Tightly Coupled Visual-Inertial-UWB Indoor Localization System With Multiple Position-Unknown Anchors

Chao Hu, Ping Huang, and Wei Wang, *Senior Member, IEEE*

***Abstract*—In this letter, we perform a tightly-coupled fusion of a monocular camera, a 6-DoF IMU, and multiple position-unknown Ultra-wideband (UWB) anchors to construct an indoor localization system with both accuracy and robustness. Prior to this, there have been several works that have achieved satisfactory results by fusing UWB ranging measurements with visual-inertial system. However, these approaches still have some limitations: 1) these approaches either require the UWB anchor position to be calibrated in advance or the UWB anchor position estimation method used is not robust enough; 2) these approaches do not allow for dynamic changes to the number of UWB anchors in a tightly coupled estimator. Our approach uses visual object detection algorithm to provide UWB anchor initial position and refine it in the factor graph, using chi-square test algorithm to identify UWB ranging outliers. Based on the above two ideas, we implement a tightly coupled estimator that dynamically adjusts the number of UWB anchors, i.e. adding them to the factor graph when their ranging measurements are available and discarding them when their ranging measurements are outliers. These ideas improve the efficiency and robustness of the fusion about UWB ranging measurements with the visual-inertial system, as well as the easy setup of UWB anchors. Experimental results show that the proposed method outperforms previous methods in terms of estimating anchor position and improving localization accuracy.**

***Index Terms*—Visual-inertial SLAM, sensor fusion, localization.**

I. Introduction

ACCURATE and consistent position estimation is essential for mobile robots. In recent years, Visual Simultaneous Localization and Mapping (Visual-SLAM) is a popular research direction in the field of robot localization. It can obtain continuous 6-DoF pose estimates using only the camera. However, the pose estimation accuracy of Visual-SLAM in weak texture environment is poor. Due to the complementary nature of camera and Inertial Measurement Unit (IMU), introducing IMU into Visual-SLAM systems can obtain more robust pose estimation results, which is called Visual-Inertial SLAM (VI-SLAM) or Visual-Inertial Odometry (VIO) system. There are many excellent open source VIO systems, filter-based systems such as OpenVINS [1], R-VIO [2] and StructVIO [3], optimization-based systems such as VINS-Mono [4], OKVIS [5] and ORB-SLAM3 [6]. However, sensor noise and calculation errors cause the system drift to accumulate with increasing distance. A natural idea is to introduce a global measurement in the VIO system to constrain its drift. Global Navigation Satellite System (GNSS) is a good choice in outdoor environments, but it is not available in indoor environments [7], [8]. Recent studies have shown that Ultra-wideband (UWB) can provide reliable global constraints for VIO system in indoor environments [9], [10].

A natural idea is to loosely couple the localization results of UWB subsystem with those of VIO subsystem to obtain more accurate position estimates [11], [12]. However, it requires at least four calibrated UWB anchors, which makes the setting of UWB anchors not flexible enough to work in special environments such as corridors and tunnels. In contrast, a system with tightly coupled camera, IMU, and UWB can fully utilize the ranging information of UWB and has a more flexible UWB anchor setup. These systems use the global UWB ranging information to constrain the drift of the visual inertial sensor, and achieve more robust and accurate localization results [13], [14], [15]. UWB anchor position estimation is a frequently mentioned module in tightly coupled systems, and a number of anchor position estimation approaches based on UWB ranging measurements have been proposed in recent years. These approaches collect UWB ranging data over a period of time and then solve for the position of the UWB anchor using linear or nonlinear methods [10], [16], [17]. More recently, a distributed UWB network anchor position initialization approach is proposed to significantly reduce the computational complexity at the robot side by distributing the computing load for anchor position estimation to the anchor's on-board computer [18]. The approach strikes a good balance between the efficiency of anchor position initialization and the computational complexity of a single node. The above approaches have made excellent contributions to the fusion of camera, IMU, and UWBs for indoor localization. However, these approaches still have the following drawbacks: 1) Anchor position initialization methods based on UWB ranging measurements do not guarantee robustness in complex environments such as non-line-of-sight (NLOS); 2) Once the UWB anchor initialization is complete, the number of anchors in the

Manuscript received 1 July 2023; accepted 13 October 2023. Date of publication 30 October 2023; date of current version 28 November 2023. This letter was recommended for publication by Associate Editor H. Saito and Editor P. Vasseur upon evaluation of the reviewers' comments. This work was supported by the National Natural Science Foundation of China under Grant 62271163. *(Corresponding author: Ping Huang.)*

The authors are with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: moore_hc@qq.com; hppmonkeyking@163.com; wangwei407@hrbeu.edu.cn).

Digital Object Identifier 10.1109/LRA.2023.3328367

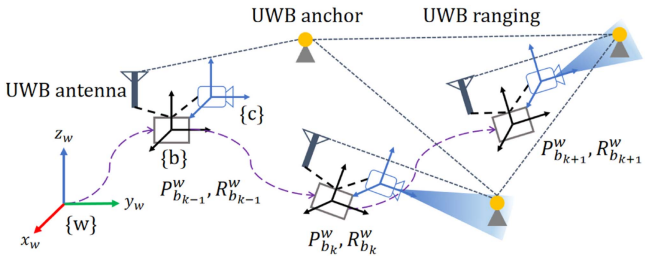

Fig. 1. Diagram of coordinate frames and measurements.

system is fixed. There is no dynamic adjustment of the number of anchors fused in the system based on the quality of the ranging measurements of the UWB anchors.

