2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

Range-focused Fusion of Camera-IMU-UWB for Accurate and Drift-Reduced Localization

Thien Hoang Nguyen, Thien-Minh Nguyen, and Lihua Xie*, *Fellow, IEEE*

***Abstract*— In this work, we present a tightly-coupled fusion scheme of a monocular camera, a 6-DoF IMU, and a single unknown Ultra-wideband (UWB) anchor to achieve accurate and drift-reduced localization. Specifically, this paper focuses on incorporating the UWB sensor into an existing state-of-the-art visual-inertial system. Previous works toward this goal use a single nearest UWB range data to update robot positions in the sliding window ("position-focused") and have demonstrated encouraging results. However, these approaches ignore 1) the time-offset between UWB and camera sensors, and 2) all other ranges between two consecutive keyframes. Our approach shifts the perspective to the UWB measurements ("range-focused") by leveraging the propagated information readily available from the visual-inertial odometry pipeline. This allows the UWB data to be used in a more effective manner: the time-offset of each range data is addressed and all available measurements can be utilized. Experimental results show that the proposed method consistently outperforms previous methods in both estimating the anchor position and reducing the drift in long-term trajectories.**

I. Introduction

Reliable and globally consistent localization is still an open research problem for many robotic applications. In recent years, visual-inertial odometry (VIO) or simultaneous localization and mapping (VI-SLAM) are popular approaches for this purpose due to the complementary nature of camera and IMU sensors. Even though state-of-the-art methods such as [1]–[3] can achieve very accurate and high-rate pose and velocity estimates, sensor noise and computation errors make the system prone to accumulated drift over time. The popular solution to this issue is to include an additional global sensor such as GPS [4], [5]. For situations where GPS is not available (indoor, tunnel, corridor etc.), UWB is an alternative option suitable for small-scale operations [6], [7].

Methods for fusing UWB data and VIO have been proposed for various applications and scenarios [8]–[11]. These approaches work in a *loosely-coupled* manner, which means the UWB ranges and camera-IMU data are first computed in separated localization systems, then the position estimates obtained by the UWB and camera-IMU subsystems are aligned and fused together afterwards. While these methods can be constructed straightforwardly, we believe the results could be improved if all sensors data are fused at once to take advantage of the correlations between available information. Furthermore, they require a setup with multiple known UWB anchors for range-based localization which can be costly and inhibit the applicability in many spatially constrained scenarios such as indoor, tunnel, corridor, etc.

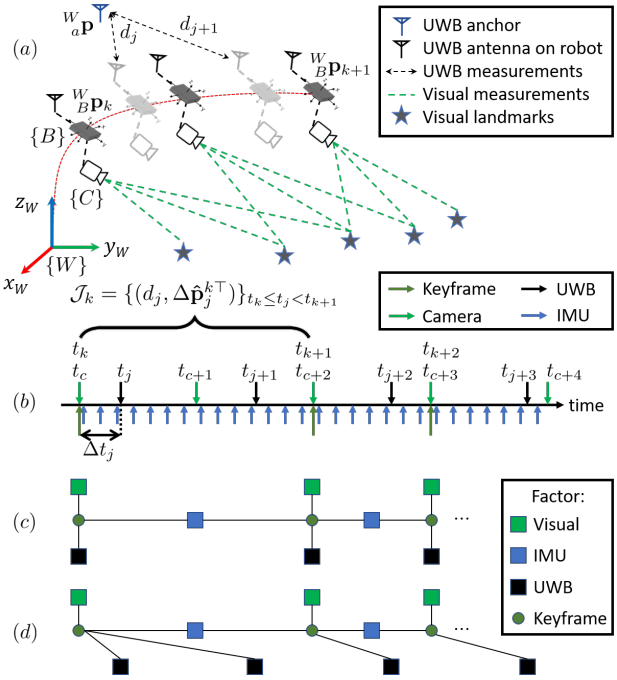

Fig. 1: a) Illustration of the reference frames and measurements. b) Timing of the sensor measurements and keyframes in our formulation; c-d) factor graphs of previous ("position-focused") and proposed ("range-focused") approaches, respectively. Note that a camera frame will be selected as keyframe only if certain conditions are met.

More recently, approaches that employ only a single UWB anchor with unknown position have been introduced [12]–[15]. Such systems would enjoy both the advantages of having drift-free range measurements for accurate localization as well as ease of usage for practical applications since no setup time to calibrate the anchor position is required. Results indicate that by tightly-coupling the UWB, camera and/or IMU measurements in a joint optimization problem, more accurate and robust localization can be achieved. However, these approaches process the UWB data in a simulation-like manner: each camera position is paired with one range measurement, and any other ranges between two consecutive camera frames are not considered. This perspective does not reflect real-life sensor systems for many reasons: real UWB sensor is independent from camera/IMU sensors, hence there will always be a time-offset between the range/image mes-

The authors are with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, 50 Nanyang Avenue.
*Email of corresponding author: elhxie@e.ntu.edu.sg

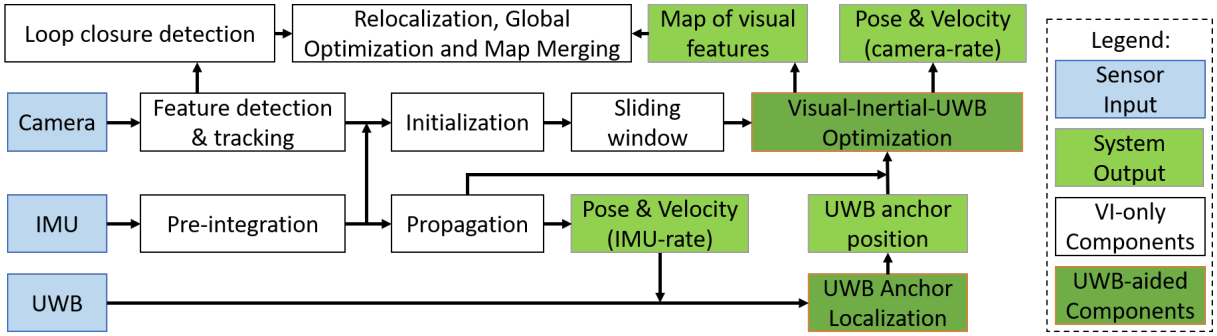

Fig. 2: Overview of the complete system. Based on the state-of-the-art VIO system VINS-Mono [1], our contributions are embedded in the UWB-aided components, which are discussed in details in Section IV.

sages; UWB range measurement rate does not conform to standard camera or IMU rate, with UWB data rate often a few times higher than that of camera; UWB data rate can vary during actual operation due to loss of line-of-sight which means the amount of UWB data available can also vary.

