Measurement 206 (2023) 112256

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

GNSS-denied UAV indoor navigation with UWB incorporated visual inertial odometry

Huei-Yung Lin [a,*], Jia-Rong Zhan [b]

[a] Department of Computer Science and Information Engineering, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Road, Taipei, 106344, Taiwan
[b] Department of Electrical Engineering, National Chung Cheng University, 168 University Road, Min-Hsiung, Chiayi, 621301, Taiwan

ARTICLE INFO

Keywords:
Visual inertial odometry
Ultra-wideband
Unmanned aerial vehicle
Indoor localization

ABSTRACT

The localization of unmanned aerial vehicles (UAVs) in GPS-denied areas is an essential issue for indoor navigation. This paper presents a technique to improve the accuracy of visual inertial odometry (VIO) by combining the ultra-wideband (UWB) positioning technology. The proposed architecture is divided into two stages. In the initial stage, the constraint on UWB short-term position change is adopted to improve the pose estimation results of the VIO system. It is also used to mitigate the translation error caused by the vibration and lack of features during the flight. In the second stage, a loose coupling approach based on nonlinear optimization is utilized to fuse the local pose estimator of the VIO system with the global constraints from the UWB positioning. At the beginning of each operation, the alignment between the VIO and UWB frames is estimated to avoid the influence of coordinate transformation due to the VIO cumulative error. It is shown that our optimization-based fusion method is able to achieve a smooth localization trajectory under the global coordinate frame. In the experiments, the performance evaluation carried out in the real-world scenes has demonstrated the effectiveness of the proposed technique.

1. Introduction

In recent years, unmanned aerial vehicles (UAV) have been widely used in many applications. In particular, the development trend of lightweight UAVs has gained the attentions in a few different fields. These drones are commonly adopted for aerial photography and map construction to help people obtain the information about hard-to-reach places. In order for robots to locate their locations, the positioning technology is the key to have the automatic navigation capability. Depending on the types of sensors used by the robots and the interactions with the environments, the positioning techniques can be roughly divided into outdoor and indoor approaches. With the advances of GNSS (Global Navigation Satellite System) technologies, outdoor positioning services have become increasingly mature. RTK-GPS (Real-Time Kinematic) can even reach the accuracy of centimeter-level.

For the robot positioning in indoor environments, GNSS is not applicable since satellite signals cannot penetrate building materials. Moreover, it is more difficult to operate aerial drones autonomously indoors due to messy environments which might cause damages from possible collision. Hence, the positioning accuracy and stability are much more important than flying in the outdoor environment. In this regard, there exist many UAV positioning approaches utilizing different sensors and sensing technologies [1].

At present, the most popular positioning method for aerial drones is via simultaneous localization and mapping (SLAM). It aims to make the robot estimate its own position and posture by using the environmental features observed in motion. The map of the environment is simultaneously constructed based on the pose estimation results. Among the existing techniques, there are methods using lasers, cameras, IMUs, and sonars, etc. The disadvantages of these approaches are the computational difficulties due to limited features detected in the scenes. This might happen for the environment with a large and open space, insufficient light, or less texture. Since there does not have the global coordinate system as a reference, the accumulated error cannot be eliminated without the help of loop closure or other global optimization techniques. There also exist some methods based on signal processing, such as using UWB, GNSS, Wi-Fi, and Bluetooth. These techniques utilize their inherent global coordinate systems, so there will be no systematic cumulative errors. They are capable of providing good positioning effects for open space or the environments without sufficient features, but the results might be scattered due to the influence of noise and discontinuous measurements [2].

In this paper, we focus on the positioning for aerial drones in indoor environments such as cluttered factories or open spaces with

* Corresponding author at: Department of Computer Science and Information Engineering, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Road, Taipei, 106344, Taiwan.
E-mail address: lin@ntut.edu.tw (H.-Y. Lin).

https://doi.org/10.1016/j.measurement.2022.112256
Received 4 September 2022; Received in revised form 2 November 2022; Accepted 21 November 2022
Available online 26 November 2022

less prominent features. Considering the sensor size and cost, the VIO-SLAM approach using a monocular camera and an IMU (inertial measurement unit) is currently the most cost-efficient configuration for estimating the state of a robot. It also provides stable performance in positioning accuracy. However, with a long-term operation or in some environments containing weak features, the SLAM system will encounter the problem of accumulated errors inevitably. If the closed-loop detection is not applicable in the environment, the drifting due to the error accumulation will greatly affect the control of the UAV. In the case of a large vibration or unstable feature points, the SLAM system will also produce large drifting errors when estimating the robot pose, and therefore results in a decreasing positioning accuracy. For outdoor environments, common solutions mainly rely on GNSS assisted positioning to reduce errors [3] In indoor environments with limited GNSS signals, it is necessary to use more expensive motion capture systems to achieve high-precision positioning [4].

Recently, the low-cost UWB positioning technology is used as an alternative in GNSS-constrained environments [5]. In this paper, a VIO-SLAM framework is utilized for the initial pose estimation, and UWB modules are then integrated to improve the positioning accuracy. The short-term variation in the UWB is used to constrain the local estimates, and reduce the stability degradation due to the influence of motion vibration or feature mismatching. In our implementation, the UWB positioning is obtained by pre-arranging several indoor anchors and placing a tag module on the UAV. Based on the non-linear optimization, the VIO pose estimation is converted into an incremental form and integrated with the global UWB positioning information. The experiments demonstrate that significant improvements on drifting can be achieved for the long-term navigation. We have verified that the proposed UWB short-term variation constraint can greatly improve the stability of VIO-SLAM pose estimate. It also eliminates the common cumulative error of the SLAM system, and reduces the system dependence of VIO on closed-loop detection.

The main contributions of this paper are as follows:

- The proposed localization approach using the UWB short-term variation constraint effectively limits the inevitable drift errors of VIO-SLAM techniques, and improves the robustness of VIO systems.
- The optimization-based fusion method is used to fuse the global UWB estimates with local VIO-SLAM estimates. It achieves smooth positioning without cumulative errors under the global coordinate frame.
- The robustness of the proposed localization technique is demonstrated using the public EuRoC dataset and the images collected in a weakly-featured environment.

2. Related work

SLAM systems usually include multiple sensors and multi-functional modules, which are mainly distinguished according to the core functions. Current methods are commonly divided into laser or image based techniques. In the laser positioning, the transmitters such as LiDAR are usually used [6]. Although the laser-based SLAM provides high accuracy and is relatively stable, it is too heavy and bulky for the lightweight aerial drone applications. Consequently, the recent configurations for aerial drone positioning are mostly based on visual SLAM. It mainly utilizes vision sensors to perceive the surrounding environment and estimate the 3D position and orientation. Some commonly adopted sensing devices include monocular, binocular, RGB-D and omnidirectional cameras, etc. [7].

In the existing literature, there are many mature techniques for laser and vision-based SLAM. GMapping proposed by Grisetti et al. is one popular SLAM approach which utilizes laser scan data as inputs [8]. It was developed based on the Rao–Blackwellised particle filter. The information of the surrounding environment is constructed through the continuous movements and observations of the robot. The mapping and movement model from the previous moment are then used to predict the current robot pose. Recently, the Cartographer [9] proposed by Google adopts the mainstream SLAM framework, including the steps of feature extraction, closed-loop detection and back-end optimization. A certain number of laser points form a submap, and then a series of subgraphs are used to form a global map. In [10], Zhang et al. presented a learning-based VIO framework for image segmentation and tracking. With the tightly coupled feature observation and IMU pre-integration measurement, the system is capable of loop closure detection and long distance navigation.

For the use of 2D images and IMU, VINS-Mono [11] and ORB-SLAM3 [12] are the well-established SLAM techniques. They are implemented based on the mainstream SLAM architecture. However, VINS-Mono, which utilizes the optical flow as its front-end, has better robustness than ORB-SLAM3. The descriptor used as the front-end of ORB-SLAM3 is more likely to lose characteristic information during fast movements. Thus, for the application scenarios with flying drones, the positioning using VINS-Mono is more stable than ORB-SLAM3. In larger scenes or moving with slower motion, ORB-SLAM3 has some advantages over VINS-Mono because the loop detection with subgraphs is more appropriate to use. Since this work aims to develop an indoor positioning technique for drones, the VINS-Mono system with relatively stable in fast motion is adopted as the local pose estimator for subsequent data fusion.

Over the last decade, there have been many indoor positioning techniques based on signal processing, such as Bluetooth, Wi-Fi, and UWB, etc. Although the methods using Wi-Fi and Bluetooth are low-cost and easy to set up, they are susceptible to interference in the environment, and have limited accuracy with about one meter error. The UWB signal is different from the 2.4G frequency pulse waves used by Wi-Fi and Bluetooth. It is transmitted by very narrow pulse waves for the 3–10 GHz bandwidth. Based on the TOA (time of arrival) or TDOA (time difference of arrival) computation, the position of the tag can be derived from the distances to multiple known positions of the reference locations (anchors). Because UWB signals have high-speed transmission, high anti-interference, and low power consumption, the underlying technology is well-suited for the development of indoor positioning system.

Recently, many approaches based on UWB positioning have been investigated. Several methods are designed to operate in the locations without GNSS signals. In [13], Bottigliero et al. presented a low-cost real-time locating system (RTLS) for the indoor environments. It was based on the one-way ranging and TDOA of non-synchronized UWB pulse sequences. Liu et al. proposed a framework for cooperative positioning based on the fusion of peer-to-peer UWB and Wi-Fi measurements with the mobile user's inertial information [14]. The localization of multiple users was achieved by combining long-range Wi-FI and short-range UWB positioning, and followed by the refinement with IMU-based dead reckoning. Different from the general UWB positioning methods, Queralta et al. believed that TDOA-based positioning would limit the mobility of the entire system [15]. Thus, they proposed to use only ToF distance estimates for positioning, and then analyze the self-correction of anchor positions for the mobile robots. How the positioning accuracy was affected by the speed, height and position of the anchor is also studied. Alternatively, Nguyen et al. investigated the issues related to the collaborative process among the drones and some other platforms [16], while Macoir et al. presented a method to navigate aerial drones in cluttered warehouses [17]. In more recent studies, the UWB localization accuracy affected by the obstacles in the signal propagation path was investigated [18]. A composite filtering method was proposed to deal with the dynamic uncertainty of quadrotor UAVs [19].

