IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Geo-LSTM: A Geometry and Temporal Feature Fusion Algorithm for Multi-Sensor 3D Localization

Kai Li[1], Le Bao[2], and Wansoo Kim[3], *Member, IEEE*

***Abstract*—Accurate three-dimensional (3D) localization is critical for robust human-robot collaboration (HRC) in dynamic indoor environments. However, realizing high-precision localization in complex scenarios still faces challenges such as multipath effects, field-of-view occlusion, etc. To address these limitations, we propose Geo-LSTM, a geometry-constrained long short-term memory (LSTM) framework that integrates ultra-wideband (UWB) sensors, inertial measurement unit (IMU), and barometric pressure (BMP) sensors. First, a Simplified Geometric Localization (SGL) algorithm is proposed, which uses dual-BMP sensors and IMU sensor to obtain precise height information and utilizes the geometric relationships between the UWB tag and anchors to compute an initial location estimate, serving as a priori input for the Geo-LSTM network. This Geo-LSTM algorithm then incorporates multi-source geometric information to extract time-series features from the UWB ranging data and the tag's a priori location, further enhancing 3D localization accuracy. The experimental results from the cluttered indoor environments, including real-world HRC tasks with occlusions, show that the Geo-LSTM algorithm achieves an average 3D localization root mean square error (RMSE) of 0.103 m, representing improvements of 38.60% and 31.20% over the weighted least squares (WLS) method and the range-based LSTM algorithm, respectively. These results demonstrate Geo-LSTM's potential for reliable multi-sensor 3D localization in HRC applications.**

***Index Terms*—Localization, Sensor Fusion, Human-Robot Collaboration (HRC), Ultra-Wideband (UWB).**

I. Introduction

HUMAN-robot collaboration (HRC) is a critical area of research in robotics, aiming to combine the strengths of humans and robots to enhance safety, productivity, and adaptability in shared work environments [1]. One fundamental requirement in HRC is reliable localization of humans and collaborative tools. When a robot has accurate knowledge of a human collaborator's position, it can plan actions more effectively, anticipate movements, and perform assistance tasks in real-time (Fig. 1). In practice, however, consistent and high-accuracy localization in three-dimensional (3D), especially in indoor environments, remains challenging due to occlusions, cluttered scenes, and varying lighting conditions.

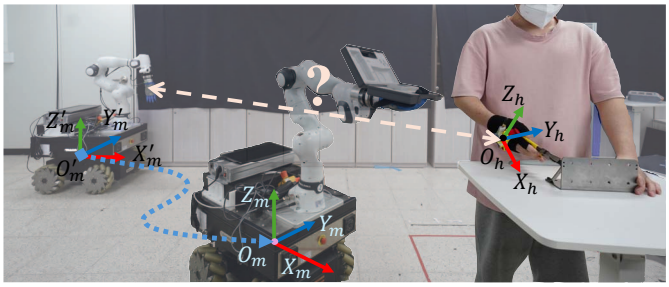

Fig. 1. Relationship in human-robot collaboration scenarios: the robot realizes the collaboration task by recognizing the human's task operation location.

Current research frequently uses techniques such as optical motion capture systems (MCS) [2] [3] and cameras [4] for localizing human assistance needs. Although these methods offer high localization accuracy, their complex deployment, high setup costs, and susceptibility to occlusions constrain their adoption in real-world HRC scenarios.

Compared to visual sensors, wireless ranging technologies such as ultra-wideband (UWB) sensors show advantages in indoor localization due to their environmental adaptability and low setup cost [5]. Classical UWB-based localization often adopts extended kalman filtering [6], [7], or particle filtering to address the nonlinearity issues [8]. While these techniques may achieve acceptable performance under line-of-sight (LOS) conditions, they can degrade significantly when the signals undergo multipath or non-line-of-sight (NLOS) conditions such as complex application scenarios with static and dynamic obstacles (furniture, walls, human body, etc.) block UWB signal transmission [9].

To improve localization accuracy under both LOS and NLOS scenarios, neural network-based methods have been introduced into the field of UWB localization, showing enhanced robustness in complex environments. In [10], a long short-term memory (LSTM) network was proposed that uses range data as input, with network parameters optimized through simulation experiments to predict user locations. However, it heavily relies on the quality of input data and lacks validation through real-world experiments. An artificial intelligence model based on an LSTM network was employed to classify channel conditions using channel impulse response (CIR) signals, thereby enhancing localization performance [11]. In [12], CIR signals were analyzed using LSTM networks to predict the ranging error of UWB sensors, minimizing errors during the

Manuscript received: March, 14, 2025; Revised May, 15, 2025; Accepted July, 11, 2025. This paper was recommended for publication by Editor C. Javier upon evaluation of the Reviewers' comments. This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1C1C1008306), and in part by the research fund of Hanyang University (HY-202400000003293), and in part by the China Scholarship Council(CSC) (No. 202308260035). *(Kai Li and Le Bao are co-first authors and contributed equally to this work.) (Corresponding author: Wansoo Kim.)*

[1]Kai Li is with the Department of Mechatronics Engineering, Hanyang University, Ansan, Gyeonggi-do 15588, Republic of Korea. likai@hanyang.ac.kr

[2]Le Bao is with the Research Institute of Engineering & Technology, Hanyang University ERICA, Ansan, Gyeonggi-do 15588, Republic of Korea. baole@hanyang.ac.kr

[3]Wansoo Kim is with the Robotics Department, Hanyang University ERICA, Ansan, Gyeonggi-do 15588, Republic of Korea. wansookim@hanyang.ac.kr

Digital Object Identifier (DOI): see top of this page.

UWB ranging phase. Beyond utilizing CIR signals, other studies have explored alternative approaches. Local spatial features were extracted from range data and received signal strength (RSS) data using a convolutional neural network, while temporal correlations between consecutive frames were captured through an LSTM network [13]. Another study embedded UWB anchor coordinates and ranging data into a deep attention-based network to extract spatial features, then used a gated recurrent unit module to learn temporal features for estimating tag locations, and a geometric loss function was designed to improve the accuracy of network training [14]. These models, which integrate features from multiple data sources, exhibit high adaptability in complex environments. Nonetheless, CIR and RSS data are often unavailable for common low-cost UWB sensors [14]. Moreover, the existing studies rarely address the Z-axis' localization accuracy, limiting their practicality in 3D tasks.

