An Accurate Filter-based Visual Inertial External Force Estimator via Instantaneous Accelerometer Update

Junlin Song, Antoine Richard, and Miguel Olivares-Mendez

***Abstract*— Accurate disturbance estimation is crucial for reliable robotic physical interaction. To estimate environmental interference in a low-cost and sensorless way (without force sensor), a variety of tightly-coupled visual inertial external force estimators are proposed in the literature. However, existing solutions may suffer from relatively low-frequency preintegration. In this paper, a novel estimator is designed to overcome this issue via high-frequency instantaneous accelerometer update.**

I. Introduction

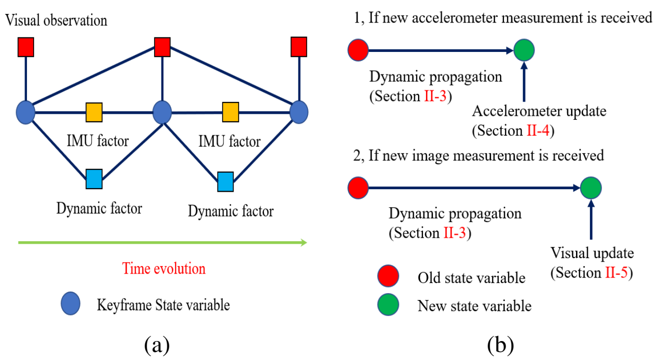

Fig. 1: (a) Factor graph framework used in [1], [2], [3], [4]. (b) Our novel filter framework.

VIMO [1] is the first approach exploiting the external force within the framework of optimization-based visual inertial odometry (VIO) [5]. Dynamic measurements are preintegrated as relative motion constraints between keyframes, refering to IMU preintegration [6]. However, the unknown external force is modeled as zero-mean Gaussian. Therefore, the estimator will be greatly affected by large or continuous external force. VID-Fusion [2] improves VIMO by exploiting the prior knowledge of external force derived from IMU and dynamic measurements. However, the external force preintegration term between adjacent keyframes is averaged to approximate the force variable. This approximation introduces estimation error. HDVIO [3] and VIEW [4] employ similar factor graph framework, as shown in the Fig. 1a.

To accurately and efficiently estimate the external force, we propose a novel filter-based framework, which is presented in the Fig. 1b. Dynamic and gyroscope measurements are utilized to propagate the state variable. Accelerometer and camera measurements are used to update the state with respective frequency. The external force is updated with instantaneous accelerometer measurements without average approximation error in factor graph framework, which is prone to suffer from relatively "long-term" dynamic preintegration. We adopt a Multi-State Constraint Kalman Filter (MSCKF) [7] framework based on OpenVINS [8]. The superiority of our approach is demonstrated in the UAV case.

This research was supported by the European Union's Horizon 2020 project SESAME (grant agreement No 101017258).

Space Robotics (SpaceR) Research Group, Int. Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg.

II. Problem Formulation

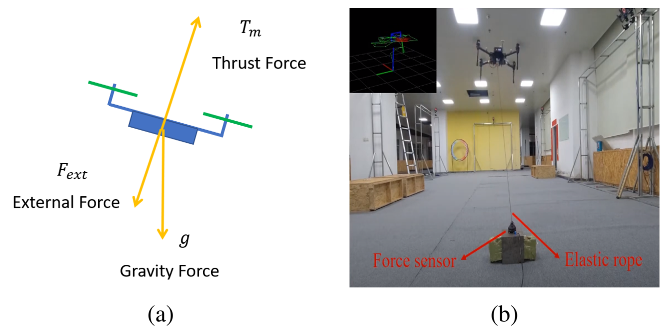

Fig. 2: (a) Mass-normalized forces of UAV. (b) Data collection platform of VID-Dataset [9].