The multi-sensor fusion algorithms are mainly divided into the optimization-based approaches [20,21], and the filtering-based approaches [22–24]. According to the types of sensors used in the system,

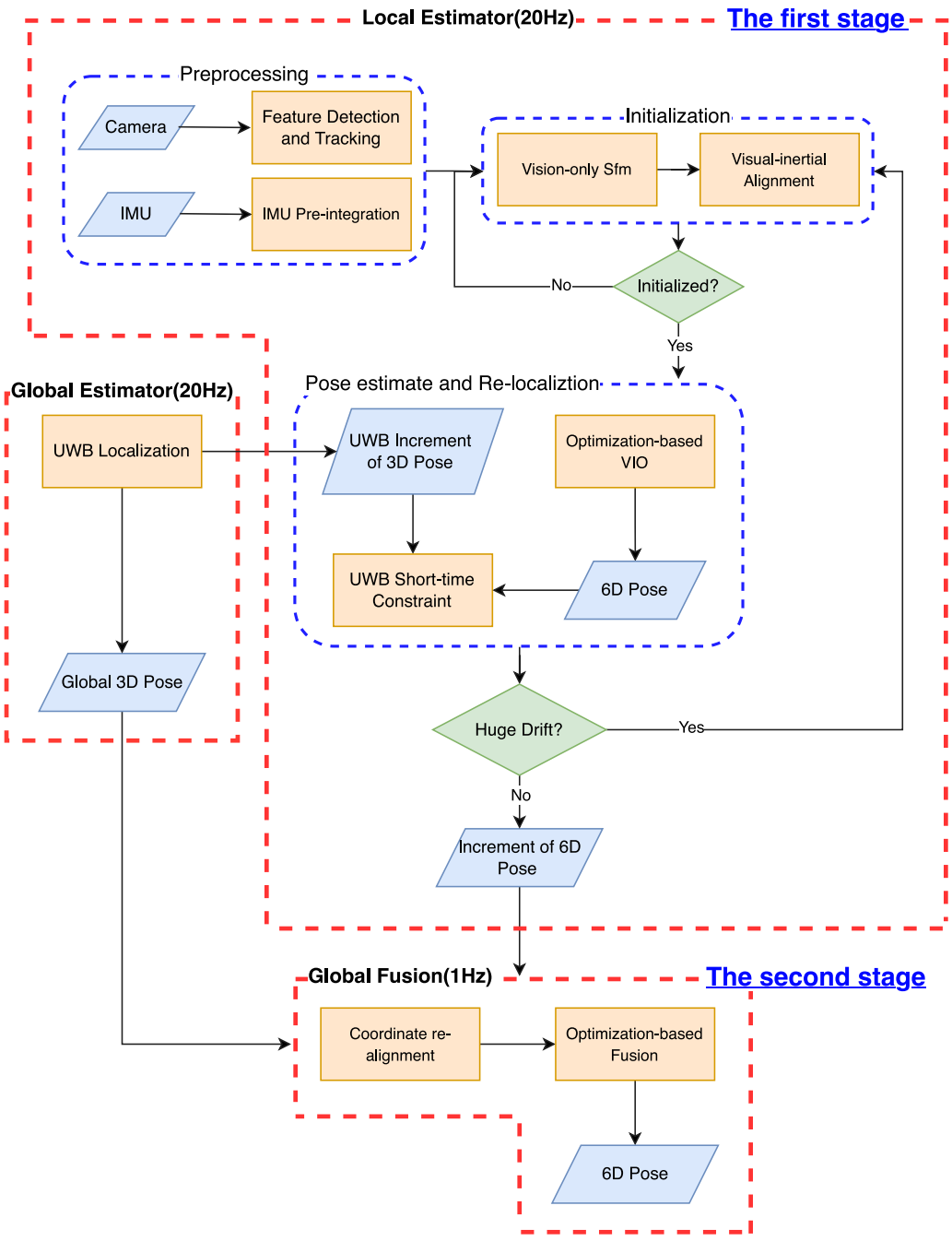


Fig. 1. The overall system architecture of the proposed technique consists of three modules: global estimation (UWB), local estimation (VIO), and global fusion.

the positioning techniques can be performed locally or globally by considering the broadness of the input source. The conventional and RGB-D cameras, IMU, LiDAR are commonly adopted in local positioning for 6 DoF state estimation, while the global positioning is usually carried out using GNSS. In general, optimization-based approaches have obvious advantages over filtering-based methods. The filtering-based such as the traditional EKF update step usually performs only a single cycle, while in an optimization-based architecture such as bundle adjustment [25] there are many visual and IMU measurement values saved in a vector. The states related to the observation are optimized through multiple iterations, and the accuracy is greatly improved. In terms of robustness, the most significant issue of the filtering-based fusion methods is time synchronization. Thus, a sorting mechanism is used to ensure the order for all measurements consistent in time. On the other hand, optimization-based methods can wait for a long time and store measurement values. The conversion is optimized for the local and global coordinates, and a higher stability is obtained after fusion.

Both the filter-based and optimization-based fusion methods can achieve highly accurate state estimation. However, due to the lack of global measurement, the cumulative drift over time is inevitable for both cases. Thus, it is necessary to incorporate the global information to mitigate the drift from accumulated errors. The global sensor measurement is absolute with respect to the world coordinate system and irrelevant to initial position estimates. However, the signals from global sensing are usually with low-frequency and noisy, and not suitable for applications demanding high accuracy. In [26,27], the fusion of global and local sensor data is used to achieve an accurate positioning in a world coordinate system. The former is based on the EKF framework by combining visual and IMU data with the GNSS information, and the latter is based on non-linear optimization to fuse local pose estimates of

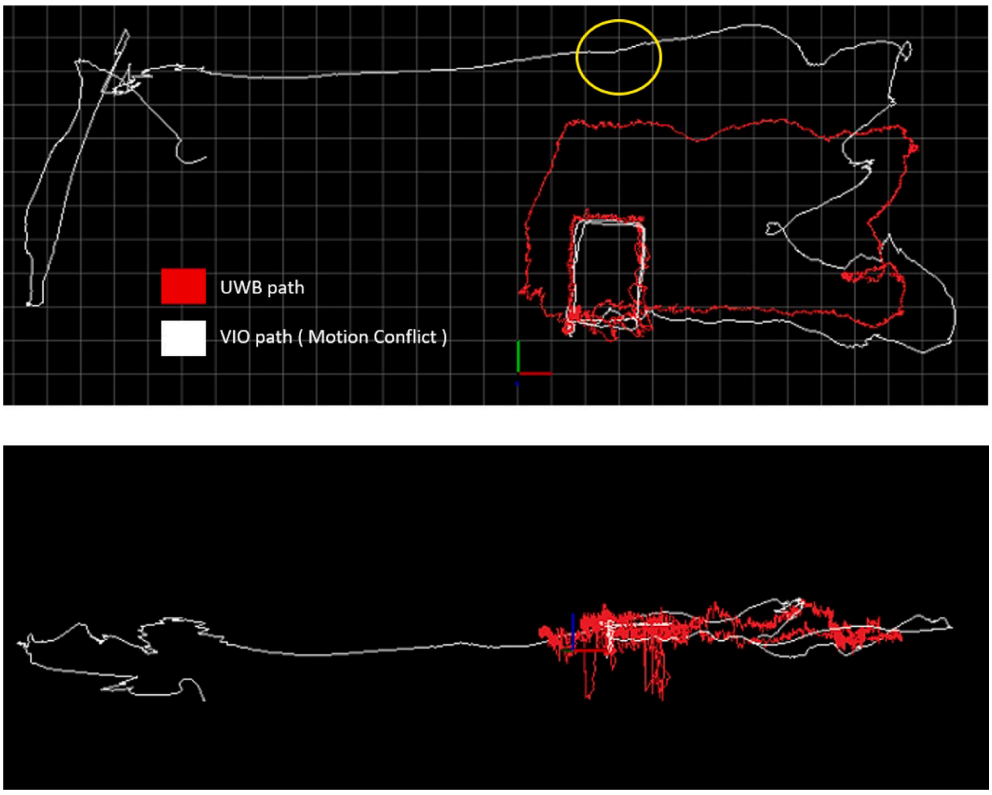


Fig. 2. The positioning drift from VINS-Mono. The yellow circle indicates where a large odometry drift occurs, the red curve is the UWB path, and white curve is the VINS-Mono path.

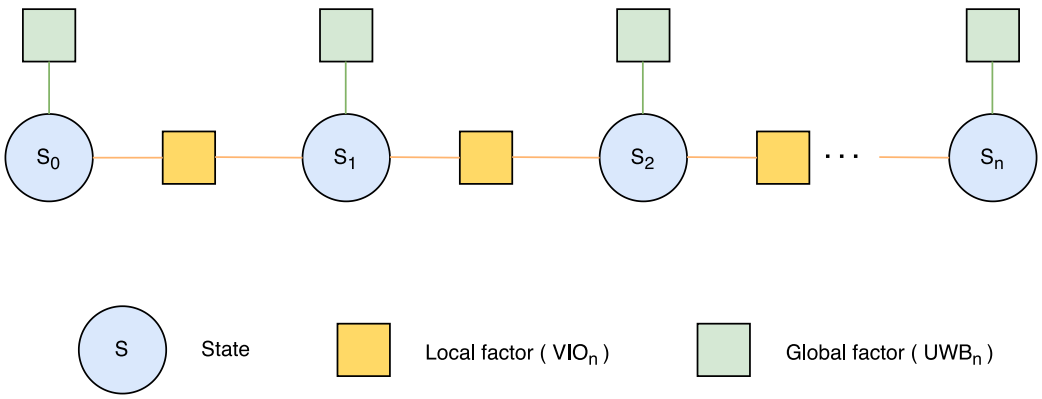


Fig. 3. The pose graph structure with optimization. The node (State) is the pose in the world coordinate system, the edge (VIO) is the local constraint of VINS-Mono pose estimation, and the edge (UWB) is the global constraint generated by the UWB global positioning. After the positioning results of VINS-Mono and UWB are aligned, the associated position changes are minimized.

VIO with the GNSS information. Both methods provide the global state estimation of cumulative drift and local accuracy through data fusion.

To integrate UWB positioning with the SLAM system, Song et al. proposed an EKF-based method [28]. The lidar data were used to improve the accuracy of UWB positioning, while the constraint posed by UWB was used to eliminate the cumulative error of laser SLAM. In [29], Gao et al. presented a low drift VIO technique by combining the UWB positioning for indoor localization. A rudimentary cost function was formulated based on the output of VIO and UWB positioning modules, followed by a nonlinear optimization procedure to refine the location measurement. Alternatively, a cost function for the integration of VIO and UWB was designed using the mutual information from the UWB [30]. By applying the UWB constraints and information residual to VIO computation, the localization accuracy was improved for the environment with few feature points. More recently, Liu et al. proposed a mobile robot localization and map building technique based on the fusion of UWB, odometry and 2-D LiDAR data [31]. Similar to our proposed system, a graph optimization was used to remove the cumulative error and derive the initial trajectory. Nevertheless, their approach focused on the map construction, and the loop-closure module was an essential part of the technique.

