FEJ-VIRO: A Consistent First-Estimate Jacobian Visual-Inertial-Ranging Odometry

Shenhan Jia, Yanmei Jiao, Zhuqing Zhang, Rong Xiong and Yue Wang

Abstract—In recent years, Visual-Inertial Odometry (VIO) has achieved many significant progresses. However, VIO methods suffer from localization drift over long trajectories. In this paper, we propose a First-Estimates Jacobian Visual-Inertial-Ranging Odometry (FEJ-VIRO) to reduce the localization drifts of VIO by incorporating ultra-wideband (UWB) ranging measurements into the VIO framework consistently. Considering that the initial positions of UWB anchors are usually unavailable, we propose a long-short window structure to initialize the UWB anchors’ positions as well as the covariance for state augmentation. After initialization, the FEJ-VIRO estimates the UWB anchors’ positions simultaneously along with the robot poses. We further analyze the observability of the visual-inertial-ranging estimators and proved that there are four unobservable directions in the ideal case, while one of them vanishes in the actual case due to the gain of spurious information. Based on these analyses, we leverage the FEJ technique to enforce the unobservable directions, hence reducing inconsistency of the estimator. Finally, we validate our analysis and evaluate the proposed FEJ-VIRO with both simulation and real-world experiments.

I. INTRODUCTION

State estimation is a fundamental capacity to enable the operation of autonomous robot systems with applications such as autonomous driving, unmanned aerial vehicles and so on. Robust and reliable state estimation is essential in real world applications as large errors in state estimation can lead to the destruction of robots. In recent years, visual inertial odometry (VIO) is attracting increasing attention because of its lightweight, accuracy, and reliability, which has been successfully applied to many real-time robotic systems [1][2][3].

There is significant estimation drift during a long period of time in onboard self-localization mode of VIO system. Therefore, loop-closure and bundle adjustment (BA) techniques are required to correct drift, but are often accompanied by drastic changes in estimates that introduce destabilization to the robot [4]. So in live operation of robotic systems, these techniques are usually disabled to ensure stable operation. To achieve accurate onboard self-localization, an alternative solution is to fuse the external localization information from GPS [5] motion capture (mocap) [6], or artificial visual markers (such as AprilTags) [7]. However, GPS can only work in open spaces and mocap requires expensive and complex setup in a fixed indoor environment, neither can be rolled out to dense urban areas. The use of artificial visual markers is inflexible and it’s difficult to arrange the markers to cover a large area. Therefore, incorporating Ultra Wideband (UWB) measurements is a promising solution which can be deployed in both indoor and outdoor environments at low cost. In this paper, we focus on the situation that the positions of UWB anchors are unknown and require to be estimated simultaneously during operation.

For the fusion of UWB, there are two lines of methods. The first line is the optimization-based methods which fuse the UWB data by constructing cost function to optimize [8][4]. However, such methods are computationally intensive. Due to the limited on-board computing resources on robotic systems, the lightweight fusion methods are required. Another lightweight line is the filter-based methods which use extended Kalman filter (EKF) for state estimation [9]. The key problem in filtering-based methods is the state modeling of UWB measurements, including the design of state vector and the evaluation of its Jacobian matrices. After conducting experiments using existing filtering-based fusion methods, we find the accuracy decrease without reasonably dealing with the uncertainty. Therefore, an explicit theory is required to guide the state modeling of fusing UWB measurements in filtering-based approaches.

In this paper, we follow the lightweight line of filter-based methods to first propose a multi-state constraint Kalman filter (MSCKF) [10] based fusion solution to VIRO (as shown in Fig.1), which is less computationally intensive than EKF based methods and satisfies the on-board computing requirements of robotic systems. In addition, we propose two techniques to solve the state modeling of UWB measure-
ments in theory. One is for robust initialization of the UWB, the other is for consistent fusion of the UWB measurements into VIRO. For the initialization problem of UWB, the long-short sliding window strategy is proposed which solves the problem of initialization which allows to initialize the UWB anchors’ positions with sufficient measurements as well as to evaluate the covariance for state augmentation. For the consistent fusion of UWB measurements, we first analyze the observability of the VIRO system with unknown UWB anchors’ position and prove that there are four unobservable directions, which correspond to the global translation and the global rotation about the gravity, despite the fusion of ranging sensors. Although the VIRO system cannot achieve drift-free estimation, we can improve estimate accuracy by fusing the UWB ranging measurements. However, we prove that fusing the UWB measurements by direct EKF update operation causes inconsistency. Therefore, we propose a FEJ-VIRO framework which linearizes the ranging measurement models at the first-estimate of state, which is known as the FEJ technique, to address the inconsistency issue. Experiments show that the reasonable initialization and consistent filter designing bring obvious improvements to estimation accuracy to FEJ-VIRO. In summary, the contributions of this paper are listed as follows:

- A lightweight MSCKF-based Visual-inertial-ranging odometry (VIRO) method is proposed and a long-short sliding window strategy is presented for initializing UWB anchors and state augmentation.
- The observability analysis of visual-inertial-ranging estimator is derived which points out four unobservable directions of VIRO existing in the ideal case.
- A FEJ-VIRO framework is proposed which extends the FEJ technique to maintain consistency, leading to obvious improvement in estimation accuracy.
- Experiments on multiple simulation and real-world datasets validate the derived observability analysis and the effectiveness of the proposed FEJ-VIRO method.

