Robotics: Science and Systems 2019
Freiburg im Breisgau, June 22-26, 2019

VIMO: Simultaneous Visual Inertial Model-based Odometry and Force Estimation

Barza Nisar*, Philipp Foehn*, Davide Falanga, Davide Scaramuzza

***Abstract*—In recent years, many approaches to Visual Inertial Odometry (VIO) have become available. However, they neither exploit the robot's dynamics and known actuation inputs, nor differentiate between desired motion due to actuation and unwanted perturbation due to external force. For many robotic applications, it is often essential to sense the external force acting on the system due to, for example, interactions, contacts, and disturbances. Adding a motion constraint to an estimator leads to a discrepancy between the model-predicted motion and the actual motion. Our approach exploits this discrepancy and resolves it by simultaneously estimating the motion and the external force. We propose a relative motion constraint combining the robot's dynamics and the external force in a preintegrated residual, resulting in a tightly-coupled, sliding-window estimator exploiting all correlations among all variables. We implement our Visual Inertial Model-based Odometry (VIMO) system into a state-of-the-art VIO approach and evaluate it against the original pipeline without motion constraints on both simulated and real-world data. The results show that our approach increases the accuracy of the estimator up to 29% compared to the original VIO, and provides external force estimates at no extra computational cost. To the best of our knowledge, this is the first approach exploiting model dynamics by jointly estimating motion and external force. Our implementation will be made available open-source.**

Resources: http://rpg.ifi.uzh.ch/vimo/index.html
Keywords: Visual-Inertial, Model, Force, Estimation

I. Introduction

A. Motivation

Recent advances in robot perception have led to a number of Visual Inertial Odometry (VIO) systems becoming more robust and accessible solutions for state estimation and navigation, such as [1, 2, 3, 4, 5, 6, 7]. Although these systems work well in most conditions, they all neglect the robot's dynamics and cannot sense forces, such as contacts and interactions, and disturbances, such as wind and other environmental influences. Additionally, these approaches do not consider the fundamental distinction between the desired motion due to actuation and unwanted perturbation due to external forces. Adding the system dynamics to a VIO system (i) allows the perception of external force acting on a robot, and (ii) adds information to the estimation problem, resulting in increased accuracy.

Applications such as inspection, grasping, manipulation, and delivery require a robot to sense interaction or forces, which are often recovered using an estimator loosely-coupled with an odometry system, as proposed in [8, 9, 10, 11, 12, 13]. Such estimators introduce latency, computational overhead, and neglect correlation among the estimated variables and their noise characteristics. This shows the necessity for joint estimation of motion and external force in a unified approach addressing both, model and sensor noise characteristics.

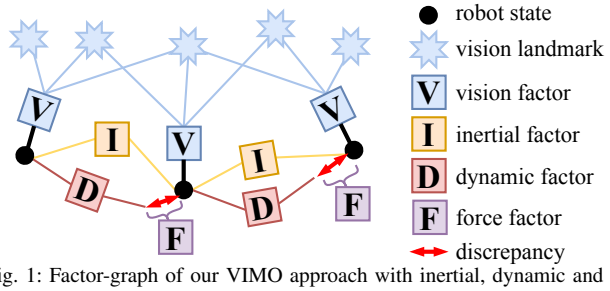


Fig. 1: Factor-graph of our VIMO approach with inertial, dynamic and force factors. The red arrows indicate the discrepancy between the dynamic and VIO factors, which is resolved by including an external force.

On the other hand, VIO approaches on Unmanned Aerial Vehicle (UAV), rely on minimal sensor configurations, typically consisting of visual and inertial sensors suffering from additive noise. Thanks to Gaussian filtering theory [14], it is known that additional knowledge and information improves the estimation performance, especially in the presence of Gaussian noise. By adding the system dynamics to a VIO estimation problem, we effectively add information. Intuitively, this additional knowledge allows us to increase the accuracy of the odometry. However, the pure addition of a motion constraint from the system dynamics does not account for any external influences, and may lead to a motion prediction deviating from the actual motion, as depicted in Fig. 1. Since this would degrade the estimator performance due to a wrong prior, it highlights the importance of including external force and jointly estimating all variables.

To the best of our knowledge, we present the first tightly-coupled approach exploiting the model dynamics while jointly estimating motion and external force. We derive the resulting motion constraint and formulate a dynamic residual. This residual is added to a pose-graph formulation of the VIO approach in [2] and is solved using numerical optimization. The resulting estimator demonstrates up to 29% increased accuracy and inherent ability to sense external force, opening the door to a number of possible future research topics and applications. As a call to the community, we want to raise awareness for the importance of contact-enabled robotics and the need for estimators to provide not only odometry information, but also leverage the robot dynamics to increase accuracy and sense external forces from contacts and interaction.

* Both authors contributed equally. This work was supported by the National Centre of Competence in Research Robotics (NCCR) through the Swiss National Science Foundation and the SNSF-ERC Starting Grant. This paper has been selected to appear in both, the RSS 2019 Proceedings and in IEEE Robotics and Automation Letters, under the same title.

B. Related Work

Previous approaches on external force estimation can be split into two groups: deterministic and probabilistic.

1) Deterministic Approaches: Deterministic approaches estimate external force by subtracting the collective thrust vector from the inertial measurements [8]. [9, 10] proposed a nonlinear force and torque observer based on the quadrotor's dynamical model, assuming that an estimate of the robot state is available from another estimator. These deterministic approaches do not consider (i) the thrust input noise, (ii) the noise in the state, and (iii) noise and unknown time-varying bias in the Inertial Measurement Unit (IMU). Hence, deterministic methods only work appropriately in practice when their inputs and outputs are carefully processed or when the signal to noise ratio of the used sensor data is very high.

2) Probabilistic Approaches: Realizing the drawbacks of deterministic force observers, [11] proposed an Unscented Kalman Filter (UKF) to account for the process and sensors noise and, consequently, improve the force estimate. Other similar filtering-based external force estimators include a Kalman filter [12] and UKF [13]. These methods can be classified as loosely-coupled, since they use the state estimate from a separate estimator [15, 13], and then fuse this estimate with their prediction from the UAV's dynamic model in a separate estimation step. Loosely-coupled estimators do not consider correlations among all estimated variables, which may lead to inaccuracies [3]. Moreover, the external force is estimated in an additional fusion step, which may introduce latency and extra computation cost.