To solve the above problems, we propose a tightly coupled visual-inertial-UWB indoor localization system with dynamically adjusts the number of UWB anchors for accurate UWB anchor position estimation and robust visual-inertial-UWB odometry. The main contributions are summarized as follows:

- a visual measurement-based UWB anchor position initialization method, i.e., first using a deep learning-based object detection algorithm to obtain the position of the UWB anchor in the image, and then calculating its position in the world frame by a multi-view triangulation equation.
- a visual-inertial-UWB tightly coupled estimator that dynamically adjusts the number of UWB anchors based on the quality of UWB anchor ranging measurements for more accurate robot position estimation.
- extensive experimental results validate the performance of the proposed UWB anchor initialization method and the visual-inertial-UWB tightly coupled estimator.

This letter is structured as follows: Firstly, Section II introduces an overview of related work. Secondly, the details of all modules in the proposed system are presented in Section III. Next, Section IV describes the real-life experimental setup and experimental results compared with state-of-the-art methods. Finally, the letter is concluded in Section V.

II. Related Work

A. UWB Anchor Initialization

Calibrating the UWB anchor position in an unknown environment is difficult. Therefore, online calibration of UWB anchors using sensor outputs on the robot will greatly improve the efficiency of UWB anchors deployment. [10], [16] proposed a linearization method to estimate the position of UWB anchors, where [16] used an information-theoretic approach to improve the UWB anchors initialization accuracy. In [17] a long and short sliding window is proposed for initializing UWB anchors and state augmentation. [18] proposed a fully distributed framework for massive UWB network initialization that uses the Fisher information matrix to determine whether a UWB anchor is observable or not, while performing anchor position solving on the anchor's on-board computer. This method not only improves the initialization accuracy of the UWB anchors but also reduce the computational burden at the robot side. All of the above methods only use UWB ranging measurements to initialize the position of UWB anchors. However, UWB signals in complex environments may be in NLOS, resulting in inaccurate UWB range measurements thus the initialization error of UWB anchor position is large [19]. Inspired by the visual-based multi-robot relative localization problem in [20] and [21], we proposed an object detection-based UWB anchor position initialization method that uses visual measurements to compute the UWB anchor position and avoids the NLOS problem.

B. Visual-Inertial-UWB Coupled Odometry

UWB can realize independent localization based on the ranging measurements, and then fuse with the localization results of other systems. In [22], [23], VIO is first performed to obtain the position of a single robot, and then UWB ranging measurements are used to constrain the drift of VO/VIO. In these methods UWB and VIO data are fused together in a loosely coupled manner. However, the correlation between multi-sensor data is not taken into account and the localization results are suboptimal. Therefore many recent studies have fused raw measurements from camera, IMU and UWBs in a tightly coupled manner, and an UWB anchor position estimation module have been included in these systems [24], [25], [26]. VIR-SLAM [14] uses the factor graph optimization framework to integrate camera, IMU and UWB data, and utilize a double-layer sliding window to enhance the constraint effect of UWB ranging measurements. In [27], Yang et al. propose an elastic Visual-Inertial-UWB tightly coupled method to achieve seamless switching between different positioning modes of the system. In [15], Nguyen et al. propose a "range-focused" Visual-Inertial-UWB tightly coupled method, which takes into account the temporal offset between the image and the UWB data, so as to improve the accuracy of multi-sensor data matching. However, once the initialization of the UWB anchor positions is complete, the number of anchors in these systems is fixed, and there is no way to dynamically adjust the anchors based on the quality of the anchor data. So we design a dynamic adjustment strategy for UWB anchors based on the quality of UWB ranging measurements.

III. Methodology

In this section, the coordinate frames and notations of our system are presented first. Then, we introduce an overview of the proposed system, including the UWB anchor position initialization module and the visual-inertial-UWB tightly coupled estimator. Finally we present the details of all the parts in the system.

A. Coordinate Frames and Notations

The coordinate frames and notations involved in this letter are explained as below. $\mathbf{c}$ denotes the camera frame, $\mathbf{b}$ denotes the IMU Frame and $\mathbf{w}$ denotes the world frame. The relationship between these coordinate frames is depicted in Fig. 1. The notations used for coordinate transformations include rotation matrix $\mathbf{R}^{\mathbf{B}}_{\mathbf{A}}$, quarternion $\mathbf{q}^{\mathbf{B}}_{\mathbf{A}}$ and translation vector $\mathbf{p}^{\mathbf{B}}_{\mathbf{A}}$. $\|\mathbf{x}\|$