Different from the commonly adopted loose coupling approaches for data fusion-based localization techniques, Nguyen et al. proposed a tightly-coupled method utilizing image, IMU and UWB inputs [32]. A range-focused methodology was considered to leverage the readily available VIO localization and time-delayed UWB positioning. In the approach presented by Magnago et al. for ground mobile robots, the relative control input and the global information obtained from UWB positioning were fused through an unscented Kalman filter to mitigate the cumulative error [33]. Perez-Grau et al. used the Monte Carlo

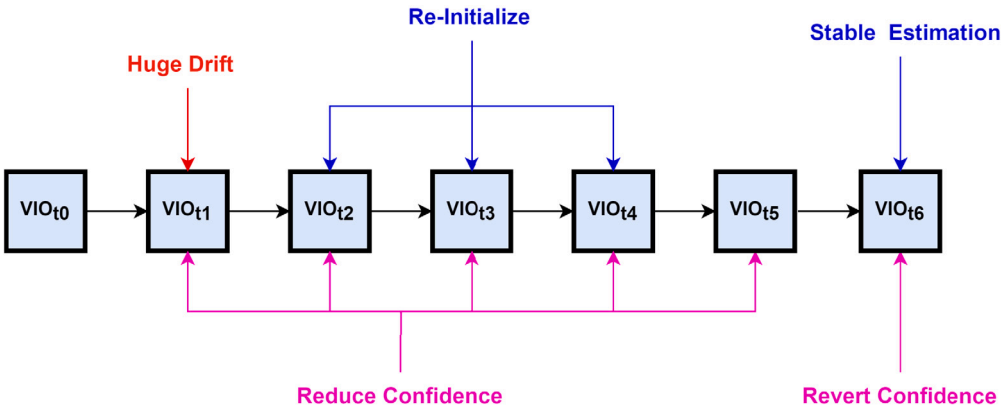
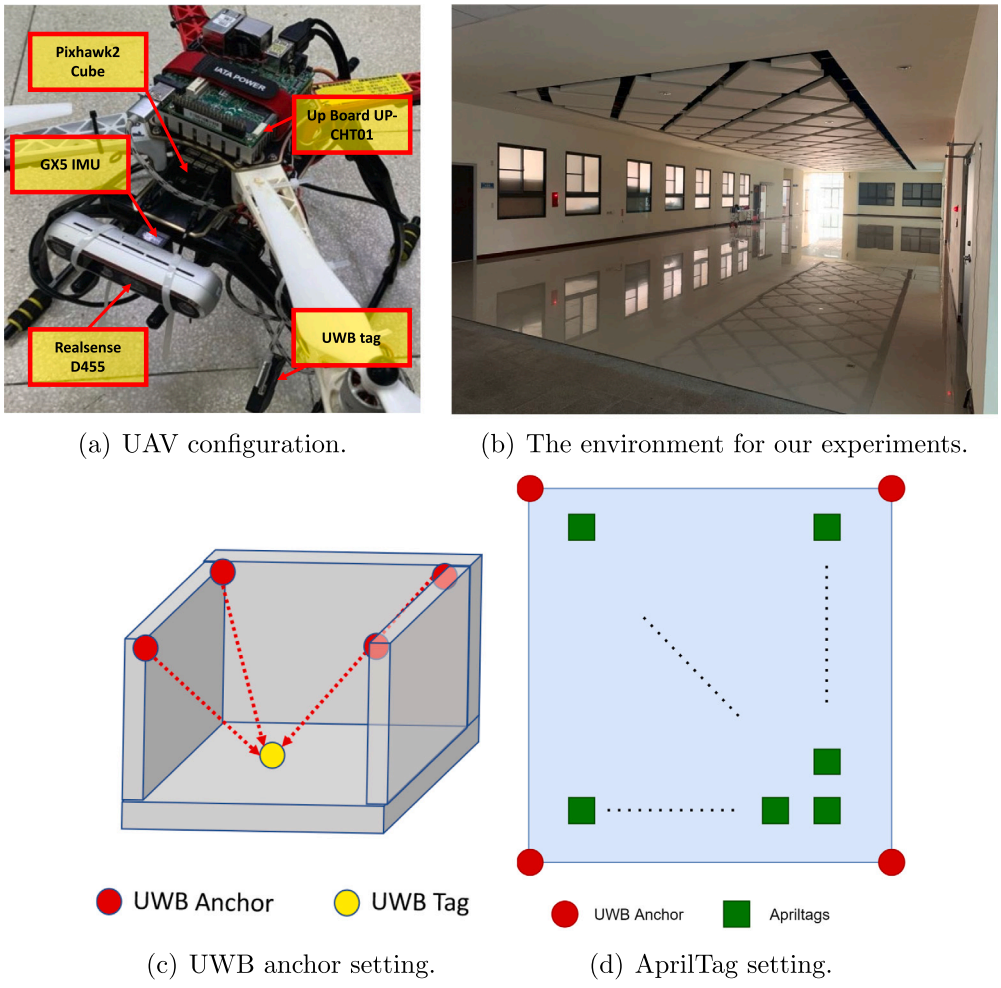


Fig. 4. The confidence adjustment of VIO when the positioning drift occurs. When encountered large errors due to the positioning drift, the VIO pose confidence for initialization will be reduced. The confidence will be reverted after the initialization process is completed.



(a) UAV configuration. (b) The environment for our experiments.

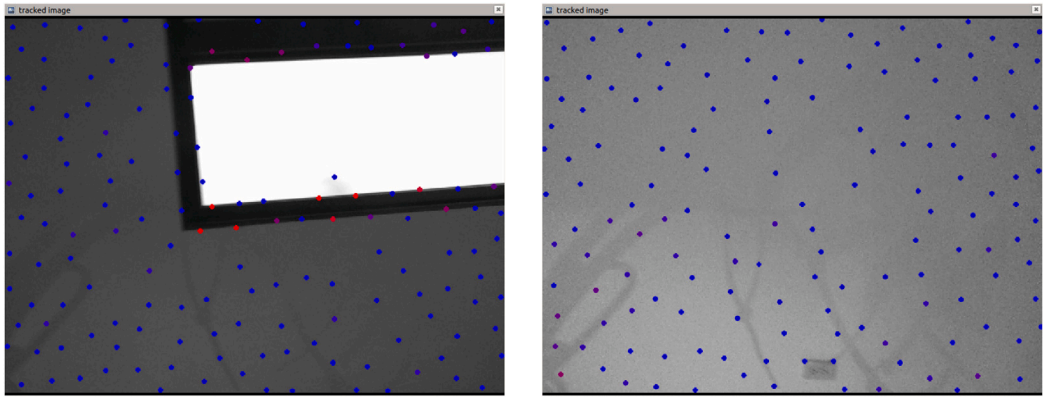
(c) UWB anchor setting. (d) AprilTag setting.

Fig. 5. The environment and settings for our indoor positioning experiments using an UAV, UWB devices and Apriltags.

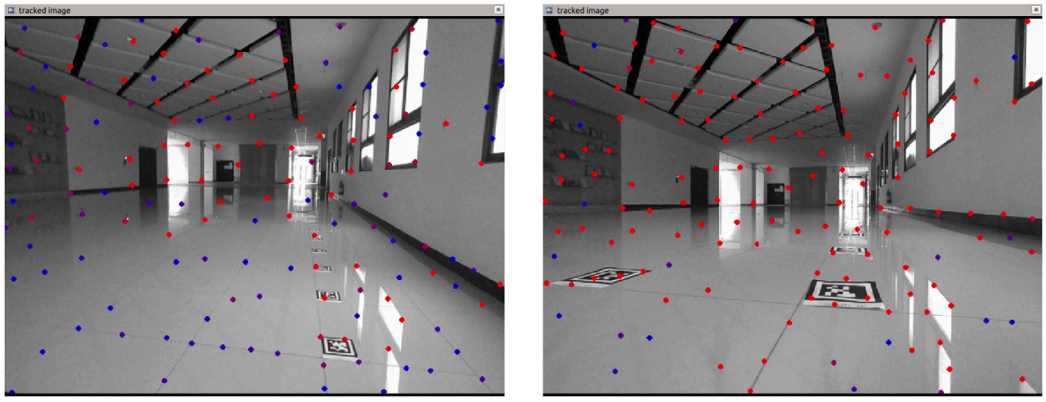
Localization (MCL) method to integrate the point cloud data obtained by an RGB-D sensor and the UWB positioning information through the IMU [34]. Wang et al. proposed a technique to replace the loop detection function obtained from multiple pre-configured UWB anchor directly [35]. Because the loop detection step was omitted, it did not only reduce the computational complexity, but also prevented the error from accumulation. In [36], Cao and Beltrame also utilized the UWB information to eliminate the cumulative errors. Different from the previous work [35], they did not omit the loop detection module. A single anchor which was not assigned to a location for positioning was adopted instead. The system mobility was also improved significantly.

3. Proposed approach

In our proposed technique, the UWB positioning constraints are applied to VINS-Mono to reduce the impact from the errors caused by the vibration or feature point instabilities. It is then followed by the VINS-Fusion method to merge the improved UWB and VIO positioning.



(a) Weak feature environment simulation



(b) Robust feature environment simulation

Fig. 6. The simulation of the visual input in a low-texture environment with weak features to verify the robustness of our proposed UWB constrained VIO technique. The red dots represent the detected stable feature points, and the blue dots are the unstable features.

The overall system architecture is illustrated in Fig. 1. Considering the expandability of the system, the loose coupling approach is adopted in this work. The global measurement is mainly derived through the pre-configured UWB anchors, and the constrained VINS-Mono positioning is used to derive the local estimates. These two positioning results are then integrated using a non-linear optimization method.

First, the same initialization as VINS-Mono is used for the front-end of the local pose estimation module. In the data pre-processing stage, the data from camera and IMU are tracked and pre-integrated. SFM (structure from motion) is then used to estimate the initial motion state of the camera, followed by the alignment with the IMU pre-integration result. We add the UWB short-term constraint obtained by the global estimation to the back-end module and determine whether abnormal shifts have occurred. If an abnormal shift is detected, a separate local estimation will be re-initialized to maintain the accuracy of the local pose estimates. In the global fusion module, we utilize the VINS-Fusion-based architecture to construct the improved VINS-Mono positioning result as a local measurement value, and take UWB as a global observation value. The information fusion is then performed with a nonlinear optimization method. At the beginning of each optimization, VINS-Mono is aligned to the UWB coordinate system to avoid excessive accumulated errors.