3) Extension to Sliding Window Smoother: A widely used state estimator for UAVs is Visual Inertial Odometry (VIO) based on sliding-window smoothing [2] with IMU preintegration [16] to make the optimization problem computationally tractable in real time. IMU preintegration was first proposed in [17] and later modified in [16] to address the manifold structure of the rotation group. High-rate IMU measurements are typically integrated between image frames to form a single relative motion constraint. IMU preintegration theory reparameterizes this constraint to remove the dependence of integrated IMU measurements on previous state estimates. This avoids repeated integration when the state estimates change during each iteration of the optimization. [18] combined the idea of incorporating dynamic factors for localization of UAVs from [19] with the preintegration scheme from [16] to develop a model-based visual-inertial state estimator similar to the one proposed in our work, but without considering external forces. [18] showed that in a smoothing-based VIO pipeline, the dynamic residual in combination with the IMU residual acts as an additional source of acceleration information, which adds robustness to state estimation, especially in slow speed flights, when accelerometer measurements have low signal-to-noise ratio. While [18] chose to model air drag but ignored external forces in the dynamic model of the quadrotor, our work includes external forces and estimates them together with the robot state. An implication of not modelling external disturbances, such as wind, in model-aided state-estimation problems was studied in [15]. In the presence of wind or external forces, the estimator from [18] can tend to wrongly adjust the IMU biases due to the mismatch between sensor measurements and vehicle dynamics and therefore only works in a disturbance-free environment, as confirmed by the authors. [20] proposed to use Dynamic Differential Programming to estimate the state, parameters, and disturbances (forces) in a synthetic planar motion example, assuming perfect data association, velocity and landmark position measurement without real world applications. Their approach is significantly simplified by modelling landmark position measurements, instead of realistic camera projection measurements. Differently from [20], our method extends an optimization-based VIO framework with motion factors to simultaneously estimate state and external force in real time on real world data. To the best of our knowledge, there is no precedent of a tightly-coupled or smoothing-based method that jointly estimates robot states and 3-dimensional external forces.

C. Contribution

This work extends an optimization-based VIO in [3, 2, 16] with a residual term integrating the dynamic model of the quadrotor. Our main contribution is the derivation of this residual term from a motion constraint enforced by the model dynamics including external force, enabling a VIO framework to jointly estimate this force in addition to the robot state and IMU bias. Our approach works as a tightly-coupled estimator, using visual-inertial measurements, and the collective thrust input. Since current smoothing-based VIO systems offer higher accuracy compared to filtering-based methods, we employ nonlinear optimization as estimation strategy.

Inspired from IMU preintegration [16], the high-rate thrust inputs are preintegrated, resulting in dynamic factors used as residuals between consecutive camera frames. A factor graph representation of the VIO problem with dynamic factors is depicted in Fig 1. The dynamic factors represent relative motion constraints similar to the IMU factors but with a different model and source of measurement. In our work, we exploit this redundant motion representation to estimate external force. The dynamic residual is implemented into VINS-mono [2], an open-source sliding-window monocular VIO framework. VINS-Mono was chosen because of its availability, real-time capability, and requirement for only one camera and an IMU. We show on real and simulated data that the proposed estimator compared to VINS-mono, not only increases the accuracy of the estimates (up to 29%) but also offers external force estimates without increasing the computation time. Our approach can be implemented analogously on other robots, such as fixed-wings, manipulators and mobile ground robots.

D. Structure of this paper

The model-based VIO problem is described in Sec. II, followed by the preintegration of the dynamic residual in Sec. III. We report our experiments in Sec. IV and the limitations in Sec. V. Finally the paper is concluded in Sec. VI.

II. Problem Formulation

A. Notation

All coordinate frames used are depicted in Fig. 2. The quadrotor pose is the body-fixed frame described in world frame. The IMU frame corresponds to the body frame, attached to the center of mass of the vehicle. The world frame is denoted by $[\ ]^w$, the body frame by $[\ ]^b$ and the camera frame by $[\ ]^c$, while a hat $\hat{[\ ]}$ represents noisy measurements. The robot state at the time t_k is defined as

$$\mathbf{x_k} = [\mathbf{p}^w_{b_k}, \mathbf{v}^w_{b_k}, \mathbf{q}^w_{b_k}, \mathbf{b}_{a_k}, \mathbf{b}_{\omega_k}], \qquad k \in [0, n] \tag{1}$$

comprised of position $\mathbf{p}^w_{b_k}$, velocity $\mathbf{v}^w_{b_k}$ and Hamilton quaternion $\mathbf{q}^w_{b_k}$ encoding the rotation of the body frame with respect to the world frame, and accelerometer and gyroscope biases $\mathbf{b}_{a_k}, \mathbf{b}_{\omega_k}$ in the IMU body frame. n is the number of the most recent keyframes in the optimization window, where the n^{th} frame is the latest frame that does not need to be a keyframe. The sliding window optimization variables are given by

$$\mathcal{X} = [l_1, \cdots, l_m, \mathbf{x}_0, \mathbf{f}^b_{e_0}, \mathbf{x}_1, \cdots, \mathbf{f}^b_{e_{n-1}}, \mathbf{x}_n] \tag{2}$$

where m is the total number of features in the sliding window and l_i is the inverse depth of the i^{th} feature as in [2]. The total mass normalised external force $\mathbf{f}^b_{e_k}$ is expressed in body frame and experienced by the quadrotor from the time of k to $k+1$ image i.e. during $[t_k, t_{k+1})$. If the duration between consecutive image frames is small, $\mathbf{f}^b_{e_k}$ will be a good approximation of the instantaneous force experienced at t_k.

B. Dynamic Residual

To include the model dynamics and external force in a nonlinear optimization, we formulate a dynamic residual $\mathbf{e}^k_d(\mathbf{x}_k, \mathbf{f}^b_{e_k}, \mathbf{x}_{k+1}, \hat{\mathbf{z}}^{b_k}_{b_{k+1}})$, with the preintegrated measurements $\hat{\mathbf{z}}^{b_k}_{b_{k+1}}$. The full nonlinear optimization problem which solves for the maximum aposteriori estimate of $\mathcal{X}$ is formulated as

$$\min_{\mathcal{X}} \sum_{k=0}^{n-1} \left\| \mathbf{e}^k_d(\mathbf{x}_k, \mathbf{f}^b_{e_k}, \mathbf{x}_{k+1}, \hat{\mathbf{z}}^{b_k}_{b_{k+1}}) \right\|^2_{\mathbf{W}^k_d} + J_{VIO}(\mathcal{X}, \hat{\mathbf{z}}^{b_k}_{b_{k+1}}) \tag{3}$$

where J_{VIO} contains the sum of prior residual $\mathbf{e}_p$, the visual residual $\mathbf{e}_v$ of all visible landmark reprojections, and the inertial residual $\mathbf{e}_s$ comprising of the preintegrated measurements. As proposed in [2] we summarize it into:

$$J_{VIO} = \sum_{k=0}^{n} \sum_{j \in \mathcal{J}_k} \rho\left(\left\| \mathbf{e}^{j,k}_v \right\|^2_{\mathbf{W}_v} \right) + \sum_{k=0}^{n-1} \left\| \mathbf{e}^k_s \right\|^2_{\mathbf{W}^k_s} + \left\| \mathbf{e}^k_p \right\|^2. \tag{4}$$

$\mathcal{J}_k$ is the set of visible landmarks in frame k, while $\mathbf{e}^k_v$ is robustified by the Huber-norm $\rho(x) = \left(\sqrt{1 + (x/\delta)^2} - 1\right)\delta^2$. The reader can refer to [2] for the derivation of J_{VIO}.

In the next section, we formulate the dynamic residual $\mathbf{e}^k_d$ as a function of the robot states and external forces at times $[t_k, t_{k+1}]$ and preintegrated thrust inputs and IMU measurements $\hat{\mathbf{z}}^{b_k}_{b_{k+1}}$. Additionally, we derive the weight $\mathbf{W}^k_d$ for the Mahalanobis norm of $\mathbf{e}^k_d$ by propagating the covariance from the measurement noise. While the formulation so far was robot-agnostic, we now focus on the quadrotor model.

