VID-Fusion: Robust Visual-Inertial-Dynamics Odometry for Accurate External Force Estimation

Ziming Ding[1,2,3], Tiankai Yang[1,2,3], Kunyi Zhang[1,2,3], Chao Xu[1,2], and Fei Gao[1,2]

Abstract— Recently, quadrotors are gaining significant attention in aerial transportation and delivery. In these scenarios, an accurate estimation of the external force is as essential as the six degree-of-freedom (DoF) pose since it is of vital importance for planning and control of the vehicle. To this end, we propose a tightly-coupled Visual-Inertial-Dynamics (VID) system that simultaneously estimates the external force applied to the quadrotor along with the six DoF pose. Our method builds on the state-of-the-art optimization-based Visual-Inertial system [1], with a novel deduction of the dynamics and external force factor extended from VIMO [2]. Utilizing the proposed dynamics and external force factor, our estimator robustly and accurately estimates the external force even when it varies widely. Moreover, since we explicitly consider the influence of the external force, when compared with VIMO [2] and VINS-Mono [1], our method shows comparable and superior pose accuracy, even when the external force ranges from neglectable to significant. The robustness and effectiveness of the proposed method are validated by extensive real-world experiments and application scenario simulation. We will release an open-source package of this method along with datasets with ground truth force measurements for the reference of the community.

I. INTRODUCTION

Micro Aerial Vehicles (MAVs), especially quadrotors, have long been paying expectancy in package delivery, aerial manipulation, and collaborative transportation. In these applications, quadrotors are required to carry heavy payloads or suspended cargoes, which apply external forces periodically to the drone. These forces are un-neglectable, thus causing large disturbances and significantly harming the flight of the quadrotor. To compensate for the disturbances, the controller and planner of a quadrotor have to precisely know the magnitude and direction of the external force in advance. Therefore, it is necessary to design a state estimator, which takes external forces into explicit consideration.

Recently, the Visual Inertial Odometry (VIO) [1, 3] system has shown its reliable capability to provide robust and accurate state estimation of a quadrotor, even within the minimalist hardware configuration, such as a monocular or stereo camera with an inertial measurement unit (IMU). To combine the external disturbance force estimation into a VIO system, VIMO [2] formulates a model-based VIO in a tightly coupled manner [1] and opens the door to simultaneously
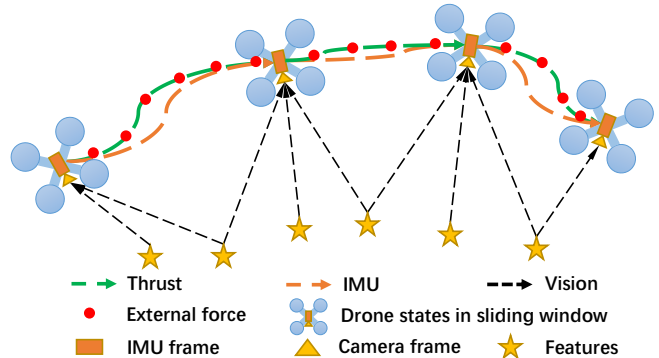
Fig. 1: This figure illustrates the relationship among thrust, external force observation, and IMU measurements in the sliding window. The integration from a combination of thrust and external force, which is equal to IMU measurements integration ideally, can both represent the movement between two continuous frames of drone states. Therefore, the thrust and external force will be added into the sliding window as an extra factor, formulating the model-based visual-inertial force estimation framework.

estimate motion and external forces with a VIO system. By considering the dynamics model of the quadrotor as a new information source, VIMO improves the accuracy of pose estimation and utilizes the dynamics to provide a clue for the external force estimation. VIMO [2] jointly estimates the motion and forces by modeling the unknown external forces as zero-mean Gaussian values, assuming the external force estimated is near zero or only occurs in a short period. Obviously, this assumption only holds for narrow scenarios and results in severe unstability or even failure to pose estimation when the drone receives large or continuous forces.

To accurately model and measure the effects of external forces, it is necessary to incorporate the quantification of external force into a VIO system. As illustrated in Fig. 1, the external force is roughly equal to the discrepancy between motions measured from IMU and thrust measurements without considering sensor noise characteristics. In this paper, instead of using the zero-mean Gaussian assumption, we provide a rough force observation by comparing IMU and thrust. We derive an external force preintegration term to represent the average force between consecutive camera frames and then combine the thrust preintegration proposed in [2] to reformulate the dynamics residual. Finally, the stacked residual is jointly minimized by using nonlinear optimization within the sliding window, as shown in Fig. 1.

In this paper, we improve the force estimation based on the model-based VIO [2] and propose an optimization-based Visual-Inertial-Dynamics (VID) framework extended from

[1, 3, 4]. Compared with VIMO [2] and VINS-Mono [1], the proposed estimator demonstrates superior pose and force estimation accuracy and attains robustness even when the external force varies widely.

To summarize our contributions:

- We derive the external force preintegration and propose state, covariance propagation, and force bias correction for discrete-time processing.
- We incorporate the external force preintegration into the state-of-the-art VIO framework and present a complete, robust, tightly-coupled Visual-Inertial-Dynamics state estimator.
- We demonstrate robust and accurate external force and pose estimation onboard the quadrotor and present a new dataset with ground truth measurements.

The rest of the paper is organized as follows. Related works are shown in Sec.II. We describe the proposed framework in detail in Sec.III. Following several experiments in Sec.IV, we show the performance of algorithms in this article. Finally, we give relevant conclusions on our algorithm and future work in Sec.V.