3.1. UWB localization

For UWB positioning, we use the TDOA (Time-Difference-of-Arrival) ranging approach. By measuring the difference of the flight times between the tag and two anchors, the distance difference can be obtained. According to the formulation, the time difference between the tag sending signal to the anchors is constant. Thus, the position of the tag is on the hyperbola with these two points as the foci. If there exist four positioning base stations, then the single intersection of the four hyperbolas is the location of the label. That is,

$$\begin{aligned} D_{ij} &= (t_i - t_j) \times c \\ &= \sqrt{(x - X_i)^2 + (y - Y_i)^2 + (z - Z_i)^2} \\ &\quad - \sqrt{(x - X_j)^2 + (y - Y_j)^2 + (z - Z_j)^2} \end{aligned} \tag{1}$$

where (X_i, Y_i, Z_i) is the coordinate of the ith known anchor, (x, y, z) is the coordinate of the tag, t_i is the arrival time of the signal from the tag to the ith anchor, c is the light speed, and $(t_i - t_j)$ is the arrival time difference. In the 3D space, at least four anchors are required to derive the position of the tag. It is given by of the solution of these equations.

The placement and settings of anchors for UWB positioning vary for different environments. If the anchors are installed at the same height, it will result in larger z-axis errors [37]. Although the accuracy in the z-axis is improved if the anchors are placed with different heights, the accuracy and stability of overall static and dynamic positioning capability will decrease. Moreover, considering the messy indoor environment, this will also increase the probability of occlusion and cause non-line-of-sight errors (NLOS). In this paper, four UWB anchors are installed at the same height, and their positioning confidences along the z-axis are reduced in the fusion algorithm. The UWB covariance matrix is given by

$$Q_{UWB} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \tau \end{bmatrix} \tag{2}$$

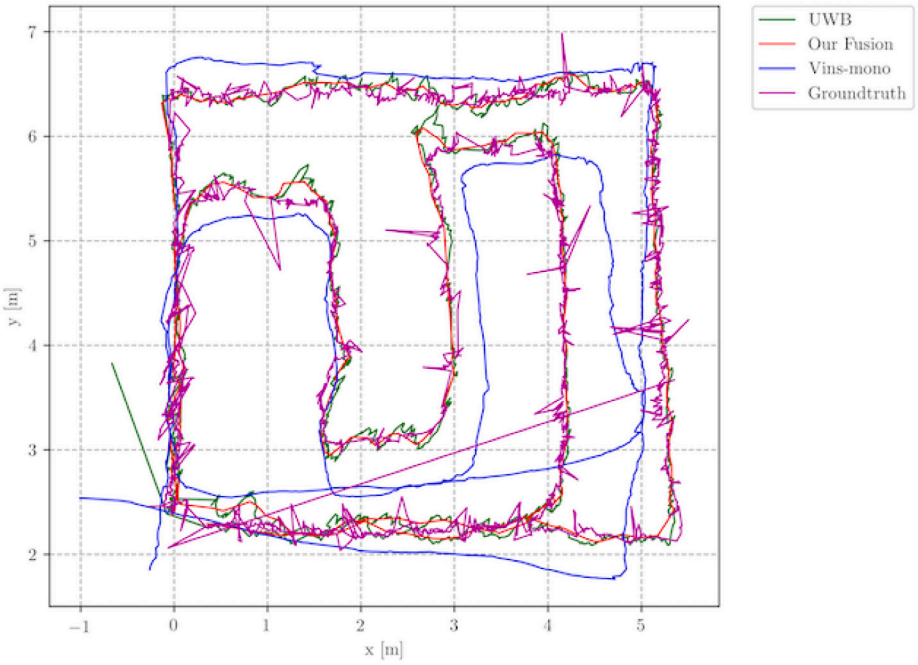


Fig. 7. The positioning results of VIO, UWB, ground truth, and using our fusion method in the (x, y) plane.

where τ represents the disturbance in the z-axis and is set as 10 in the implementation. It is then followed by using the local estimates derived from VIO to compensate the z-axis error of UWB positioning.

3.2. Visual-inertial odometry

We adopt VINS-Mono for local pose estimation. The overall architecture is divided into front-end preprocessing, initialization, tightly coupled monocular VIO, relocation, and the global modules such as pose graph optimization. Since our objective is to reduce the dependence of the VIO system on closed-loop detection and eliminate the accumulated errors, only the non-relocation result is used as a local estimate. It is then converted into an incremental form for fusion with the UWB positioning results.

In the pre-processing stage, we first use a feature detector is used to extract the corner points, followed by the KLT optical flow algorithm for tracking [38]. The IMU pre-integration is then calculated between the two consecutive frames [39]. Since IMU provides acceleration and angular velocity, the objective of IMU integration is to obtain the pose transformation during a period of time by integrating the IMU measurement values. In the initialization stage, a good initial value is needed to fuse the two measurement results because the monocular camera is not able to directly observe the scale. A loosely coupled fusion method is adopted by VINS-Mono to align these two types of sensor data to derive the initial values required by the tightly coupled monocular VIO. After estimating the initial camera motion state, the IMU pre-integration results are aligned using SFM to restore the scale, gravity, velocity and gyroscope zero offset of the IMU.

In our back-end optimization, VINS-Mono is mainly performed based on the sliding window method in tightly coupled manner. By constructing the residual term based on the visual and IMU measurements into a problem of solving maximum likelihood estimation (MLE), the optimal solution of the entire optimization can be regarded as accurate state estimation. The state to be estimated is defined by

$$\begin{aligned} X_l &= [x_0, x_1, \dots x_n, \lambda_0, \lambda_1, \dots \lambda_n] \\ x_k &= [p_k^l, v_k^l, q_k^l, b_a, b_g], \quad k \in [0, n] \end{aligned} \tag{3}$$

It contains all of the camera states in the sliding window. The parameter x_k is the kth frame IMU state, which includes the position p_k of the IMU center relative to the camera coordinate system, velocity v_k, direction q_k, accelerometer offset b_a, and gyroscope offset b_g. The subscript n is the total number of feature points in the sliding window, and λ is the inverse depth from the IMU pose observation value of the first frame to the lth feature.

The entire maximum likelihood estimation can be expressed as a nonlinear least squares problem

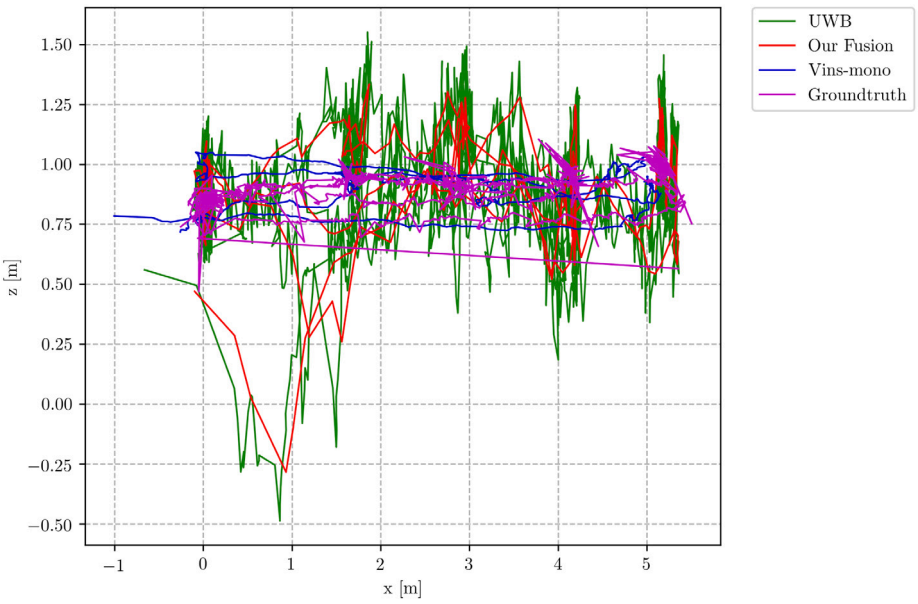
$$\begin{aligned} \min_X \{ \|r_p - H_p X\|^2 + \sum_{k\in\beta} \|r_\beta(\hat{z}_{b_{k+1}}^{b_k}, X)\|^2_{P_{b_{k+1}}^{b_k}} \\ + \sum_{(l,j)\in C} \rho(\|r_c(\hat{z}_l^{c_j}, X)\|^2_{P_l^{c_j}}) \} \end{aligned} \tag{4}$$

and solved using Ceres Solver [40]. In the first term, r_p and H_p are the prior information associated with the last sliding window marginalization. The second term $r_\beta(\hat{z}_{b_{k+1}}^{b_k}, X)$ represents the IMU residual. In the third term, $r_c(\hat{z}_l^{c_j}, X)$ denotes the residual of visual measurements, and ρ is the Huber norm [41]. The obtained solution is then used as the input for the subsequent fusion local pose estimation.

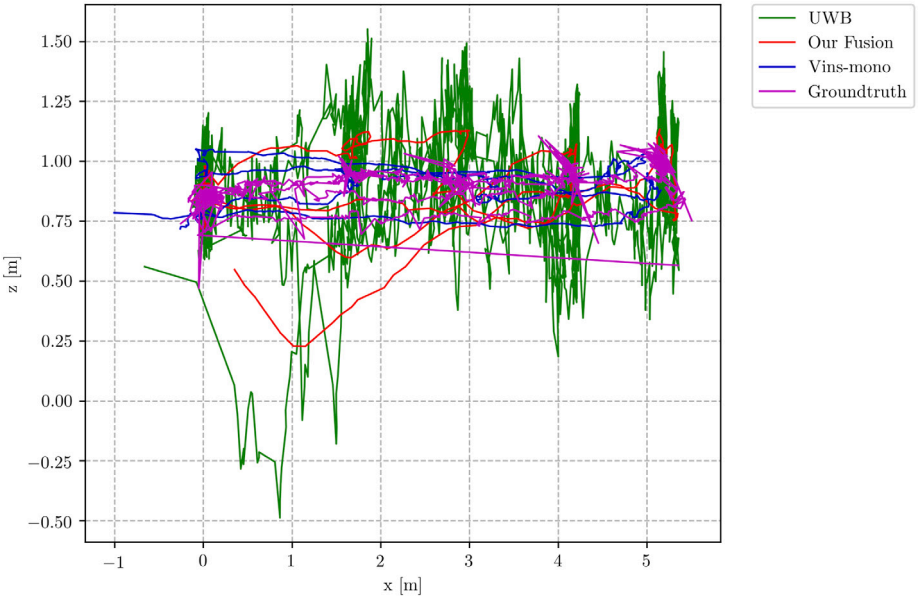
Although the state estimation constructed from VINS-Mono has high accuracy in a static environment with stable features. According to the experiment, the state estimates derived from VINS-Mono have huge deviations when the vibration is large or the feature points are weak. The shift will seriously reduce the accuracy of the entire partial estimation values. Therefore, we introduce UWB short-term variation constraints to improve the VINS-Mono state estimation, and obtain more accurate and stable local pose estimates.