II. RELATED WORK

There are extensive works on visual-inertial odometry (VIO) [3] [10] [11] and visual-inertial SLAM (VI-SLAM) [1] [2]. The main difference between VIO and VI-SLAM is whether to use global information: VI-SLAM gains better accuracy by performing mapping (and thus loop closure), while the localization errors of VIO grows unbounded due to the lack of global information [12]. However, it is well known that the computational complexity of VI-SLAM is significantly higher than VIO due to the iterative non-linear optimization, which makes it unsuitable to apply VI-SLAM to on-board devices [10] [3]. As a result, there are two lines of methods to achieve highly accurate estimation with bounded computational complexity: by reducing the complexity of VI-SLAM [13] [14] [15] and by incorporating global information into VIO framework [5] [4]. In this paper, we focus on the multi-sensor fusion methods to reduce localization drifts of VIO by incorporating ranging measurements from multiple unknown UWB anchors.

Over the last decades, many researchers have investigated in the use of UWB measurements for localization [9] [4] [8] [16]. Some early works [17] [18] [19] assume prior knowledge of anchor positions and achieve drift-free estimation of the robots. However, these methods in practice require an offline calibration of the initial position of robots, which is difficult in dynamic and large-scale environments. In recent years, researchers propose to estimate state and UWB anchor positions simultaneously. [9] propose a EKF-based framework to initialize anchor positions and estimate them jointly with the robot position. However, the initialization accuracy of this method is not sufficiently high and the algorithm fail easily due to the ill-conditioned of matrix. [20] propose to generate an initial guess of the UWB anchor position with a variant of the Levenberg-Marquardt method and simultaneously estimate the position with the robot state. [4] propose an estimator to fuse visual-inertial-ranging-lidar measurements where the UWB anchor positions are also assumed unknown and need to be estimated online, based on which [8] further propose mapping and loop-closure with these measurements. Note that most recent methods [20] [4] [8] are based on non-linear optimization and relies on global bundle adjustment to achieve high accuracy, which is computational intensive and hard to run on-board in real-time. In this paper, we propose a filter-based visual-inertial-ranging system which incorporates multi-sensor measurements in an efficient MSCKF framework. We also assume unknown UWB anchor position and propose a method to initialize UWB anchor positions as well as its initial variance and covariance with the state.

The observability and consistency of VIO has been extensively studied. [21] [22] proved that visual-inertial estimator has four unobservable directions which correspond to the global translation and the global rotation about the gravity. They further study the inconsistency of EKF-based VIO and propose a first-estimated Jacobian technique to address the inconsistency issue. [21] proved the key cause of inconsistency is the gain of spurious information along unobservable directions, which results in the over-confidence of the estimator. Extensive works studied this problem and proposed many techniques to address it. We refer the readers to [23] for detailed review. In this paper, we prove the general VIRO system also has four unobservable directions despite of the fusion UWB measurements. Although we cannot achieve drift-free estimation with VIRO, we can reduce the localization drifts of VIO by incorporating ranging measurements. However, from the unobservable directions of the VIRO, we find that fusing the ranging measurements with standard EKF-update results in inconsistency, which is caused by the estimator’s over-confidence (as [21]). Based on these analyses, we propose leveraging FEJ technique to address the inconsistency issue.

III. ESTIMATOR DESIGN

In this section, we describe in detail the filter based VIRO, which fuses measurements from UWB, cameras, and IMU in a MSCKF framework. We assume that there are
three UWB anchors available in the environment, and we need to estimate their global positions simultaneously. An illustration of our system can be seen in 1, where the dash lines represents sensor measurements, and we write the states needs to be estimated with blue font.

Before describing our estimator, we first present some notations. There are three coordinate frames in this paper, which are the global frame $\mathcal{F}_G$, the IMU frame $\mathcal{F}_I$ and the camera frame $\mathcal{F}_C$. Fig.2 illustrates the frames and sensors used in the VIRO.

A. State Vector

The state vector of the proposed VIRO at time $t_k$ consists of the current IMU states, the position of SLAM features, the position of UWB anchors, and a short window and a long window which contain cloned IMU poses corresponding to the past images:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{I,k}^T & \mathbf{G} \mathbf{p}_{f,k}^T & \mathbf{x}_{a,k}^T & \mathbf{x}_{c,l,k}^T & \mathbf{x}_{e,k}^T \end{bmatrix}^T$$

(1)

$$\mathbf{x}_{I,k} = \begin{bmatrix} \mathbf{l}_{\mathcal{F}_G}^T \mathbf{b}_g^T \mathbf{G} \mathbf{v}_{I,k}^T \mathbf{b}_a^T \mathbf{G} \mathbf{p}_{I,k}^T \end{bmatrix}^T$$

(2)

$$\mathbf{x}_{a,k} = \begin{bmatrix} \mathbf{G} \mathbf{p}_{a_1}^T \mathbf{G} \mathbf{p}_{a_2}^T \mathbf{G} \mathbf{p}_{a_3}^T \end{bmatrix}^T$$

(3)

$$\mathbf{x}_{c,l,k} = \begin{bmatrix} \mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{f,l_1}^T \cdots \mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{f,k-m+1}^T \end{bmatrix}^T$$

(4)

$$\mathbf{x}_{e,k} = \begin{bmatrix} \mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{f,l_1}^T \cdots \mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{f,l_M}^T \end{bmatrix}^T$$

(5)

where $\mathbf{G} \mathbf{v}_I$ is the IMU velocity in the global frame; $\mathbf{b}_g$ and $\mathbf{b}_a$ denote the gyroscopic and accelerometric biases; $\mathbf{G} \mathbf{p}_{f,k}$ is the feature’s position expressed in the global frame; $\mathbf{G} \mathbf{p}_{a_i}$ ($i = 1, 2, 3$) are the position of UWB anchors expressed in the global frame; $\{\mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{l,k-i} \}$ ($i = 0, \cdots, m - 1$) are the cloned IMU poses at time $t_k - i$; and $\{\mathbf{l}_{\mathcal{G}}^T \mathbf{G} \mathbf{p}_{l,j} \}$ ($j = 1, \cdots, M$) are the long window for UWB anchor's initialization, which will be introduced in Sec.IV.