II. RELATED WORKS

A. System model and identification

For better external force estimation, we must distinctly comprehend the source of the driven force of MAVs. [5] describes an online Kalman filter-based method to identify a quadrotor system, especially geometrical and inertial parameters. However, dynamics parameters such as the thrust coefficient still need to be measured by experiments. In [6], the authors give us almost all the details corresponding to a multi-copter system, but it demands offline experiments with a specialized device to obtain the relation between rpm and aerodynamic force, including thrust.

B. Motion state estimation

Before disturbance estimation, the motion state is the hot spot in the relevant field. Especially, visual-inertial odometry has achieved a lot of excellent achievements and extensive applications. MSCKF [7], VINS-Mono [1], Open-VINS [8] and so on are some brilliant open-source works. All algorithms above work well in most conditions due to effective, robust and online features, but they neglect the dynamics of the robots, focusing only on the ego-motion state estimation with as little as possible sensors. So in complex environments [9]–[11] with contacts, interactions and disturbances, we can not achieve an ideal goal only with an unreliable setting value of disturbance.

C. External force estimation

Roughly, we can obtain the external force directly from the collective thrust and acceleration measurements [12]. A nonlinear observer such as [13]–[15] is designed for disturbance observation, but they require good dynamics models and enough measurements. The filter-based method is usually used for denoising. [16]–[19] use a loosely-coupled filter-based method to estimate a disturbance but neglecting

the interaction of different sensor data, resulting in a draft or bias of the estimator. VIMO [2] bridges the gap between motion estimation and external force estimation within an optimization-based VIO system, but the unknown external force is modeled as a zero-mean Gaussian. So the estimator will be highly affected and even leads to failure of the system by large or continuous forces.

III. METHODOLOGY

In this section, we will describe the proposed Visual-Inertial-Dynamics framework in detail. Fig. 2 shows the definition of the coordinate system, where body frame, IMU frame and the geometric frame are assumed coincided, denoted by $[\cdot]^b$. Otherwise, $[\cdot]^w$ denotes the world frame, and $[\hat{\ }]$ represents the value with measurement.

A. External Force Preintegration

We formulate our external force preintegration inspired by acceleration preintegration in [20] and [4], and extend preintegration of dynamics factor in [2] to be a dynamics and external force preintegration factor.

1) External Force Quantification: The quantification of the external force exerted on a drone needs to combine accelerometer data from inertial measurement unit (IMU) and the rotor thrust from the rotor rotating speed measurement unit (RMU), in which the raw acceleration $\hat{\boldsymbol{a}}^b$ and the mass normalized thrust acceleration $\hat{\boldsymbol{T}}^b$ are modeled as:

$$\hat{\boldsymbol{a}}^b = \boldsymbol{a}^b + \boldsymbol{b}_a + \mathbf{R}_w^b \boldsymbol{g}^w + \boldsymbol{n}_a$$
$$\hat{\boldsymbol{T}}^b = \frac{1}{m} \sum_{i=1}^{n_{rotor}} \left(\tau_{f_i} \omega_i^2 \right) + \boldsymbol{n}_T, \quad (1)$$

where the noise of acceleration and thrust measurements are assumed as Gaussian white noise:

$$\boldsymbol{n}_a \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_a^2), \quad \boldsymbol{n}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_T^2). \quad (2)$$

$\boldsymbol{a}^b$ is true value of acceleration in IMU frame, and $\boldsymbol{b}_a$ is the acceleration measurement bias, whose derivative is also assumed as Gaussian white noise with

$$\dot{\boldsymbol{b}}_a = \boldsymbol{n}_{b_a}, \quad \boldsymbol{n}_{b_a} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{b_a}^2). \quad (3)$$

$\mathbf{R}_w^b$ is the rotation matrix from the world frame to body frame, requiring the rotation evolution provided by IMU rotation model $\dot{\boldsymbol{q}}_b^w = \boldsymbol{q}_b^w \otimes [0, \boldsymbol{\omega}^b]^T$, where $\boldsymbol{q}_b^w$ is the quaternion for the rotation, and $\otimes$ is the quaternion multiplication and $\boldsymbol{\omega}^b$ is the body angular velocity, and the gyroscope measurements with noise and bias from IMU is

$$\hat{\boldsymbol{\omega}}^b = \boldsymbol{\omega}^b + \boldsymbol{b}_\omega + \boldsymbol{n}_\omega, \quad (4)$$

$\boldsymbol{n}_\omega$ and the derivative of $\boldsymbol{b}_\omega$ are also considered as Gaussian white noise:

$$\dot{\boldsymbol{b}}_\omega = \boldsymbol{n}_{b_\omega}, \quad \boldsymbol{n}_{b_\omega} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_{b_\omega}^2), \quad \boldsymbol{n}_\omega \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_\omega^2). \quad (5)$$

The gravity vector is $\boldsymbol{g}^w = [0, 0, -9.79] m/s^2$. Besides, m stands for mass, and n_{rotor} stands for the number of rotors, and τ_{f_i} is thrust coefficient for ith rotor of any arbitrary
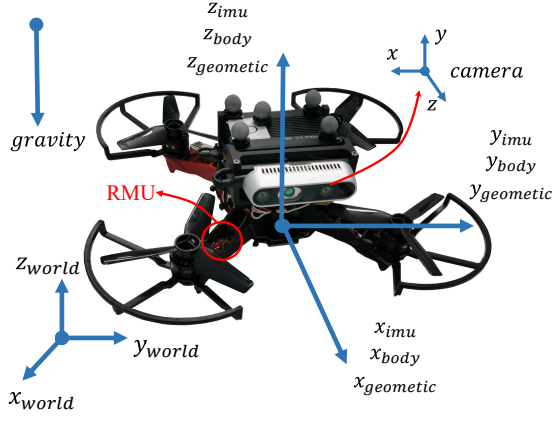
Fig. 2: The frames defined in our system, where body frame, IMU frame and the geometric frame are assumed coincided.