3.3. UWB constraints VIO

For the VIO positioning system adopted in practical applications, in addition to the general sensor noise, the cumulative error source will also cause serious positioning drift problems when the instantaneous acceleration of the IMU is large and the feature points are unstable. A typical example is illustrated in Fig. 2, with the yellow circle shows where the positioning drift occurs. Because the huge offset error in this situation will greatly affect the result of the subsequent fusion, we apply the short-term variation of the UWB positioning result to constrain the VIO system. When the VIO short-term positioning change is greater than the UWB positioning constraints, we perform the pose



(a) Before UWB marginalization.



(b) After UWB marginalization.

Fig. 8. The improvement on cumulative errors in the z direction.

re-estimation for the entire VIO, and the pose before the positioning drift is retained. After the VIO re-initialization is completed, the estimation resumes from the pose before the positioning drift. This process is regarded as the relocation.

The equations

$$\begin{aligned} UWB &= (x_1, y_1, z_1) \\ VIO &= (x_2, y_2, z_2) \end{aligned} \tag{5}$$

and

$$\begin{aligned} UWB_{diff} &= \left\| UWB_t - UWB_{t+s} \right\|^2 \\ VIO_{diff} &= \left\| VIO_t - VIO_{t+s} \right\|^2 \end{aligned} \tag{6}$$

define the 3-D positions of UWB and VIO systems, as well as the associated short-term changes. In Eq. (6), UWB_{diff} and VIO_{diff} are the changes of the two positioning systems over time derived from the Euclidean distance of the 3-vectors, s is an adjustable variable decided by the feature point robustness and the IMU bias. s will increase if the feature point is weak or the IMU bias is large, otherwise it will decrease.

Algorithm 1: UWB Odometry Constraint

```
th ← threshold;
VIO_state ← (x1, y1, z1);
UWB_state ← (x2, y2, z2);
if VIO_diff > UWB_diff + th then
    VIO_state ReInitialize;
end
```

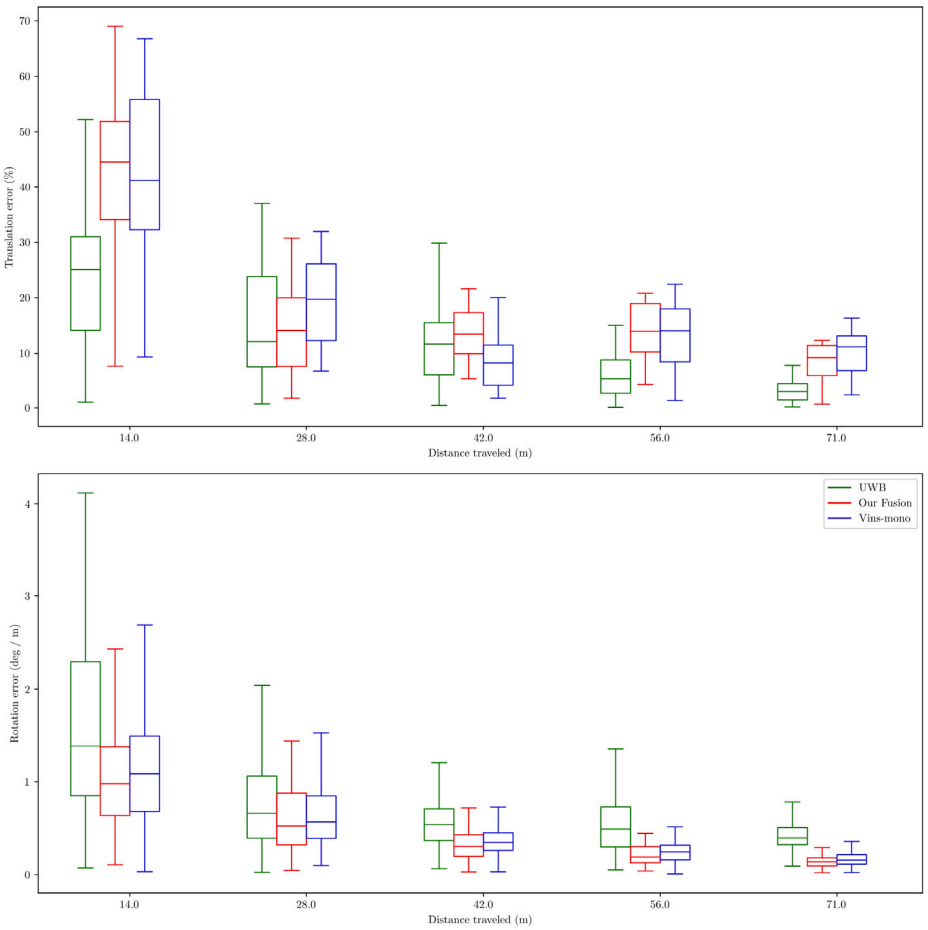


Fig. 9. The relative errors of our fusion result with z-axis marginalization. Top: the translation error. Bottom: the rotation error.

Algorithm 1 (UWB Odometry Constraint) is the process of using UWB to constrain VIO, where *threshold* is a custom parameter to avoid the influence of extreme values for UWB positioning. It is used to detect whether VIO possesses a huge deviation. This value is determined based on the positioning quality of UWB, the amount of image features, and the extent of vibration of the UAV. After obtaining the positioning results and short-term variation between them, the VIO positioning is verified. If it is greater than the sum of the short-term variation of UWB and the threshold, the VIO system is re-initialized.

3.4. UWB and VIO fusion

Accurate state estimation is an essential problem for mobile robots. To achieve locally accurate and globally unbiased state estimation, multiple sensors with complementary characteristics are usually adopted to derive the robot pose. Local sensors (such as camera, IMU, LiDAR, etc.) can provide accurate pose estimates in a small range, while global sensors (such as GNSS, magnetometer, barometer, etc.) can provide noisy but globally unbiased positioning in a large environment. The alignments on the coordinates is an important step in the fusion algorithm. The extent of alignment will directly affect the accuracy of the subsequent fusion.

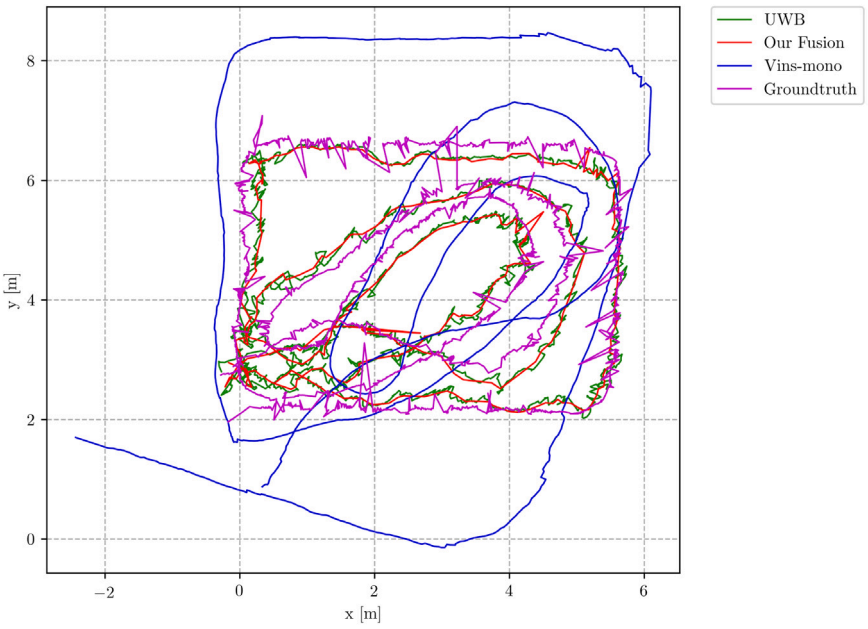
Current fusion strategies are mainly divided into local and global time alignment. We utilize the global alignment method in our fusion algorithm. Before each optimization process, the conversion matrix between the local coordinate system (VIO) and the global coordinate system (UWB) is estimated and used for alignment. This approach performs the alignment for each optimization instead of only taking the partial time alignment. It can greatly enhance the robustness of the system, and avoid the coordinate shift during the fusion caused by the VIO when excessive errors occur.

We refer to the architecture of VINS-Fusion [27], and utilize the nonlinear optimization to fuse two independent positioning results. As shown in Fig. 3, a sensor fusion framework based on an existing VIO algorithm is used to derive the local pose. It is then combined with the global sensing data as the input to derive the global pose map. They are transformed to uniform factors to construct the optimization problem, and the global estimator will obtain locally accurate and globally unbiased 6 DoF pose results.

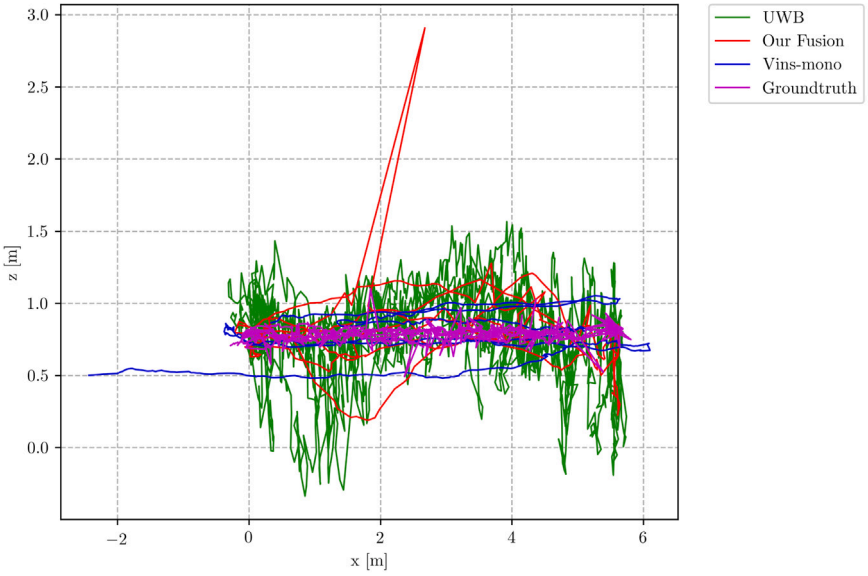
In terms of local factor, the relative poses of the two streams output by VINS-Mono are used to construct the residual term. Likewise, the global factor is constructed by the UWB positioning result. It is obtained by taking the difference of the UWB positioning result and the origin of the global coordinates at the same time. Since UWB is used as the global sensor in this work, we only need to pre-set the UWB anchor positions, and use the TDOA algorithm to calculate the position of the UWB tag. The local and global factors are constructed by the equations

$$\begin{aligned} VIO_{tn} &= (\chi, \omega) \\ UWB_{tn} &= \chi \\ VIO_{diff} &= VIO_{ti} - VIO_{tj} \\ UWB_{diff} &= UWB_{ti} - UWB_{tj} \\ \min &\|VIO_{diff} - UWB_{diff}\|^2_Q \end{aligned} \tag{7}$$

where χ represents the 3D position in the global coordinate frame, ω denotes the camera pose (roll, pitch, yaw) expressed as a quaternion,



(a) The trajectory in (x, y) plane.