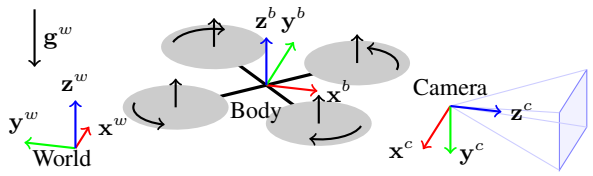


Fig. 2: Quadrotor scheme with world, body and camera frame indicated.

III. Preintegration of Quadrotor Dynamics

A. Model Dynamics

In the dynamical model we consider the evolution of position and velocity of the quadrotor subject to three forces: collective rotor thrust $\mathbf{T}^b_t$, external forces $\mathbf{f}^b_{e_t}$, and gravity $\mathbf{g}^w = [0, 0, -9.81]^T \mathrm{m\,s^{-2}}$. The translational dynamics of the quadrotor is given by the following equations:

$$\dot{\mathbf{p}}^w_{b_t} = \mathbf{v}^w_{b_t} \qquad \dot{\mathbf{v}}^w_{b_t} = \mathbf{R}(\mathbf{q}^w_{b_t})\left(\mathbf{T}^b_t + \mathbf{f}^b_{e_t}\right) + \mathbf{g}^w \tag{5}$$

where $\mathbf{R}(\mathbf{q}^w_{b_t})$ is the rotation matrix corresponding to the rotation from body to world frame. Since we do not know the dynamics of external force, we assume it to be a Gaussian variable $\mathbf{f}_{e_t} = \mathcal{N}(\mathbf{0}, \sigma_f^2)$. This allows the framework to distinguish between slowly walking accelerometer biases and incidental external forces.

Preintegration of the system dynamics requires separation of the residual terms dependent on optimization variables from the terms dependent on the measurement. The rotational dynamics of the quadrotor is not considered here, since the control torques can not be separated from their dependency on the optimization variables rendering preintegration ineffective.

B. Preintegration of Dynamic Factors

In this section we derive the preintegration of the dynamic factors. The integration of (5) requires the evolution of rotation, which is provided by the IMU's rotation model $\dot{\boldsymbol{q}}^w_{b_t} = \frac{1}{2}\mathbf{q}^w_{b_t} \otimes [0, \boldsymbol{\omega}^b_t]^\intercal$ where $\otimes$ is the quaternion multiplication and $\boldsymbol{\omega}^b$ is the angular velocity of the body expressed in the body frame. The involved noisy measurements are the biased angular velocity $\hat{\boldsymbol{\omega}}^b_t = \boldsymbol{\omega}^b_t + \mathbf{b}_{\omega_t} + \boldsymbol{\eta}_\omega$ from the IMU and the collective rotor thrust $\hat{\mathbf{T}}^b_t = \mathbf{T}^b_t + \boldsymbol{\eta}_T$. As in [2], the gyroscope noise is considered as Gaussian $\boldsymbol{\eta}_\omega \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2_\omega)$ and its bias as random walk $\dot{\mathbf{b}}_{\omega_t} = \boldsymbol{\eta}_{b_\omega}$ with $\boldsymbol{\eta}_{b_\omega} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2_{b_\omega})$. Since neither the magnitude nor the direction of the actual thrust is known precisely, we assume Gaussian noise in the thrust as $\boldsymbol{\eta}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2_T)$. The vehicle state can be propagated between two frames over time interval $\Delta t_k = t_{k+1} - t_k$ by integrating the thrust and gyroscope measurements:

$$\begin{aligned}
\mathbf{p}^w_{b_{k+1}} &= \mathbf{p}^w_{b_k} + \mathbf{v}^w_{b_k}\Delta t_k + \frac{1}{2}\mathbf{g}^w \Delta t_k^2 \\
&\quad + \int\int_{t_k}^{t_{k+1}} \mathbf{R}^w_{b_\tau}\left(\hat{\mathbf{T}}^b_\tau + \mathbf{f}^b_{e_\tau} - \boldsymbol{\eta}_T\right) d\tau^2 \\
\mathbf{v}^w_{b_{k+1}} &= \mathbf{v}^w_{b_k} + \mathbf{g}^w \Delta t_k + \int_{t_k}^{t_{k+1}} \mathbf{R}^w_{b_\tau}\left(\hat{\mathbf{T}}^b_\tau + \mathbf{f}^b_{e_\tau} - \boldsymbol{\eta}_T\right) d\tau \\
\mathbf{q}^w_{b_{k+1}} &= \mathbf{q}^w_{b_k} \otimes \int_{t_k}^{t_{k+1}} \frac{1}{2}\boldsymbol{\Omega}\left(\hat{\boldsymbol{\omega}}^b_\tau - \mathbf{b}_{\omega_\tau} - \boldsymbol{\eta}_\omega\right)\mathbf{q}^{b_k}_{b_\tau}\, d\tau
\end{aligned} \tag{6}$$

$$\text{where: } \mathbf{\Omega}(\omega) = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & -\omega_z & \omega_y \\ \omega_y & \omega_z & 0 & -\omega_x \\ \omega_z & -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{7}$$

To make the integration of the measurements independent of the states at frame k, we group the terms containing measurements in $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$, $\hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k}$, and change the reference frame from world to body frame as done in IMU preintegration [16]:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} &= \int\int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \left(\hat{\mathbf{T}}_\tau^b - \boldsymbol{\eta}_T \right) d\tau^2 \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} &= \int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k} \left(\hat{\mathbf{T}}_\tau^b - \boldsymbol{\eta}_T \right) d\tau \\ \hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k} &= \int_{t_k}^{t_{k+1}} \frac{1}{2} \mathbf{\Omega} \left(\hat{\boldsymbol{\omega}}_\tau^b - \mathbf{b}_{\omega_\tau} - \boldsymbol{\eta}_\omega \right) \hat{\gamma}_{b_\tau}^{b_k} \, d\tau. \end{aligned} \tag{8}$$

We then derive the prediction of the terms in (8) from the model equations in (6) to form the factors

$$\begin{aligned} \boldsymbol{\alpha}_{b_{k+1}}^{b_k} &= \mathbf{R}_w^{b_k} \left(\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w - \mathbf{v}_{b_k}^w \Delta t_k - \frac{1}{2} \mathbf{g}^w \Delta t_k^2 \right) - \frac{1}{2} \mathbf{f}_{e_k}^b \Delta t_k^2 \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} &= \mathbf{R}_w^{b_k} \left(\mathbf{v}_{b_{k+1}}^w - \mathbf{v}_{b_k}^w - \mathbf{g}^w \Delta t_k \right) - \mathbf{f}_{e_k}^b \Delta t_k \\ \boldsymbol{\gamma}_{b_{k+1}}^{b_k} &= \mathbf{q}_w^{b_k} \otimes \mathbf{q}_{b_{k+1}}^w. \end{aligned} \tag{9}$$

C. Dynamic Residual

Now we can combine (8) and (9) into the dynamic residual between frames b_k and b_{k+1}, which also includes the zero-mean prior on external forces.

$$\mathbf{e}_d^k = \begin{bmatrix} \boldsymbol{\alpha}_{b_{k+1}}^{b_k} - \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} - \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \\ \mathbf{f}_{e_k}^b \end{bmatrix} \quad \mathbf{W}_d^k = \begin{bmatrix} \mathbf{P}_{b_{k+1}[0:5]}^{b_k -1} & \mathbf{0} \\ \mathbf{0} & w_f \mathbf{I} \end{bmatrix} \tag{10}$$

Finally, the weight of the residual can be formulated by the inverse of the covariance in $\hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k}$ and $\hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k}$ extracted from $\mathbf{P}_{b_{k+1}}^{b_k}$ (derived in Sec. III-D) and a diagonal weight w_f for the external force zero-mean prior.