To address these issues, in this letter we propose a more efficient approach to fuse visual, inertial and UWB measurements. In essence, we leverage the existing state propagation process in the VIO pipeline to efficiently formulate an UWB error term for each range data. Our main contributions include:

- a so-called "range-focused" perspective of processing UWB measurements by leveraging the propagated data computed by the VIO pipeline, which adequately addresses the time-offset between UWB-camera sensors and allows all available UWB data to be consumed;
- a tightly-coupled fusion scheme of a monocular camera, a 6-DoF IMU and a single UWB anchor to provide drift-reduced odometry, with a built-in UWB anchor localization module to estimate the unknown anchor positions;
- extensive experimental results to verify the performance of each of the proposed UWB-aided system components.

This letter is structured as follows: Firstly, Section II presents the works related to the topic. Section III then outlines the system as well as pertinent concepts of VIO. Section IV starts with the idea of the so-called "range-focused" perspective of processing UWB measurements, which is then used for the two tasks of UWB anchor localization and tightly-coupled fusion of camera-IMU-UWB sensors for odometry. Next, Section V offers real-life and simulation results and comparison with state-of-the-art methods. The letter is finally concluded in Section VI.

II. Literature Review

A. UWB-aided localization and mapping

Different methods of incorporating UWB into existing localization systems have been put forth [16]. UWB ranges can be used in an independent localization method which is then combined with: monocular camera [17]–[19], IMU [20], [21], RGB-D camera [22], IMU and RGB-D [9], LiDAR [23], etc., to improve the accuracy and robustness of the SLAM system. Obtaining a unique solution for 3D range-only localization requires either: 1) a minimum of four UWB anchors with known positions, or 2) three known anchors and height data of the robot [24], [25]. This assumption limits the applicable scenario for the system since the operating area needs to accommodate the UWB anchors setup and every new environment requires additional time and effort to calibrate the anchor's positions. To alleviate this requirement, recent methods try to estimate the map of anchors during the operation given that the robot has access to metric-scale odometry with additional inter-anchor ranges [23], or just metric-scale odometry [15], [22], or even up-to-scale odometry only [17], [18]. These solutions, however, still process the UWB data in a sub-optimal manner which is explained in Section II-B. In this work, we explore the combination of camera-IMU and UWB with only a single anchor at an unknown position. Such setup would combine the benefits of having the VIO pipeline for accurate short-term odometry and the most flexible UWB anchor configuration. Furthermore, visual information is essential for many high-level researches and applications in the era of deep learning.

B. Visual-Inertial-Range localization and mapping

Most related to this letter are the works that use camera-IMU-UWB sensors for the localization and mapping tasks. While most approaches employ VIO for onboard localization and separately use UWB for range-based relative localization [26]–[29], recent works have demonstrated that it is possible to fuse visual, inertial and UWB data to simultaneously obtain the anchor position estimate and improve pose estimates, with [15] proposing an EKF solution and [14] adopting a pose-graph optimization framework. While differ in final goal and fusion approach, these methods share the same fundamental underlying principle of using UWB data: the residual is formulated from the point of view of the position in the state vector. This view leads to the following issues: 1) one position is paired with one nearest UWB measurement with the time-offset between the camera frame and range data ignored, 2) all other ranges between two

consecutive camera frames are discarded. In contrast, the proposed system formulates the UWB residual according to the range measurement's timestamp which allows the range data to be used at the precision of the sensor. By leveraging the results of the IMU state propagation process in the VIO pipeline, the UWB residual is derived for each range measurement thus the time-offset issue is accounted for and all available ranges can be utilized.

III. Preliminaries

In this section, we provide an overview of the system, describe the notation and the most relevant concepts of the VIO that will be used in Section IV-B and IV-C. Due to length constraints, for the visual-inertial components interested readers are directed to VINS-Mono [1] for more details.

A. System overview

Fig. 2 illustrates the overview of the proposed system. A mobile robot is equipped with a monocular camera, a 6-DoF IMU and an UWB sensor rigidly attached to the body frame, with all intrinsic and extrinsic parameters calibrated. Range measurements to a single UWB anchor placed at an unknown location are available. In this work, the two-way time-of-flight (TW-ToF) UWB sensor is employed since it does not require clock synchronization between sensors and thus would be more appropriate for many application scenarios.

The system operates in two phases:

1) *UWB anchor localization based on VIO* (Section IV-B): initially only the camera and IMU will be used to provide accurate short-term odometry which will be combined with range measurements to estimate the UWB anchor position. The anchor position estimate is considered fixed once the uncertainty falls below a certain threshold.
2) *Visual-Inertial-Range odometry* (Section IV-C): once the UWB anchor position is found, subsequent range measurements will be tightly fused together with visual and inertial data in a joint keyframe-based optimization to obtain accurate and drift-reduced long-term odometry.

In theory, the two phases can be combined into one (i.e., extending the state vector of the VIO optimization with the UWB anchor position). Nevertheless, such system is unlikely to work in practice for multiple reasons:

- The sliding window for VIO is designed to contain data in a short time span, which is not favorable for the UWB anchor localization task which relies strongly on having spatially diverse position data in 3D. Extending the window size is not viable since a significant computation burden will be imposed on the computer due to the exponential increase of visual measurements.
- The optimization would struggle to produce any satisfactory results if a good initial guess for the anchor position is not provided. Since this initial guess is typically measured manually and can change for each operation, it is subject to human errors and should be avoided.