(b) The trajectory in z-axis.

Fig. 10. The improvement on cumulative errors for the environment mainly containing only weak features. Note that the big jump in Fig. 10(b) was due to the hardware computation capability issue. It caused a delay in the alignment of axes and produced a large drift on VIO.

VIO_{in} is the state of VIO at the time instant i, VIO_{diff} denotes the relative pose difference between VINS-Mono at two time instants, UWB_{diff} is the absolute position difference at the same time instants, and Q is the covariance matrix that determines the confidence of the data. The purpose is to align the positioning results from VINS-Mono and UWB, and then minimize the associated position difference so as to eliminate the cumulative error of VINS-Mono.

The overall fusion process is shown in Algorithm 2 (UWB VIO Fusion). Since VIO is not able to output the pose during the initialization, it will take some time after the initialization is completed for the entire positioning result to stabilize. Thus, it is necessary to reduce the confidence on VIO pose estimates during the initialization period before the stabilization. Fig. 4 illustrates the process of the confidence level adjustment. Since we set UWB anchors at the same height and the positioning stability of the z-axis is low, it is particularly required to reduce the confidence along the UWB z-axis positioning in the overall fusion algorithm.

4. Experiments

Currently, there are no public datasets for UWB and visual SLAM positioning to provide the accuracy comparison. Thus, we adopt the EuRoC MAV dataset [45] to compare our method with other state-of-the-art SLAM techniques. Due to the lack of UWB data in EuRoC, the noise with a Gaussian distribution is added to the ground truth to generate the UWB localization input. The noise level is set as ± 10 cm in the x and y directions and 50 cm in z-axis. The performance evaluation in terms of ATE (absolute trajectory error) is shown in Table 1 for

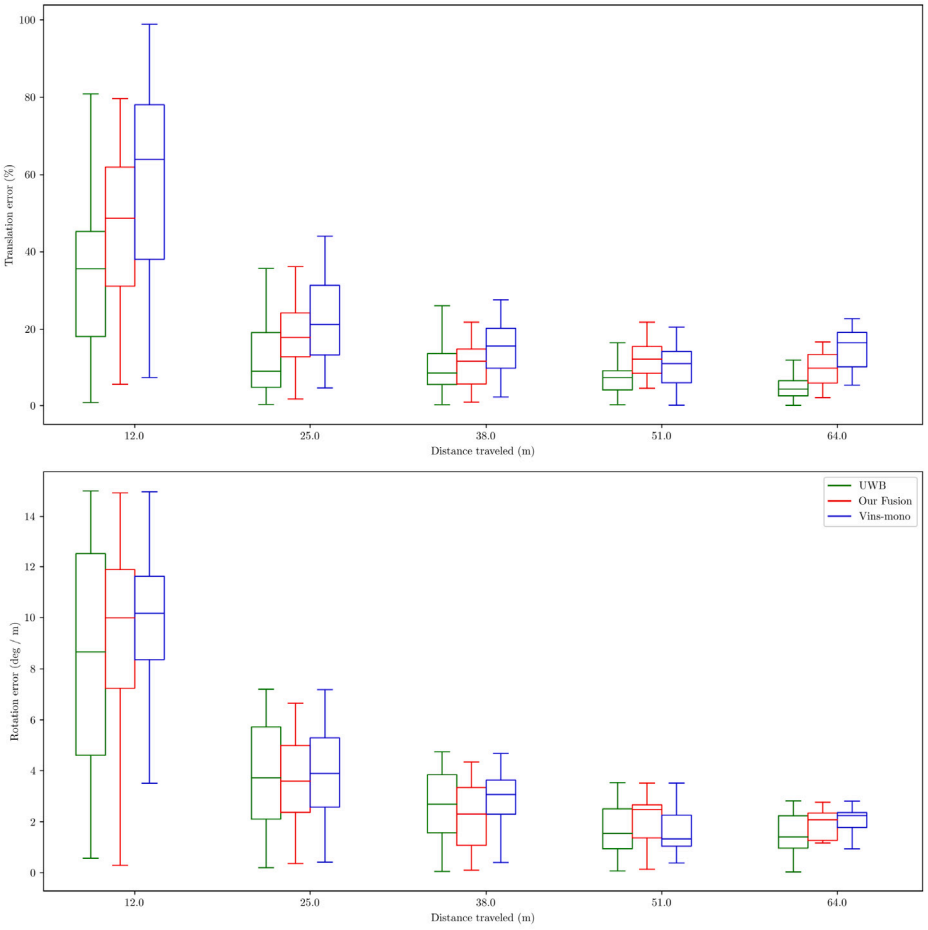


Fig. 11. The comparison of relative pose errors derived in the weak feature environment. It shows the stability of our technique compared to UWB and VINS-Mono.

Algorithm 2: UWB VIO Fusion

$VIO_{state} \leftarrow (x, y, z)$;
$UWB_{state} \leftarrow (x, y, z)$;
if $Huge\ Drift$ **then**
  $VIO_{state}(Confidence)Reduce$;
end
if $VIO_{state}EstimationStable$ **then**
  $VIO_{state}(Confidence)Revert$;
end
$AlignCoordinate(VIO_{state}, UWB_{state})$;
$Fusion(VIO_{state}, UWB_{state})$;

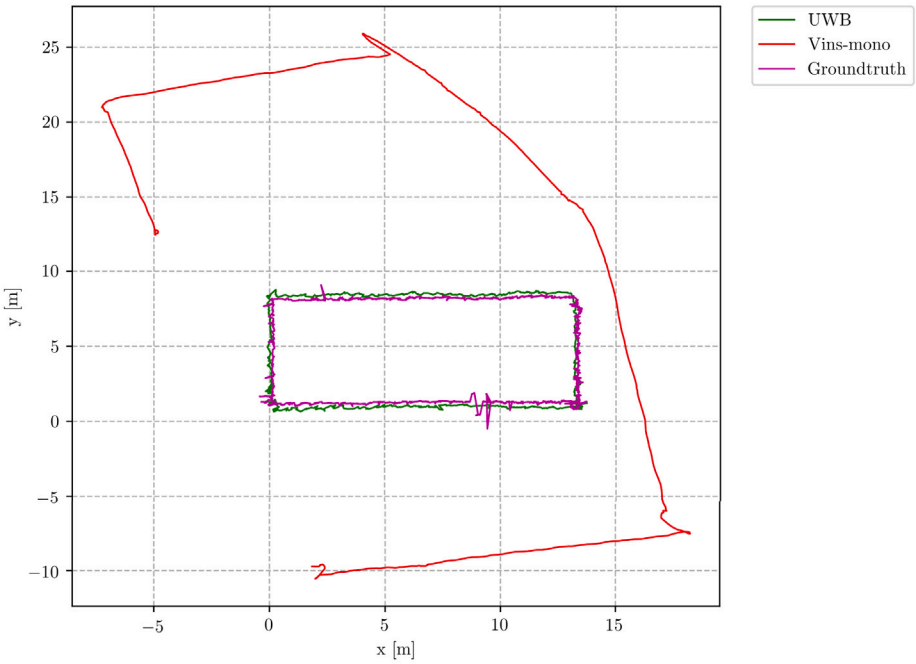
Table 1
The performance evaluation using the EuRoC dataset in ATE (absolute trajectory error in *m*). Here we show the results reported by the authors of each system.

Method	MH_01	MH_02	MH_03	MH_04	MH_05	Avg.
MCSKF [42]	0.420	0.450	0.230	0.370	0.480	0.390
OKVIS [43]	0.160	0.220	0.240	0.340	0.470	0.286
ROVIO [44]	0.210	0.250	0.250	0.490	0.520	0.344
ORB-SLAM3 [12]	*0.062*	*0.037*	*0.046*	*0.075*	*0.057*	*0.055*
VINS-Mono [11]	0.084	*0.105*	0.074	*0.122*	0.147	*0.106*
VINS-Mono-NL[a]	0.239	0.214	0.280	0.366	0.405	0.300
Ours	**0.063**	**0.061**	**0.122**	**0.213**	**0.185**	**0.128**

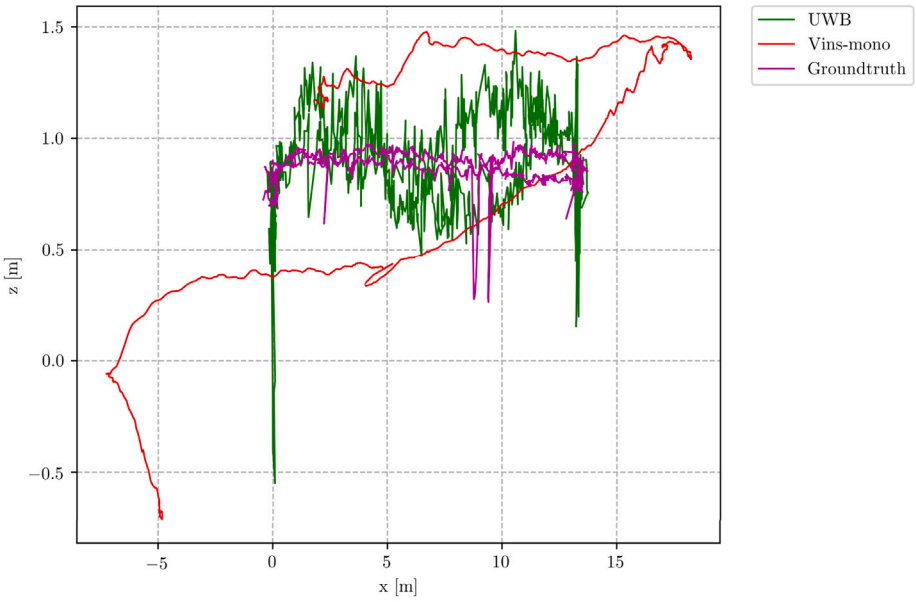
[a]VINS-Mono-NL : VINS-Mono without the loop closure.

different techniques, including MCSKF [42], OKVIS [43], ROVIO [44], ORB-SLAM3 [12], VINS-Mono [11], and our approach. It is tested using the image sequences MH_01 to MH_05 recorded in a factory environment. As shown in the table, although our method does not provide the most accurate results, we are able to eliminate the accumulated error without relying on the loop closure. Except for ORB-SLAM3 and VINS-Mono with loop closure enabled, the proposed technique outperforms the rest of localization algorithms. We also compare with the accuracy of VINS-Mono without the loop closure (VINS-Mono-NL) to demonstrate the effect of cumulative errors. The result indicates that more accurate trajectories in all cases can be obtained by our method. Since the collision of UAVs will cause serious consequences, the real-time precise positioning is required. In many practical scenarios, it is generally necessary to consider the environment where the loop closure cannot be adopted.