To expand the application of UWB localization system in HRC, this letter proposes the Geo-LSTM algorithm, a 3D localization algorithm that integrates geometric information with the LSTM model. Unlike error compensation methods that rely on channel information, the Geo-LSTM algorithm constructs geometric data based on UWB ranging data and prior location estimates, and extracts multi-source temporal features through the LSTM model, enabling high-precision localization even when channel information is unreliable. In particular, we design a simplified geometric localization (SGL) method based on UWB, inertial measurement unit (IMU), and barometric pressure (BMP) sensors to provide a 3D location reference for tags, and verify its superiority over classical methods. The LSTM network then learns temporal features from the reference locations and UWB sensors' ranging data, further enhancing the accuracy of location estimation. Experimental results validate that the Geo-LSTM algorithm achieves high stability and accuracy under both LOS and NLOS scenarios, effectively meeting the practical requirements of HRC applications.

The main contributions of the letter are as follows:

- The SGL algorithm reduces computational complexity by fusing sensor data and utilizing the geometric relationship between tag and anchors, serving as a foundational method for lightweight localization.
- The Geo-LSTM algorithm effectively extracts the temporal features from both the SGL's prior coordinates and UWB ranging data while preserving data independence, thereby enhancing the accuracy and stability of 3D localization.
- The applicability of Geo-LSTM algorithm in HRC scenarios is verified by comparing the performance of multiple methods in complex environments.

II. System Hardware Framework

The proposed 3D localization system comprises two main components: the tag system and the anchor system. The anchor system mainly consists of four UWB anchors (LinkPG3.9 module), which estimate the distance to the tag based on the time-of-flight of wireless signals. The BMP sensor (MS5611 module) in the anchor system provides accurate altitude and temperature measurements.

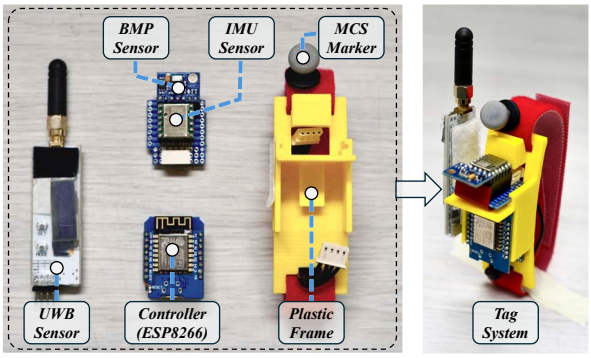

Fig. 2. Tag system consisting of UWB sensor, IMU sensor and BMP sensor.

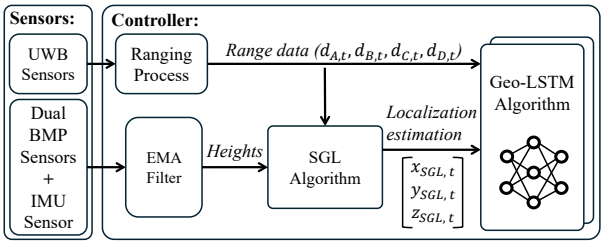

Fig. 3. The whole methodology of the localization system.

TABLE I
Sensors' Specific Parameters in Localization System

Sensor	Sampling rate	Accuracy	Resolution
UWB	20 Hz	0.1 m	0.001 m
IMU	20 Hz	0.05 g (Acceleration)	0.0005 g/LSB
BMP	20 Hz	1.5 mbar	0.012 mbar

As illustrated in Fig. 2, the tag system integrates three types of sensors: UWB tag, IMU, and BMP, which can be used to estimate the sea level altitude and the distance between the tag and the anchors. The UWB tags and BMP sensors are identical to those used in the anchor system. The IMU sensor (JY901 module) provides Z-axis velocity and acceleration data, enhancing height estimation accuracy. The specific parameters of the sensors are shown in Table I.

The following section details how these sensor inputs feed into the algorithm to complete the localization.

III. Methodology

The flowchart of the entire methodology is shown in Fig. 3. It covers the key steps, including tag's height estimation, tag's a priori localization estimation and Geo-LSTM algorithm to provide complete and reliable location information for HRC.

A. IMU-based Dynamic Height Estimation for Dual-BMP Sensors

In HRC applications, precise localization is related to the robot's correct judgment of human collaboration areas. In [2], robots must be able to determine the precise location of the human hand to perform tasks effectively. In addition, height

accuracy directly influences HRC safety, which is particularly critical in real-world scenarios.

In indoor environments, the difference in temperature variation is small, but factors such as building sealing and ventilation conditions can lead to fluctuations in indoor air pressure [15]. Our previous research [16] verified that the air pressure drift tends to be consistent within the same indoor space. By using the height difference calculated from the BMP sensors installed on both the anchor system and the tag system, the drift problem can be effectively offset. The height of the tag can be calculated by the deformation (1) of the barometric formula,

$$h = \frac{T_0}{L} \cdot \left(1 - \left(\frac{P_0}{P}\right)^{\frac{RL}{gM}}\right), \tag{1}$$

where T_0 is the standard atmospheric temperature, L is the temperature lapse rate, and P_0 is the standard atmospheric pressure. R is the gas constant, g is the gravitational acceleration, and M is the molar mass of air.