B. IMU Propagation

The state is propagated forward with the IMU linear acceleration and angular velocity:

$$\omega_m(t_k) = \dot{\mathbf{A}} \omega(t_k) + \mathbf{b}_g(t_k) + \mathbf{n}_g(t_k)$$

(6)

$$\mathbf{a}_m(t_k) = \dot{\mathbf{G}} \mathbf{R}(t_k) (\mathbf{G} \mathbf{a}(t_k) + \mathbf{G} \mathbf{g}) + \mathbf{b}_a(t_k) + \mathbf{n}_a(t_k)$$

(7)

where $\omega_m$ and $\mathbf{a}_m$ are the raw inertial measurement data, $\mathbf{G} \mathbf{g}$ is the gravitational acceleration expressed in $\mathcal{F}_G$, and $\mathbf{n}_g$ and $\mathbf{n}_a$ are zero-mean white Gaussian noise. We propagate the state estimate and the covariance from time $t_k$ to $t_{k+1}$ based on the inertial kinematic model $f(\cdot)$ [10]:

$$\mathbf{x}_{k+1|k} = f(\mathbf{x}_{k|k}, \mathbf{a}_m(t_k : t_{k+1}), \omega_m(t_k : t_{k+1}, 0, 0))$$

(8)

$$\mathbf{P}_{k+1|k} = \Phi(t_{k+1}, k) \mathbf{P}_{k|k} \Phi(t_{k+1}, k)^T + \mathbf{Q}_{a,k}$$

(9)

where the zero-mean noise vectors are represented by the last two 0 entries of (8), and $\Phi_k$ and $\mathbf{Q}$ are the state transition matrix and discrete noise covariance.

C. Camera measurement model

In this subsection, we review concisely the visual observation model with the calibrated perspective camera assumption. Specifically, at time $t_k$, the position of a tracked feature point in the camera frame is $\mathbf{G} p_f = [x_j y_j z_j]^T$. The feature measurement can be obtained by projecting its 3D pose onto the image plane with the projection model. For $j = k + 1, \cdots, k - m$, a range-bear measurement model can be written as:

$$\mathbf{z}_{n,j} = \Pi(\mathbf{G} p_f) + \mathbf{n}_z = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} + \mathbf{n}_z$$

(10)

$$c_j p_f = c_i R_f^T R (\mathbf{G} p_f - \mathbf{G} p_{l,j}) + c_p$$

(11)

where $\mathbf{G} p_f$ is the 3D position of the tracked feature in $\mathcal{F}_G$; and $\mathbf{n}_z$ and $\mathbf{n}_o$ are the zero-mean white Gaussian measurement noise of range and bear measurements. We refer readers to [10] for detailed presentation of visual update and null-space projection.

D. Ranging Update

We leverage UWB ranging measurements to update the robot pose and the global position of UWB anchor nodes. Following [16], the UWB ranging measurement from the i-th UWB anchor node, $d_{r,i}$, can be modeled as:

$$d_{r,i} = h(x_I, \mathbf{G} p_{a_i})$$

$$= \| \mathbf{G} p_f + \mathbf{G} R_{f}^T \mathbf{p}_r - \mathbf{G} p_{a_i} \| + d_{bias}$$

(12)

where $\{\mathbf{G} R, \mathbf{G} p_f \}$ is the IMU pose; $\mathbf{G} p_f$ is the position of UWB ranging node expressed in the IMU frame, which can be calibrated offline; $\mathbf{G} p_{a_i}$ is the global position of the i-th UWB anchor node; and $d_{bias}$ is the bias of distance measurements, which can be calibrated offline [24].

The distance measurements between UWB anchor nodes are available to the robot by robot-node communication. The
distance measurement between the i-th UWB anchor and the j-th UWB anchor, \( d_{e_{ij}} \), can be modeled as:

\[
d_{e_{ij}} = \|G_{p_{ai}} - G_{p_{aj}}\| + d_{bias}
\]

(13)

where \( G_{p_{ai}} \) and \( G_{p_{aj}} \) are the global position of the i-th UWB anchor node and the j-th UWB anchor node; and \( d_{bias} \) is the bias of distance measurements, which can be calibrated offline [24].

We can derive a concise expression of Jacobian by defining \( d_{2r_{ij}} = (d_{e_{ij}} - d_{bias})^2 = \|G_{p_{i}} + f^{\prime}_G R^T p_r - G_{p_{j}}\|^2 \). Then we perturb \( d_{2r_{ij}} \) to derive the measurement Jacobians:

\[
\frac{\partial d_{2r_{ij}}}{\partial G_{p_{ai}}} = 2 \cdot (G_{p_{i}} - G_{p_{ai}})^T
\]

(14)

\[
\frac{\partial d_{2r_{ij}}}{\partial G_{p_{aj}}} = 2 \cdot (G_{p_{j}} - G_{p_{aj}})^T
\]

(15)

\[
\frac{\partial d_{2r_{ij}}}{\partial f_{G}} = 2 \cdot (G_{p_{i}} - G_{p_{j}})^T f_{G} R^T [1 p_r \times]
\]

(16)

where \( G_{p_{i}} = G_{p_{f}} + f^{\prime}_G R^T p_r \) is the position of UWB ranging node represented in the global frame; and \( f_{G} \) is calibrated offline.