It is important to note that these preintegrated terms still depend on the gyroscope bias. This means that each time an optimization iteration changes the bias estimate slightly, we need to repropagate the measurements. To avoid this computationally expensive repropagation, we will adopt the solution proposed in [16], and explained in Sec. III-E.

D. Propagation Algorithm

We start the propagation from an initial condition of $\hat{\boldsymbol{\alpha}}_{b_k}^{b_k} = \hat{\boldsymbol{\beta}}_{b_k}^{b_k} = \mathbf{0}_{3\times1}$ and $\hat{\boldsymbol{\gamma}}_{b_k}^{b_k} = [1, \mathbf{0}_{3\times1}]$. The Euler integration over timestep δt_i is computed by

$$\hat{\boldsymbol{\alpha}}_{i+1}^{b_k} = \hat{\boldsymbol{\alpha}}_i^{b_k} + \hat{\boldsymbol{\beta}}_i^{b_k} \delta t_i + \frac{1}{2} \mathbf{R}(\hat{\boldsymbol{\gamma}}_i^{b_k}) \mathbf{T}_i^b \delta t_i^2 \tag{11}$$

$$\hat{\boldsymbol{\beta}}_{i+1}^{b_k} = \hat{\boldsymbol{\beta}}_i^{b_k} + \mathbf{R}(\hat{\boldsymbol{\gamma}}_i^{b_k}) \mathbf{T}_i^b \delta t_i \tag{12}$$

$$\hat{\boldsymbol{\gamma}}_{i+1}^{b_k} = \hat{\boldsymbol{\gamma}}_i^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} (\boldsymbol{\omega}_{m_i} - \bar{\mathbf{b}}_{\omega k}) \delta t_i \end{bmatrix} \tag{13}$$

To achieve optimal linearization accuracy, the algorithm is run at the rate of the fastest available measurement, typically the IMU rate. The covariance $\mathbf{P}_{b_{k+1}}^{b_k}$ is derived by linearizing the error $\delta \mathbf{z} = [\delta\boldsymbol{\alpha}, \delta\boldsymbol{\beta}, \delta\boldsymbol{\theta}, \delta\mathbf{b}_\omega]^\intercal$ and noise $\boldsymbol{\eta} = [\boldsymbol{\eta}_T, \boldsymbol{\eta}_\omega, \boldsymbol{\eta}_{b_\omega}]^\intercal$ dynamics between integration steps as

$$\boldsymbol{z}_{i+1}^{b_k} = \mathbf{A}_i \boldsymbol{z}_i^{b_k} + \mathbf{G}_i \begin{bmatrix} \boldsymbol{\eta}_T \\ \boldsymbol{\eta}_\omega \\ \boldsymbol{\eta}_{b_\omega} \end{bmatrix} \quad \boldsymbol{\gamma}_t^{b_k} \approx \hat{\boldsymbol{\gamma}}_t^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} \delta\boldsymbol{\theta}_t^{b_k} \end{bmatrix} \tag{14}$$

where $\delta\boldsymbol{\theta}$ is the minimal perturbation around the mean of $\boldsymbol{\gamma}$. Finally, $\mathbf{P}_{b_{k+1}}^{b_k}$ is linearly propagated from $\mathbf{P}_{b_k}^{b_k} = \mathbf{0}$ by

$$\mathbf{P}_{i+1}^{b_k} = \mathbf{A}_i \mathbf{P}_i^{b_k} \mathbf{A}_i^T + \mathbf{G}_i \mathbf{Q} \mathbf{G}_i^T \tag{15}$$

with the linearization $\boldsymbol{A}_i = \frac{\partial \boldsymbol{z}_{i+1}^{bk}}{\partial \boldsymbol{z}_i^{bk}}$ and $\boldsymbol{G}_i = \frac{\partial \boldsymbol{z}_{i+1}^{bk}}{\partial \boldsymbol{\eta}}$.

E. Bias Correction

The first-order Jacobian matrix $\mathbf{J}_{b_{k+1}}$ of $\mathbf{z}_{b_{k+1}}^{b_k}$ with respect to $\mathbf{z}_{b_k}^{b_k}$ can be computed recursively by $\mathbf{J}_{i+1} = \mathbf{A}_i \mathbf{J}_i$ starting from the initial Jacobian of $\mathbf{J}_{b_k} = \mathbf{I}$. The preintegrated terms can then be corrected by their first order approximation with respect to the change in gyroscope bias $\delta\mathbf{b}_{\omega_k} = \mathbf{b}_{\omega_k} - \bar{\mathbf{b}}_{\omega_k}$ from the initial estimate $\bar{\mathbf{b}}_{\omega_k}$ as follows:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} &\leftarrow \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_\omega}^{\alpha} \delta\mathbf{b}_{\omega_k} \qquad \mathbf{J}_{b_\omega}^{\alpha} = \frac{\partial \boldsymbol{\alpha}_{b_{k+1}}^{b_k}}{\partial \mathbf{b}_{\omega_k}} \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} &\leftarrow \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_\omega}^{\beta} \delta\mathbf{b}_{\omega_k} \qquad \mathbf{J}_{b_\omega}^{\beta} = \frac{\partial \boldsymbol{\beta}_{b_{k+1}}^{b_k}}{\partial \mathbf{b}_{\omega_k}}. \end{aligned} \tag{16}$$