In order to reduce the influence of BMP sensors by environmental high-frequency noise, it is difficult to accurately reflect the dynamic changes of the tag in the vertical direction by using only a fixed-parameter filter. This study introduces a dynamic exponential moving average (EMA) filter with variable weighting parameters. When the tag moves in the vertical direction, we adjust the filter's weight according to the IMU's Z-axis velocity v_z. This allows the BMP sensor to respond quickly to the height changes and improves the overall accuracy of the tag system. The relevant formulas are shown in equation (2):

$$\begin{aligned} h_{tag}(t) &= \omega_{tag} \cdot h_{tag} + (1 - \omega_{tag}) \cdot h_{tag}(t-1), \\ \omega_{tag} &= \begin{cases} 0.2 & \text{if } v_z > 0, \\ 0.05 & \text{if } v_z = 0, \end{cases} \end{aligned} \tag{2}$$

where $h_{tag}(t)$ is the filtered estimate of the BMP sensor in the tag system at time t, and $h_{tag}(t-1)$ is the filtered estimate at time t-1. h_{tag} is the estimated height of the BMP sensor at time t. ω_{tag} is the filtered weight value in the tag system, and it is dynamically adjusted based on the vertical velocity component. A higher weight (0.2) is used when vertical movement is detected to allow faster response to height changes, while a lower weight (0.05) is used during horizontal movement or stationary periods to provide more stable filtering.

The location of the BMP sensor in the anchor system is constant, so the filter's weight value ω_{base} is fixed at 0.05, and the filtering formula for the anchor system is shown below:

$$h_{base}(t) = \omega_{base} \cdot h_{base} + (1 - \omega_{base}) \cdot h_{base}(t-1). \tag{3}$$

Therefore, the tag height $H_{tag}(t)$ calculated by sensor fusion at time t is expressed as:

$$H_{tag}(t) = h_{tag}(t) - h_{base}(t) + H_{base}, \tag{4}$$

where H_{base} is the actual height of the anchor system's BMP sensor location.

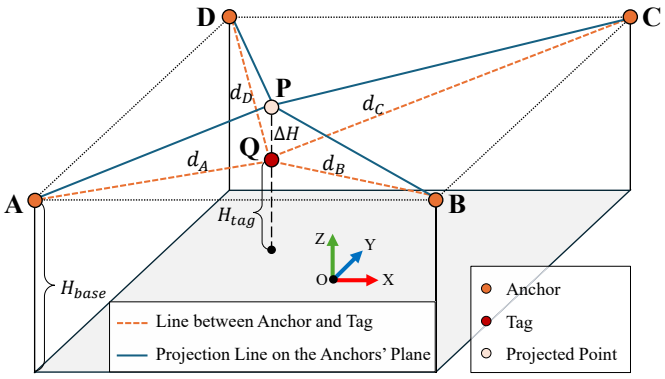

Fig. 4. Geometric distribution of the localization system. The location of tag Q is calculated based on its geometrical relationship with anchors A, B, C and D. Point P is the projection point of tag Q on the anchors' plane.

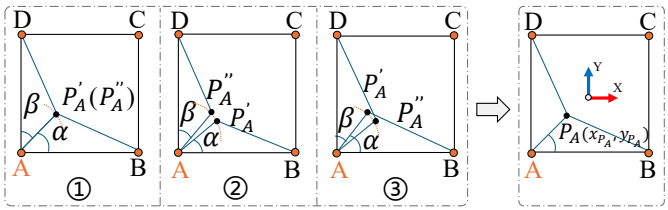

Fig. 5. Anchor A-based triangulation analysis under the UWB sensors' ranging errors. ① represents the ideal ranging condition, while ② and ③ correspond to underestimated and overestimated ranging errors, respectively.

B. Simplified Geometric Localization Algorithm

Based on the range data from UWB sensors, the traditional trilateration and least squares (LS) methods [17] [18] are the most commonly used methods in UWB localization. However, in the 3D environment, the LS method needs to address the non-linear characteristics of the ranging equation. While linearization can simplify the solution process, it introduces error amplification, resulting in decreased localization accuracy. In this letter, we propose an SGL algorithm, based on geometric relations and trigonometric functions, to offer a more efficient and lightweight solution for UWB localization.

For realistic application considerations, in indoor HRC scenarios, it is common to install UWB anchors at the same height (e.g., along the upper part of walls) due to the ease of hardware installation, maintenance cost, and environmental layout. When all anchors are placed at the same height, accurate localization in the 2D plane can be ensured [13] [19]. Meanwhile, in this study, the height estimation obtained based on sensor fusion ensures the tag's localization accuracy in the vertical direction. Therefore, all subsequent localization will be carried out under the condition of coplanar anchor placement.

As depicted in Fig. 4, the height difference ΔH between the tag's position and the anchors' setting plane can be determined using the anchor's setting height H_{base} and the tag's estimated height $H_{tag}(t)$. Assuming that the projection point of the tag Q onto the anchor setting plane is point P, the distances between point P and the four anchors can be calculated using the Pythagorean theorem as AP, BP, CP, and DP.

Geometric analysis is conducted with anchor A serving as

a reference point in Fig. 5. An arc is constructed with point A as the center and AP as the radius. BP, DP and the arc are compared to points $P_A^{'}$ and $P_A^{''}$ respectively. However, due to the ranging error of the UWB sensors, $P_A^{'}$ of ΔABP and $P_A^{''}$ of ΔADP may not necessarily coincide. Therefore, the angles α and β are derived to represent $angePAB$ and $angePAD$, respectively, and calculate the average value of $\angle PAB$, denoted as $\bar{\alpha}$,

$$\begin{cases} \alpha = \arccos\left(\dfrac{AP^2 + AB^2 + BP^2}{2 \cdot AP \cdot AB}\right) \\ \beta = \arccos\left(\dfrac{AP^2 + AD^2 + DP^2}{2 \cdot AP \cdot AD}\right) \\ \bar{\alpha} = \dfrac{\alpha + (90^\circ - \beta)}{2} \end{cases} . \tag{5}$$

The point P_A (x_{P_A}, y_{P_A}) base on anchor A (x_A, y_A) and AP can be calculated,

$$\begin{cases} x_{P_A} = x_A + (AP \cdot \cos\bar{\alpha}) \\ y_{P_A} = y_A + (AP \cdot \sin\bar{\alpha}) \end{cases} . \tag{6}$$