Also in the same way, by defining \( d_{2e_{ij}} = (d_{e_{ij}} - d_{bias})^2 \), we have:

\[
\frac{\partial d_{2e_{ij}}}{\partial G_{p_{ai}}} = 2 \cdot (G_{p_{i}} - G_{p_{ai}})^T
\]

(17)

\[
\frac{\partial d_{2e_{ij}}}{\partial G_{p_{aj}}} = 2 \cdot (G_{p_{j}} - G_{p_{aj}})^T
\]

(18)

Now we can perform EKF update with the square-unbiased ranging measurements.

**E. Data synchronization**

In this work, we interpolate the ranging measurements to synchronize the visual and ranging measurements. Assume that we receive an image (or stereo images) at times \( t_k \) and the adjacent ranging measurements are received at \( t_\alpha \) and \( t_\beta \), where \( t_\alpha < t_k < t_\beta \), the aligned ranging measurement is computed by:

\[
d_{r_{,k}} = d_{r_{,\alpha}} + \frac{t_k - t_\alpha}{t_\beta - t_\alpha}(d_{r_{,\beta}} - d_{r_{,\alpha}})
\]

(19)

where \( d_{r_{,\alpha}} \) and \( d_{r_{,\beta}} \) are ranging measurements received at \( t_\alpha \) and \( t_\beta \). Note that both \( d_{r_{,\alpha}} \) and \( d_{r_{,\beta}} \) should be from the same UWB anchor node. In practice, if the time offsets \( t_k - t_\alpha \) and \( t_\beta - t_k \) are bigger than a threshold, we discard this pair of ranging measurements.

**IV. UWB INITIALIZATION**

We design a long-short-window structure to estimate the global position as well as the initial covariance. The short window is the same as the MSCKF sliding window, while the long window contains a list of key-frames selected according to the distance interval. Once the length of the long window is over a threshold, we leverage the poses in the long window and corresponding ranging measurements to estimate all UWB anchors’ global positions.

\[
\begin{array}{c|c|c|c|c}
\hline
\text{state} & x_f & \ell & x_t & \ell_p \\hline
\text{cov.} & P_{cc} & P_{ca} & P_{ac} & P_{aa} \\hline
\end{array}
\]

Fig. 3: Illustration of the state augmentation step after the UWB anchors’ initialization. We evaluate the covariance matrix blocks related to the UWB anchor positions and augment the state vector and covariance matrix after that.

Assume that there are \( n \) poses in the long-short window, we build the linear problem following [9]:

\[
\begin{bmatrix}
-2x_1(t_1) & -2y_1(t_1) & -2z_1(t_1) & 1 \\
-2x_2(t_2) & -2y_2(t_2) & -2z_2(t_2) & 1 \\
\vdots & \vdots & \vdots & \vdots \\
-2x_n(t_n) & -2y_n(t_n) & -2z_n(t_n) & 1 \\
\end{bmatrix} \begin{bmatrix}
\Delta(t_1) \\
\Delta(t_2) \\
\vdots \\
\Delta(t_n) \\
\end{bmatrix} = \begin{bmatrix}
D_{x_1} \\
D_{y_1} \\
D_{z_1} \\
D_{a_i} \\
\vdots \\
D_{a_n} \\
\end{bmatrix}
\]

(20)

with

\[
D_{a_i} = x_{a_i}^2 + y_{a_i}^2 + z_{a_i}^2
\]

(21)

\[
\Delta = (d_{r_{,i}} - d_{bias})^2 - (x_r^2 + y_r^2 + z_r^2)
\]

(22)

where \( G_{p_{r}} = G_{p_{f}} + f^{\prime}_G R^T \) \( p_r \) is the global position of the UWB ranging node; \( G_{p_{a_i}} = [x_{a_i}, y_{a_i}, z_{a_i}]^T \) is the i-th anchor’s global position; \( d_{r_{,i}} \) is the raw measurement from the i-th UWB anchor node; and \( d_{bias} \) is the bias of ranging measurements. By solving (20), we can roughly estimate the global position of the i-th UWB anchor node.

Since the matrix A in (20) is usually ill-conditioned, we jointly optimize all anchor nodes’ global position by minimizing the following cost function:

\[
\sum_{p_{a_i}} \sum_{k=1}^{n} \frac{(d_{r_{,k}} - d_{bias} - \|G_{p_{a_i}} - G_{p_{r_{,k}}}\|)^2}{\sigma_r^2} + \sum_{p_{a_i}} \sum_{d_{c} \in \Omega_{ij}} \frac{(d_c - d_{bias} - \|G_{p_{a_i}} - G_{p_{a_j}}\|)^2}{\sigma_c^2}
\]

(23)

where \( d_{r_{,k}} \) is the ranging measurement from the i-th UWB anchor node at time-step \( k \); \( G_{p_{r_{,k}}} = G_{p_{f}} + f^{\prime}_G R^T \) \( p_r \) is the global position of the UWB ranging node; \( \sigma_r^2 \) is the noise density of ranging measurements; \( d_c \) is the ranging measurements between two UWB anchor nodes; \( \Omega_{ij} \) is the set of measurements between the i-th and the j-th UWB anchor nodes during the last n time-step; \( \sigma_c^2 \) is the noise density of ranging measurements between UWB anchor nodes.