F. Marginalization

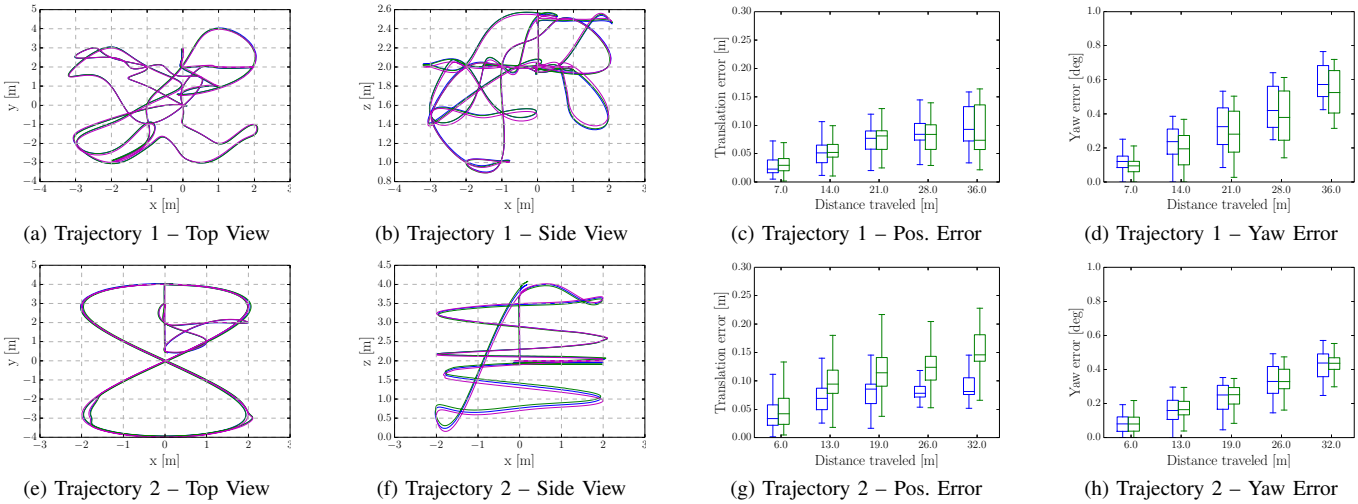
We adapt the marginalization strategy proposed in [2], such that when the second last frame in the window is a keyframe, we marginalize out the oldest keyframe's state and external force $\mathbf{f}_{\mathbf{e_0}}$. The corresponding visual, inertial, and thrust measurements of the marginalized states are converted into a prior. If the second last frame is not a keyframe, its state, external force and corresponding visual measurements are dropped, while the preintegrated IMU and thrust measurements are kept and continued to be preintegrated till the last frame.

IV. Experiments

We perform 3 types of experiments: IV-A: simulation based experiments; IV-B: evaluation on the Blackbird dataset [21] with real pose, inertial, and rotor speed measurements but synthetic camera frames; IV-C real-world experiments.

A. Simulation

Experiment Setup: To generate repeatable data in a fully controlled environment, we used the RotorS simulator from [22], a Micro-Aerial Vehicle Simulator using Gazebo in ROS. We used a forward looking camera with 752×480 image resolution. The base simulation vehicle was *Hummingbird* from [22] according to which the onboard IMU was corrupted with noise of $\sigma_\omega = 0.004\text{rad/s}\sqrt{\text{Hz}}$ for the gyroscope, $\sigma_a = 0.1\text{m/s}^2\sqrt{\text{Hz}}$ for the accelerometer, and a bias random walk of $\sigma_{b_\omega} = 0.000038\text{rad/s}^2\sqrt{\text{Hz}}$ for the gyroscope, and $\sigma_{b_a} = 0.00004\ [\text{m/s}^3\sqrt{\text{Hz}}]$ for the accelerometer. The tuning parameters $\boldsymbol{\sigma}_T$ and w_f were hand-tuned and then kept the



(a) Trajectory 1 – Top View (b) Trajectory 1 – Side View (c) Trajectory 1 – Pos. Error (d) Trajectory 1 – Yaw Error

(e) Trajectory 2 – Top View (f) Trajectory 2 – Side View (g) Trajectory 2 – Pos. Error (h) Trajectory 2 – Yaw Error

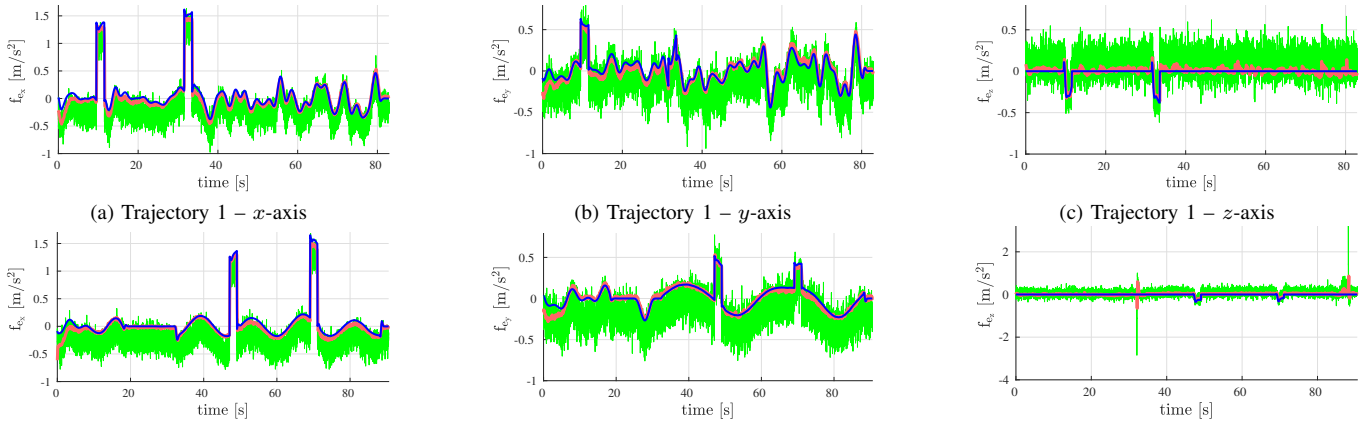
Fig. 3: Comparison between VINS — (green), VIMO — (ours, blue) and ground truth — (purple) on a random trajectory (top) and a helical eight trajectory (bottom) at $2.5\,\mathrm{m\,s^{-1}}$ with external forces applied. This configuration depicts the worst performance of VIMO compared with VINS-Mono. The two left columns show the estimated trajectories aligned with the ground truth. The two right columns summarize the relative translation and yaw error statistics over trajectory segments. Boxes indicate the middle two quartiles while whiskers denote upper and lower quartiles and the center line indicates the median.

same across all of the experiments. The dynamic residual was implemented in VINS-Mono with a maximum number of 150 features tracked per frame. For a fair comparison, no loop closure was applied. The estimator is run on a 2.5 GHz Intel Core i7 CPU. VINS-Mono processes frames and provides estimates at 10 Hz, with IMU measurements sampled at 900 Hz, thrust inputs at 150 Hz, and camera images at 40 Hz. An external force is applied programmatically in the simulation, therefore its ground truth is known. We acquired simulated datasets for two trajectory shapes: trajectory 1 is 73.7 m long and is generated by arbitrarily choosing waypoints (Figs. 3a and 3b); trajectory 2 is helical eight (Figs. 3e and 3f) given by formulation $\mathbf{p}(\theta) = [l_x \sin 2\theta, l_y \cos\theta, \frac{h}{2\pi}(\sin\theta - \theta)]$ with $l_x = 2\,\mathrm{m}$ $l_y = 4\,\mathrm{m}$ and height $h = 3.2\,\mathrm{m}$. In the first set of experiments, the quadrotor flies undisturbed at speeds of $[1, 2, 2.5, 4, 5]\mathrm{m\,s^{-1}}$, while in the second set external forces act on the vehicle flying at $[1, 2, 2.5]\mathrm{m\,s^{-1}}$. In all the experiments, the reference heading was set to sinusoidally change with a magnitude of 30°. We first compare the performance of our approach (VIMO) against VINS-Mono in terms of accuracy and computation times. Finally, we compare the quality of the external force estimate against the estimate obtained from a naive approach.