1) Notation: There, let us define the gravity aligned world coordinate frame as $\{W\}$, the geometric central coordinate frame as $\{B\}$, the centroid frame as $\{M\}$, and the IMU frame as $\{I\}$. Frame $\{B\}$, $\{M\}$ and $\{I\}$ are assumed to be coincident, like [1], [2], [3]. Hereafter, only $\{I\}$ is used to refer to the body coordinate frame of UAV.

${}^{W}(\bullet)$ represents a physical quantity in frame $\{W\}$. The position of I in frame $\{W\}$ is expressed as ${}^{W}p_I$. The velocity of I in frame $\{W\}$ is expressed as ${}^{W}v_I$. A unit quaternion, ${}^{I}_{W}q$, represents the attitude of frame $\{I\}$ with respect to frame $\{W\}$, and its corresponding rotation matrix is ${}^{I}_{W}R$. The transpose of a matrix is $[\bullet]^T$.

The force analysis of a UAV is shown in the Fig. 2a. The resultant force other than gravity and thrust is termed as external force. The mass-normalized thrust and external force are denoted as T_m and F_{ext} respectively. These two forces are expressed in frame $\{I\}$. g is gravity vector, which is expressed in frame $\{W\}$.

2) State Vector: The MSCKF state vector includes the current robot state, N augmented historical pose clones and L augmented visual features:

$$\begin{aligned} &x = \begin{bmatrix} x_I^T & x_c^T & x_f^T \end{bmatrix}^T \\ &x_I = \begin{bmatrix} {}^{I}_{W}q^T & {}^{W}p_I^T & {}^{W}v_I^T & b_\omega^T & b_a^T & F_{ext}^T \end{bmatrix}^T \\ &x_c = \begin{bmatrix} x_{c_1}^T & \cdots & x_{c_N}^T \end{bmatrix}^T x_{c_i} = \begin{bmatrix} {}^{I_i}_{W}q^T & {}^{W}p_{I_i}^T \end{bmatrix}^T \\ &x_f = \begin{bmatrix} {}^{W}p_{f_1}^T & \cdots & {}^{W}p_{f_L}^T \end{bmatrix}^T \end{aligned} \tag{1}$$

Where x_I is the current robot state, including the robot pose, velocity, the bias of IMU and external force. x_{c_i} is the augmented robot pose, which is obtained by cloning the first two physical quantities of x_I at different camera times. N is known as the sliding window size, a fixed parameter. The pose clones in the sliding window are used for the triangulation of environmental visual feature points. ${}^{W}p_{f_j}$ is augmented feature, or SLAM feature [8].

3) Dynamic Propagation: The nonlinear continuous-time process model of the system can be expressed as:

$$\begin{aligned} {}_{W}^{I}\dot{q} &= \frac{1}{2}\Omega\left(\omega_m - b_\omega\right){}_{W}^{I}q \\ {}^{W}\dot{p}_I &= {}^{W}v_I \\ {}^{W}\dot{v}_I &= {}_{W}^{I}R^T\left(T_m + F_{ext}\right) + g \\ \dot{b}_\omega &= n_{b_\omega}, \dot{b}_a = n_{b_a}, \dot{F}_{ext} = n_F \end{aligned} \tag{2}$$

Where the definition of $\Omega\left(\omega\right)$ can be found in [7], [8]. $n_{[\bullet]}$ represents the zero mean Gaussian noise of $[\bullet]$. x_I is driven by angular velocity measurement ω_m, and thrust measurement T_m, which is sampled as IMU frequency.

4) Accelerometer Measurement Update: Accelerometer measurement a_m can observe external force.

$$a_m = h\left(x_I\right) = h\left(b_a, F_{ext}\right) = T_m + F_{ext} + b_a + n_a \tag{3}$$

Unlike factor graph framework in the Fig. 1a, the external force is updated with accelerometer frequency. We found this is an easier and more effective way than preintegration.

5) Visual Measurement Update: We follow OpenVINS [8] for the technical details of visual update.