Similarly, the coordinates of point P can be calculated based on anchor B, C and D respectively as P_B (x_{P_B}, y_{P_B}), P_C (x_{P_C}, y_{P_C}) and P_D (x_{P_D}, y_{P_D}). The coordinates of P are calculated as P (x_P, y_P) by solving for the mean value, so the equation for the coordinates of the tag Q $(x_{SGL}, y_{SGL}, z_{SGL})$ is

$$\begin{cases} x_{SGL} = x_P = \dfrac{x_{P_A} + x_{P_B} + x_{P_C} + x_{P_D}}{4} \\ y_{SGL} = y_P = \dfrac{y_{P_A} + y_{P_B} + y_{P_C} + y_{P_D}}{4} \\ z_{SGL} = H_{tag}(t) \end{cases} . \tag{7}$$

The SGL algorithm uses a localization mechanism based on spatial projection and geometric angle derivation, which does not rely on matrix construction and inverse, and thus has a computational complexity of $O(1)$ for each moment. In contrast, the traditional LS method requires constructing and solving a system of linear equations with a computational complexity of $O(n^3)$, where n is the number of unknown variables. Although the ranging error may still have an impact on the localization accuracy, compared with the traditional LS methods, this algorithm offers advantages in terms of computational efficiency and implementation difficulty, and has wider applicability in practical applications.

C. Geo-LSTM Algorithm based on Tag's Geometric Information and LSTM Networks

LSTM networks are recurrent neural networks with time-series modeling capability, which dynamically adjusts the storage and update of information at each discrete time step through a gating mechanism. The structure of LSTM network consists of forget gates, input gates and output gates [20].

The LSTM network using only UWB ranging data can utilize time series for data processing, but it is difficult to learn accurate 3D coordinates directly [10]. This limitation is primarily due to the absence of spatial structure in the ranging data and their susceptibility to environmental complexities.

To address this problem, we propose a Geo-LSTM network that uses the ranging data from UWB sensors and the coordinates provided by the SGL algorithm as inputs. The UWB ranging data has temporal correlation, which can reflect the distance change between the tag and the anchors over time, but it lacks complete spatial information. On the other hand, the 3D coordinates provided by the SGL algorithm are calculated through multi-sensor fusion at each time step, which can provide a priori spatial location information, but these coordinates may contain noise, especially in dynamic environments.

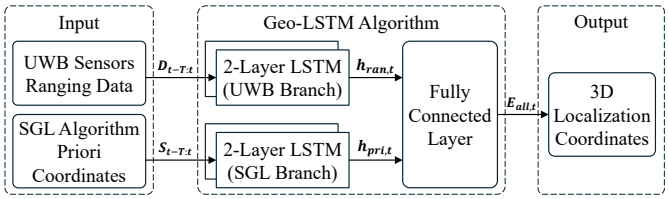

Fig. 6. The structure of the Geo-LSTM algorithm. The UWB range data and the a priori coordinates provided by the SGL algorithm are the dual inputs, the temporal features are extracted by two 2-layer LSTMs, and the final 3D localization results are output.

Given the significant differences in physical meaning, data dimensionality, and numerical scale between the two input types, directly concatenating them can lead to feature interference and hinder stable model training. The proposed Geo-LSTM algorithm uses two LSTM networks to process different types of input data, effectively extracting their temporal features, which can better retain the independence of the data and improve the localization accuracy of the fused data.

The structure of the Geo-LSTM algorithm is shown in Fig. 6. The normalized time series of the ranging values from the UWB sensors, $\mathbf{D}_{t-T:t}$, is represented as $[d_{A,t}, d_{B,t}, d_{C,t}, d_{D,t}]$, where $d_{X,t}$ $(X \in \{A, B, C, D\})$ represents the time-series ranging data of the anchor X and T denotes the time step of the network. Similarly, the normalized time series of the 3D coordinates provided by the SGL algorithm, $\mathbf{S}_{t-T:t}$, is represented as $[x_{SGL,t}, y_{SGL,t}, z_{SGL,t}]$. These time series are fed into a two-layer LSTM network to extract temporal features from the range data and the 3D coordinates, respectively:

$$\mathbf{h}_{ran,t} = \text{LSTM}(\mathbf{D}_{t-T:t}), \tag{8}$$

$$\mathbf{h}_{pri,t} = \text{LSTM}(\mathbf{S}_{t-T:t}). \tag{9}$$

Each layer of the LSTM network follows the classical formulation:

$$\mathbf{f}_t = \sigma\left(W_f \mathbf{x}_t + R_f \mathbf{h}_{t-1} + b_f\right), \tag{10}$$

$$\mathbf{i}_t = \sigma\left(W_i \mathbf{x}_t + R_i \mathbf{h}_{t-1} + b_i\right), \tag{11}$$

$$\mathbf{o}_t = \sigma\left(W_o \mathbf{x}_t + R_o \mathbf{h}_{t-1} + b_o\right), \tag{12}$$

$$\mathbf{g}_t = \tanh\left(W_g \mathbf{x}_t + R_g \mathbf{h}_{t-1} + b_g\right), \tag{13}$$

$$\mathbf{c}_t = \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \circ \mathbf{g}_t, \tag{14}$$

$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{c}_t), \tag{15}$$

where W, R, b are the weight matrices and bias terms, and $\sigma, \tanh$ denote the activation functions, respectively.