Comparison with VINS-Mono: Fig. 3 shows plots comparing simulation performance of VINS-Mono with VIMO on the two trajectory shapes flown at 2.5 m/s top speed and disturbed with external forces. This scenario represents the worst performance of VIMO in comparison with VINS-Mono on trajectory 1 and an average performance for trajectory 2, as visible from Tab. I. The plots were generated and the absolute and relative errors were computed using the open source trajectory evaluation toolbox for VIO pipelines [23]. For all the experiments, we align all the estimated states to the ground truth using *posyaw* trajectory alignment method of the toolbox. The top and side view of the estimated trajectories by VINS-Mono and VIMO almost overlap and are very close

TABLE I: Comparison between performance of VINS and VIMO.

		trans. RMSE (m)			**rot. RMSE (deg)**			**avg solve time (ms)**		**max solve time (ms)**	
	top speed (m/s)	VINS	VIMO	% decrease	VINS	VIMO	% decrease	VINS	VIMO	VINS	VIMO
Trajectory 1: 73.7 m without external forces	1.0	0.066	**0.039**	40.9	1.40	**0.57**	59.3	42.0	40.9	52.1	54.7
	2.0	0.093	**0.073**	21.5	0.69	**0.64**	7.2	39.9	39.9	61.8	63.3
	2.5	0.085	**0.076**	10.6	0.60	**0.56**	6.7	38.5	38.7	50.	49.7
	4.0	0.038	**0.033**	13.2	0.49	**0.36**	26.5	37.9	38.0	49.1	50.5
	5.0	0.068	**0.062**	8.8	0.66	**0.47**	28.8	38.3	38.3	51.1	53.8
Trajectory 1: 73.7 m with external forces	1.0	0.105	**0.089**	15.2	1.81	**0.75**	58.6	42.0	40.7	52.2	54.2
	2.0	0.057	**0.051**	10.5	0.75	**0.61**	18.7	39.6	39.7	50.8	55.5
	2.5	**0.055**	0.059	- 7.3	0.71	**0.69**	2.8	39.3	38.8	59.7	51.0
Trajectory 2: 65.8 m without external forces	1.0	0.228	**0.189**	17.1	1.45	**1.12**	22.8	40.7	40.9	54.0	60.7
	2.0	0.147	**0.143**	2.7	0.67	**0.42**	37.3	39.7	39.1	52.6	51.8
	2.5	0.203	**0.158**	22.2	0.74	**0.48**	35.1	39.3	38.5	77.2	54.4
	4.0	0.085	**0.068**	20.0	0.81	**0.65**	19.8	38.3	38.0	50.5	57.4
	5.0	0.073	**0.061**	16.4	0.72	**0.48**	33.3	38.2	38.0	51.8	61.6
Trajectory 2: 65.8 m with external forces	1.0	0.162	**0.154**	4.9	1.29	**1.00**	22.5	40.8	40.9	55.6	61.8
	2.0	0.157	**0.136**	13.4	0.74	**0.62**	16.2	40.2	38.8	84.5	58.7
	2.5	0.094	**0.061**	35.1	0.64	**0.52**	18.8	39.5	38.5	52.1	61.7



(a) Trajectory 1 – x-axis (b) Trajectory 1 – y-axis (c) Trajectory 1 – z-axis

(d) Trajectory 2 – x-axis (e) Trajectory 2 – y-axis (f) Trajectory 2 – z-axis

Fig. 4: Comparison between external force estimates from VIMO — (pink), the naive approach — (green) and calculated ground truth — (blue) on the random trajectory (top, a - c) and the helical-eight trajectory (bottom, d - f). The external force estimate consists of air drag in body x- and y-axis and 2 external forces applied at $t = 10\,\mathrm{s}$ and $t = 32\,\mathrm{s}$ for the top experiment and $t = 47\,\mathrm{s}$ and $t = 68\,\mathrm{s}$ for the bottom experiment.

to the ground truth. For this worst-case scenario, the relative translation error for VIMO is less than or similar to the error for VINS-Mono, while the relative yaw errors for VIMO is slightly higher than VINS-Mono. We report all measured RMSE and computation time for VINS-Mono and VIMO in Table I, together with the percentage decrease in RMSE of VIMO compared to VINS-mono. The maximum increase in accuracy is $\sim 40\%$, experienced at a speed of 1 m/s in random trajectory, without external forces, while one outlying experiment (trajectory 1, with forces at $2.5\,\mathrm{m\,s^{-1}}$) showed a decrease of accuracy. Overall, we achieve a decrease in translational RMSE of $\sim 15\%$, and a decrease in rotational RMSE $\sim 25\%$ in the simulated experiments. In general, the addition of dynamic residuals excels especially in scenarios of low signal-to-noise ratio in the IMU data, which occur at low accelerations. While we could tune the parameters $\boldsymbol{\sigma}_T$ and w_f to increase the accuracy of individual experiments, we wanted to fairly evaluate our estimator's performance without tuning between scenarios to accurately represent real-world applications. In addition to increasing the accuracy, it can be observed in Tab. I that our approach does not increase the average solving time, but keeps it nearly equal to VINS-Mono.

Evaluation of External Force Estimate: In this section we compare VIMO's external force estimate against the estimate obtained from a naive approach and the ground truth. We compute a naive deterministic estimate as $\hat{\mathbf{f}}_{e_t} = \hat{\mathbf{a}}_t^b - \hat{\mathbf{T}}_t^b$ by simply subtracting the mass normalised thrust $\hat{\mathbf{T}}_t^b$ from the accelerometer measurements $\hat{\mathbf{a}}_t^b$. Fig 4 shows plots of force estimates obtained for the different trajectory shapes flown at 2.5 m/s top speed. In both the experiments, we disturb the quadrotor at its center of mass by 2 external forces for 2 seconds each, one after the other, in all three body axes. The ground truth of the external force is computed as a sum of mass normalised external disturbance measured by the force sensor and the drag force. Since RotorS does not provide ground truth of the drag force, we approximate it offline using the linear drag model $-diag([d_x, d_y, d_z])R_b^{w\,T} v_b^w$ [24], and the ground truth rotation, velocity and mass normalized drag coefficients d_x, d_y, d_z from the simulator. We assume $d_z = 0$ because the drag in body z axis is very small. From the plots it is evident that the naive deterministic estimate needs additional filtering and bias removal steps, whereas our estimator implicitly takes into account the noise characteristics of the IMU, its bias, the noise in the state estimates, and the noise of the commanded thrust. Hence, our estimate lies closer to the computed ground truth force. The plots also show that the force estimates take time to converge at the beginning, as long as the IMU bias estimate is not converged (first $\sim 8 - 10\mathrm{s}$). One peculiarity visible in Fig 4(f) are the peaks in the estimate at $t = 32\,\mathrm{s}$ and $t = 88\,\mathrm{s}$, which are not visible in the ground truth. This is the result of a high change in commanded thrust, while the actual thrust has latency introduced by the motors and speed controllers.