multi-rotors, and ω_i is the ith rotor rotation speed from RMU. More details and system identification about thrust coefficient will be discussed in III-C. Here, we assume that the IMU frame is the body frame, and the geometric center of the drone is coincident with the drone centroid and IMU center, as shown in Fig. 2. They are all called the body frame in our framework. We consider external force as a resultant force except for rotor thrust and aircraft gravity, acting on the mass center. So we can derive the relationship between $\hat{a}^b$ and $\hat{T}^b$:

$$
\begin{aligned}
\boldsymbol{a}^b &= \hat{\boldsymbol{T}}^b + \boldsymbol{f}_{ex}^b - \mathbf{R}_w^b \boldsymbol{g}^w - \boldsymbol{n}_T \\
&= \hat{\boldsymbol{a}}^b - \boldsymbol{b}_a - \mathbf{R}_w^b \boldsymbol{g}^w - \boldsymbol{n}_a.
\end{aligned} \tag{6}
$$

$\boldsymbol{f}_{ex}^b$ represents the external force observation, combined by rotor drag and other force. And now we can measure the external force with simple operation as follow:

$$
\boldsymbol{f}_{ex}^b = \left(\hat{\boldsymbol{a}}^b - \boldsymbol{b}_a - \hat{\boldsymbol{T}}^b \right) - \left(\boldsymbol{n}_a - \boldsymbol{n}_T \right). \tag{7}
$$

2) External Force Preintegration: The external force $\boldsymbol{f}_{ex}^b$ calculated from (7) have the same dimension as acceleration expressed in body frame, so we can use the acceleration measurement preintegration method to integrate external force between keyframes b_k and b_{k+1} and represent it at frame b_k:

$$
\hat{\boldsymbol{J}}_{b_{k+1}}^{b_k} = \int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \boldsymbol{f}_{ex}^{b_\tau} d\tau \tag{8}
$$

$$
\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k} = \frac{1}{\Delta t_k} \hat{\boldsymbol{J}}_{b_{k+1}}^{b_k}. \tag{9}
$$

$\hat{\boldsymbol{J}}_{b_{k+1}}^{b_k}$ represents the integration of external force over time interval $\Delta t_k = t_{k+1} - t_k$ referencing frame b_k. Then we use $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ to represent the average force between frames b_k and b_{k+1}. During $[t_k, t_{k+1}]$, $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ can be considered as a constant force exerting in b_k frame. $\mathbf{R}_{b_\tau}^{b_k}$ is provided by the integration of rotation $\hat{\boldsymbol{\gamma}}_{b_\tau}^{b_k}$, which will be mentioned in (11).

We add external force preintegration derived above into preintegration of dynamics factor in [2], and adjust to three preintegration terms containing measurements: $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}, \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$,

$\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$. These terms are integrated from IMU and thrust measurements expressed in frame b_k, which are independent of the world frame.

$$
\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} = \iint_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \left(\hat{\boldsymbol{T}}^{b_\tau} - \boldsymbol{n}_T \right) d\tau^2
$$

$$
\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} = \int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \left(\hat{\boldsymbol{T}}^{b_\tau} - \boldsymbol{n}_T \right) d\tau \tag{10}
$$

$$
\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k} = \frac{1}{\Delta t_k} \int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \boldsymbol{f}_{ex}^{b_\tau} d\tau.
$$

$\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$ represents the contribution to distance by the thrust during $[t_k, t_{k+1}]$, $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$ represents the velocity contributed by thrust, and $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ represents the average external force in the period.

3) External Force Propagation: As for discrete-time processing, we employ first-order hold numerical integration method as done in [1]. In the period of $[t_k, t_{k+1}]$, every initial value $\hat{\boldsymbol{\alpha}}_{b_k}^{b_k}, \hat{\boldsymbol{\beta}}_{b_k}^{b_k}$ of propagation at t_k is set to $\mathbf{0}$. At the period of $[t_k, t_{k+1})$, the external force $\hat{\boldsymbol{\mathcal{F}}}_{b_k}^{b_k}$ propagates in the same way as $\hat{\boldsymbol{\alpha}}_{b_k}^{b_k}$ and $\hat{\boldsymbol{\beta}}_{b_k}^{b_k}$. At the moment of t_{k+1}, $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ will be divided by Δt_k, calculating the average force in the time interval $[t_k, t_{k+1}]$. There is a requirement of rotation matrix $\mathbf{R}_i^{b_k}$ during propagation, so quaternion preintegration $\hat{\boldsymbol{\gamma}}_i^{b_k}$ is employed to propagate from frame b_k to current timestamp t_i of IMU measurements with an initial state of identity quaternion. During the propagation period $[t_k, t_{k+1}]$, the bias of acceleration and gyroscope measurements are considered as constant value, which are the estimated result at time t_k, $\boldsymbol{b}_{a_k}$ and $\boldsymbol{b}_{\omega_k}$. The propagation is presented by