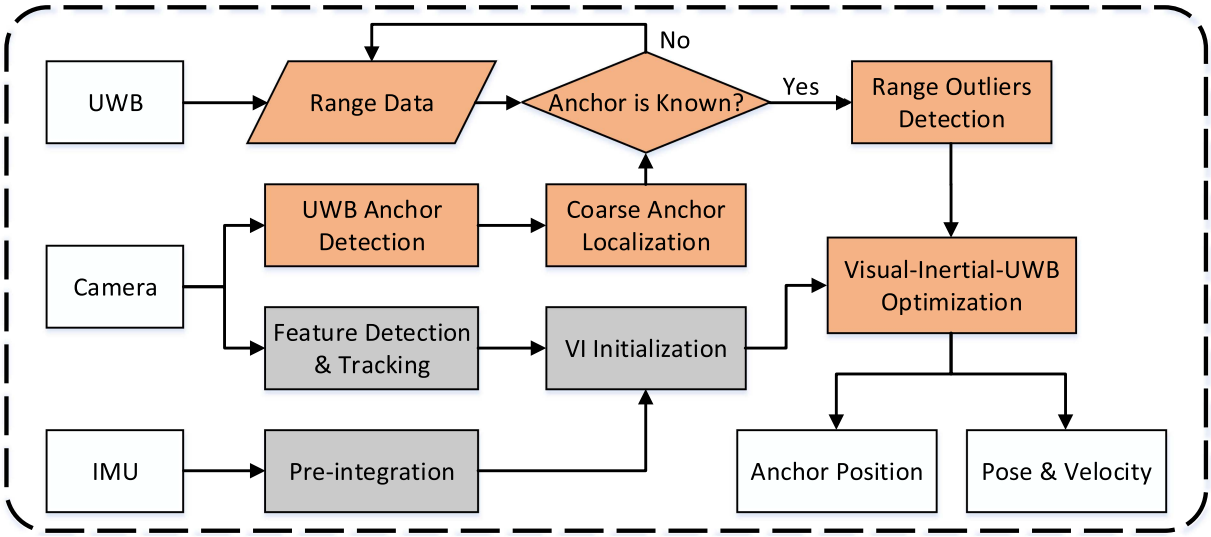

Fig. 2. Overview of the proposed system. Based on the state-of-the-art open-source VIO system VINS-mono [4], the orange boxes indicate the modules for the contribution of this letter.

TABLE I
GLOSSARY OF NOTATION

Symbol	Meaning	Symbol	Meaning
$\mathbf{c}$	Camera Frame	$\mathbf{b}$	IMU Frame
$\mathbf{w}$	World Frame	$r(\tilde{\mathbf{z}}, \chi)$	measurement residual
$\mathbf{R}_{\mathbf{A}}^{\mathbf{B}}$	Rotation Matrix from Frame $\mathbf{A}$ to Frame $\mathbf{B}$	$\mathbf{q}_{\mathbf{A}}^{\mathbf{B}}$	Quarternion from Frame $\mathbf{A}$ to Frame $\mathbf{B}$
$\mathbf{p}_{\mathbf{A}}^{\mathbf{B}}$	Translation Vector from Frame $\mathbf{A}$ to Frame $\mathbf{B}$	$\|\mathbf{x}\|$	the Euclidean Norm of $\mathbf{x}$

denotes the Euclidean norm of a vector. $r(\tilde{\mathbf{z}}, \chi)$ denotes the measurement residual in the cost function.

For the convenience of reading, all involved coordinate frames and some important notations used in this letter are summarized in Table I.

B. System Overview

Fig. 2 illustrates the overview of the proposed system. The orange boxes indicate the modules for the contribution of this letter. First, the VIO is run normally after the visual-inertial initialization is completed. When the camera detects a UWB anchor, the rough position of the anchor is calculated by the coarse anchor localization module. Then, the similarity between UWB ranging measurements and the body-anchor distance is calculated to determine the ID of the anchor. After that, the position and ranging measurements of the anchor are added to the Visual-Inertial-UWB optimization module to optimize the anchor position and finally fix it. In addition, the range outlier detection module detects the range outliers of an anchor, and when the number of outliers exceeds a threshold, the measurements and states associated with this anchor are removed from the factor graph until its ranging measurements are normal, and then it will be added to the factor graph again, thus realizing the dynamic adjustment of the UWB anchors.

C. UWB Anchor Localization Based on Object Detection

Manual calibration of the positions of UWB anchors is tedious and time consuming, so it is necessary to estimate the positions of anchors online based on sensor data. Recently, a distributed UWB network anchor position initialization approach is proposed to realize online initialization of anchor positions for massive UWB networks [18]. This approach significantly reduce the computational complexity at the robot side by distributing the computing load for anchor position estimation to the anchor's on-board computer. However this method using only UWB ranging measurements does not guarantee that all anchors are in LOS conditions, commonly found in complex indoor environments [19]. Ranging measurements under NLOS condition can significantly increase the initialization error of the anchor positions. Inspired by visual-based methods for relative localization of multiple robots, we use a camera on the robot to detect UWB anchors to ensure that the robot and UWB anchors are at LOS condition.

The visual object detection algorithm we use is TPH-YOLOv5 [28], which performs very well in terms of detection speed and accuracy, so it meets the real-time requirements of SLAM. Firstly, we collected a dataset containing 1000 UWB anchor pictures to train the UWB anchor detection model. Then we take the RGB image of the camera as the input of the model to obtain the detection result and the pixel location of the anchor. The dataset labeled images and anchor detection results are shown in Fig. 3.

As shown in Fig. 4, accurate body pose in a short time can be obtained through VIO, so multiple frames of anchor detection results and their camera poses can be collected. Therefore, Using the multi-view triangulation model, we can derive the

Fig. 3. (a) Labeled images for the UWB anchor dataset. (b) Detection result of UWB anchor by TPH-YOLOv5.

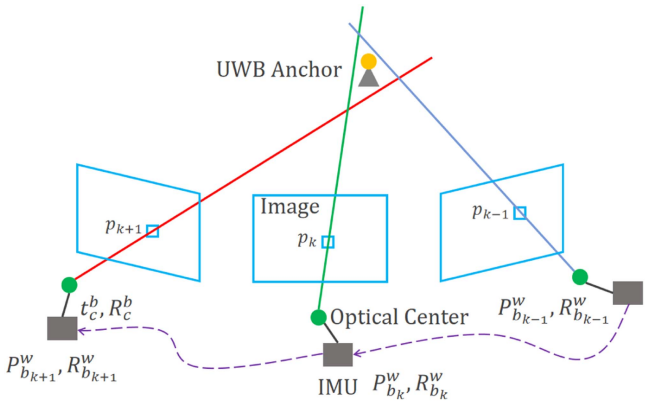

Fig. 4. Multi-frame anchor projection.