B. Blackbird Dataset

Experiment Setup: Additionally, we evaluate the performance of VIMO and VINS-Mono on the Blackbird dataset from [21], which uses a motion capture system for closed-loop control of a UAV along fast trajectories, while rendering photorealistic images of synthetic scenes synchronized with onboard IMU and rotor thrust measurements. We use the two sequences *star* and *picasso* at speeds from 1 to $4\,\mathrm{m\,s^{-1}}$ with the camera forward-facing for the *star* sequence and at a fixed yaw for the *picasso* sequence. Since this dataset does not include any applied external forces, we only evaluate pose estimation as direct comparison on the public available dataset for reproducibility. Since the dataset contains IMU measurements at 100 Hz, we downsample the images, which are available at a faster rate of 120 Hz, to 30 Hz to allow proper IMU preintegration. We use the rotor thrust measurements at the provided ~ 190 Hz.

Evaluation: Also for the Blackbird dataset [21], we use the trajectory alignment toolbox from [23] with the *posyaw* alignment. Even though this dataset does not include sequences with applied (and measured) external forces, we could measure a slight performance increase as shown in Table II. Different from most available datasets (Sec. V-B), the Blackbird dataset includes the rotor speed measurements which we exploit

TABLE II: Blackbird Dataset Evaluation

	trans. RMSE (m)			rot. RMSE (deg)		
	VINS	VIMO	%decrease	VINS	VIMO	%decrease
star 1m/s	0.102	**0.088**	13.7	**0.46**	0.48	-4.3
star 2m/s	0.133	**0.082**	38.2	0.67	**0.60**	10.5
star 3m/s	0.235	**0.183**	22.1	0.96	**0.88**	8.7
picasso 1m/s	0.097	**0.055**	43.5	**0.67**	0.77	-14.9
picasso 2m/s	0.043	**0.040**	7.8	0.46	**0.43**	9.1
picasso 3m/s	0.045	**0.043**	2.9	0.34	**0.30**	14.6
picasso 4m/s	0.056	**0.049**	11.9	0.67	**0.53**	21.7

(a) Snapdragon Flight Quadrotor (b) Force sensor ATI Mini40

Fig. 6: Experimental quadrotor platform (a) and an force/torque sensor (b).

through the known system dynamics and achieve superior accuracy in nearly all test sequences. One can observe that in the *star* trajectory the translational errors are generally higher and the highest tested speed is $3\,\mathrm{m\,s^{-1}}$. This is because of the high yaw rate and the resulting high optical flow, rendering the estimation problem more difficult, and causing the system to fail at $4\,\mathrm{m\,s^{-1}}$ without significant retuning.

C. Real-World Validation

Experiment Setup: To fully validate our approach, we provide a real world experiment where we record data consisting of camera frames, IMU data, commanded collective thrust, force measurements and quadrotor state ground truth. For the quadrotor, we used an ARM-based platform with a monochrome global-shutter VGA resolution camera at 30 Hz synchronized with an IMU providing inertial data at 400 Hz, based on the Qualcomm Snapdragon Flight as depicted in Fig. 6a. For the experiment we used our inhouse-developed flight stack. To provide ground truth data, we employed an OptiTrack motion capture system. As a force ground truth, we used an ATI Mini40-SI-20-1 force and torque sensor (Fig, 6b) also tracked in our motion capture system to recover the direction of the force. We evaluate disturbance-free figure-eight trajectory flight with $l_x = 2.25\,\mathrm{m}$, $l_y = 1.5\,\mathrm{m}$, $h = 0\,\mathrm{m}$, and disturbing the vehicle in hover with $\sim 3\,\mathrm{N}$ by pushing it with the force-measurement pole.

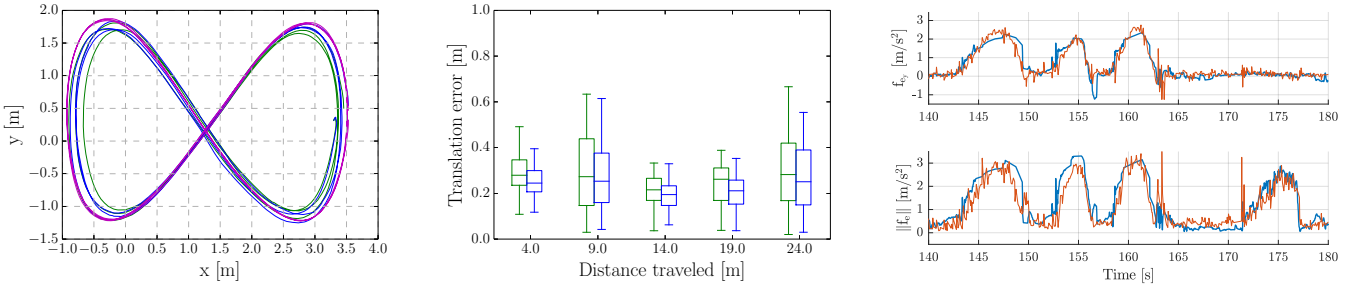
Evaluation: As a simple validation of our approach, we depict the top view on the position estimate of VINS-Mono, VIMO and ground truth in Fig. 5a, indicating a very similar performance of both approaches. We evaluate a translational RMSE of $0.1069\,\mathrm{m}$ for VIMO and $0.1497\,\mathrm{m}$ for VINS-Mono, corresponding to 29% error reduction, while the rotational RMSE is at $4.95°$ for VIMO and $5.15°$ for VINS-Mono, corresponding to 4% error reduction. Fig. 5b reports the error statistics on the real world data computed with the trajectory evaluation toolbox [23]. Additionally, we disturbed the vehicle with $\sim 3\,\mathrm{N}$ while in hover, as shown in Fig. 5c. The estimate is accurate, while noisy due to high vibrations on the used vehicle.

V. Discussion

A. Limitations due to Measurement Modality

While our approach offers the benefits of improving state estimates and estimating external force, it also comes with two limitations due to its measurement modality.

First, we consider the acceleration measurement $\mathbf{a}_k^b$ in body frame $\mathbf{a}_k^b - \mathbf{b}_{a_k}^b = \mathbf{T}_k^b + \mathbf{f}_{e_k}^b = \mathbf{T}_k^b + \bar{\mathbf{f}}_{e_k}^b + \mathbf{f}_{d_k}^b$ where we have separated $\mathbf{f}_{e_k}^b$ from (5) into the true external force $\bar{\mathbf{f}}_{e_k}^b$ and the aerodynamic drag $\mathbf{f}_{d_k}^b$ force. While $\mathbf{a}_k^b$ and $\mathbf{T}_k^b$ are measured quantities, all other quantities have to be estimated. Due to the additive nature of external ($\bar{\mathbf{f}}_{e_k}^b$) and drag ($\mathbf{f}_{d_k}^b$) force, one can only estimate the sum of both ($\mathbf{f}_{e_k}$, as done in this paper) if one does not add any additional assumption or model the aerodynamic drag. Furthermore, the same additive nature introduces an ambiguity between external force (i.e. summed $\mathbf{f}_{e_k}^b$) and the bias $\mathbf{b}_{a_k}^b$. But contrary to force and drag, summed external force and bias can be discriminated by their different dynamics, implemented as an additional prior. Due to the nature of IMU bias, we have to assume a random walk prior by $\dot{\mathbf{b}}_{a_k} = \mathcal{N}(\mathbf{0}, \sigma_{b_a}^2)$ with $\mathcal{N}$ as the Gaussian distribution. In contrast, we assume the external forces to be zero-mean Gaussian (10), since we are mainly interested in detecting incidental changes in the force. Any constant component in the external force will be estimated as



(a) Real world x-, y- position comparison between VINS — (green), VIMO — (ours, blue) and ground truth — (purple) in top view.