B. Notation

Depicted in Fig. 1a are the coordinate frames used in this work, including the body frame $\{B\}$ which corresponds to the IMU frame, the camera frame $\{C\}$ and the world frame $\{W\}$. The homogeneous transformation matrix $\mathbf{T} \in SE(3)$, which consists of the rotation matrix ${}^{A}\mathbf{R}$ and position vector ${}^{A}\mathbf{p}$, represents a 3D pose in frame $\{A\}$:

$$ {}^{A}\mathbf{T} \coloneqq \begin{bmatrix} {}^{A}\mathbf{R} & {}^{A}\mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix}, {}^{A}\mathbf{R} \in SO(3), {}^{A}\mathbf{p} \in \mathbb{R}^3. \tag{1} $$

The quaternion representation of ${}^{A}\mathbf{R}$ is ${}^{A}\mathbf{q} = [q_x, q_y, q_z, q_w]^\top$ (q_w being the scalar part). We use the compact form of ${}^{A}\mathbf{T}$ as $[{}^{A}\mathbf{p}^\top, {}^{A}\mathbf{q}^\top]^\top$ during computation.

Denote t_c, t_i, and t_j as the timestamp of the camera, IMU, UWB range measurements, respectively. t_k is the timestamp of one of the keyframes in the sliding window, which is set to be $t_k \coloneqq t_c$ without loss of generality. It should be noted that camera frames are only selected to be keyframes if certain criteria are satisfied [1]. This can lead to two consecutive keyframes to be separated by multiple camera frames. As a result, the number of UWB measurements available between two keyframes is often not uniform across the sliding window. Fig. 1b) demonstrates an example of this observation.

C. Optimization-based Monocular Visual-Inertial Odometry

The sliding window $\mathcal{X}$ consists of the visual feature states $\mathcal{X}_L$ and IMU states $\mathcal{X}_B$:

$$ \begin{aligned} \mathcal{X} &= \{\mathcal{X}_L, \mathcal{X}_B\}, \\ \mathcal{X}_B &= [\mathbf{x}_1, \cdots, \mathbf{x}_k, \cdots, \mathbf{x}_K], \\ \mathbf{x}_k &= [{}^{W}_{B}\mathbf{p}_k, {}^{W}_{B}\mathbf{q}_k, {}^{W}_{B}\mathbf{v}_k, \mathbf{b}_{a_k}, \mathbf{b}_{g_k}], k \in [1, K], \end{aligned} \tag{2} $$

where K is the number of keyframes. Each IMU state $\mathbf{x}_k$ at time t_k consists of position, orientation, velocity of the IMU in the world frame, as well as the biases of accelerometer and gyroscope in the body frame. The VIO pipeline uses a bundle adjustment formulation and minimizes the cost function:

$$ E_{\boldsymbol{VI}}(\mathcal{X}) = \sum_{(n,m)\in\mathcal{C}} \rho(\|\mathbf{e}_{\mathbf{v}}^{n,m}\|^2) + \sum_{k=1}^{K} \left\|\mathbf{e}_{\mathbf{i}}^{k}\right\|^2 + \|\mathbf{e}_{\mathbf{p}}\|^2, \tag{3} $$

which consists of the visual residuals $\mathbf{e}_{\mathbf{v}}^{n,m}$, the IMU residuals $\mathbf{e}_{\mathbf{i}}^{k}$ and marginalization residuals $\mathbf{e}_{\mathbf{p}}$ [30]. $\mathcal{C}$ is the set of landmarks that have been observed in both n-th and m-th keyframes. $\|*\|$ denotes the Euclidean norm of a vector. The Huber norm ρ [31] is applied to the visual residuals to reduce the effect of tracking outliers.

D. IMU state propagation

The IMU pre-integration methods [32] allow the high-frequency linear acceleration and angular velocity to be pre-integrated into pseudo-measurements $\boldsymbol{\alpha}_{i+1}^{i}, \boldsymbol{\beta}_{i+1}^{i}, \boldsymbol{\gamma}_{i+1}^{i}$ [1] using IMU measurements from t_i to t_{i+1}. These terms correspond to the pre-integrated inertial and relative-orientation

measurement, respectively. Every time a new keyframe is created at t_k, all propagated states are reset:

$$ {}^W_B\hat{\mathbf{p}}_i = {}^W_B\mathbf{p}_k;\ {}^W_B\hat{\mathbf{v}}_i = {}^W_B\mathbf{v}_k;\ {}^W_B\hat{\mathbf{q}}_i = {}^W_B\mathbf{q}_k. \tag{4} $$

For every subsequent new IMU measurement at t_{i+1}, the states are recursively propagated following the formulation [1]:

$$ \begin{bmatrix} {}^W_B\hat{\mathbf{p}}_{i+1} \\ {}^W_B\hat{\mathbf{v}}_{i+1} \\ {}^W_B\hat{\mathbf{q}}_{i+1} \end{bmatrix} = \begin{bmatrix} {}^W_B\hat{\mathbf{p}}_i + {}^W_B\hat{\mathbf{v}}_i\Delta t - \frac{1}{2}\mathbf{g}\Delta t^2 + {}^W_B\hat{\mathbf{q}}_i\boldsymbol{\alpha}^i_{i+1} \\ {}^W_B\hat{\mathbf{v}}_i - \mathbf{g}\Delta t + {}^W_B\hat{\mathbf{q}}_i\boldsymbol{\beta}^i_{i+1} \\ {}^W_B\hat{\mathbf{q}}_i\boldsymbol{\gamma}^i_{i+1} \end{bmatrix} \tag{5} $$

to produce the IMU-rate state estimates. Leveraging this operation, at t_j when a new range measurement is received ($t_j > t_k$), the propagated state $({}^W_B\hat{\mathbf{p}}_j, {}^W_B\hat{\mathbf{v}}_j, {}^W_B\hat{\mathbf{q}}_j)$ is readily available. From that, the IMU-based prediction of the change in position in the world frame from t_k to t_j is simply:

$$ \Delta\hat{\mathbf{p}}^k_j = {}^W_B\hat{\mathbf{p}}_j - {}^W_B\mathbf{p}_k. \tag{6} $$

IV. Range-focused Fusion of Camera-IMU-UWB

In this section, we present the main contributions of this work, from the "range-focused" formulation of UWB residuals to the implementation of UWB anchor localization and keyframe-based visual-inertial-range odometry.