By solving (23) with Gaussian-Newton method, we can obtain refined global positions of all UWB anchor nodes.
$Gp_a_i (i = 1, 2, 3)$. The last step of our UWB initialization is to estimate the initial covariance matrix. In this work, we estimate the covariance of the $i$-th anchor node with all of the $p$ ranging measurements: $d_r = [d_{r,1}^T d_{r,2}^T \ldots d_{r,n}]^T$, where $d_{r,k} = h_k(x_r, Gp_a_i) + n_{r,k}$ is the ranging measurement from the $i$-th anchor node to the $k$-th time-step. Following [25], we linearize (12):

$$r_i = d_{r,i} - h(\tilde{x}_f, Gp_a_i)$$

$$\simeq H_x \tilde{x}_f + H_a Gp_a_i + n_i$$

(24)

(25)

Since the number of ranging measurements is larger than the dimension of $Gp_a_i$, we perform Givens-rotation to zero-out the excess rows in $H_a$:

$$\begin{bmatrix} r_{11} \\ r_{12} \end{bmatrix} = \begin{bmatrix} H_{x1} \\ H_{x2} \end{bmatrix} \tilde{x}_f + \begin{bmatrix} H_{a1} \\ 0 \end{bmatrix} Gp_a_i + \begin{bmatrix} n_{11} \\ n_{12} \end{bmatrix}$$

(26)

As a result, we evaluate the covariance matrix by:

$$P_{aa} = H_{a1}^{-1}(H_{x1}P_{xx}H_{x1}^T + \sigma_r^2 I)H_{a1}^{-T}$$

$$P_{xa} = -P_{xx}H_{x1}^T H_{a1}^{-T}$$

(27)

(28)

where $\sigma_r$ is the noise density of ranging measurements. Finally we augment the state and the covariance matrix.

V. CONSISTENCY ANALYSIS

Without loss of generality, we study the observability based on a general EKF state vector with one visual feature and two UWB anchor nodes:

$$x = \{Gf, Gv, Gp_f, Gp_i, Gp_j, Gp_a, Gp_{a1}\}^T$$

(29)

where $\{Gf, Gp_f\}$ is the IMU pose in the global frame; $Gv$ is the IMU velocity; $b_g$ and $b_a$ are the IMU bias; $Gf$ is the visual feature’s global position; and $Gp_a_i, (i = 1, 2, 3)$ is the global position of UWB anchor node. Note that by considering two UWB anchor nodes, we can take ranging measurements between anchors into our analysis, and this setting can be easily generalized to the cases with multiple UWB anchors.

A. Estimator Observability

During the IMU propagation described in Sec.III-B, the state transition matrix from the starting time-step $t_1$ to $t_k$ has the following form:

$$\Phi_{k,1} = \begin{bmatrix} \Phi_{t_k,t_1} \\ 0_{9 \times 15} \\ 0_{9 \times 9} \end{bmatrix}$$

(30)

$$\Phi_{t_k,t_1} = \begin{bmatrix} \phi_{t_1} \phi_{t_2} 03 & 01 & 03 \\ 01 & 03 & 03 \\ 03 & 03 & 03 \end{bmatrix}$$

(31)

where the analytical form of $\Phi_{t_k,t_1}$ can be found in [21]. From Sec.III-D, the Jacobian matrix of UWB ranging measurements is:

$$H_{a_k} = \begin{bmatrix} \frac{\partial Gp_a_i}{\partial x} R^T [p_i \times] & 0_{15 \times 9} \\ \frac{\partial Gp_a_i}{\partial x} R^T [p_i \times] & 0_{15 \times 9} \end{bmatrix} + \begin{bmatrix} -\zeta_i & 0 & 0 \\ 0 & -\zeta_i & 0 \end{bmatrix}$$

(32)

$H_{a_k}$ is a $3 \times 24$ matrix, and

$$\zeta_i = 2 \cdot (Gp_a_i - Gp_{a,k} - \frac{1}{2}R^T [p_i \times])^T$$

(33)

$$\Lambda_a = 2 \cdot (Gp_a_i - Gp_{a,k})^T$$

(34)

Note that the first two block rows of (32) corresponds to the ranging measurements between UWB anchor nodes and the UWB ranging node, while the last block row corresponds to the ranging measurements between two UWB anchor nodes. By combining (30) and (32), we have:

$$H_{k,u} \Phi_{k,1} =$$

$$\begin{bmatrix} \Gamma_{11} & \Gamma_{12} - \zeta_i \phi_{33} - \zeta_i \phi_{44} - \zeta_i 0 & \zeta_i 0 \\ \Gamma_{21} & \Gamma_{22} - \zeta_i \phi_{33} - \zeta_i \phi_{44} - \zeta_i 0 & 0 \zeta_i \end{bmatrix}$$

(35)

where $\Gamma_{11} = 2 \cdot I_k R^T [p_i \times] \phi_{11} - \zeta_i \phi_{51}$, $i = 1, 2$ (36)

$$\Gamma_{22} = 2 \cdot I_k R^T [p_i \times] \phi_{12} - \zeta_i \phi_{52}$, $i = 1, 2$ (37)

By study the analytical form of $\Theta$, we have the following theorem:

Theorem 1. (Ideal observability proprieties of VIRO) The right null-space $N_o$ of the observability matrix $\Theta$ of the linearized visual-inertial-ranging estimator

$$\Theta \cdot N_o = 0$$

(38)

is spanned by the following four directions:

$$N_o = \begin{bmatrix} 0_3 & I_k R \cdot Gg \\ 0_3 & 0_3 \\ 0_3 & -Gv \times ] \cdot Gg \\ I_3 & -Gp_f \times Gg \\ I_3 & -Gp_a \times Gg \\ I_3 & -Gp_{a1} \times Gg \end{bmatrix}$$

(39)

Proof. See Appendix of [26].