(b) Translational errors over trajectory segments of VINS — (green), VIMO — (ours, blue) on real world data as statistical box plot.

(c) External force estimate f_{e_y} (top) and $\|\mathbf{f}_e\|$ (bottom) on real world data compared between VIMO — and a force sensor — with a disturbance of $\sim 3\,\mathrm{N}$ magnitude.

Fig. 5: Real world experiments flying a figure 8 trajectory at $1.5\,\mathrm{m\,s^{-1}}$ depicted in top view x,y-plot (Fig. (a)) and statistical box plots (Fig. (b)). Fig. (c) shows the force estimate and ground-truth (obtained with a force sensor) of a disturbance of $\sim 2\,\mathrm{N}$.

accelerometer bias, since the bias is the only estimated variable without cost on its magnitude, effectively forming a low-pass filter. Further evaluations of the observability of visual-inertial localization can be found in [25]. Finally, we use commanded thrust in the dynamic model whereas the accelerometer detects acceleration due to the actual rotor thrust. Therefore, our estimator comprehends the difference between commanded and actual thrust, if large enough, as external force. This is observed as mentioned before in Sec. IV-A as peaks in Fig 4(f), indicating that VIMO also has the capability to detect model inaccuracy as external force. This difference between commanded and actual rotor thrust could be mitigated by using advanced motor speed controllers with feedback on the actual rotor speed or by modelling the motor dynamics.

B. Other Datasets

Several existing UAV visual-inertial datasets, such as EuRoC MAV [26], UPenn fast flight [27], Zurich Urban MAV [28], have been used extensively for evaluating the performance of VIO. Although these datasets include synchronized camera and IMU data with accurate ground truth, we could not use them to evaluate our approach since they do not provide rotor speed measurements or commanded thrust.

VI. Conclusion

This paper extends a visual inertial estimator by adding a motion constraint derived from the dynamic model including external forces. The resulting tightly-coupled system is shown to accurately estimate vehicle's motion, IMU biases, and external forces, from visual and inertial measurements and commanded thrust inputs. Thereby, our approach enables differentiation between actuation and disturbance by the detected external forces. Inspired from IMU preintegration, the high-rate collective rotor thrust is preintegrated into relative motion constraints, implemented as residuals into an existing VIO pipeline (VINS-Mono). Synthetic and real world experiments, conducted in the presence of external disturbances, illustrate that, compared to VINS-Mono, our estimator not only improves odometry accuracy up to 29% on real world data, but also estimates time-varying external forces without increasing the computation time. Our unified state and force estimator enables a robot to sense motion and external forces, opening the door to a number of possible future research works and applications. As a call to the community, we want to raise awareness for the importance of contact-enabled robotics and the need for estimators to provide not only odometry information, but also leverage the robot dynamics to increase accuracy and sense external forces from contacts and interaction.

References

[1] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza. SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Trans. Robot.*, 33(2):249–265, 2017.

[2] T. Qin, P. Li, and S. Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. In *IEEE Trans. Robot.*, July 2018.

[3] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual-inertial SLAM using nonlinear optimization. *Int. J. Robot. Research*, 2015.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Trans. Robot.*, 31(5):1147–1163, 2015.

[5] J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. 2016. URL http://arxiv.org/pdf/1607.02565.pdf.

[6] G. Loianno, C. Brunner, G. McGrath, and V. Kumar. Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and IMU. *IEEE Robot. Autom. Lett.*, 2017.

[7] J. Delmerico and D. Scaramuzza. A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. *IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018.

[8] T. Tomic and S. Haddadin. A unified framework for external wrench estimation, interaction control and collision reflexes for flying robots. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2014.

[9] B. Yüksel, C. Secchi, H. Bülthoff, and A. Franchi. A nonlinear force observer for quadrotors and application to physical interactive tasks. In *IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, 2014.

[10] F. Ruggiero, J. Cacace, H. Sadeghian, and V. Lippiello. Impedance control of vtol uavs with a momentum-based external generalized forces estimator. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014.

[11] C. D. McKinnon and A. P. Schoellig. Unscented external force and torque estimation for quadrotors. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2016.

[12] F. Augugliaro and R. D'Andrea. Admittance control for physical human-quadrocopter interaction. In *IEEE Eur. Control Conf. (ECC)*, 2013.

[13] A. Tagliabue, M. Kamel, S. Verling, R. Siegwart, and J. Nieto. Collaborative transportation using mavs via passive force control. In *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2017.

[14] R. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82:35–45, 1960.

[15] D. Abeywardena, Z. Wang, G. Dissanayake, S.L. Waslander, and S. Kodagoda. Model-aided state estimation for quadrotor micro air vehicles amidst wind disturbances. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2014.

[16] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Trans. Robot.*, 33(1):1–21, 2017.

[17] T. Lupton and S. Sukkarieh. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. *IEEE Trans. Robot.*, 2012.

[18] Amado Antonini. Pre-integrated dynamics factors and a dynamical agile visual-inertial dataset for UAV perception. Master's thesis, Massachusetts Institute of Technology, 2018.

[19] D. Ta, M. Kobilarov, and F. Dellaert. A factor graph approach to estimation and model predictive control on unmanned aerial vehicles. In *IEEE Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, 2014.

[20] M. Koliraov, D. Ta, and F. Dellaert. Differential dynamic programming for optimal estimation. *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015.

[21] A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, and S. Karaman. The blackbird dataset: A large-scale dataset for UAV perception in aggressive flight. In *Int. Symp. Experimental Robotics (ISER)*, 2018.

[22] F. Furrer, M. Burri, M. Achtelik, and R. Siegwart. RotorS—a modular gazebo MAV simulator framework. In *Studies in Computational Intelligence*, pages 595–625. Springer, 2016.

[23] Z. Zhang and D. Scaramuzza. A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018.

[24] M. Faessler, A. Franchi, and D. Scaramuzza. Differential flatness of quadrotor dynamics subject to rotor drag for accurate tracking of high-speed trajectories. *IEEE Robot. Autom. Lett.*, 3(2):620–626, April 2018.

[25] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Camera-IMU-based localization: Observability analysis and consistency improvement. *Int. J. Robot. Research*, 33(1):182–201, 2014.

[26] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart. The EuRoC micro aerial vehicle datasets. *Int. J. Robot. Research*, 35:1157–1163, 2015.

[27] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception. *IEEE Robot. Autom. Lett.*, 3(3):2032–2039, July 2018.

[28] A. L. Majdik, C. Till, and D. Scaramuzza. The Zurich urban micro aerial vehicle dataset. *Int. J. Robot. Research*, 2017.