A. Position-focused vs. range-focused UWB residual

Fig. 1a-b) visualizes an example of the time-offset between camera and UWB measurements. Let $e_\mathbf{r}$ be the UWB residual. The formulation of $e_\mathbf{r}$ in previous works (e.g., [14], [15], [33]), which is referred to as "position-focused" in this letter, associates each position in the sliding window ${}^W_B\mathbf{p}_k$ with one range measurement at the nearest timestamp:

$$ e^k_\mathbf{r} = d_j - \left\| {}^W_B\mathbf{p}_k - {}^W_a\mathbf{p} \right\|, \quad \forall {}^W_B\mathbf{p}_k. \tag{7} $$

However, this formulation does not account for real-life issues as discussed in Section II-B. In this work, we propose the so-called "range-focused" formulation which associates each range data with a position at the same timestamp:

$$ e^j_\mathbf{r} = d_j - \left\| {}^W_B\mathbf{p}_j - {}^W_a\mathbf{p} \right\|, \quad \forall d_j, \tag{8} $$

by adapting the position to each range data at t_j $({}^W_B\mathbf{p}_j)$, which is further explained in Section IV-B and IV-C.

One can argue that (7) and (8) are identical if the following conditions are met concurrently ($\Delta t_j = t_j - t_k$):

1) there is no time-offset between ${}^W_B\mathbf{p}_k$ and d_j ($\Delta t_j = 0$), i.e. the camera and UWB sensors are synchronized, and
2) either the UWB and camera sensors have the same data rate or all other extra range measurements between two camera frames are ignored.

Indeed, simulation (Section V-A) shows that the two formulations have similar results under these conditions. However, in reality the camera and UWB sensors always work independent of one another, hence the time-offset issue is inherent and should not be ignored. Furthermore, the standard UWB data rate is often many times higher than that of the camera. As a result, previous methods would always discard a significant portion of available range data which means the UWB sensor is still underutilized.

B. UWB Anchor Localization based on VIO data

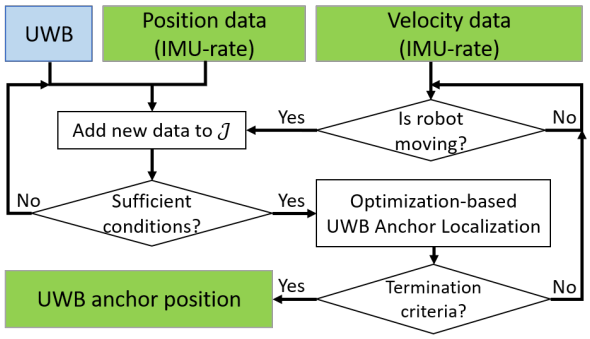

Fig. 3: Overview of the UWB anchor localization component of the system.

1) Problem Formulation: To estimate the UWB anchor position ${}^W_a\mathbf{p}$ in the world frame (Fig. 1a) using short-term accurate VIO data (pose and velocity), we formulate an optimization problem over a data window:

$$ \mathcal{J} = \{(d_j, {}^W_B\hat{\mathbf{p}}^\top_j)\}_{t_j>0}, \tag{9} $$

which consists of range data and corresponding IMU-rate output position ${}^W_B\hat{\mathbf{p}}_j$ from the VIO pipeline. Fig. 3 outlines the module which can be developed as a standalone system. The cost function to be minimized is

$$ E_r({}^W_a\mathbf{p}) = \sum_{j\in\mathcal{J}} \rho(e^j_\mathbf{r}), \tag{10} $$

with ρ as the Huber loss. The UWB residual $e^j_\mathbf{r}$, as introduced in Equ. (8), can be computed with the IMU-rate position

$$ {}^W_B\mathbf{p}_j \coloneqq {}^W_B\hat{\mathbf{p}}_j. \tag{11} $$

We remark that for a low cost system with noisy IMU data, the IMU-rate state propagation can be unreliable. As such, ${}^W_B\mathbf{p}_j$ can be computed instead from the more stable camera-rate state estimates as ${}^W_B\mathbf{p}_j \coloneqq {}^W_B\mathbf{p}_k + {}^W_B\mathbf{v}_k\Delta t_j$.

2) Sufficient conditions: The observability of the problem has been established in [29] which states that the robot should not move directly towards the anchor. In practice, the estimation result would depend on the trajectory covering all 3D axes as well as how far is the anchor relative to the movement radius. Denote the sample variance of position on the x axis as $\boldsymbol{S}^2_x$ (similarly for y and z axes), which is recursively updated for each new position data when the robot is moving. To ensure the performance of the optimization, the following conditions are checked to start or skip the optimization process:

- $\left\| {}^W_B\hat{\mathbf{v}}_j \right\| > v_{\text{min}}$: checking whether the robot is moving, since if the robot is static new data would be the same and the optimization result would not improve,

- $\min \boldsymbol{S}^2_{xyz} > \boldsymbol{S}^2_{\text{min}}$: checking whether the sample variance of the position for each axis is sufficiently large, i.e. the robot's movement covers all directions,

with v_{min} and $\boldsymbol{S}^2_{\text{min}}$ being user-defined parameters. The first condition is checked for every new VIO data until the termination criterion (Section IV-B.3) is met. The second condition is checked until it is satisfied for the first time. By enforcing these conditions, we found that the estimation can achieve satisfactory results with an effortless initial guess of the anchor position ($^W_a\mathbf{p}_0 = [0,0,0]^\top$), which is the setup in all of our experiments. Nonetheless, the convergence time can be improved with a good initial guess and can be taken into account to improve performance in practice.

