under human-body and obstacle occlusion scenarios to validate the effectiveness of the proposed approach. Experimental results demonstrate that the proposed method consistently outperforms conventional UWB localization, KF, and weighted least squares methods in terms of both ranging and positioning accuracy. In static NLOS scenarios,

the ranging accuracy is improved by up to 53.7% under human-body occlusion and 47.22% under wooden-board occlusion. In dynamic indoor environments, the proposed framework achieves positioning accuracy improvements of 38.64%, 28.95%, and 12.9% compared with conventional UWB localization, KF, and weighted least squares methods, respectively.

These results indicate that the proposed framework is robust against NLOS-induced errors and is well suited for practical UWB indoor localization in complex environments. However, its effectiveness still depends on the size and diversity of the training dataset, the geometry of anchor placement, and the consistency between the deployment environment and the training scenarios. When the environment becomes substantially larger or more complex in terms of layout, obstacle distribution, and propagation conditions, the CIR feature distribution may change, which can degrade the performance of the trained model. In such cases, additional calibration or retraining may be required to maintain reliable performance. Future work will therefore focus on improving the robustness and generalization capability of the proposed framework under larger datasets, different anchor configurations, and more diverse indoor environments.

Acknowledgments

This work was supported in part by the Research Project of Zhejiang Provincial Department of Education (Grant No. Y202558237) and the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province (Grant Nos. 2023C01030, 2026C02A1245, and 2026LDC01011(JT)), in part by the National Natural Science Foundation of China (Grant No. 62071212 and 62371218).

Data Availability Statement

The source code and the dataset used in this study are available from the corresponding author upon reasonable request.

References

[1] SVECOVA, M., KOCUR, D. Time of arrival complementing method for cooperative localization of a target by two-node UWB sensor network. *Radioengineering*, 2016, vol. 25, no. 3, p. 602–611. DOI: 10.13164/re.2016.0602

[2] MUQAIBEL, A. H., ALAWSH, S. A., BINMAKHASHEN, G. M. Under-sampled UWB NLOS/LOS channel classification using machine learning. *Arabian Journal for Science and Engineering*, 2024, vol. 50, no. 8, p. 6095–6108. DOI: 10.1007/s13369-024-09785-x

[3] SANTAMARIA-PEDRÓN, J. C., BERKVENS, R., MIRALLES, I., et al. Machine learning integration in ultra-wideband-based indoor positioning systems: A comprehensive review. *Electronics (Basel)*, 2025, vol. 15, no. 1, p. 1–55. DOI: 10.3390/electronics15010181

[4] MAJEED, A. F., ARSAT, R., BAHARUDIN, M. A., et al. Accurate multiclass NLOS channels identification in UWB indoor positioning system-based deep neural network. *IEEE Access*, 2024, vol. 12, p. 179431–179448. DOI: 10.1109/ACCESS.2024.3509343

[5] REN, H. Y., GUO, C. X., YANG, R. F. Research on improving UWB ranging accuracy by Kalman filter (in Chinese). *Electronic Measurement Technology*, 2024, vol. 44, no. 18, p. 111-115.

[6] CHEN, S. Y., YIN, D., NIU, Y.F. Research and implementation of improved SS-TWR method based on UWB (in Chinese). *Application Research of Computers*, 2023, vol. 38, no. 11, p. 3398 to 3402. DOI: 10.19734/j.issn.1001-3695.2021.04.0116

[7] TIEN LE MINH, DUNG TRINH XUAN Applying Kalman filter to UWB positioning with DS-TWR method in LOS/NLOS scenarios. In *2021 International Symposium on Electrical and Electronics Engineering (ISEE)*. Ho Chi Minh City (Vietnam), 2021, p. 95–99. DOI: 10.1109/ISEE51682.2021.9418707

[8] WANG, G., LI, S., CHENG, P., et al. ToF-based NLoS indoor tracking with adaptive ranging error mitigation. *IEEE Transactions on Signal Processing*, 2024, vol. 72, p. 4855–4870. DOI: 10.1109/TSP.2024.3468467

[9] GUO, A. J. Joint positioning method of PDOA and TOF in coal mines based on UWB (in Chinese). *Journal of Mine Automation*, 2023, vol. 49, no. 3, p. 137–141. DOI: 10.13272/j.issn.1671-251x.18078

[10] GUVENC, I., CHONG, C. C., WATANABE F., et al. NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistic. *EURASIP Journal on Advances in Signal Processing*, 2007, no. 1, p. 1–14. DOI: 10.1155/2008/271984

[11] MARANO, S., GIFFORD, W. M., WYMEESCH, H., et al. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE Journal on Selected Areas in Communications*, 2010, vol. 28, no. 7, p. 1026–1035. DOI: 10.1109/JSAC.2010.100907

[12] DJOSIC, S., STOJANOVIC, I., JOVANOVIC, M., et al. Multi-algorithm UWB-based localization method for mixed LOS/NLOS environments. *Computer Communications*, 2022, vol. 181, p. 365 to 373. DOI: 10.1016/j.comcom.2021.10.031

[13] XIAO, N., SHI, S. A real-time identification and suppression algorithm of non-line-of-sight error based on TOA (in Chinese). *Guangdong Communication Technology*, 2020, vol. 40, no. 5, p. 67–70.