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{i+1}^{b_k} &= \hat{\boldsymbol{\alpha}}_i^{b_k} + \hat{\boldsymbol{\beta}}_i^{b_k} \delta t_i + \frac{1}{2} \mathbf{R}_i^{b_k} \hat{\boldsymbol{T}}^{b_i} \delta t_i^2 \\
\hat{\boldsymbol{\beta}}_{i+1}^{b_k} &= \hat{\boldsymbol{\beta}}_i^{b_k} + \mathbf{R}_i^{b_k} \hat{\boldsymbol{T}}^{b_i} \delta t_i \\
\hat{\boldsymbol{\mathcal{F}}}_{i+1}^{b_k} &= \hat{\boldsymbol{\mathcal{F}}}_i^{b_k} + \mathbf{R}_i^{b_k} \left(\hat{\boldsymbol{a}}^{b_i} - \boldsymbol{b}_{a_k} - \hat{\boldsymbol{T}}^{b_i} \right) \delta t_i \\
\hat{\boldsymbol{\gamma}}_{i+1}^{b_k} &= \hat{\boldsymbol{\gamma}}_i^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} (\boldsymbol{\omega}^{b_i} - \boldsymbol{b}_{\omega_k}) \delta t_i \end{bmatrix},
\end{aligned} \tag{11}
$$

where δt_i is time interval between two continuous IMU measurements i and $i+1$. Here, $\hat{\boldsymbol{T}}^{b_i}$ updates at the rate of RMU, while $\hat{\boldsymbol{a}}^{b_i}$ and $\boldsymbol{\omega}^{b_i}$ at IMU rate. Propagation runs at the faster measurement rate. When the faster measurement update, the slower one will keep the latest measurement, and both of them will be propagated together.

As for propagation of the discrete-time dynamics during $[t_k, t_{k+1}]$, the error state $\delta \boldsymbol{z} = [\delta \boldsymbol{\alpha}^T \ \delta \boldsymbol{\beta}^T \ \delta \boldsymbol{\mathcal{F}}^T \ \delta \boldsymbol{b}_a^T \ \delta \boldsymbol{b}_\omega^T]^T$ and the state covariance matrix $\mathbf{P}_i^{b_k}$ will be propagated by a linearized model with initialization of $\delta \boldsymbol{z}_{b_k}^{b_k} = \mathbf{0}, \mathbf{P}_{b_k}^{b_k} = \mathbf{0}$, while the first-order Jacobian matrix $\mathbf{J}_i$ of $\boldsymbol{z}_i^{b_k}$ with respect to $\boldsymbol{z}_{b_k}^{b_k}$ can be also propagated with initialization of $\mathbf{J}_{b_k} = \mathbf{I}$:

$$
\delta \boldsymbol{z}_{i+1}^{b_k} = \mathbf{F}_i \delta \boldsymbol{z}_i^{b_k} + \mathbf{G}_i \boldsymbol{n}, \tag{12}
$$

$$
\mathbf{P}_{i+1}^{b_k} = \mathbf{F}_i \mathbf{P}_{i+1}^{b_k} \mathbf{F}_i^T + \mathbf{G}_i \mathbf{Q} \mathbf{G}_i^T, \tag{13}
$$

$$
\mathbf{J}_{i+1} = \mathbf{F}_i \mathbf{J}_i, \tag{14}
$$

where $\mathbf{F}_i$ is the linearized state-transition matrix of the error state, while $\mathbf{G}_i$ is the noise matrix of the error state, and

$\boldsymbol{n}$ is the noise vector $\boldsymbol{n} = [\boldsymbol{n}_T^T \; \boldsymbol{n}_\omega^T \; \boldsymbol{n}_{b_\omega}^T \; \boldsymbol{n}_a^T \; \boldsymbol{n}_{b_a}^T]^T$. Using the recursive propagation method, we can get the covariance matrix $\mathbf{P}_{b_{k+1}}^{b_k}$ and Jacobian $\mathbf{J}_{k+1}$ at time t_{k+1}.

4) Bias Correction: As we propagate $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ with an assumption that the bias $\boldsymbol{b}_{a_k}$, $\boldsymbol{b}_{\omega_k}$ are constant, a approximate correction will be employed when the bias changes slightly:

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_\omega}^\alpha \delta \boldsymbol{b}_{\omega_k} \\
\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_\omega}^\beta \delta \boldsymbol{b}_{\omega_k} \\
\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k} + \frac{1}{\Delta t_k}\big(\mathbf{J}_{b_a}^\mathcal{F} \delta \boldsymbol{b}_{a_k} + \mathbf{J}_{b_\omega}^\mathcal{F} \delta \boldsymbol{b}_{\omega_k}\big),
\end{aligned}
\tag{15}
$$

where $\mathbf{J}_{b_\omega}^\alpha$, $\mathbf{J}_{b_\omega}^\beta$, $\mathbf{J}_{b_a}^\mathcal{F}$, $\mathbf{J}_{b_\omega}^\mathcal{F}$ are the corresponding block in Jacobian $\mathbf{J}_{k+1}$. And $\delta \boldsymbol{b}_{\omega_k}$, $\delta \boldsymbol{b}_{a_k}$ are the bias variations between the initial estimated state and the latest estimated state in t_k of gyroscope and acceleration, respectively.

B. Tightly Coupled VIO with External Force Estimation

We employ a sliding window-based tightly-coupled VIO framework as in VINS-Mono [1] and VIMO [2].

1) System Formulation: In the sliding window, the full state of the optimization is defined as following:

$$
\begin{aligned}
\boldsymbol{\mathcal{X}} &= [\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \lambda_0, \lambda_0, \cdots, \lambda_m], \\
\boldsymbol{x}_k &= [\boldsymbol{p}_{b_k}^w, \boldsymbol{v}_{b_k}^w, \boldsymbol{q}_{b_k}^w, \boldsymbol{b}_a, \boldsymbol{b}_\omega, \boldsymbol{F}_{ex}^{b_k}], \; k \in [0, n],
\end{aligned}
\tag{16}
$$

where $\boldsymbol{x}_k$ is the body state at the kth keyframe time with the composition of position $\boldsymbol{p}_{b_k}^w$, velocity $\boldsymbol{v}_{b_k}^w$, orientation represented in Hamilton quaternion $\boldsymbol{q}_{b_k}^w$, IMU measurement bias $\boldsymbol{b}_a$, $\boldsymbol{b}_\omega$, and the average external force $\boldsymbol{F}_{ex}^{b_k}$ in body frame during $[t_k, t_{k+1}]$. Here, n is the size of the sliding window, m is the total number of features in the sliding window, and λ_l is the inverse depth of the lth feature from its first observation in the sliding window as in [1].