3) Termination criterion: Once started, the cost function (10) can be optimized using the standard Levenberg-Marquardt algorithm [34] with the Ceres solver [35]. Since the system is performed online, a termination criterion is introduced to determine the uncertainty of the solution:

$$\sigma_{\text{max}} < \sigma_p \tag{12}$$

with σ_{max} being the maximum singular value of the covariance matrix and σ_p a given threshold. Once the criterion is met, the UWB anchor localization operation is finished and $^W_a\mathbf{p}$ is fixed. An example of the convergence can be seen in Fig. 4. Since the computation of σ_{max} is not essential but might be time consuming due to the rank deficient check and inversion of the Jacobian matrix, this termination check can be run in a separate thread at a slower rate (every 10 new UWB ranges, for example). As a result, the optimization might run a few extra times but the processing time will not be affected.

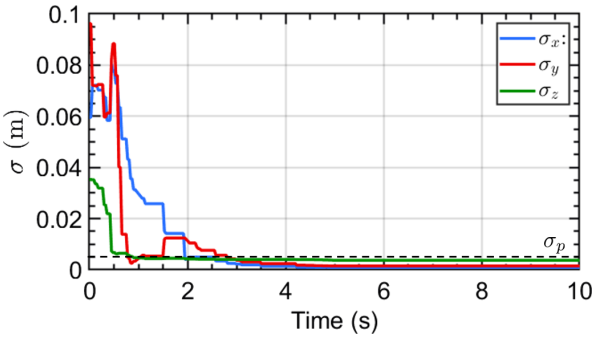

Fig. 4: Example of the covariance of the anchor position estimates in one of the real-life experiments. The anchor position is considered fixed when the maximum singular value of the covariance matrix is smaller than a certain threshold.

C. Keyframe-based Visual-Inertial-Range Odometry

1) Problem Formulation: Once the UWB anchor position has been found, the camera-IMU-UWB sensors can be tightly-fused to provide a more robust, accurate and drift-reduced odometry. To this end, we propose augmenting the VI-only cost function $E_{\boldsymbol{VI}}(\chi)$ (3) with the "range-focused" UWB residuals (8) to form the overall cost function:

$$E_{\boldsymbol{VIR}}(\chi) = E_{\boldsymbol{VI}}(\chi) + E_{\boldsymbol{R}}(\chi), \tag{13}$$

$$E_{\boldsymbol{R}}(\chi) = \gamma_r \sum_{k=1}^{K} \sum_{j \in \mathcal{J}_k} \rho(e^j_{\mathbf{r}}), \tag{14}$$

where $\mathcal{J}_k$ is the set of range data and corresponding predicted change of position (computed from (6)) between two keyframes at t_k and t_{k+1} in the sliding window:

$$\mathcal{J}_k = \{(d_j, \Delta\hat{\mathbf{p}}^{k\top}_j)\}_{t_k \le t_j < t_{k+1}}. \tag{15}$$

Fig. 1c-d) depicts the factor graph of the proposed system compared to previous "position-focused" methods. The measurements in $\mathcal{J}_k$ are used to create the UWB factors connected to the state $\mathbf{x}_k$. The UWB residual is re-weighted by a pre-defined factor of γ_r to amplify its impact on the optimization. Next, the UWB residual $e^j_{\mathbf{r}}$ is derived from the main idea of (8) with modifications to enhance the performance.

2) Range-focused UWB Factor: The position at time t_j can be associated to the state $\mathbf{x}_k$ by one of the following methods:

$$^W_B\mathbf{p}_j \coloneqq {}^W_B\mathbf{p}_k + \Delta\hat{\mathbf{p}}^k_j, \tag{16}$$

$$^W_B\mathbf{p}_j \coloneqq {}^W_B\mathbf{p}_k + {}^W_B\mathbf{v}_k \Delta t_j. \tag{17}$$

Using (16), the propagated position from VIO pipeline based on IMU data only ($\Delta\hat{\mathbf{p}}^k_j$, computed from (6)) is used to predict the position at any timestamp. However, since only one position state ($^W_B\mathbf{p}_k$) is directly linked to multiple UWB measurements, the solution might overfit to noisy UWB and/or IMU data. On the other hand, using (17) allows both position ($^W_B\mathbf{p}_k$) and velocity ($^W_B\mathbf{v}_k$) states to be coupled with the range measurements. Nonetheless, the constant velocity model might not work well for agile maneuver, especially when Δt_j can be longer than the period between two camera frames.

In this work, we propose combining (16) and (17) to take advantage of both prediction based on IMU data as well as motion model:

$$^W_B\mathbf{p}_j \coloneqq {}^W_B\mathbf{p}_k + \frac{1}{2}\Delta\hat{\mathbf{p}}^k_j + \frac{1}{2}\,{}^W_B\mathbf{v}_k \Delta t_j. \tag{18}$$

Following the marginalization strategy of VINS-Mono [1], the UWB factors connected to the first keyframe are transformed together with visual and inertial factors into a linearized prior whenever this keyframe is marginalized .

V. Experimental Results

The system is evaluated on both real-life experiments and simulation for each module (IV-B and IV-C) individually. The comparison is done mainly between the proposed "range-focused" and previous "position-focused" approaches, with results of VINS-Mono as baseline to access improvement.

A. UWB Anchor Localization

We first evaluate only the UWB anchor localization component with real-life experiments. The hardware includes an Intel Realsense T265 camera which provides stereo images at 30Hz, IMU data at 200Hz and UWB data at 38Hz from two separate anchors (no inter-anchor ranges) at 2m away from the starting position. Ground truth is provided by a Vicon system. The system runs in VIO-only mode (left camera and IMU data are used for odometry, UWB is not involved). The distance error of the estimated anchor position e_p is used for comparison.