reprojection residual equations of the anchor position as follows:

$$\begin{cases} \mathbf{r}_a(\tilde{\mathbf{z}}_a^{c_1}, \mathbf{p}_a^{\mathbf{w}}) & = \mathcal{K}(\mathbf{R}_{\mathbf{b}}^{\mathbf{c}}(\mathbf{R}_{\mathbf{w}}^{\mathbf{b}_1}(\mathbf{p}_a^{\mathbf{w}} - \mathbf{p}_{\mathbf{b}_1}^{\mathbf{w}}) - \mathbf{p}_{\mathbf{c}}^{\mathbf{b}})) - p_1, \\ \mathbf{r}_a(\tilde{\mathbf{z}}_a^{c_2}, \mathbf{p}_a^{\mathbf{w}}) & = \mathcal{K}(\mathbf{R}_{\mathbf{b}}^{\mathbf{c}}(\mathbf{R}_{\mathbf{w}}^{\mathbf{b}_2}(\mathbf{p}_a^{\mathbf{w}} - \mathbf{p}_{\mathbf{b}_2}^{\mathbf{w}}) - \mathbf{p}_{\mathbf{c}}^{\mathbf{b}})) - p_2, \\ & \vdots \\ \mathbf{r}_a(\tilde{\mathbf{z}}_a^{c_j}, \mathbf{p}_a^{\mathbf{w}}) & = \mathcal{K}(\mathbf{R}_{\mathbf{b}}^{\mathbf{c}}(\mathbf{R}_{\mathbf{w}}^{\mathbf{b}_j}(\mathbf{p}_a^{\mathbf{w}} - \mathbf{p}_{\mathbf{b}_j}^{\mathbf{w}}) - \mathbf{p}_{\mathbf{c}}^{\mathbf{b}})) - p_j, \end{cases} \tag{1}$$

where $\mathcal{K}(\cdot)$ is the camera projection function, $\mathbf{R}_{\mathbf{b}}^{\mathbf{c}}$ is the camera-IMU extrinsic rotation matrix, $\mathbf{p}_{\mathbf{b}}^{\mathbf{c}}$ is the camera-IMU extrinsic translation vector, and p_j is the pixel location of the center of the anchor detection box. By constructing the anchor projection residuals for multiple image frames, we can calculate the coarse position of the anchor using least square method. Finally, the similarity between UWB ranging measurements and the body-anchor distance is calculated to determine the ID of the anchor. Since the robot and UWB anchors are in LOS condition, it is guaranteed that correct anchor ID matching results are obtained.

When the coarse position of an anchor is determined, the anchor and its ranging measurements are added to the Visual-Inertial-UWB optimization framework. Then the anchor position is refined in the iteration and fixed when the change is less than a threshold. In this letter, we set the threshold to 0.01 meters.

D. Visual-Inertial-UWB Odometry With Dynamic Adjustment of Number of Anchors

1) Status in Our System: The proposed system uses a sliding window optimization manner. Therefore, the states of the

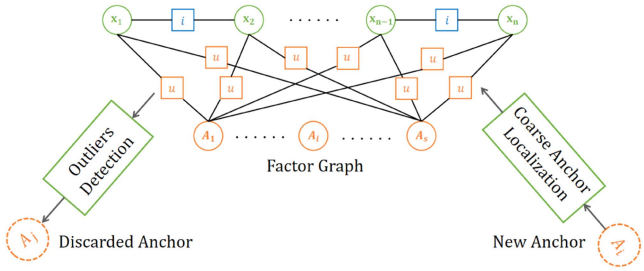

Fig. 5. Dynamic adjustment process for UWB anchors.

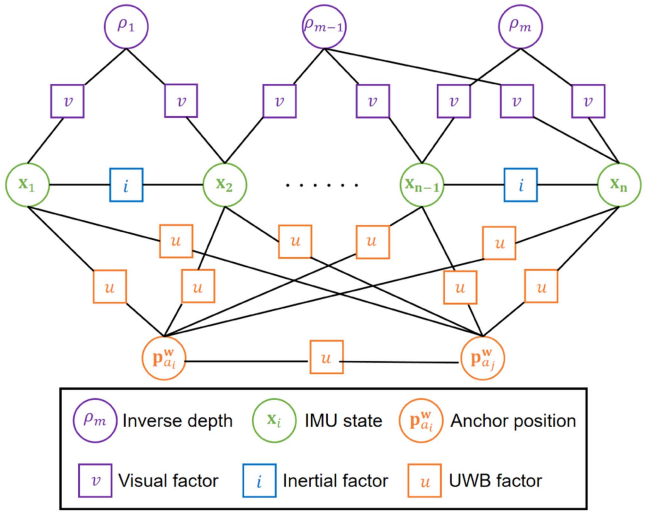

Fig. 6. Factor graph representation of the optimization problem in our system, where the circles represent states and the boxes represent factors.

proposed system within the sliding window is summarized as follows:

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \rho_1, \rho_2, \ldots, \rho_m, \mathbf{p}_{a_1}^{\mathbf{w}}, \mathbf{p}_{a_2}^{\mathbf{w}}, \ldots, \mathbf{p}_{a_s}^{\mathbf{w}}], \\ \mathbf{x}_k &= [\mathbf{p}_{\mathbf{b}_k}^{\mathbf{w}}, \mathbf{v}_{\mathbf{b}_k}^{\mathbf{w}}, \mathbf{q}_{\mathbf{b}_k}^{\mathbf{w}}, \mathbf{b}_a, \mathbf{b}_g], k \in [1, n], \end{aligned} \tag{2}$$

where $\mathbf{p}_{\mathbf{b}_k}^{\mathbf{w}}$ is the position of the body in the world frame, $\mathbf{q}_{\mathbf{b}_k}^{\mathbf{w}}$ is the orientation of the body in the world frame, $\mathbf{v}_{\mathbf{b}_k}^{\mathbf{w}}$ is the velocity, $\mathbf{b}_a$ is the accelerometer bias, $\mathbf{b}_g$ is the gyroscope bias, ρ is the inverse depth for each feature, $\mathbf{p}_{a_j}^{\mathbf{w}}$ is the position of the jth anchor in the world frame, n is the window size, m is the number of feature points in the sliding window and s is the number of UWB anchors in the factor graph.