In order to obtain the maximum posterior estimation of all residuals concerning $\boldsymbol{\mathcal{X}}$, we employ the nonlinear optimization problem formulation in [1] and its extension, which considers the dynamics residual in [2]. Crucially, we reformulate the dynamics residual in [2] by using our proposed external force preintegration. The cost function is:

$$
\min_{\boldsymbol{\mathcal{X}}} \left\{ \|\boldsymbol{r}_p\|^2 + \sum_{k \in \mathcal{B}} \|\boldsymbol{r}_\mathcal{B}^k\|_{\mathbf{W}_\mathcal{B}^k}^2 + \sum_{(l,j) \in \mathcal{C}} \rho\big(\|\boldsymbol{r}_\mathcal{C}^{l,j}\|_{\mathbf{W}_\mathcal{C}^{l,j}}^2\big) \right.
$$
$$
\left. + \underbrace{\sum_{k \in \mathcal{D}} \left\| \boldsymbol{r}_\mathcal{D}^k\big(\boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \hat{\boldsymbol{z}}_{b_{k+1}}^{b_k}\big) \right\|_{\mathbf{W}_\mathcal{D}^k}^2}_{\text{the external force residual}} \right\},
\tag{17}
$$

where the first three residuals are the VIO residual terms of prior residual, the sum of the inertial residuals, and the sum of visual residuals with Huber-norm, which can be found in more details in [1], the last one is the sum of Mahalanobis norm of our proposed external force residual, in which, $\mathcal{D}$ is the set of thrust and external force measurements in the sliding window. $\mathbf{W}_\mathcal{D}^k$ is the weight matrix of Mahalanobis

norm for the kth state in the sliding window. $\boldsymbol{r}_\mathcal{D}^k$ is the corresponding residual at t_k which considers the state $\boldsymbol{x}_k$, $\boldsymbol{x}_{k+1}$ and external force preintegration $\hat{\boldsymbol{z}}_{b_{k+1}}^{b_k}$.

2) External Force Residual: We combine $\boldsymbol{\alpha}_{b_{k+1}}^{b_k}$, $\boldsymbol{\beta}_{b_{k+1}}^{b_k}$, $\boldsymbol{\mathcal{F}}_{b_{k+1}}^{b_k}$, the the prediction of preintegration term $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ in (10), to formulate our proposed residual between frame b_k and b_{k+1}.

$$
\boldsymbol{r}_\mathcal{D}^k\big(\boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \hat{\boldsymbol{z}}_{b_{k+1}}^{b_k}\big) = \begin{bmatrix} \boldsymbol{\alpha}_{b_{k+1}}^{b_k} - \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} - \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \\ \boldsymbol{\mathcal{F}}_{b_{k+1}}^{b_k} - \hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k} \\ \boldsymbol{b}_{a_{k+1}} - \boldsymbol{b}_{a_k} \end{bmatrix}, \mathbf{W}_\mathcal{D}^k = \mathbf{P}_{b_{k+1}}^{b_k - 1},
\tag{18}
$$

where

$$
\begin{aligned}
\boldsymbol{\alpha}_{b_{k+1}}^{b_k} &= \mathbf{R}_w^{b_k}\Big(\boldsymbol{p}_{b_{k+1}}^w - \boldsymbol{p}_{b_k}^w - \boldsymbol{v}_{b_k}^w \Delta t_k - \frac{1}{2}g^w \Delta t_k^2\Big) \\
&\quad - \frac{1}{2}\boldsymbol{F}_{ex}^{b_{k+1}} \Delta t_k^2 \\
\boldsymbol{\beta}_{b_{k+1}}^{b_k} &= \mathbf{R}_w^{b_k}\Big(\boldsymbol{v}_{b_{k+1}}^w - \boldsymbol{v}_{b_k}^w - g^w \Delta t_k\Big) - \boldsymbol{F}_{ex}^{b_{k+1}} \Delta t_k \\
\boldsymbol{\mathcal{F}}_{b_{k+1}}^{b_k} &= \boldsymbol{F}_{ex}^{b_{k+1}}.
\end{aligned}
\tag{19}
$$

The residual $\delta\boldsymbol{\alpha}, \delta\boldsymbol{\beta}, \delta\boldsymbol{\mathcal{F}}, \delta\boldsymbol{b}_a$ represents the difference between the prediction model and the measurement integration of the contribution of distance by the thrust, the velocity contributed by thrust, the average force and the measurement bias deviation, respectively. $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\mathcal{F}}}_{b_{k+1}}^{b_k}$ are the corrected measurements preintegration in (15). Instead of a manual weight for the external force residual in VIMO, $\mathbf{W}_\mathcal{D}^k$ is the weight of Mahalanobis norm, and the covariance $\mathbf{P}_{b_{k+1}}^{b_k}$ is propagated by (13).

As for the other part of the motion state estimation, we use the same marginalization and initialization strategy in [1, 2].