[14] ZENG, L., PENG, C., LIU, H. UWB indoor positioning algorithm based on NLOS identification (in Chinese). *Computer Applications*, 2018, vol. 36, no. 6, p. 1557–1561.

[15] PARK, J. W., NAM, S. C., CHOI, H. B., et al. Improving deep learning-based UWB LOS/NLOS identification with transfer learning: An empirical approach. *Electronics*, 2020, vol. 9, no. 10, p. 1–13. DOI: 10.3390/electronics9101714

[16] CORTESI, S., DREHER, M., MAGNO, M. Design and implementation of an RSSI-based Bluetooth low energy indoor localization system. In *17th International Conference on Wireless and Mobile Computing, Networking and Communications*. 2021, p. 163–168. DOI: 10.1109/WiMob52687.2021.9606272

[17] YAO, X. L., XU, Z. J., QIANG, F. High-precision indoor localization via dual-modal AOA/TOA fusion with deep learning and particle filters. *Radioengineering*, 2025, vol. 34, no. 4. p. 624 to 640. DOI: 10.13164/re.2025.0624

[18] GUVENC, I., CHONG, C. C. A survey on TOA based wireless localization and NLOS mitigation techniques. *IEEE Communications Surveys & Tutorials*, 2009, vol. 11, no. 3. p. 107 to 124. DOI: 10.1109/SURV.2009.090308

[19] LIU, Q. L, WAN, Z. P., ZHOU, W. M., et al. An improved indoor localization method based on multi-source information fusion (in Chinese). *Telecommunications Technology*, 2021, vol. 61, no. 12, p. 1526–1533.

[20] WANG, W., HUANG, J., CAI, S., et al. Design and implementation of synchronization-free TDOA localization system based on UWB. *Radioengineering*, 2019, vol. 27, no. 1, p. 320–330. DOI: 10.13164/re.2019.0320

[21] ZHANG, D., XU, H. B., ZHAN, L., et al. Accurate joint estimation of position and orientation based on angle of arrival and two-way ranging of ultra-wideband technology. *Electronics*, 2025, vol. 14, no. 3, p. 1–17. DOI: 10.3390/electronics14030429

[22] LIU, M. Q. *NLOS Identification and Error Reduction Method for Ultra-Wideband Positioning*. Diploma Theses. Jiangsu, China University of Mining and Technology, 2023.

[23] GUAN, B., WANG, D. H., SHU, D., et al. Data-driven casting defect prediction model for sand casting based on random forest classification algorithm. *China Foundry*, 2024, vol. 21, no. 2, p. 137–146. DOI: 10.1007/s41230-024-3090-1

[24] YANG, S., CHOU, L., LI, H., et al. Hierarchical self-distillation with attention for class-imbalanced acoustic event classification in elevators. *Sensors*, 2026, vol. 26, no. 2, p. 1–27. DOI: 10.3390/s26020589

[25] KHODARAHMI, M., MAIHAMI, V. A review on Kalman filter models. *Archives of Computational Methods in Engineering*, 2023, vol. 30, no. 1, p. 727–747. DOI: 10.1007/s11831-022-09815-7

About the Authors ...

Xiaoli YAO (corresponding author) was born in Anhui Province, China in 1991. She received her Ph.D. in Electronic Science and Technology from Hangzhou Dianzi University in 2020. She is currently a lecturer at Zhejiang Polytechnic University of Mechanical and Electrical Engineering, and her research focuses on deep learning and fusion positioning.

Zhijiang XU (corresponding author) was born in 1973. He received his Ph.D. in Information and Communication Engineering in 2005 from Zhejiang University, China. He was appointed Associate Professor in the College of Information Engineering at Zhejiang University of Technology, China, from 2007 to 2019. Since 2019, he has joined the Zhejiang Polytechnic University of Mechanical and Electrical Engineering in the School of Automation and was promoted to full Professor in 2020. His research interests include digital communications over fading channels, channel modeling, coding, digital synchronization, etc.

Guojun LIU was born in Zhejiang Province, China in September 1980. He is currently Manager and Shareholder of Zhejiang Zhenhang Industrial Group Co., Ltd. A technical expert and standard-setting participant in the field of hydraulic filtration, he has long been engaged in filter R&D, production management and standardization, driving technological innovation within the enterprise and the implementation of industry standards.