The evaluation using the EuRoC dataset is carried out with the images captured in a cluttered factory environment full of detectable features. To verify the robustness of our localization technique, our experiments are conducted in a weakly-featured environment. As illustrated in Fig. 5(a), a UAV equipped with a camera (Intel Realsense D455) and an IMU (LORD Sensing 3DM-GX5-25) is used in our experiments. The UWB modules are LinkTrack S launched by Nooploop, and the UP Squared development version is used to collect the data. In the experiments, four UWB anchors are placed at preset locations to derive the 3D coordinates of the UAV. For the drone control, Pixhawk2 Cube with an STM32 processor and R9DS receiver is adopted. An indoor environment with the dimension about $20 \times 10 \times 5$ m^3 as depicted in Fig. 5(b) is used for the experiments. To mitigate the error induced by occlusions, four UWB anchors are placed at the same height as



(a) The trajectory in (x, y) plane.



(b) The trajectory in z-axis.

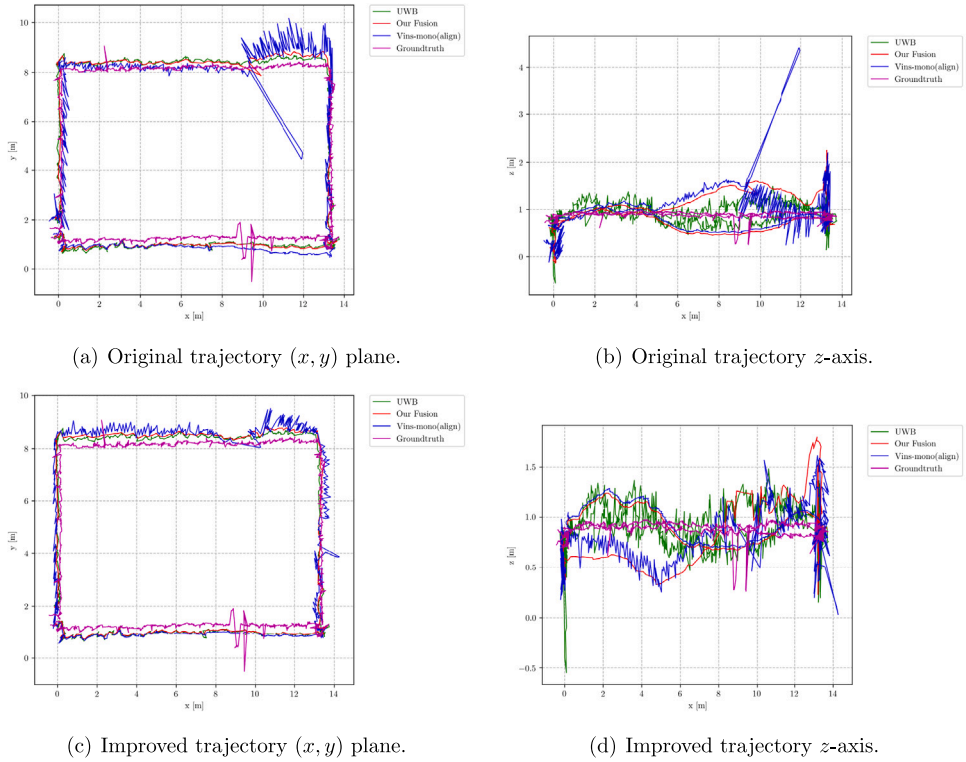
Fig. 12. The VIO error caused by the positioning drift. The ATEs for slight and large drifts are derived as 11.874 m and 170.075 m, respectively.

illustrated in Fig. 5(c). In the proposed loose coupling system, the time synchronization is accomplished by providing a threshold for VIO and UWB positioning. The results will be synchronized if the difference is within the threshold.

We use AprilTag to collect the positioning data for benchmarking [46]. The evaluation is carried out using the method presented by Zhang and Scaramuzza [47]. We evaluate the accuracy in the testing environment, and the errors in the x, y and z axes are reported as 0.031 m, 0.037 m and 0.051 m, respectively. Since the error is less than 6 cm, it is considered as a negligible amount compared to the SLAM systems carried out in a large environment. Fig. 5(d) illustrates the setting for the AprilTags we used to collect the ground truth.

In the experiments, the Apriltag images are collected by a downward facing camera installed on the drone. The computation of relative orientation and translation is aligned with the UWB coordinate system to obtain the true 6-DoF pose of the drone, and then followed by using RMSE (Root Mean Square Error) to calculate ATE (absolute trajectory error). In addition, as illustrated in Fig. 6, we also simulate the visual input in the low-texture environment with weak features through the flight direction of the drone to verify the robustness of our proposed UWB constrained VIO method. The flight speed of the drone for data acquisition is at about 0.5 m/s.

In the first experiment, we evaluate the improvement of our fusion method over the general cumulative error of VIO, and the effect of



(a) Original trajectory (x, y) plane. (b) Original trajectory z-axis.

(c) Improved trajectory (x, y) plane. (d) Improved trajectory z-axis.

Fig. 13. The fusion results with slight positioning drifts and the improvement after adding our relocation method. The blue curves represent the output of VINS after axis alignment with VIO fusion for comparison.

Table 2
The improvement on cumulative errors by UWB marginalization in z-axis.

Method	ATE
UWB	0.289 m
VINS-Mono	0.519 m
Our fusion without z-axis marginalization	0.223 m
Our fusion with z-axis marginalization	0.179 m

Table 3
The accuracy of VINS-Mono and our fusion results before and after adding the relocation method using Algorithm 1 for improvement.

Method	ATE
VINS-Mono (before improvement)	0.718 m
VINS-Mono (after improvement)	0.508 m
Our fusion (before improvement)	0.594 m
Our fusion (after improvement)	0.434 m

UWB positioning drift and marginalization in the z-axis. We first use the data with relatively stable feature points and less vibration to evaluate the algorithm. Fig. 7 illustrates the common cumulative error problem of a general VIO. It can be seen that using UWB as a global estimation have greatly improved the drift effect caused by the VIO cumulative error. The VIO pose increment can also eliminate the noise of UWB positioning. Since our UWB setting takes the occlusion of the environment and the overall positioning stability into account, the anchors are placed at the same height. To reduce the noise in the z-axis, we adopt the strategy of marginalizing the UWB z-axis data and then fusing with the VIO results. Fig. 8 and Table 2 illustrate the influence of marginalization in the z-axis after fusion. The stability of z-axis values is greatly improved after marginalization using our algorithm. The comparison of relative pose errors after the z-axis marginalization is shown in Fig. 9. It can be seen that our fusion method provides stable performance in both the translation and rotation precision after a long range navigation.

In the second experiment, we investigate the improvements on the cumulative errors and positioning stability in the cases of large vibration amplitudes and weak features. Fig. 10 shows the positioning results with improvements. The ATEs of UWB, VINS-Mono and our fusion method are 0.463 m, 1.506 m and 0.414 m, respectively. It illustrates that our fusion result has good accuracy performance when the feature points are weak and the VIO errors are large. Compared with VINS-Mono, the error of our fusion is reduced about one meter. Although the difference in accuracy is not very obvious compared to UWB positioning, the smoothness of the overall path is significantly improved as shown in figures. Fig. 11 shows the comparison of relative pose errors derived in the weak feature environment. The plot illustrates the stability of our technique compared to UWB and VINS-Mono.

In the last experiment, we demonstrate the impact of positioning drifts on accuracy before and after fusion, followed by evaluating the degree of improvements on the positioning drifts with the UWB short-term constraints using the proposed method. Fig. 12 illustrates the errors caused by the positioning drift. The ATEs for slight and large drifts are derived as 11.874 m and 170.075 m, respectively. The result indicates that, if the relocation is not adopted when the positioning drift occurs, the local positioning before fusion will be almost unreliable. Fig. 13 shows the fusion results with the positioning drifts and the improvement after adding our relocation method. Table 3 tabulates our improvement in terms of ATE. The UWB short-term constrained relocation has improved about 20 cm in both the local pose increment and the overall accuracy. Moreover, the improvement for the high positioning drift is large and the accuracy can achieve the same level as the slight positioning drift case.

5. Conclusion and discussion

The indoor environments are usually messy or narrow, and the high positioning accuracy will be necessary for a drone to fly smoothly.

During a long-range navigation, the cumulative errors are inevitable for a SLAM system and have great effects on the controls of UAVs. In addition, a large positioning drift will be produced due to the vibration or unstable features, and result in a substantial decrease in the overall accuracy. In this paper, we present an optimization-based fusion technique for localization improvement. The relocation method using UWB short-term variation constraint effectively limits the drift of the VIO-SLAM systems. Our fusion approach is able to achieve smooth positioning results without cumulative errors under the global coordinate frame. The experiments carried out in a real-world environment and using the public EuRoC dataset have demonstrated the effectiveness and robustness of the proposed technique.

Since the proposed technique for VIO relocation is based on the short-term positioning results of UWB, it may fail when there are non-line-of-sight errors. This will in term reduce the robustness of the system. Our current fusion strategy utilizes an online coordinate alignment method to integrate the outputs from VIO and UWB. However, the relocation module does not exclusively consider the VIO coordinate consistency. In the future work, the optimization of VIO back-end for localization drift will be further investigated to improve the relocation accuracy.

CRediT authorship contribution statement

Huei-Yung Lin: Conceptualization, Methodology, Investigation, Supervision, Validation, Project administration, Funding acquisition, Writing – review & editing. **Jia-Rong Zhan:** Data curation, Software, Methodology, Visualization, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 109-2221-E-194-037-MY3 is gratefully acknowledged.