While the experiments are done for two anchors, the estimation for each anchor's position is done in separate threads in order not to affect one another. Fig. 5 illustrates the result of estimating two anchors simultaneously. Both methods reach the same final error of less than 0.1m, but it is clear that the proposed method provides much faster convergence time. The reason is that our "range-focused" approach improves the solution for every new range measurement whereas the "position-focused" counterpart is stuck in a local minimum before sufficient data are collected to make a breakthrough.

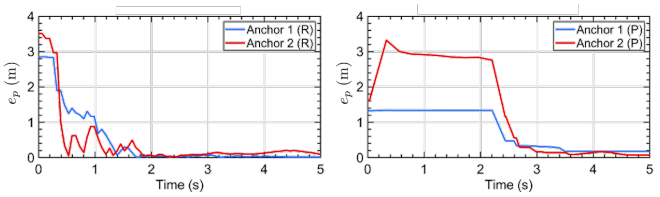

Fig. 5: Comparison of convergence time for the UWB anchor localization problem in real-life experiment with 2 anchors. Left: The proposed "range-focused" method (R). Right: The previous "position-focused" method (P).

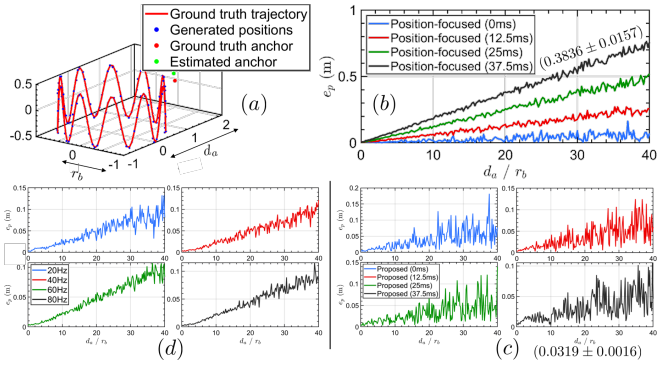

Fig. 6: Simulation results for the UWB anchor localization. a) Simulation setup. b-c) Results of the previous "position-focused" and proposed "range-focused" methods with different time-offset values and distance to anchor. The mean and standard error over 200 simulations of the 37.5ms case are shown in brackets. d) Result of the proposed method with different UWB data frequency (similar camera frequency) and distance to anchor. d_a is the distance from the center of movement to the anchor, r_b is the radius of movement and e_p is the final anchor position estimation error.

Simulations are carried out to validate and compare the effect of: 1) the time-offset between the UWB and camera data, 2) the distance to the anchor (d_a), relative to the "radius" of the trajectory (r_b), and 3) the UWB data frequency relative to camera data frequency. Fig. 6a shows the simulation setup. The position and velocity data are generated at 20Hz and one UWB anchor provides range data at 20Hz. All data are corrupted with Gaussian noise $\eta \sim \mathcal{N}(0, 0.02)$. For each simulation, the estimation is carried out until the stopping criteria of the Levenberg-Marquardt algorithm are met. If the condition (12) is satisfied, we stop the simulation and obtain the anchor position estimates as well as the position error. Otherwise, the estimation continues with the next data point and the final estimates are used to measure the error.

Firstly, the effect of time-offset are verified by providing both previous and proposed methods the same UWB and position data but varying time-offset ($0 - 37.5\text{ms}$). Fig. 6b-c shows the results of the previous and proposed methods, respectively. Secondly, we evaluate the performance with increasing UWB data frequency, given the same time-offset of 37.5ms and other settings. Fig. 6d shows the results of our method. The previous "position-focused" method does not use extra UWB data and hence is not included. It can be seen from the results depicted in Fig. 6b-d that:

- While the "position-focused" method suffers when the time-offset increases (Fig. 6b), the "range-focused" method always performs well (Fig. 6c). On the other hand, when there is no time-offset the performance of the two formulations is on-par with one another, which follows the analysis in Section IV-A.
- For both methods, as the UWB anchor is placed further away from the robot, the error increases (Fig. 6b-c). However, since the error of the proposed method is not affected by the time-offset, it would be more favorable for real applications.
- Nevertheless, there is no noticeable improvement in the performance when the UWB rate increases (Fig. 6d) for the proposed system. For this task, the result strongly relies on large movement and the anchor being in relatively close proximity to achieve desirable outcome.

B. Visual-Inertial-Range Odometry

The absolute trajectory error (ATE) is used to evaluate the odometry performance, which is a standard method for the evaluation of SLAM systems [36]. The estimated and ground truth trajectories are first aligned before the absolute pose differences are calculated to compute the overall ATE.

1) Simulation with EuRoC dataset: For the EuRoC dataset, one UWB anchor is simulated at the origin of the SLAM frame. We follow the description of VIR-SLAM [14] to make sure that the simulation setting is identical for comparison. Since all of the "MH" sequences starts with a handheld bootstrap movement, the UWB anchor position is estimated before the actual flight. With the number of available ranges between two consecutive keyframes increases as more camera frames are skipped until a new keyframe, we have $|\mathcal{J}_k| \approx 2 - 4$ for most sequences (with the camera

ATE (m)	VINS-Mono (no UWB)	VIR-SLAM (position-focused)	Proposed (range-focused)			Other prediction models for ${}^{W}_{B}\mathbf{p}_j$	
			$\lvert\mathcal{J}_k\rvert = 1$	$\lvert\mathcal{J}_k\rvert = 2$	Any $\lvert\mathcal{J}_k\rvert$	IMU-based (16)	Model-based (17)
MH_01	0.186	0.178	0.142	0.110	**0.079**	0.132	0.137
MH_02	0.240	0.188	0.078	0.070	**0.061**	0.138	0.162
MH_03	0.271	0.260	0.161	0.155	**0.101**	0.142	0.111
MH_04	0.402	0.366	0.148	0.138	**0.124**	0.310	0.227
MH_05	0.388	0.291	0.217	0.155	**0.134**	0.220	0.164

TABLE I: Comparison of ATE (m) on EuRoC dataset. Results of VINS-Mono and VIR-SLAM are extracted from [14] verbatim. The full proposed system corresponds to the "Any $|\mathcal{J}_k|$" column (all UWB data in the sliding window are consumed). Experiments with other formulations for ${}^{W}_{B}\mathbf{p}_j$ use similar settings as the full proposed system. The best results are in **bold**.