C. Thrust Coefficient Identification

Since we consider the external force acting on the center of mass and ignore the disturbance on torque, what we need to identify for the system are the mass of the drone and the thrust coefficients of the propellers. The mass can be easily obtained, but thrust coefficient identification is often required in the force sensor, which is expensive. Here, we employ a simple method to estimate the thrust coefficients with a given mass. We consider the centroid and the geometry center are coincident, and each propeller from the quadrotor is equally distributed $mg/4$ gravity when the quadrotor is hovering. Then we estimate the thrust coefficient for each propeller by the Recursive Least Squares method.

IV. RESULTS

In this section, we evaluate our proposed algorithm by real-world experiments in different scenes, focusing on the accuracy and robustness of both odometry and force. Within the experiments, some comparisons among the proposed, VIMO and VINS-Mono will be presented within the same condition and the same parameters (sliding window size,
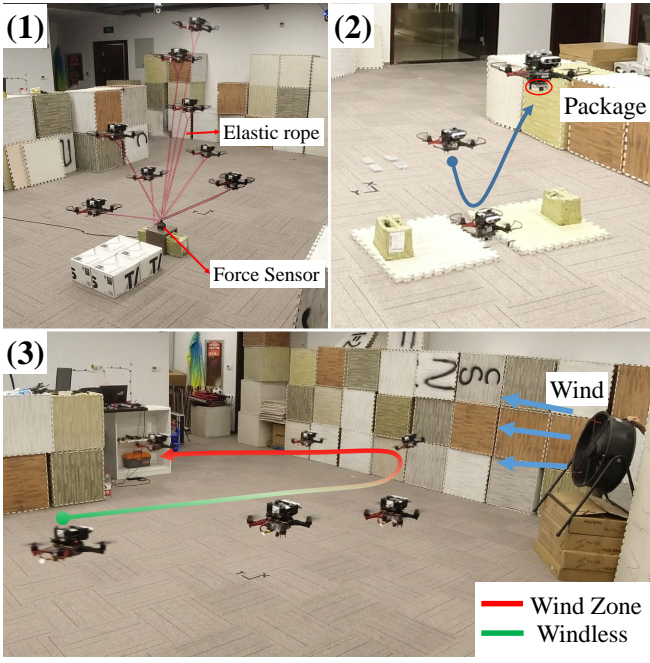
Fig. 3: (1) shows the roped flight experiment. The drone flies with an elastic rope. (2) shows the flying scale application, the drone flies down and picks up the package, hovering for weighing. (3) shows the flight in wind, the drone is blown away in the red part.
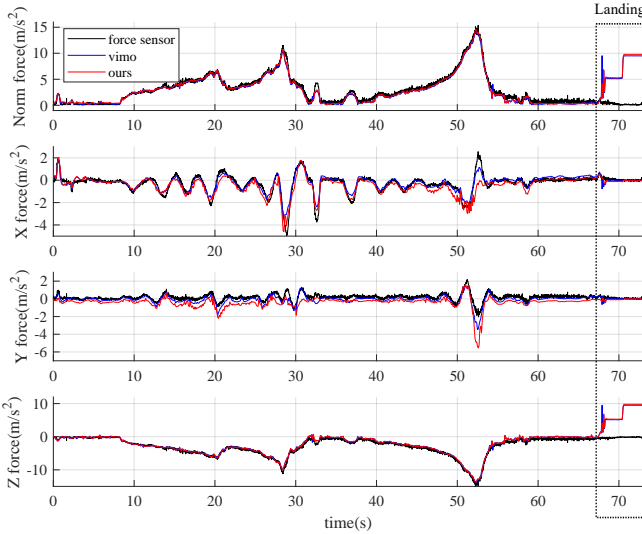

Fig. 4: Force from elastic rope in norm, x, y, z, with RMSE = $[0.998, 0.488, 0.751, 0.442]$. The drone lands after 66.1s.

noise parameters, and so on), while the weight for the external force residual in VIMO is set as the recommended one in the open-source file.

A. Real-world Experiments

1) Hardware Setup: As for validation in real-world experiments, we use the quadrotor platform shown in Fig. 2, with an onboard computer (i7-8550U), a stereo camera (Intel Realsense D435), a DJI N3 controller providing IMU data. As for the homemade RMU, we measure the changing of rotor phases by a voltage comparator and an ARM-based micro computing unit (MCU), and then we calculate the rotor
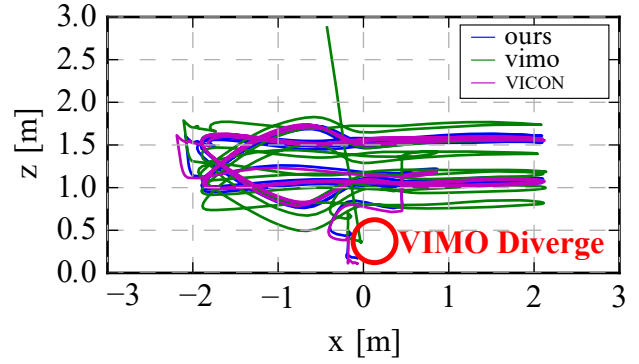

Fig. 5: Trajectories 2 of flight with loads. The odometry estimation of VIMO is affected by the external force from loads. When the drone lands, the external force is the normal force, causing the divergence of VIMO, as illustrated above.
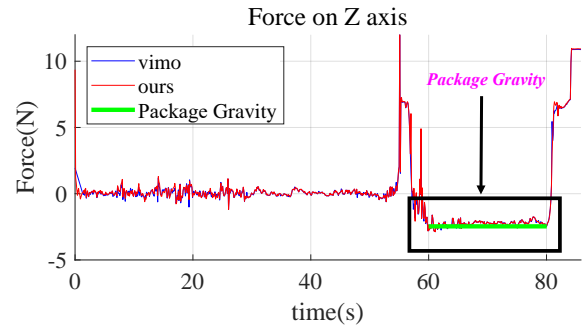

Fig. 6: Force estimation in flying scale experiment. The drone weighs the package during 60s and 80s. The peak before weighing is caused by the landing for the package. The RMSE between the gravity of the package and the estimation is about 0.29N.

rpm by fast Fourier transform (FFT) [21]. We obtain the IMU data at 400Hz, imagery data at 30Hz, RMU data at 100Hz. We identify the thrust coefficient by the proposed method in III-C with a given mass of the drone.