Subsequently, the temporal features $(\mathbf{h}_{ran,t}, \mathbf{h}_{pri,t})$ extracted from the two LSTM networks are concatenated, and

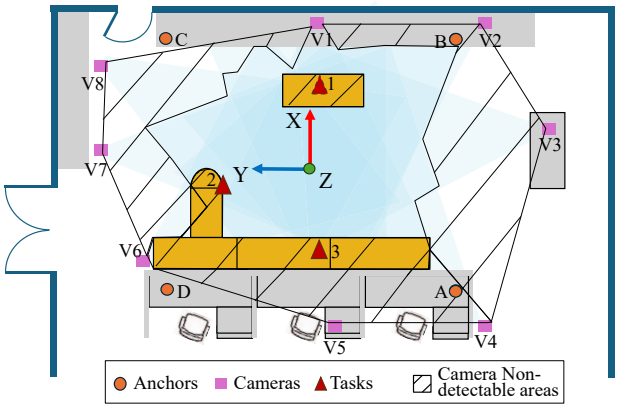

Fig. 7. Laboratory scenarios and experimental setups. The experimental scenario contains the distribution of UWB anchors, MCS cameras, and task points. The shaded area is the undetectable region of the MCS.

the final 3D localization estimate $\mathbf{E}_{all,t}$ is generated through a fully connected layer and Rectified Linear Unit (ReLU) activation function. The calculation formulas are as follows:

$$\mathbf{h}_{all,t} = \text{concat}(\mathbf{h}_{ran,t}, \mathbf{h}_{pri,t}), \tag{16}$$

$$\mathbf{E}_{all,t} = \text{ReLU}(W_{all}\mathbf{h}_{all,t} + b_{all}). \tag{17}$$

The Geo-LSTM algorithm fully utilizes the ranging data obtained from the UWB sensors and the a priori location estimation based on multi-sensor fusion. Compared to the LSTM network using only ranging input, this structure compensates for limitations in directionality and 3D structural information in the ranging data, improving both accuracy and stability. By integrating multi-source data inputs, the Geo-LSTM algorithm effectively overcomes the limitations of single-input LSTM networks.

IV. Experiments

A. Experimental Setup

Indoor HRC applications typically require centimeter-level localization accuracy. To evaluate the localization accuracy and reliability of the proposed method, experiments were conducted in an indoor environment at the HARCO Laboratory of Hanyang University, with an experimental space size of $7 \times 5 \times 2.6$ meters. As shown in Fig. 7, the experimental setup includes four UWB anchors and an MCS consisting of eight cameras. The four UWB anchors in the localization system were set at A $(-2.485m, -2.485m, 1.914m)$, B $(2.485m, -2.485m, 1.914m)$, C $(2.485m, 2.485m, 1.914m)$, and D $(-2.485m, 2.485m, 1.914m)$. The MCS with eight cameras was used to provide the reference coordinates of the tags during the experiment.

The Geo-LSTM network was implemented using TensorFlow/Keras in Python. The optimizer used Adam's algorithm. The training process employed a batch size of 30, a hidden layer size of 32, and a total of 100 epochs. The time step T was set to 10. The mean square error (MSE) was used as the loss function for the experiments. The ground truth data used for supervised training were obtained from the MCS, which provides high-precision 3D locations of the tag. The entire

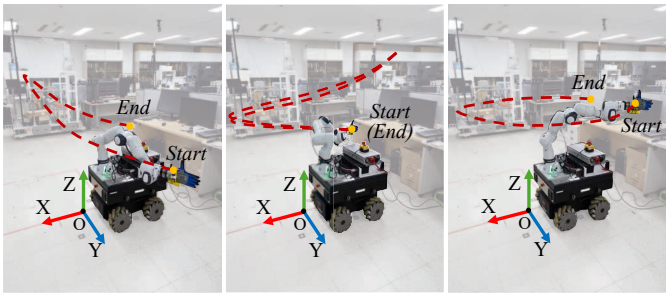

Fig. 8. Three different pre-set trajectories of the robotic arm used in the experiment.

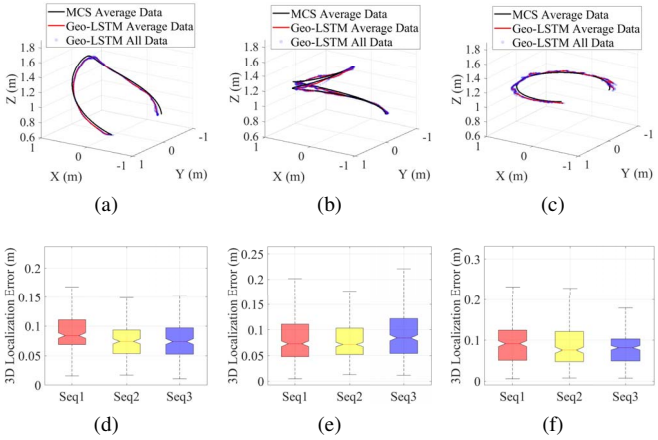

Fig. 9. 3D localization trajectory and error comparison results: (a)(b)(c) shows the results of Geo-LSTM algorithm in different localization trajectories, (d)(e)(f) shows the localization error distribution of nine trajectories in the three sets of experiments.

network was trained on an NVIDIA GeForce RTX 3060 GPU. The algorithm's output rate was 18 Hz.

To evaluate the stability and accuracy of the experimental results, the 3D localization root mean square error (RMSE) was used as the evaluation metric to quantify the difference between the estimated locations of the algorithm and the reference locations provided by the MCS.

This study compares the proposed method with three existing UWB localization methods: a traditional localization method [21], a filtering-based localization method [6], and a range-based LSTM localization method [10]. Both the traditional and filtering methods are combined with the multi-sensor fusion height estimation proposed in this letter to solve the problem of height estimation error of UWB anchors in the same plane setting. The range data-based LSTM method relies solely on UWB sensors' range data for location estimation. For simplicity, these three methods are referred to as WLS method, KF method, and LSTM method, respectively.

B. Performance Evaluation of Geo-LSTM Algorithm Using a Robotic Arm

The first experiment evaluates the effectiveness and reliability of the proposed algorithm for dynamic 3D localization. We use a 7-degree-of-freedom FRANKA robotic arm capable of executing reproducible and predefined trajectories in a 3D

workspace. As shown in Fig. 8, the tag system is mounted at the end effector of the robotic arm, which moves along three different trajectories, each repeated three times.

The experimental results are shown in Fig. 9, where the blue scatters in Fig. 9(a)(b)(c) are the set of localization data computed by the proposed Geo-LSTM algorithm, the red line denotes the average localization estimation of the algorithm, and the black line corresponds to the reference trajectory recorded by the MCS. The results demonstrate that the mean 3D trajectory closely aligns with the reference trajectory, with localization RMSE values of 0.090 m, 0.089 m, and 0.092 m for the three trajectories respectively, confirming the high accuracy of the proposed algorithm's 3D localization estimation. The blue scatter points in each trajectory set show a concentrated error distribution, with localization error standard deviations ranging approximately from 0.038 m to 0.056 m across the three repeated experiments, indicating the consistency of the experimental results. Additionally, the results under varying trajectory complexities indicate that the proposed algorithm maintains stable localization accuracy.