Theorem 2. (Actual observability proprieties of VIRO) For the actual case where the ranging measurement Jacobians are evaluated at the estimated states. The right null-space $N_o$ of the observability matrix $\Theta$ of the linearized visual-inertial-ranging estimator is spanned by the following three directions:

$$N_o = [0_3, 0_3, 0_3, I_3, I_3, I_3]^T$$

(40)

Proof. See Appendix of [26].

From theorem 1, we can not achieve drift-free state estimate with visual-inertial-ranging estimator, because the global translation and yaw rotation are unobservable. But we can improve the estimation accuracy by multi-sensor fusion algorithms. The theorem 2 states that if we linearize the ranging measurement models at the latest estimated state, the VIRO system gains spurious information about the global yaw direction, which results in the inconsistency of the estimator.
Fig. 4: Normalized estimation error squared (NEES) results of OpenVINS, VIRO, FEJ-VIRO-S and FEJ-VIRO tested on three different simulation datasets. Where the FEJ-VIRO-S is the FEJ-VIRO without long-window. We enabled FEJ for OpenVINS and for the visual-inertial part of VIRO, so the inconsistency of VIRO only results from state update with UWB ranging measurements.

B. Consistent Filter Design

To design consistent state estimator, we need to force the unobservable subspace of the estimated model (the actual case) to be the same as that of the ideal case, hence preventing spurious information gain and reducing inconsistency. The inconsistency issue also happens in the EKF-based VIO, and several approaches has been proposed to design consistent filter, such as the first-estimate Jacobian (FEJ) idea [27] and the OC methodology [28].

In this paper, we extend the traditional FEJ to the VIRO system by always evaluating the measurement Jacobian matrices at the first-estimated robot pose and UWB anchors’ positions. And we follow [27] to perform FEJ for the IMU propagation and visual update. With these changes, the null space of the observability matrix $\mathbf{O}$ in the actual case keeps the same as that in the ideal case.

VI. EXPERIMENTS

We implement the proposed consistent visual-inertial-ranging filter based on OpenVINS [3], which is the state-of-the-art filter based visual-inertial estimator. We incorporate UWB ranging measurements in it and also implemented the UWB initialization described in Sec.IV. The non-linear optimization problem (23) is solved by Ceres-Solver [29].

A. Simulation Experiments

The main purpose of simulation measurements is to validate the inconsistency issue described in V-B and further evaluate the consistency of the proposed VIRO. We expanded the simulation utility provided by Open-VINS to additionally simulate ranging measurements from three UWB anchor nodes with noise density of 0.15m and bias of -0.75m. In order to validate the inconsistency issue described in V-B, we tested the proposed VIRO on three different simulation datasets, and evaluate the normalized estimation error squared (NEES, see p31 of [30]) of them.

We compare the NEES of OpenVINS, VIRO and FEJ-VIRO to validate the inconsistency issue described in V-B and evaluate the consistency of the proposed FEJ-VIRO. The NEES results are listed in Fig.4 Note that we enabled FEJ for OpenVINS, and we only disabled VIRO’s FEJ when updating state with ranging measurements, so the inconsistency of VIRO results only from state update with UWB ranging measurements.

Fig.5 shows that the standard updates with UWB ranging measurements results in large NEES. The dotted line in Fig.4 represents the 3 degree of freedoms of NEES. The OpenVINS exceeds the dotted line sometimes because the visual-inertial estimator degrades when the platform is moving with constraints (e.g. planar motion). By leveraging FEJ technique, the inconsistency issue is solved and results in lower NEES of FEJ-VIRO. To further evaluate the proposed long-short window structure, we compare it with the method with short-only sliding window for UWB initialization (see FEJ-VIRO-S in Fig.4). In the FEJ-VIRO-S,
the long-window poses in (5) are marginalized out from the state, so we can not estimate the covariance by performing (24) to (26). The FEJ-VIRO-S initializes the $P_{zz}$ in (27) with the covariance provided by Cere-Solver and initialize the cross-covariance $P_{xz}$ with zero. Note that the estimator fails if we initialize the UWB anchors' positions with only poses in the short-window.

We know from Sec.V-B that when we linearize the ranging measurement models at the current state estimates, the number of unobservable directions decreases from four to three, where the rotation about the gravity becomes (wrongly) observable. To further validate our theory, we plot the 3-$\sigma$ bounds of VIRO and FEJ-VIRO (shown in Fig.5) tested on three simulation datasets. Fig.5 shows that the 3-$\sigma$ (uncertainty) of the z-axis rotation estimated by VIRO reduces rapidly after the UWB anchors are initialized. However, we have proved in Sec.V-A that the rotation about the gravity is unobservable for visual-inertial-ranging estimator, so the VIRO is over-confident about the estimated z-axis rotation. The over-confidence is the main cause of inconsistency.

Fig.5 shows that the 3-$\sigma$ bounds of z-axis rotation estimated by FEJ-VIRO do not degrade. By linearize the model and ranging measurement models at the first estimated state, the number of unobservable directions of the estimated system is the same as that of the ideal system. The FEJ-VIRO is not over-confident about the estimated z-axis rotation, which is the key to the consistency and low NEES shown in Fig.4.