ATE (m)	T265 SLAM (no UWB)	VINS-Mono (no UWB)	Position-focused	Proposed (Any $\lvert\mathcal{J}_k\rvert$)
Loop 01	0.238	0.270	0.174	**0.137**
Loop 02	0.249	0.212	0.175	**0.106**
Loop 03	0.455	0.263	0.366	**0.131**
Open 01	1.227	0.253	0.206	**0.186**
Open 02	1.179	0.583	0.308	**0.185**
Open 03	1.458	0.438	0.351	**0.282**

TABLE II: Comparison of ATE (m) in real-life experiments. The best results are highlighted in **bold**.

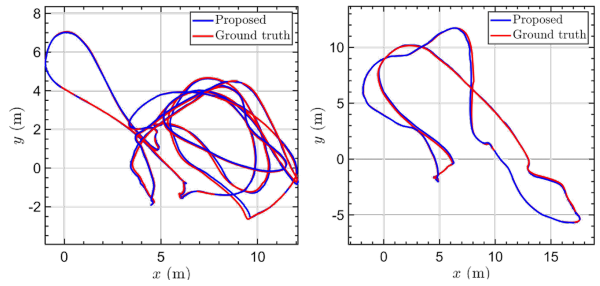

Fig. 7: Top view of the estimated (from the proposed system) and ground truth trajectories with MH_03 (left) and MH_04 (right) sequences in the EuRoC datset.

and simulated UWB having the same data frequency of 20Hz). In Table I, the full proposed system corresponds to the column "Any $|\mathcal{J}_k|$" (all available ranges are used). We further evaluate the system by limiting the maximum number of range measurements associated to one keyframe ($|\mathcal{J}_k| = 1, 2$). Fig. 7 depicts some of the estimated and ground truth trajectories.

It is clear from Table I that although VIR-SLAM - a "position-focused" method with complex formulation - improves upon the original VINS-Mono, our method outperforms when using only the same number of range measurements ($|\mathcal{J}_k| = 1$). As more ranges become available, the system is even further enhanced. However, in some cases the ATE improvement is only minor when more ranges data are included, which indicates that more UWB data does not necessarily warrant an enhancement in terms of performance. Additionally, the other formulations of ${}^{W}_{B}\mathbf{p}_j$ that can be used for the "range-focused" UWB residual (8) are tested. The selected solution (18) surpasses both IMU-based (16) and model-based (17) predictions in all of the dataset, while between (16) and (17) there is no clear distinction in terms of performance.

2) Real-life experiment: The hardware system consists of a platform equipped with the sensors as previously introduced (V-A). The experiments are designed to test the drift of the odometry: the "Loop" tests include several minutes of continuous movement in a $6\text{m} \times 6\text{m}$ area with Vicon motion capture system for ground truth, while the "Open" tests consists of various trajectories in a $30\text{m} \times 10\text{m}$ outdoor area with ground truth provided by a Leica MS60 laser tracking system. In all of the experiments, one UWB anchor is placed at an unknown position and relocated in every new test. The anchor position is estimated online during the operation.

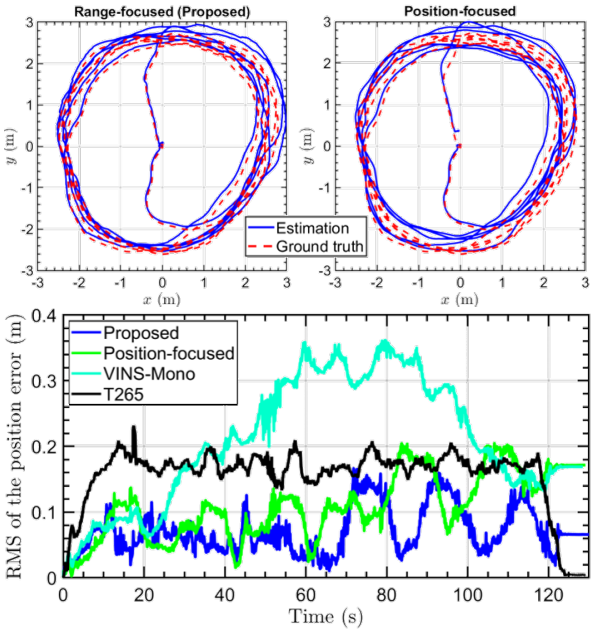

Fig. 8: Results of the visual-inertial-range odometry with the *Loop_02* author-collected dataset. Top: top view of the trajectories. Bottom: RMS of the position error.

Table II reports the ATE results. Comparison is done between the T265 VI-SLAM (stereo cameras + IMU), VINS-Mono (mono camera + IMU), our implementation for the "position-focused" system and the proposed "range-focused" method (mono camera + IMU + one UWB anchor). Fig. 8 shows the overview of the trajectories and the root mean

square (RMS) of the position error in the *Loop_02* dataset. For visual-inertial only methods (VINS-Mono and T265 VI-SLAM), the accumulated drift were not corrected over time which led to larger error. When UWB data is involved with the "position-focused" method, the performance is noticeably improved. Nonetheless, the proposed solution evidently excels in all of the experiments.

VI. Conclusions

In this letter, a new "range-focused" approach for fusion of camera-IMU-UWB sensors is presented. We leverage the propagated data readily available from the VIO system to compensate for the time-offset between UWB and camera sensors and allow all available UWB data to be used. This idea is incorporated into the two UWB-aided components: a UWB anchor localization module and a tightly-coupled optimization-based fusion of visual-inertial-range data to provide accurate and drift-reduced odometry in long-term operations. For both components, real-life and simulated experimental results verify that the proposed system outperforms previous "position-focused" approach. Extension to multi-robots scenarios is the main future research direction. Specifically, we wish to leverage the ranging data between the robots to not only improve each system's odometry but also combine the individual maps that do not share any common visual loop closures.