As shown in Fig. 5, when the initialization of an anchor is complete, the rough position of this anchor and its ranging measurements are added to the factor graph. In contrast, when an anchor's ranging measurement is detected as an outlier several consecutive times, the position and ranging measurements of this anchor are discarded from the factor graph.

2) Factor Graph: The factor graph representation of our system is plotted in Fig. 6, which includes visual factor v, inertial factor i and UWB factor u. Because visual factor and inertial factor are covered in detail in VINS-Mono [4], this letter only gives a brief introduction to them. Visual factor represents the reprojection error of the same feature point in different image

frames. These feature points are extracted from the image frames and tracked by the KLT sparse optical flow algorithm [29]. Inertial factor represents the error between the IMU pre-integration measurement and the prediction obtained based on the states of adjacent frames, where the IMU pre-integration technique is described in [30].

In this letter, we use two-way ranging (TWR) to measure the distance between two UWB sensors because it does not require time synchronization between sensors and is therefore suitable for more scenarios. We first consider the use of body-anchor ranging measurements to optimize the body position and anchor position. However, the body position continues to drift over time, and this drift will be transmitted to the anchor position through the UWB factor, which eventually leads to a large error in the anchor position estimation. Therefore we add the anchor-anchor ranging measurements to the UWB factor. The anchor-anchor ranging measurements represent the constant distance information between anchors, which can effectively reduce the impact of body drift on anchor position estimation. Thus the UWB factor in our system includes two kinds of residuals: body-anchor ranging residual and anchor-anchor ranging residual. These two kinds of residuals can be formulated as (3) and (4), respectively.

$$r_{u_b}(\tilde{\mathbf{z}}_{b_k}^{a_j}, \chi) = ||\mathbf{p}_{\mathbf{b}_k}^{\mathbf{w}} + \mathbf{R}_{\mathbf{b}_k}^{\mathbf{w}}\mathbf{p}_{\mathbf{r}}^{\mathbf{b}} + \Delta\hat{\mathbf{p}}_j^k - \mathbf{p}_{a_j}^{\mathbf{w}}|| - d_{b_k}^{a_j}, \quad (3)$$

$$r_{u_a}(\tilde{\mathbf{z}}_{a_j}^{a_i}, \chi) = ||\mathbf{p}_{a_j}^{\mathbf{w}} - \mathbf{p}_{a_i}^{\mathbf{w}}|| - d_{a_j}^{a_i}, \quad (4)$$

In the above formulation, $\mathbf{p}_{\mathbf{r}}^{\mathbf{b}}$ denotes the translation vector for UWB receiver in the body frame, $d_{b_k}^{a_j}$ denotes body-anchor ranging measurements, $d_{a_j}^{a_i}$ denotes anchor-anchor ranging measurements and $\Delta\hat{\mathbf{p}}_j^k$ denotes the body position change caused by time offset between image frame and UWB measurements. The body position change is predicted by the IMU pre-integration, which can be formulated as (5).

$$\Delta\hat{\mathbf{p}}_j^k = \iint_{t\in[t_k,tj]} \mathbf{R}_{b_t}^{b_{t_k}}(\tilde{\mathbf{a}}_t - \mathbf{b}_{a_t})dt^2, \quad (5)$$

In our Visual-Inertial-UWB optimization framework, we consider the time offset between image frame and UWB measurement as well as the anchor-anchor ranging measurements, which improves the accuracy of the anchor and body position estimation.

3) Outliers Detection: In this letter, Chi-Squared Test is used to detect UWB ranging outliers. For each UWB ranging measurement, we can calculate its residual using (3), where the body position is obtained through VIO and the anchor position is known. A sliding window consisting of the UWB measurement residuals is maintained in the system. r_{u_k} is the current residual, which is combined with the residuals in the sliding window to form a new residual vector $\mathbf{r}_u$ as shown in (6). Then the covariance matrix $\mathbf{S}_u$ corresponding to the residual vector is calculated by (7), where $\mathbf{J}$ is the Jacobian of the residual against the body position, $\mathbf{P}$ is the covariance matrix of the position estimation and $\mathbf{R}_u$ is the covariance matrix of UWB measurement noise. The $\mathbf{P}$ corresponding to the current UWB measurement moment is obtained using IMU pre-integration.

$$\mathbf{r}_u = [r_{u_1}, r_{u_2}, r_{u_3}, \ldots, r_{u_n}, r_{u_k}]^T, \quad (6)$$

$$\mathbf{S}_u = \mathbf{J}^T\mathbf{P}\mathbf{J} + \mathbf{R}_u, \quad (7)$$

Since the residual vector $\mathbf{r}_u$ follows a zero-mean Gaussian distribution, the normalized sum of squares of its values should follow a Chi-Squared (χ^2) distribution with degrees of freedom equal to the dimension of the residual vector. However, the UWB measurement outliers will break this assumption, so we can define the Chi-Squared distribution statistic and discriminant criteria as follows:

$$l_u = \mathbf{r}_u^T\mathbf{S}^{-1}\mathbf{r}_u, \quad (8)$$

$$l_u \underset{H_1}{\overset{H_0}{\lessgtr}} \chi^2(0.95), \quad (9)$$

In this letter, the threshold for the Chi-Squared test was set to 95%, the dimension of the residual vector is 6, so the corresponding threshold is 12.59. Values above this threshold will be marked as outliers, the corresponding r_{u_k} will not be added to the sliding window of the UWB measurement residuals, and the corresponding UWB ranging measurement will not be added to the cost function.