### B. Real-world Experiments

We further evaluate our method based on a challenging real-world dataset, VIRAL [16], which provides measurements from extensive sensors, such as IMU, stereo cameras, and three UWB anchor nodes.

VIRO and FEJ-VIRO are tested on six trajectories available in the VIRAL dataset. The only difference between VIRO and FEJ-VIRO is the linearization point of UWB ranging measurement models. We use 11 clones and 200 features for the real-time estimation. Ranging measurements from three UWB anchors and echo ranging measurements between each pair of UWB anchor nodes are used. For UWB initialization, we select key-frame poses every 0.3m to build the long window, and perform UWB initialization (see Sec.IV) when there are over 50 poses in the long window.

We compare VIRO and FEJ-VIRO with the original OpenVINS to validate that our methods improve the accuracy and reduce localization drifts of VIO in long distance trajectories. We also compare our methods with VINS-Fusion [31], which is the state-of-the-art VI-SLAM estimator based on sliding-window smoothing. As discussed in Sec.II, the VINS-Fusion leverages global information (loop-closure) and global non-linear optimization to achieve bounded error with significantly higher computational complexity. We use the configure files provided by [32] with for experiments.

We record the trajectory of these methods and evaluate with the methods recommended by [16]. The OpenVINS configurations are available on the home page of VIRAL [32]. ATE results are listed in Tab.I. VINS-Fusion outputs the optimized path after loop-closure and the real-time pose estimates. Since poses of the optimized path utilize future information to improve accuracy and smoothness, which is non-causal and hardly used by motion controllers in practice, we compare our methods with the real-time pose estimates and only list the accuracy of optimized path for reference. The column of VINS-Fusion (RT) in Tab.I represents the real-time pose estimates, while the column of VINS-Fusion (LC) represents the optimized path after loop-closure. We set to bold the best accuracy among OpenVINS, VINS-Fusion (RT), VIRO and the proposed FEJ-VIRO. We also set the VINS-Fusion (LC) to italic if it outperforms FEJ-VIRO although it is listed only for reference. The estimated trajectories by FEJ-VIRO as well as the groundtruth are plotted in Fig.6. From I, our FEJ-VIRO is accurate and is nearly as accurate as the optimized path of VINS-Fusion. Note that our methods only requires an UWB anchor initialization at the beginning and an FEJ-EKF update during the running. On the other hand, although the optimized path of VINS-Fusion is in high accuracy, the real-time pose estimates suffer from sudden changes due to the loop-closure adjustment. While

![Fig. 6: Trajectories estimated by the proposed FEJ-VIRO (blue) and the corresponding groundtruth (red) of three runs available in the VIRAL dataset.](image)

| TABLE I: Average ATE (m) of six runs in VIRAL Dataset |
|----------|----------|----------|----------|----------|
| OpenVINS | VINS-Fusion (RT / LC) | VIRO | VIRO |
| eee_01   | 0.647    | 0.564 / 0.364 | 0.824 | **0.479** |
| eee_02   | 0.604    | 0.582 / 0.252 | 0.506 | **0.340** |
| eee_03   | 0.417    | 0.540 / 0.390 | 0.350 | **0.348** |
| nya_01   | 0.741    | 0.538 / 0.235 | 0.552 | **0.505** |
| nya_02   | 0.478    | 0.375 / 0.331 | 0.368 | **0.202** |
| nya_03   | 0.692    | 0.748 / 0.370 | 0.467 | **0.290** |

We compare VIRO and FEJ-VIRO with the original OpenVINS to validate that our methods improve the accuracy and reduce localization drifts of VIO in long distance trajectories. We also compare our methods with VINS-Fusion [31], which is the state-of-the-art VI-SLAM estimator based on sliding-window smoothing. As discussed in Sec.II, the VINS-Fusion leverages global information (loop-closure) and global non-linear optimization to achieve bounded error with significantly higher computational complexity. We use the configure files provided by [32] with for experiments.

We record the trajectory of these methods and evaluate with the methods recommended by [16]. The OpenVINS configurations are available on the home page of VIRAL [32]. ATE results are listed in Tab.I. VINS-Fusion outputs the optimized path after loop-closure and the real-time pose estimates. Since poses of the optimized path utilize future information to improve accuracy and smoothness, which is non-causal and hardly used by motion controllers in practice, we compare our methods with the real-time pose estimates and only list the accuracy of optimized path for reference. The column of VINS-Fusion (RT) in Tab.I represents the real-time pose estimates, while the column of VINS-Fusion (LC) represents the optimized path after loop-closure. We set to bold the best accuracy among OpenVINS, VINS-Fusion (RT), VIRO and the proposed FEJ-VIRO. We also set the VINS-Fusion (LC) to italic if it outperforms FEJ-VIRO although it is listed only for reference. The estimated trajectories by FEJ-VIRO as well as the groundtruth are plotted in Fig.6. From I, our FEJ-VIRO is accurate and is nearly as accurate as the optimized path of VINS-Fusion. Note that our methods only requires an UWB anchor initialization at the beginning and an FEJ-EKF update during the running. On the other hand, although the optimized path of VINS-Fusion is in high accuracy, the real-time pose estimates suffer from sudden changes due to the loop-closure adjustment. While
our method does not result in jump of the real-time pose estimation, thus it is a better approach to on-board estimator that collaborate with motion controllers.