Fig. 9(d)(e)(f) depict the distribution of localization errors for the nine trajectories. The median localization errors of different trajectories range from 0.074 m to 0.091 m, indicating that the algorithm consistently maintains low average errors, and this relatively concentrated error range reflects limited variation among the trajectories. The interquartile range (IQR) of the experimental localization error distributions are all within 0.042 m, further demonstrating the overall consistency of the experimental data distribution and verifying the stability and reliability of the proposed algorithm in repeated experiments.

C. 3D Localization Performance in Complex Operational Scenarios

The second experimental scenario contained multiple pieces of furniture to simulate complex operational scenarios in a real-world environment. The MCS has a restricted field of view in some areas due to obstacles such as furniture blocking and equipment limitations. The specific setup of the scenario, along with the MCS's restricted visibility, is depicted in Fig. 7.

The experiment was conducted three times, each time containing the following three tasks:

- Task 1: The experimenter operated in a squatting position and took out the electric drill from under the table with the right hand along the X-axis direction, keeping the hand height unchanged during the process;
- Task 2: The experimenter operated in a standing position, holding the tool in the specified location to complete the tasks, simulating the regular work scenario;
- Task 3: The experimenter used a ladder and held the tool along the Z-axis to reach the target height and complete the task.

During the experiment, the tag system was securely fixed to the back of the experimenter's hand. Due to furniture, human body occlusion, and task execution, the UWB sensors encountered LOS and NLOS conditions, which can fully evaluate the localization performance of the proposed localization algorithm in complex scenarios. Table II shows the distribution of UWB sensor conditions for different tasks.

TABLE II
DISTRIBUTION OF ANCHORS' LOS/NLOS SCENARIOS IN DIFFERENT TASKS

Task	Anchor A	Anchor B	Anchor C	Anchor D
Task 1	LOS	LOS/NLOS	LOS/NLOS	LOS/NLOS
Task 2	LOS/NLOS	LOS	LOS	LOS/NLOS
Task 3	LOS/NLOS	LOS	LOS	LOS

TABLE III
COMPARISON OF 3D LOCALIZATION RMSE AND ALGORITHM PERFORMANCE IMPROVEMENT

Number	Algorithm	3D RMSE (m)	Improvement
Seq 1	Geo-LSTM (Ours)	0.105	-
	WLS	0.176	+40.39%
	LS-KF	0.200	+47.41%
	SGL (Ours)	0.169	+37.91%
	LSTM	0.150	+30.00%
Seq 2	Geo-LSTM (Ours)	0.104	-
	WLS	0.158	+34.22%
	LS-KF	0.170	+38.82%
	SGL (Ours)	0.162	+35.74%
	LSTM	0.150	+30.73%
Seq 3	Geo-LSTM (Ours)	0.100	-
	WLS	0.170	+41.20%
	LS-KF	0.186	+46.19%
	SGL (Ours)	0.154	+35.06%
	LSTM	0.149	+32.89%

Fig. 10(a) shows one of the sets of 3D trajectory results, where the proposed Geo-LSTM algorithm achieves a 3D localization RMSE of 0.105 m, outperforming the LSTM model with an RMSE of 0.150 m. Due to the NLOS conditions encountered during the Task 1 process, the localization estimation error of the LSTM model, which relies on UWB sensor range data, is significantly increased. In contrast, the proposed Geo-LSTM algorithm incorporates a priori location inputs and fully exploits the temporal features of geometric information, achieving a reduction in mean localization error by 0.046 m, resulting in enhanced localization performance.

Fig. 10(b) shows the experimental data along the X, Y, and Z-axis, respectively. The green shaded area indicates the regions where the tag is undetectable by the MCS during the experiment, which can not obtain the reference data by MCS. In these areas, trajectory references were supplemented based on task settings. Compared with other algorithms, the Geo-LSTM algorithm shows higher accuracy on individual axes, with RMSE values of 0.086 m on the X-axis, 0.043 m on the Y-axis, and 0.044 m on the Z-axis, which is closer to the reference data. In addition, the UWB sensor-based methods provide stable localization data in the undetectable area of the MCS, indicating a significant advantage in compensating for the limitations of the visual sensors.

Fig. 11 and Table III collectively showcase the localization performance of various algorithms. In the comparison of the two base methods, the RMSE results show that the SGL algorithm outperforms the WLS method in localization accuracy. Additionally, the average IQR for the SGL algorithm and WLS