IV. Experimental Results

In this section, each module of our system is evaluated separately by public dataset and real-life experiments. VINS-Mono, VIR-SLAM and DC-VIRO [18] are used as baselines to compare with our system in localization performance.

A. UWB Anchor Localization

We first evaluated the UWB anchor localization module through real-life experiments. As shown in Fig. 7(a), The hardware is a multi-sensor helmet with an Intel RealSense D455 camera, a DW1000 UWB receiver, and an Ublox ZED-F9P GNSS receiver mounted on it. In the experiments of this letter, RealSense D455 camera provides mono RGB images at 30 Hz, IMU data at 200 Hz and DW1000 UWB receiver provides ranging data at 20 Hz. Ublox ZED-F9P GNSS receiver is not used in this letter. Some static test points are measured by the total station as ground truth.

A UWB anchor is placed at an unknown position and the system runs in VIO-only mode to estimate the anchor position. The distance error of the estimated anchor position is used for comparison, and the reference position is provided by the total station. We run the system with three different trajectories respectively, the anchor position estimation results are shown in Fig. 8, and the corresponding trajectories are shown in Fig. 9. Our method calculates the initial position of the anchor through the coarse anchor localization module within 1.5 seconds, and the initial position is very close to the reference position. Over the next few seconds the anchor positions are adjusted in the factor graph. We compare the anchor localization error estimated by the proposed method with VIR-SLAM, and the results are shown in Table II. Our method achieves a smaller localization error, which is due to the fact that the coarse anchor localization module provides the initial position of the anchor, and therefore it is easier to converge near the reference position in the subsequent optimization process compared to

Fig. 7. Experimental equipment (a) and environment (b) used in real-life experiments.

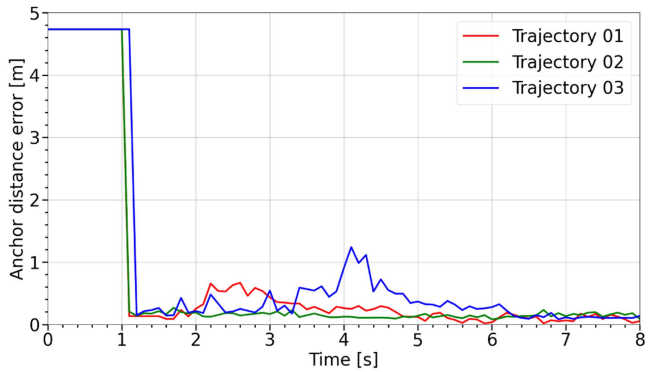

Fig. 8. Experimental results for the UWB anchor localization.

VIR-SLAM. If the DC-VIRO authors open the source code of their distributed UWB anchor initialization method, our method will be compared with them as well.

B. *Visual-Inertial-UWB Odometry*

In this subsection, the performance of the odometry is evaluated with public dataset and real-life experiment. Absolute trajectory error (ATE) is used to evaluate the performance of the odometry, which is a standard method for evaluating SLAM systems. All estimated trajectories were obtained by running these systems on an HP-OMEN-7 laptop.

1) Public Dataset: We first evaluate our approach on the VIRAL public dataset [31], which contains data from multiple sensors such as stereo cameras, IMU, three UWB anchors, etc.

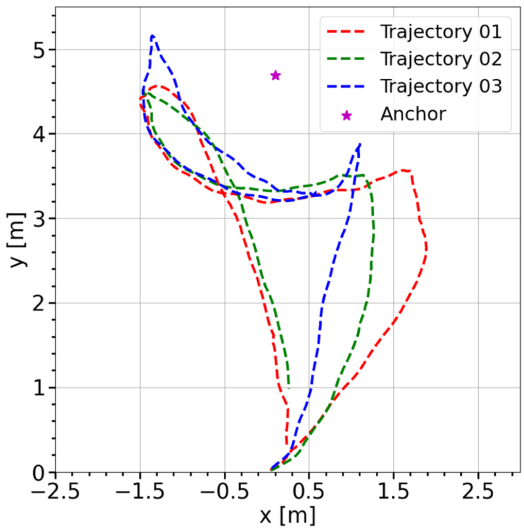

Fig. 9. System running trajectories corresponding to the anchor localization results.

TABLE II
COMPARISON OF UWB ANCHOR DISTANCE ERRORS (M) FOR DIFFERENT METHODS

Distance error (m)	VIR-SLAM	Proposed
Trajectory 01	0.230	**0.053**
Trajectory 02	0.368	**0.087**
Trajectory 03	0.770	**0.102**

The best results are highlighted in **bold**.

TABLE III
COMPARISON OF ATE RMSE (M) FOR DIFFERENT APPROACHES ON VIRAL DATASET

ATE RMSE (m)	VINS-Mono	VIR-SLAM	Proposed	DC-VIRO [18]
eee_01	1.305	N/A	0.781	**0.524**
eee_02	0.854	1.640	0.678	**0.382**
eee_03	1.065	N/A	**0.315**	0.331
nya_01	0.915	1.338	0.604	**0.412**
nya_02	0.554	0.843	**0.212**	0.217
nya_03	1.445	1.743	0.513	**0.263**

The best results are highlighted in **bold**, the failed results are denoted by N/A.