Acknowledgement

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, under the Wallenberg-NTU Presidential Postdoctoral Fellowship Program.

References

[1] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[2] V. Usenko, N. Demmel, D. Schubert, J. Stueckler, and D. Cremers, "Visual-Inertial Mapping with Non-Linear Factor Recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, 2020.

[3] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A Research Platform for Visual-Inertial Estimation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020. [Online]. Available: https://github.com/rpng/open_vins

[4] G. Cioffi and D. Scaramuzza, "Tightly-coupled Fusion of Global Positional Measurements in Optimization-based Visual-Inertial Odometry," pp. 5089–5095, 2020.

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

[6] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra-wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[7] M. R. Mahfouz, C. Zhang, B. C. Merkl, M. J. Kuhn, and A. E. Fathy, "Investigation of high-accuracy indoor 3-D positioning using UWB technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 6, pp. 1316–1330, 2008.

[8] T.-M. Nguyen, T. H. Nguyen, M. Cao, Z. Qiu, and 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.

[9] C. Wang, H. Zhang, T.-M. Nguyen, and 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.

[10] H. E. Nyqvist, M. A. Skoglund, G. Hendeby, and F. Gustafsson, "Pose estimation using monocular vision and inertial sensors aided with ultra-wideband," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015, pp. 1–10.

[11] M. W. Mueller, M. Hamer, and R. D'Andrea, "Fusing ultra-wideband range measurements with accelerometers and rate gyroscopes for quadrocopter state estimation," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1730–1736.

[12] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Tightly-Coupled Ultra-wideband-Aided Monocular Visual SLAM with Degenerate Anchor Configurations," *Autonomous Robots*, pp. 1–16, 2020.

[13] T. M. Nguyen, Z. Qiu, M. Cao, T. H. Nguyen, and L. Xie, "Single landmark distance-based navigation," *IEEE Transactions on Control Systems Technology*, pp. 1–8, 2019.

[14] Y. Cao and G. Beltrame, "VIR-SLAM: Visual, Inertial, and Ranging SLAM for single and multi-robot systems," *arXiv preprint arXiv:2006.00420*, 2020.

[15] 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 *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4289–4296.

[16] P. N. T. M. Nguyen, "Ranging-based adaptive navigation for autonomous micro aerial vehicles," Ph.D. dissertation, Nanyang Technological University, 2020.

[17] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Tightly-Coupled Single-Anchor Ultra-wideband-Aided Monocular Visual Odometry System," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[18] Q. Shi, X. Cui, W. Li, Y. Xia, and M. Lu, "Visual-UWB Navigation System for Unknown Environments," in *Proceedings of the 31st International Technical Meeting of The Satellite Division of the Institute of Navigation (ION-GNSS+ 2018)*. Institute of Navigation, Oct. 2018.

[19] D. Hoeller, A. Ledergerber, M. Hamer, and R. D'Andrea, "Augmenting ultra-wideband localization with computer vision for accurate flight," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12 734–12 740, 2017.

[20] A. Benini, A. Mancini, and S. Longhi, "An IMU/UWB/Vision-based Extended Kalman Filter for Mini-UAV Localization in Indoor Environment using 802.15.4a Wireless Sensor Network," *Journal of Intelligent & Robotic Systems*, vol. 70, no. 1-4, pp. 461–476, 2013.

[21] J. Li, Y. Bi, K. Li, K. Wang, F. Lin, and B. M. Chen, "Accurate 3D Localization for MAV Swarms by UWB and IMU Fusion," in *2018 14th IEEE International Conference on Control and Automation (ICCA)*, 2018, pp. 100–105.

[22] 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 *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3495–3502.

[23] Y. Song, M. Guan, W. P. Tay, C. L. Law, and C. Wen, "UWB/LiDAR Fusion For Cooperative Range-Only SLAM," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6568–6574.

[24] B. T. Fang *et al.*, "Simple solutions for hyperbolic and related position fixes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 5, pp. 748–753, 1990.

[25] E. Fernando, O. De Silva, G. K. Mann, and R. G. Gosine, "Observability Analysis of Position Estimation for Quadrotors With Modified Dynamics and Range Measurements," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 2783–2788.

[26] H. Xu, L. Wang, Y. Zhang, K. Qiu, and S. Shen, "Decentralized Visual-Inertial-UWB Fusion for Relative State Estimation of Aerial Swarm," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[27] T.-M. Nguyen, Z. Qiu, T. H. Nguyen, M. Cao, and L. Xie, "Persistently excited adaptive relative localization and time-varying formation of robot swarms," *IEEE Transactions on Robotics*, 2019.

[28] F. Molina Martel, J. Sidorenko, C. Bodensteiner, M. Arens, and U. Hugentobler, "Unique 4-DOF Relative Pose Estimation with Six Distances for UWB/V-SLAM-Based Devices," *Sensors*, vol. 19, no. 20, p. 4366, 2019.

[29] S. van der Helm, M. Coppola, K. N. McGuire, and G. C. de Croon, "On-board range-based relative localization for micro air vehicles in indoor leader–follower flight," *Autonomous Robots*, pp. 1–27, 2019.
[30] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.
[31] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics*. Springer, 1992, pp. 492–518.
[32] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-Manifold Preintegration for Real-Time Visual-Inertial Odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.
[33] J. Tiemann, A. Ramsey, and C. Wietfeld, "Enhanced UAV indoor navigation through SLAM-augmented UWB localization," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1–6.
[34] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.
[35] S. Agarwal and K. Mierle, *Ceres Solver: Tutorial & Reference*, Google Inc.
[36] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 573–580.