VII. CONCLUSIONS

In this paper, we analyze the observability of visual-inertial-ranging estimator and propose a consistent filter incorporating measurements from cameras, IMU, and multiple ultra-widebands (UWB). In particular, we prove that four unobservable directions exist in the VIRO system, which means that we can not design drift-free odometry by fusing measurements from UWB, cameras and IMU. We further study the structure of the estimated system to show that there are only three unobservable directions when we evaluate the Jacobians at the latest state estimates during every time step, which result in inconsistency. Based on these analyses, we leverage the FEI technique to fuse UWB measurements consistently in a tightly-coupled MSCKF framework. Finally, we validate our analysis and the proposed system with both simulation and real-world experiments.

For future works, we would try other methods to address the inconsistency issue such as invariant filter or observability constraint (OC) technique to achieve higher accuracy.

REFERENCES

[1] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, “Orb-slam3: An accurate open-source library for visual-inertial, and multi-tap slam,” IEEE Transactions on Robotics, 2021.

[2] 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.

[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/rgp/openvins

[4] T.-M. Nguyen, M. Cao, S. Yuan, Y. Lyu, T. H. Nguyen, and L. Xie, “Viral-fusion: A visual-inertial-ranging-lidar sensor fusion approach,” IEEE Transactions on Robotics, 2021.

[5] W. Lee, K. Eckenhoff, P. Geneva, and G. Huang, “Intermittent gps-aided vio: Online initialization and calibration,” in Proc. of the IEEE International Conference on Robotics and Automation, Paris, France, 2020.

[6] A. Kushlevy, D. Mellinger, C. Powers, and V. Kumar, “Towards a swarm of agile micro quadrotors,” Autonomous Robots, vol. 35, no. 4, pp. 287–300, 2013.

[7] J. P. d. A. Barbosa, S. S. Dias, and D. A. d. Santos, “A visual-inertial navigation system using apriltag for real-time mav applications,” in 2018 25th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2018, pp. 1–7.

[8] T.-M. Nguyen, S. Yuan, M. Cao, T. H. Nguyen, and L. Xie, “Viral slam: Tightly coupled camera-imu-ulb-lidar slam,” arXiv preprint arXiv:2105.03296, 2021.

[9] 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.

[10] A. Mourikis and S. Roumeliotis, “A multi-state constraint kalman filter for vision-aided inertial navigation,” in 2007 IEEE International Conference on Robotics and Automation, ICRA ’07, 2007, pp. 3565–3572.

[11] C. Förster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, “Svo: Semidirect visual odometry for monocular and multicamera systems,” IEEE Transactions on Robotics, vol. 33, no. 2, pp. 249–265, 2016.

[12] G. Huang, “Visual-inertial navigation: A concise review,” in 2019 international conference on robotics and automation (ICRA). IEEE, 2019, pp. 9572–9582.

[13] 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.

[14] M. Kaess, A. Ranganathan, and F. Dellaert, “isam: Incremental smoothing and mapping,” IEEE Transactions on Robotics, vol. 24, no. 6, pp. 1365–1378, 2008.

[15] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, “isam2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering,” in 2011 IEEE International Conference on Robotics and Automation. IEEE, 2011, pp. 3281–3288.

[16] T.-M. Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, “Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint,” International Journal of Robotics Research (accepted, to appear), 2021.

[17] 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). IEEE, 2017, pp. 1602–1609.

[18] 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). IEEE, 2018, pp. 1–6.

[19] 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). IEEE, 2017, pp. 3495–3502.

[20] 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). IEEE, 2020, pp. 665–672.

[21] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, “Consistency analysis and improvement of vision-aided inertial navigation,” IEEE Transactions on Robotics, vol. 30, no. 1, pp. 158–176, 2014.

[22] M. Li and A. I. Mourikis, “High-precision, consistent ekf-based visual-inertial odometry,” The International Journal of Robotics Research, vol. 32, no. 6, pp. 690–711, 2013.

[23] C. Cadenas, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” IEEE Transactions on robotics, vol. 32, no. 6, pp. 1309–1332, 2016.

[24] B. van der Heijden, A. Ledergerber, R. Gill, and R. D’Andrea, “Inertive bias estimation for an ultra-wideband localization system,” IFAC-PapersOnLine, vol. 53, no. 2, pp. 1391–1396, 2020.

[25] M. Li, Visual-inertial odometry on resource-constrained systems. University of California, Riverside, 2014.

[26] S. Jia, Y. Jiao, Z. Zhang, R. Xiong, and Y. Wang, “Fej-viro: A consistent first-estimate jacobian visual-inertial-ranging odometry,” arXiv preprint arXiv:2207.08224, 2022.

[27] G. Huang, A. Mourikis, and S. Roumeliotis, “A first-estimates jacobian ekf for improving slam consistency,” in Experimental Robotics - The Eleventh International Symposium, ser. Springer Tracts in Advanced Robotics, Dec. 2009, pp. 373–382, 11th International Symposium on Experimental Robotics, ISER 2008 ; Conference date: 13-07-2008 through 16-07-2008.

[28] ——, “Observability-based rules for designing consistent ekf slam estimators,” International Journal of Robotics Research, vol. 29, no. 5, pp. 502–528, Apr. 2010, copyright: Copyright 2010 Elsevier B.V., All rights reserved.

[29] S. Agarwal, K. Mierle, and Others, “Ceres solver,” http://ceres-solver.org.

[30] K. J. K. Mark L. Psikak. Model-based estimation notes. [Online]. Available: http://www.mit.edu/~kircher/mlb.pdf

[31] T. Qin, J. Pan, S. Cao, and S. Shen, “A general optimization-based framework for local odometry estimation with multiple sensors,” 2019.

[32] viral-dataset-site. [Online]. Available: https://ntu-aris.github.io/ntu_viral_dataset/