Since the appearance of the UWB anchor could not be simulated in the VIRAL dataset for detection by anchor detection module, the UWB anchor position was set as known in the system and only the visual-inertial-UWB odometry performance was evaluated. Our system is compared with VINS-Mono, VIR-SLAM and DC-VIRO, the loop closure module was disabled for all systems, the Root Mean Square Error (RMSE) of ATE for all trajectories is depicted in Table III. It is worth noting that DC-VIRO's results were extracted from their paper, and their experiments used a stereo camera. Fig. 10 depicts some of the

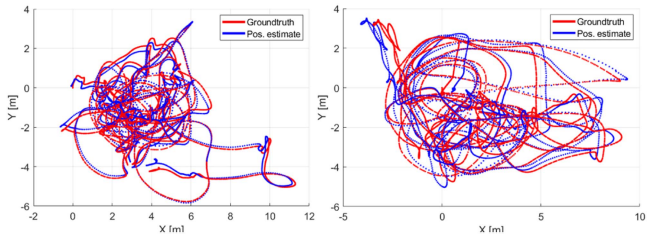

Fig. 10. Top view of the estimated (proposed system) and ground truth trajectories for the nya_02 (left) and nya_03 (right) sequences in the VIRAL dataset.

TABLE IV
COMPARISON OF ATE RMSE (M) FOR DIFFERENT APPROACHES IN REAL-LIFE EXPERIMENTS

Approaches	Sequence 01	Sequence 02	Sequence 03
VINS-Mono	1.004	0.831	0.572
VIR-SLAM	0.815	0.754	0.588
Proposed(2)	0.268	0.758	0.486
Proposed(3)	0.256	0.698	0.411
Proposed(full)	**0.214**	**0.655**	**0.246**

Proposed(*) represents the number of UWB anchors used in the proposed system. The best results are highlighted in **bold**.

estimated trajectories of the proposed system and the ground truth trajectories.

It is clear from Table III that DC-VIRO and the proposed system achieve better localization performance compared to VINS-Mono. This is because the absolute ranging measurements provided by the UWB sensor alleviates the position drift of the VIO system. However, VIR-SLAM suffers from poor robustness to UWB ranging outliers, resulting in a degradation of its localization performance or even localization failure. Although our approach uses a monocular camera while DC-VIRO uses a stereo camera, our method achieves localization performance close to that of DC-VIRO. This is because we included anchor-anchor ranging measurements in the UWB factors and carefully handled the temporal deviation between the image keyframes and the UWB data. Moreover, our UWB ranging outlier detection module also improves the robustness of the proposed system.

2) Real-Life Experiment: The multi-sensor helmet hardware system was previously presented in Section IV-A. In this experiment, three continuous movement trajectories were collected in an 11 m $\times$ 11 m hall area to test the drift of the odometry. As shown in Fig. 7(b), the experimental environment is a weakly textured indoor hall, where the localization performance of the visual-inertial system will degrade. Therefore, the effect of UWB ranging measurements on system performance can be better proved. Due to the limitation of the experimental equipment, we were unable to obtain the complete ground truth trajectory by motion capture system or laser tracking system as in most literature. Therefore, we adopted the method proposed in [32]: a static test point is set on the movement path of the multi-sensor equipment, and the position of the static test point is precisely measured by the total station. The multi-sensor equipment will pass through the test point several times and stay at the point for a few seconds each time, so that the position error of the estimated trajectory at this position can be obtained and the odometry drift can be verified. In all experiments, four UWB anchors are placed at unknown locations in the hall and each new test requires an online estimation of the anchor positions. To verify the performance of the system when the robot does not observe one or multiple UWB anchors, we manually limited the number of UWB anchors to 2 or 3 during initialization and compared their localization accuracy to the full anchors setup.

In this experiment, the localization performance of VINS-Mono, VIR-SLAM and the proposed system with different setups for the number of anchors are compared, and the ATE RMSE for all sequences are recorded in Table IV. Fig. 11 shows an overview of the estimated trajectories of VINS-Mono, VIR-SLAM and our full system in the *Sequence 01* and *Sequence 03* sequences, as well as the ATE results at the static test points for all sequences. For VINS-Mono, the cumulative error of the visual-inertial system is not corrected, which eventually leads to large trajectory drift. The localization performance of VIR-SLAM with the addition of UWB ranging measurements is improved, however, the proposed system is proven to have the best localization accuracy in all of the experiments. In addition, the performance of the proposed system is shown to improve with the increase in the number of UWB anchors.

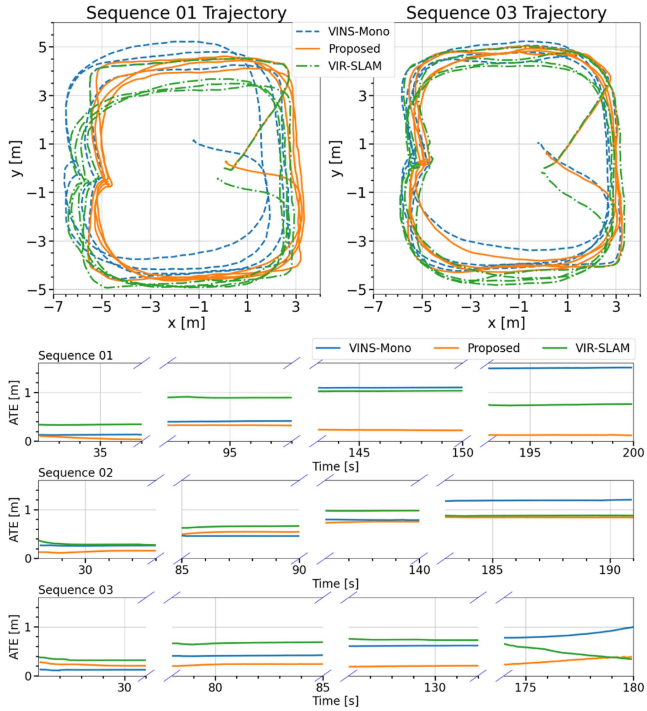

Fig. 11. Comparison results of VINS-Mono, VIR-SLAM and our full system in real-life experiments. Top: top view of the estimated trajectories in the *Sequence 01* and *Sequence 03* sequences. Bottom: ATE at the static test point for all sequences.