References

[1] H.Y. Lin, J.L. Hsu, A sparse visual odometry technique based on pose adjustment with keyframe matching, IEEE Sens. J. 21 (10) (2021) 11810–11821, http://dx.doi.org/10.1109/JSEN.2020.3015922.

[2] Y. Xu, Y.S. Shmaliy, T. Shen, D. Chen, M. Sun, Y. Zhuang, INS/UWB-Based quadrotor localization under colored measurement noise, IEEE Sens. J. 21 (5) (2021) 6384–6392, http://dx.doi.org/10.1109/JSEN.2020.3038242.

[3] W. Lee, K. Eckenhoff, P. Geneva, G. Huang, Intermittent GPS-aided VIO: Online initialization and calibration, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 5724–5731, http://dx.doi.org/10.1109/ICRA40945.2020.9197029.

[4] A. Masiero, F. Fissore, R. Antonello, A. Cenedese, A. Vettore, A comparison of UWB and motion capture UAV indoor positioning, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 42 (2019) 1695–1699.

[5] J. Li, G. Yang, Q. Cai, H. Niu, J. Li, Cooperative navigation for UAVs in GNSS-denied area based on optimized belief propagation, Measurement 192 (2022) 110797.

[6] S. Guo, Z. Rong, S. Wang, Y. Wu, A LiDAR SLAM with PCA-based feature extraction and two-stage matching, IEEE Trans. Instrum. Meas. 71 (2022) 1–11, http://dx.doi.org/10.1109/TIM.2022.3156982.

[7] H.Y. Lin, Y.C. Chung, M.L. Wang, Self-localization of mobile robots using a single catadioptric camera with line feature extraction, Sensors 21 (14) (2021).

[8] G. Grisetti, C. Stachniss, W. Burgard, Improved techniques for grid mapping with rao-blackwellized particle filters, IEEE Trans. Robot. 23 (1) (2007) 34–46, http://dx.doi.org/10.1109/TRO.2006.889486.

[9] W. Hess, D. Kohler, H. Rapp, D. Andor, Real-time loop closure in 2D LIDAR SLAM, in: 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 1271–1278, http://dx.doi.org/10.1109/ICRA.2016.7487258.

[10] C. Zhang, L. Chen, S. Yuan, ST-VIO: Visual-inertial odometry combined with image segmentation and tracking, IEEE Trans. Instrum. Meas. 69 (10) (2020) 8562–8570, http://dx.doi.org/10.1109/TIM.2020.2989877.

[11] T. Qin, P. Li, S. Shen, VINS-mono: A robust and versatile monocular visual-inertial state estimator, IEEE Trans. Robot. 34 (4) (2018) 1004–1020, http://dx.doi.org/10.1109/TRO.2018.2853729.

[12] C. Campos, R. Elvira, J.J.G. Rodríguez, J.M. M. Montiel, J. D. Tardós, ORB-SLAM3: An accurate open-source library for visual, visual–Inertial, and multimap SLAM, IEEE Trans. Robot. (2021) 1–17, http://dx.doi.org/10.1109/TRO.2021.3075644.

[13] S. Bottigliero, D. Milanesio, M. Saccani, R. Maggiora, A low-cost indoor real-time locating system based on TDOA estimation of UWB pulse sequences, IEEE Trans. Instrum. Meas. 70 (2021) 1–11, http://dx.doi.org/10.1109/TIM.2021.3069486.

[14] R. Liu, C. Yuen, T.-N. Do, M. Zhang, Y.L. Guan, U.-X. Tan, Cooperative positioning for emergency responders using self IMU and peer-to-peer radios measurements, Inf. Fusion 56 (2020) 93–102.

[15] J.P. Queralta, C. Martínez Almansa, F. Schiano, D. Floreano, T. Westerlund, UWB-based system for UAV localization in GNSS-denied environments: Characterization and dataset, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 4521–4528, http://dx.doi.org/10.1109/IROS45743.2020.9341042.

[16] T.-M. Nguyen, T.H. Nguyen, M. Cao, Z. Qiu, L. Xie, Integrated UWB-vision approach for autonomous docking of UAVs in GPS-denied environments, in: 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 9603–9609, http://dx.doi.org/10.1109/ICRA.2019.8793851.

[17] N. Macoir, J. Bauwens, B. Jooris, B. Van Herbruggen, J. Rossey, J. Hoebeke, E. De Poorter, Uwb localization with battery-powered wireless backbone for drone-based inventory management, Sensors 19 (3) (2019) 467.

[18] J. Wei, H. Wang, S. Su, Y. Tang, X. Guo, X. Sun, NLOS identification using parallel deep learning model and time-frequency information in UWB-based positioning system, Measurement 195 (2022) 111191.

[19] F. Lazzari, A. Buffi, P. Nepa, S. Lazzari, Numerical investigation of an UWB localization technique for unmanned aerial vehicles in outdoor scenarios, IEEE Sens. J. 17 (9) (2017) 2896–2903, http://dx.doi.org/10.1109/JSEN.2017.2684817.

[20] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual–inertial odometry using nonlinear optimization, Int. J. Robot. Res. 34 (3) (2015) 314–334.

[21] R. Mur-Artal, J.D. Tardós, Visual-inertial monocular SLAM with map reuse, IEEE Robot. Autom. Lett. 2 (2) (2017) 796–803, http://dx.doi.org/10.1109/LRA.2017.2653359.

[22] A.I. Mourikis, S.I. Roumeliotis, A multi-state constraint Kalman filter for vision-aided inertial navigation, in: Proceedings 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 3565–3572, http://dx.doi.org/10.1109/ROBOT.2007.364024.

[23] M. Li, A.I. Mourikis, 3-d motion estimation and online temporal calibration for camera-imu systems, in: 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 5709–5716, http://dx.doi.org/10.1109/ICRA.2013.6631398.

[24] M. Bloesch, S. Omari, M. Hutter, R. Siegwart, Robust visual inertial odometry using a direct EKF-based approach, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 298–304, http://dx.doi.org/10.1109/IROS.2015.7353389.

[25] K. Konolige, M. Agrawal, FrameSLAM: From bundle adjustment to real-time visual mapping, IEEE Trans. Robot. 24 (5) (2008) 1066–1077, http://dx.doi.org/10.1109/TRO.2008.2004832.

[26] S. Lynen, M.W. Achtelik, S. Weiss, M. Chli, R. Siegwart, A robust and modular multi-sensor fusion approach applied to mav navigation, in: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 3923–3929, http://dx.doi.org/10.1109/IROS.2013.6696917.

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

[28] Y. Song, M. Guan, W.P. Tay, C.L. Law, C. Wen, UWB/LiDAR fusion for cooperative range-only SLAM, in: 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 6568–6574, http://dx.doi.org/10.1109/ICRA.2019.8794222.

[29] B. Gao, B. Lian, D. Wang, C. Tang, Low drift visual inertial odometry with UWB aided for indoor localization, IET Commun. (2022).

[30] S. Shin, E. Lee, J. Choi, H. Myung, MIR-vio:Mutual information residual-based visual inertial odometry with UWB fusion for robust localization, in: 2021 21st International Conference on Control, Automation and Systems (ICCAS), 2021, pp. 91–96, http://dx.doi.org/10.23919/ICCAS52745.2021.9649888.

[31] R. Liu, Y. He, C. Yuen, B.P.L. Lau, R. Ali, W. Fu, Z. Cao, Cost-effective mapping of mobile robot based on the fusion of UWB and short-range 2-D LiDAR, IEEE/ASME Trans. Mechatronics 27 (3) (2022) 1321–1331, http://dx.doi.org/10.1109/TMECH.2021.3087957.

[32] T.H. Nguyen, T.-M. Nguyen, L. Xie, Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization, IEEE Robot. Autom. Lett. 6 (2) (2021) 1678–1685, http://dx.doi.org/10.1109/LRA.2021.3057838.

[33] V. Magnago, P. Corbalán, G.P. Picco, L. Palopoli, D. Fontanelli, Robot localization via odometry-assisted ultra-wideband ranging with stochastic guarantees, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 1607–1613, http://dx.doi.org/10.1109/IROS40897.2019.8968019.

[34] F.J. Perez-Grau, F. Caballero, L. Merino, A. Viguria, Multi-modal mapping and localization of unmanned aerial robots based on ultra-wideband and RGB-d sensing, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 3495–3502, http://dx.doi.org/10.1109/IROS.2017.8206191.

[35] C. Wang, H. Zhang, T.-M. Nguyen, L. Xie, Ultra-wideband aided fast localization and mapping system, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 1602–1609, http://dx.doi.org/10.1109/IROS.2017.8205968.

[36] Y. Cao, G. Beltrame, VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems, 2020, arXiv preprint arXiv:2006.00420.

[37] M. Delamare, R. Boutteau, X. Savatier, N. Iriart, Static and dynamic evaluation of an UWB localization system for industrial applications, Science 2 (2) (2020) 23.

[38] J. Shi, Tomasi, Good features to track, in: 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600, http://dx.doi.org/10.1109/CVPR.1994.323794.

[39] S. Shen, N. Michael, V. Kumar, Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 5303–5310, http://dx.doi.org/10.1109/ICRA.2015.7139939.

[40] S. Agarwal, K. Mierle, et al., Ceres Solver, http://ceres-solver.org.

[41] P.J. Huber, Robust estimation of a location parameter, in: Breakthroughs in Statistics, Springer, 1992, pp. 492–518.

[42] A.I. Mourikis, S.I. Roumeliotis, A multi-state constraint Kalman filter for vision-aided inertial navigation, in: Proceedings 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 3565–3572, http://dx.doi.org/10.1109/ROBOT.2007.364024.

[43] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visual–inertial odometry using nonlinear optimization, Int. J. Robot. Res. 34 (3) (2015) 314–334.

[44] M. Bloesch, M. Burri, S. Omari, M. Hutter, R. Siegwart, Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback, Int. J. Robot. Res. 36 (10) (2017) 1053–1072.

[45] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M.W. Achtelik, R. Siegwart, The EuRoC micro aerial vehicle datasets, Int. J. Robot. Res. (2016) http://dx.doi.org/10.1177/0278364915620033.

[46] M. Krogius, A. Haggenmiller, E. Olson, Flexible layouts for fiducial tags, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 1898–1903, http://dx.doi.org/10.1109/IROS40897.2019.8967787.

[47] Z. Zhang, D. Scaramuzza, A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 7244–7251, http://dx.doi.org/10.1109/IROS.2018.8593941.