2) Trajectory Flight: We record the data from three flight trajectories: flight without loads, flight with loads fixed on the drone, and flight in wind. The external force of flight without loads is the aerodynamic drag force. It is a combined force from loads and aerodynamic drag in flight with loads trajectories. The last one tests in a scene where the drone tries to fly in front of a fan, and it is finally blown away by the wind, as illustrated in Fig. 3 (3), and the estimated external force will be shown in the proposed video. As shown in Fig. 5 and Fig. 7, the proposed approach is more robust in odometry estimation when flight with loads.

To be mentioned, VIMO usually fails when the force is unneglectable large, especially in the landing period, as shown in Fig. 5, because the large force violates the zero-means force assumption, and the cost of the optimization cost function becomes abnormally large. The weight of the external force term in VIMO should be adjusted as the force's magnitude. Therefore, results shown in Fig. 7 ∼ 8 and Tab. I are the segments that VIMO does not diverge, for better comparing the robustness and accuracy of the proposed.

3) Roped Flight: To evaluate the estimated force's accuracy, we perform a roped flight test, connecting the drone
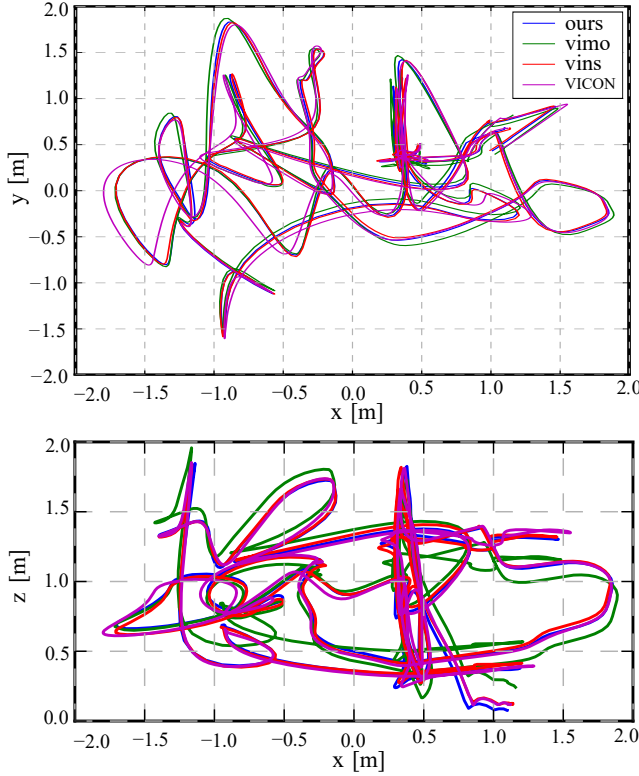
Fig. 7: Trajectories 1 of flight with loads. Ours keeps the accuracy compared with VINS-Mono, while VIMO is strongly affected by the external force, especially z axis.

TABLE I: Real-world Experiment Evaluation

Trajectory	trans. RMSE(m)			rot. RMSE(deg)		
	VINS	VIMO	Ours	VINS	VIMO	Ours
1 w/o loads	0.058	0.058	**0.057**	4.03	**3.36**	3.86
2 w/o loads	0.201	0.240	**0.198**	4.43	**4.32**	4.84
1 w/ loads	0.132	0.285	**0.120**	3.37	4.09	**3.22**
2 w/ loads	**0.102**	0.239	0.103	**4.28**	4.40	**4.28**
In wind	0.091	**0.076**	0.090	3.81	3.90	3.89
Flying scale	**0.069**	0.091	0.088	4.44	**4.28**	4.58

and force sensor by an elastic rope, as shown in Fig. 3 (1). The force estimation comparison is illustrated in Fig. 4. The result shows that our proposed method can provide accurate force estimation whenever the force is slight or large.

4) Application: We apply our proposed method in a simulated scene of package delivery as an application of weighing package in the air, just like a flying scale, as shown in Fig. 3 (3). The force estimation shown in Fig. 6 represents the accurate force estimation performance of the proposed.