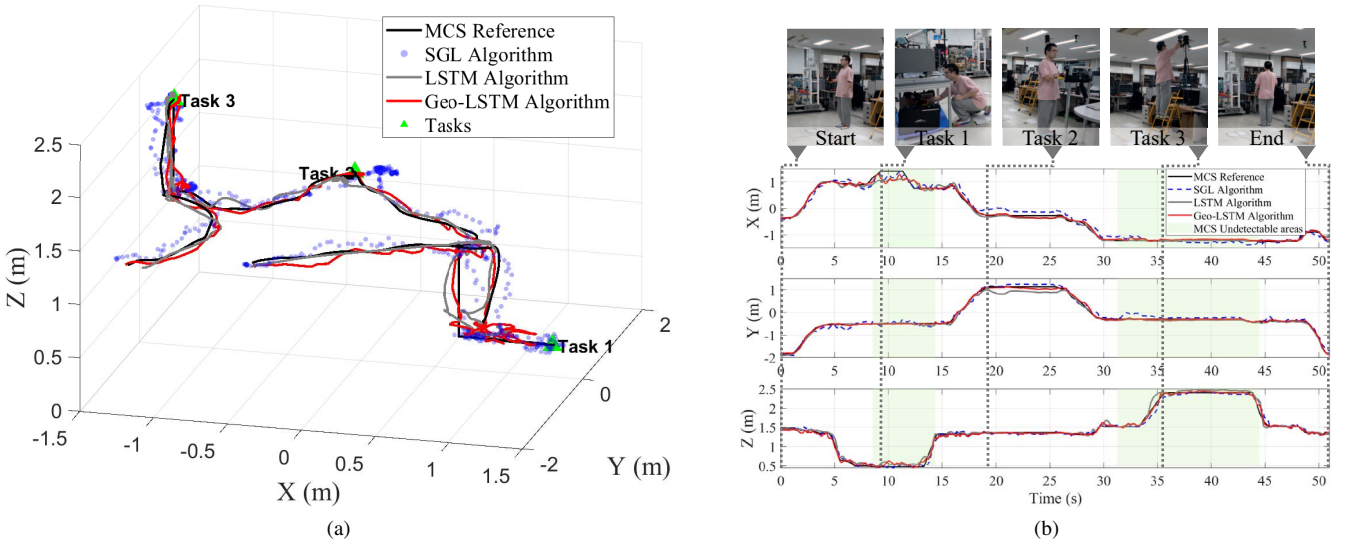

Fig. 10. 3D Localization results of different algorithms in complex task scenarios: (a) shows the comparison of 3D localization trajectories, and (b) depicts the single-axis localization results of each algorithm and a snapshot of task execution.

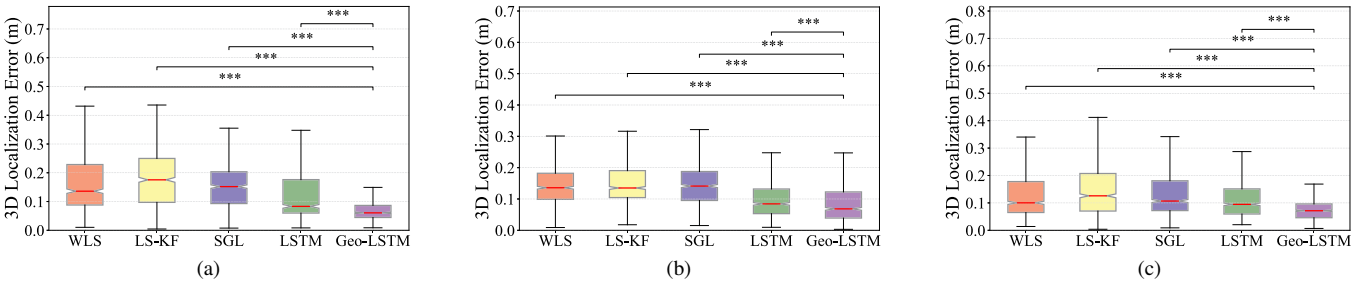

Fig. 11. Comparison and significance analysis of different algorithms on 3D localization error: (a), (b), and (c) represent the localization results of different algorithms for the three sets of experiments, respectively. (Note: "$***$" indicates p-value < 0.005.)

method is 0.104 m and 0.113 m, respectively, which proves that the SGL algorithm has lower dispersion in localization errors. The SGL algorithm reduces the average time cost of the whole testing sequence by 50% compared to the WLS method, which lowers the computational complexity of the localization estimation of the system while maintaining the localization accuracy. The two basic methods have significantly larger localization estimation errors for complex tasks, showing the limitations of UWB sensors under non-ideal conditions. The LS-KF method has the worst localization accuracy because of the NLOS condition of the UWB sensor. The localization estimation obtained by the base method has a large error, and the localization accuracy is not improved even with filtering and optimization.

When comparing neural network-based algorithms, the LSTM network relying solely on ranging data input demonstrates instability in location estimation across three experiments, with an average IQR of 0.096 m. In contrast, the proposed Geo-LSTM algorithm has a smaller error distribution, with an average IQR of 0.059 m. From the 3D localization RMSE and box plots, it is evident that the Geo-LSTM algorithm outperforms the LSTM model based solely on ranging data in both localization accuracy and error dispersion. The experimental results validate the effectiveness of introducing a priori positional inputs, which enables the LSTM network to better capture the temporal characteristics of diverse data. The Geo-LSTM algorithm employs two LSTM branches to process the a prior coordinates and UWB ranging sequences independently, effectively extracting temporal features from each input while preserving data independence and avoiding feature interference. This design enhances the fusion performance and improves the accuracy and stability of 3D localization.

Based on the results and analyses, the proposed Geo-LSTM algorithm achieves significant improvements in average localization RMSE over the three experiments, outperforming the WLS, LS-KF, and LSTM methods by 38.60%, 44.14%, and 31.20%, respectively. To evaluate the statistical significance of the results, the Wilcoxon rank sum test was used to compare

TABLE IV
COMPARISON OF THE 3D LOCALIZATION ACCURACY WITH REFERENCES

Ref.	Methods	Experimental Space* (m)	RMSE (m)
[22]	BLE	3 × 10 × 3	0.360
[17]	UWB	5 × 5 × 3	0.145
[15]	UWB, BMP	6 × 5 × 3	0.175
[23]	UWB, VIO	7 × 8 × 3.5	0.161
Ours	UWB, BMP, IMU	7 × 5 × 2.6	0.103

* Length × Width × Height

the localization errors of the Geo-LSTM algorithm against other methods. The results showed that the value for all comparisons were less than 0.005.

To evaluate the performance of the proposed method, we selected several mainstream indoor 3D localization algorithms from recent years for comparison, covering localization methods based on different sensing modalities, including UWB, visual-inertial odometry (VIO), BMP, and Bluetooth Low Energy (BLE). These methods are representative in terms of system architecture and application scenarios. Considering the differences in experimental environments across various references, we prioritized the selection of algorithms conducted in spatially comparable experimental settings, thereby ensuring a certain level of fairness and comparability in the evaluation. As shown in Table IV, the Geo-LSTM algorithm achieves the best performance among the compared methods, with an average 3D localization RMSE of 0.103 m, outperforming the overall results of the other algorithms. These findings demonstrate that the Geo-LSTM algorithm delivers high accuracy and stability in LOS and NLOS environments, providing reliable 3D localization estimation support for HRC applications.