V. Conclusion

In this letter, we present an indoor localization system that fuses a visual-inertial system with multiple position-unknown anchors. First, we use a visual object detection algorithm to provide UWB anchor coarse position, which is subsequently refined in the factor graph, and this method can improve the accuracy of UWB anchor position estimation. Secondly, we design a dynamic adjustment strategy for the number of UWB anchors to choose whether to keep them in the factor graph or not, based on the results of the chi-square test for UWB ranging measurements. This strategy improves the robustness of the visual-inertial-UWB tightly-coupled estimator in complex NLOS environments. Public dataset and real-life experiments demonstrate that the proposed system outperforms previous methods in terms of estimating anchor position and improving localization accuracy. In the future, we would like to fuse GNSS raw measurements with our visual-inertial-UWB system to construct a seamless indoor and outdoor localization system. In particular, we expect the fusion system to still have high localization accuracy in complex indoor-outdoor transition environments.

References

[1] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A research platform for visual-inertial estimation," in *Proc. IEEE Int. Conf. Robot. Automat*., 2020, pp. 4666–4672.

[2] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 6319–6326.

[3] D. Zou, Y. Wu, L. Pei, H. Ling, and W. Yu, "StructVIO: Visual-inertial odometry with structural regularity of man-made environments," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 999–1013, Aug. 2019.

[4] T. Qin, P. Li, and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[5] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[6] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.

[7] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," 2019, *arXiv:1901.03642*.

[8] S. Cao, X. Lu, and S. Shen, "GVINS: Tightly coupled GNSS–visual–inertial fusion for smooth and consistent state estimation," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2004–2021, Aug. 2022.

[9] M. W. Mueller, M. Hamer, and R. D'Andrea, "Fusing ultra-wideband range measurements with accelerometers and rate gyroscopes for quadrocopter state estimation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 1730–1736.

[10] K. Hausman, S. Weiss, R. Brockers, L. Matthies, and G. S. Sukhatme, "Self-calibrating multi-sensor fusion with probabilistic measurement validation for seamless sensor switching on a UAV," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 4289–4296.

[11] C. Wang, H. Zhang, T.-M. Nguyen, and L. Xie, "Ultra-wideband aided fast localization and mapping system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 1602–1609.

[12] J. Tiemann, A. Ramsey, and C. Wietfeld, "Enhanced UAV indoor navigation through SLAM-augmented UWB localization," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2018, pp. 1–6.

[13] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Tightly-coupled single-anchor ultra-wideband-aided monocular visual odometry system," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 665–671.

[14] Y. Cao and G. Beltrame, "VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems," *Auton. Robots*, vol. 45, pp. 905–917, 2021.

[15] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 1678–1685, Apr. 2021.

[16] J. Blueml, A. Fornasier, and S. Weiss, "Bias compensated UWB anchor initialization using information-theoretic supported triangulation points," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 5490–5496.

[17] S. Jia, Y. Jiao, Z. Zhang, R. Xiong, and Y. Wang, "FEJ-VIRO: A consistent first-estimate Jacobian visual-inertial-ranging odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 1336–1343.

[18] S. Jia, R. Xiong, and Y. Wang, "Distributed initialization for visual-inertial-ranging odometry with position-unknown UWB network," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 6246–6252.

[19] S. Zheng et al., "UWB-VIO fusion for accurate and robust relative localization of round robotic teams," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 11950–11957, Oct. 2022.

[20] Y. Wang, X. Wen, L. Yin, C. Xu, Y. Cao, and F. Gao, "Certifiably optimal mutual localization with anonymous bearing measurements," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 9374–9381, Oct. 2022.

[21] Z. Xun, J. Huang, Z. Li, C. Xu, F. Gao, and Y. Cao, "Crepes: Cooperative relative pose estimation towards real-world multi-robot systems," 2023, *arXiv:2302.01036*.

[22] T.-M. Nguyen, Z. Qiu, M. Cao, T. H. Nguyen, and L. Xie, "Single landmark distance-based navigation," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 5, pp. 2021–2028, Sep. 2020.

[23] B. Yang, J. Li, and H. Zhang, "UVIP: Robust UWB aided visual-inertial positioning system for complex indoor environments," in *Proc. IEEE Int. Conf. Robot. Automat*, 2021, pp. 5454–5460.

[24] Q. Shi, X. Cui, W. Li, Y. Xia, and M. Lu, "Visual-UWB navigation system for unknown environments," in *Proc. 31st Int. Tech. Meeting Satell. Division Inst. Navigation*, 2018, pp. 3111–3121.

[25] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Tightly-coupled ultra-wideband-aided monocular visual SLAM with degenerate anchor configurations," *Auton. Robot.*, vol. 44, no. 8, pp. 1519–1534, 2020.

[26] F. J. Perez-Grau, F. Caballero, L. Merino, and A. Viguria, "Multi-modal mapping and localization of unmanned aerial robots based on ultra-wideband and RGB-D sensing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst*., 2017, pp. 3495–3502.

[27] B. Yang, J. Li, and H. Zhang, "Resilient indoor localization system based on UWB and visual–inertial sensors for complex environments," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 8504014.

[28] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOV5: Improved YOLOV5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis*., 2021, pp. 2778–2788.

[29] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.

[30] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, Feb. 2017.

[31] T.-M. Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, "NTU VIRAL: A visual-inertial-ranging-LiDAR dataset, from an aerial vehicle viewpoint," *Int. J. Robot. Res.*, vol. 41, no. 3, pp. 270–280, 2022.

[32] W. Jiang, Z. Cao, B. Cai, B. Li, and J. Wang, "Indoor and outdoor seamless positioning method using UWB enhanced multi-sensor tightly-coupled integration," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10633–10645, Oct. 2021.