B. Discussion

As shown in Fig. 8 and Tab. I, our proposed method shows similar even better performance than VINS-Mono in pose estimation, even though it performs slightly worse than VINS-Mono in some scenarios due to the noisy thrust measurements. In Tab. I, VIMO shows similar or better performance in some indexes and performs superiorly in force estimation shown in Fig. 4 and Fig. 6, but it performs much worse in pose estimation under a large external force. These experiments highlight that our proposed VID-Fusion
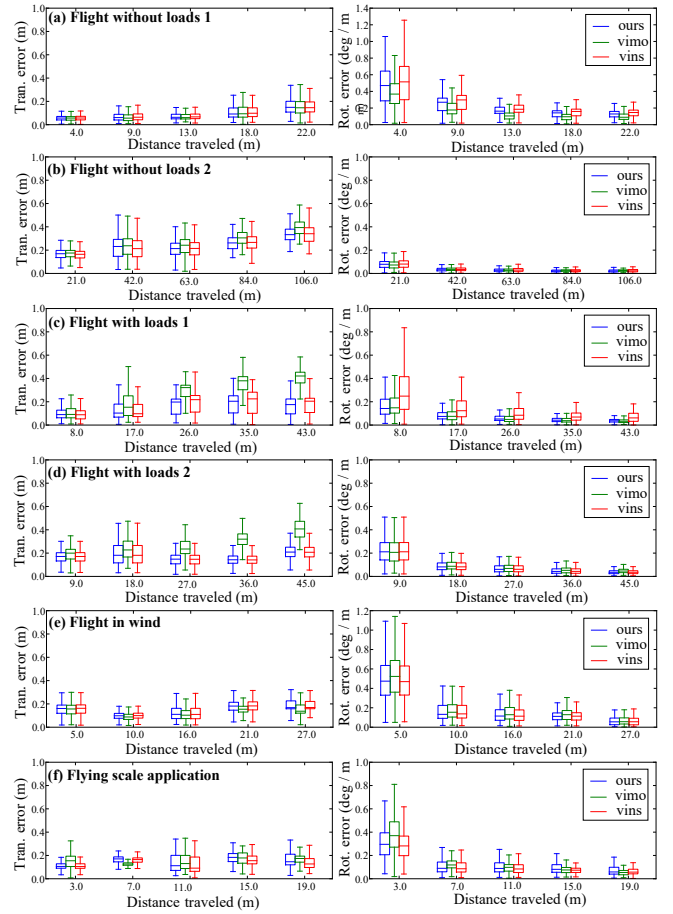

Fig. 8: RMSE of translation and rotation in the experiments.

performs robustly and accurately on both odometry and external force estimation, even the external force ranges from neglectable to significant.

V. CONCLUSION

This paper proposes an approach that offers robust and accurate odometry for external force estimation based on tightly-coupled optimization-based VIO framework [1]. Our proposed estimator increases the robustness in odometry estimation and keeps accurate external force and pose estimation, whenever the force is large or lasts for a long time.

The proposed visual-inertial-dynamics estimator framework can be easily extended into a stereo visual-inertial system such as VINS-Fusion [22], and it can run without model measurements, going back to a VIO system, and the external force will be the combination of normal force and acceleration. Moreover, compared with VINS-Mono [1], we keep the accuracy of odometry estimation and even provide an increase in some scenarios, with an accurate external force estimation additionally. Since we use a simplified dynamics model in this paper, the thrust torque is not yet utilized and is left as our future work. Finally, as a return to the community, the source code and our customized datasets will be released[1].

---

[1]Code, and the dataset will be released after the acceptance of this paper at https://github.com/ZJU-FAST-Lab/VID-Fusion

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] B. Nisar, P. Foehn, D. Falanga, and D. Scaramuzza, "Vimo: Simultaneous visual inertial model-based odometry and force estimation," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2785–2792, 2019.

[3] 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.

[4] 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.

[5] V. Wuest, V. Kumar, and G. Loianno, "Online estimation of geometric and inertia parameters for multirotor aerial vehicles," 05 2019, pp. 1884–1890.

[6] Q. Quan, *Introduction to Multicopter Design and Control*. Springer Singapore, 2017.

[7] 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.

[8] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4666–4672.

[9] A. Tagliabue, A. Paris, S. Kim, R. Kubicek, S. Bergbreiter, and J. P. How, "Touch the wind: Simultaneous airflow, drag and interaction sensing on a multirotor," *ArXiv*, vol. abs/2003.02305, 2020.

[10] H. Seo, D. Lee, C. Y. Son, C. J. Tomlin, and H. J. Kim, "Robust trajectory planning for a multirotor against disturbance based on hamilton-jacobi reachability analysis," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[11] J. Ji, X. Zhou, C. Xu, and F. Gao, "Cmpcc: Corridor-based model predictive contouring control for aggressive drone flight," *ArXiv*, vol. abs/2007.03271, 2020.

[12] T. Tomi and S. Haddadin, "A unified framework for external wrench estimation, interaction control and collision reflexes for flying robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.

[13] B. Yuksel, C. Secchi, H. H. Bulthoff, and A. Franchi, "A nonlinear force observer for quadrotors and application to physical interactive tasks," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2014.

[14] F. Ruggiero, J. Cacace, H. Sadeghian, and V. Lippiello, "Impedance control of vtol uavs with a momentum-based external generalized forces estimator," in *IEEE International Conference on Robotic and Automation*, 2014.

[15] S. Kim, S. Choi, H. Kim, J. Shin, H. Shim, and H. J. Kim, "Robust control of an equipment-added multirotor using disturbance observer," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 4, pp. 1524–1531, 2018.

[16] C. D. Mckinnon and A. P. Schoellig, "Unscented external force and torque estimation for quadrotors," 2016.

[17] F. Augugliaro and R. D'Andrea, "Admittance control for physical human-quadrocopter interaction," in *Control Conference*, 2013.

[18] A. Tagliabue, M. Kamel, S. Verling, R. Siegwart, and J. Nieto, "Collaborative object transportation using mavs via passive force control," 2016.

[19] D. Abeywardena, Z. Wang, G. Dissanayake, and S. L. Waslander, "Model-aided state estimation for quadrotor micro air vehicles amidst wind disturbances," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.

[20] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft mavs," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5303–5310.

[21] H. J. Nussbaumer, "The fast fourier transform," in *Fast Fourier Transform and Convolution Algorithms*. Springer, 1981, pp. 80–111.

[22] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," *arXiv preprint arXiv:1901.03638*, 2019.