V. Conclusion

This letter presents a Geo-LSTM localization algorithm designed to provide effective and accurate localization information for HRC applications. By fusing IMU sensor and BMP sensors and combining with dynamic EMA filtering, it achieves rapid response and high-accuracy height estimation. Based on this, an SGL localization algorithm is proposed to effectively reduce the computational complexity of tag localization. This algorithm also serves as a standalone high-precision localization approach, and experimental results verify its superiority over traditional methods. By combining the range data from the UWB sensors and the a priori position estimation provided by the SGL algorithm, the Geo-LSTM algorithm leverages the LSTM network to extract temporal features from multi-source data, resulting in more accurate localization. The stability and superiority of this 3D localization algorithm are verified through multiple repetitive experiments and complex task experiments in real-world environments, and the 3D localization RMSE is improved by 31.20% compared to an LSTM network using only range data from UWB sensors.

Future work will focus on applying the algorithms to specific robot collaboration tasks and adding more UWB anchors and tags. We will further optimize the design of the localization system to support larger-scale HRC applications. Specifically, this involves investigating the influence of distance and angle between tags and anchors, addressing potential anchor failures, and extending the system to multi-room indoor environments and multi-agent localization scenarios.

References

[1] Z. Liao, M. Lorenzini, M. Leonori, F. Zhao, G. Jiang, and A. Ajoudani, "An ergo-interactive framework for human-robot collaboration via learning from demonstration," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 359–366, Jan 2024.

[2] W. Kim, P. Balatti, E. Lamon, and A. Ajoudani, "Moca-man: A mobile and reconfigurable collaborative robot assistant for conjoined human-robot actions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 10 191–10 197.

[3] W. Zhong, L. Zhang, Z. Sun, M. Dong, and M. Zhang, "Ui-mocap: An integrated uwb-imu circuit enables 3d positioning and enhances imu data transmission," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1034–1044, 2024.

[4] J. Yan, Y. Zhang, B. Kang, W.-P. Zhu, and D. P.-K. Lun, "Multiple binocular cameras-based indoor localization technique using deep learning and multimodal fusion," *IEEE Sensors Journal*, vol. 22, no. 2, pp. 1597–1608, Jan 2022.

[5] L. Barbieri, M. Brambilla, A. Trabattoni, S. Mervic, and M. Nicoli, "Uwb localization in a smart factory: Augmentation methods and experimental assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–18, 2021.

[6] Y. Yang, X. Wang, D. Li, D. Chen, and Q. Zhang, "An improved indoor 3-d ultrawideband positioning method by particle swarm optimization algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[7] Y. Hua, Z. Zhu, G. Zhou, and G. Shen, "Chain state monitoring for a heavy scraper conveyor using uwb-based extended kalman filter technique with range constraint selection method," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.

[8] Q. Tian, K. I.-K. Wang, and Z. Salcic, "A resetting approach for ins and uwb sensor fusion using particle filter for pedestrian tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5914–5921, Aug 2020.

[9] H. Lv, J. Feng, H. Shou, J. Zhang, T. Cui, and Z. Mei, "Uwb localization based on dual-channel neural network and total least square method," *IEEE Sensors Journal*, vol. 24, no. 3, pp. 3477–3487, 2024.

[10] A. Poulose and D. S. Han, "Uwb indoor localization using deep learning lstm networks," *Applied Sciences*, vol. 10, no. 18, 2020.

[11] D.-H. Kim, A. Farhad, and J.-Y. Pyun, "Uwb positioning system based on lstm classification with mitigated nlos effects," *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1822–1835, Jan 2023.

[12] D. Gao, X. Zeng, J. Wang, and Y. Su, "Application of lstm network to improve indoor positioning accuracy," *Sensors*, vol. 20, no. 20, 2020.

[13] B. Yang, J. Li, Z. Shao, and H. Zhang, "Robust uwb indoor localization for nlos scenes via learning spatial-temporal features," *IEEE Sensors Journal*, vol. 22, no. 8, pp. 7990–8000, April 2022.

[14] K. Tang, B. Yang, and K. Ding, "Deep attention-based network combing geometric information for uwb localization in complex indoor environments," *IEEE Access*, vol. 12, pp. 31 488–31 497, 2024.

[15] D. Sun, G. He, Z. Wang, and H. Wang, "A 3-d indoor positioning method fusing relative barometric pressure and uwb measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–8, 2025.

[16] L. Bao, K. Li, J. Lee, W. Dong, W. Li, K. Shin, and W. Kim, "An enhanced indoor three-dimensional localization system with sensor fusion based on ultra-wideband ranging and dual barometer altimetry," *Sensors*, vol. 24, no. 11, 2024.

[17] J. Ma, X. Duan, C. Shang, M. Ma, and D. Zhang, "Improved extreme learning machine based uwb positioning for mobile robots with signal interference," *Machines*, vol. 10, no. 3, 2022.

[18] S.-H. Bach and S.-Y. Yi, "Constrained least-squares trilateration for indoor positioning system under high gdop condition," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4550–4558, March 2024.

[19] M. Delamare, R. Boutteau, X. Savatier, and N. Iriart, "Static and dynamic evaluation of an uwb localization system for industrial applications," *Sci*, vol. 2, no. 2, 2020.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] D. Feng, C. Wang, C. He, Y. Zhuang, and X.-G. Xia, "Kalman-filter-based integration of imu and uwb for high-accuracy indoor positioning and navigation," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3133–3146, April 2020.

[22] C. Huang, Z. Tian, W. He, K. Liu, and Z. Li, "Spotlight: A 3-d indoor localization system in wireless sensor networks based on orientation and rssi measurements," *IEEE Sensors Journal*, vol. 23, no. 21, pp. 26 662–26 676, 2023.

[23] A. Goudar, W. Zhao, and A. P. Schoellig, "Range-visual-inertial sensor fusion for micro aerial vehicle localization and navigation," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 683–690, 2024.