III. Results

To the best of our knowledge, in the UAV community, VID-Dataset [9] is the first and only dataset that contains real-world visual-inertial measurement, thrust measurement, pose groundtruth and external force groundtruth. Sequence *17* and *18* are chosen for validation, as only these two sequences have the groundtruth of external force. Data collection platform is shown in the Fig. 2b. The external force of UAV is dominated by the tension force of elastic rope. Therefore, the tension force measured by the force sensor can be regarded as the external force of UAV, namely F_{ext}.

Estimation results of F_{ext} are shown in the Fig. 3. The quantified RMSE results are provided in the Table I. Compared with VID-Fusion [2], our method has significant accuracy improvement. The quantified absolute trajectory error (ATE) results are also shown in the Table I. Our method achieves better pose estimation than VID-Fusion.

IV. Conclusion

In this work, a novel visual inertial external force estimator is proposed to address the issue of unknown environmental interference for the robot. Compared with SOTA, the estimation accuracy is significantly improved. The methodology itself is generic and can be applied to other robotic platform than UAV, like legged robot [4], by modifying the dynamic model in Section II-.3. Accurate knowledge about external force would provide strong support for downstream applications, such as control and planning [10], [11].

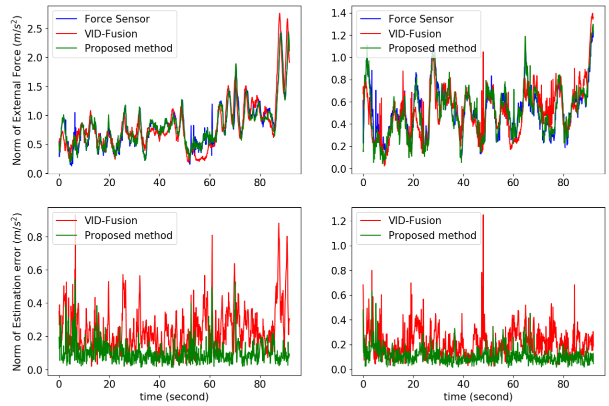

Fig. 3: External force estimation: *17* (left) and *18* (right).

TABLE I: External force estimation RMSE (m/s^2) and ATE (m) of different algorithms (VID-Fusion, Ours) are evaluated with each sequence (seq.) of VID-Dataset.

Seq.	External force estimation RMSE			ATE		
	VID-Fusion	Ours	Decrease	VID-Fusion	Ours	Decrease
17	0.206	**0.072**	65.05%	0.0456	**0.0362**	20.61%
18	0.160	**0.074**	53.75%	0.0788	**0.0499**	36.68%

References

[1] 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.

[2] Z. Ding, T. Yang, K. Zhang, C. Xu, and F. Gao, “Vid-fusion: Robust visual-inertial-dynamics odometry for accurate external force estimation,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14469–14475, IEEE, 2021.

[3] G. Cioffi, L. Bauersfeld, and D. Scaramuzza, “Hdvio: Improving localization and disturbance estimation with hybrid dynamics vio,” in *Robotics: Science and Systems*, 2023.

[4] J. Kang, H. Kim, and K.-S. Kim, “View: Visual-inertial external wrench estimator for legged robot,” *IEEE Robotics and Automation Letters*, 2023.

[5] 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.

[6] 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.

[7] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint kalman filter for vision-aided inertial navigation,” in *Proceedings 2007 IEEE international conference on robotics and automation*, pp. 3565–3572, IEEE, 2007.

[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)*, pp. 4666–4672, IEEE, 2020.

[9] K. Zhang, T. Yang, Z. Ding, S. Yang, T. Ma, M. Li, C. Xu, and F. Gao, “The visual-inertial-dynamical multirotor dataset,” in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 7635–7641, IEEE, 2022.

[10] Y. Wu, Z. Ding, C. Xu, and F. Gao, “External forces resilient safe motion planning for quadrotor,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8506–8513, 2021.

[11] Y. Wang, J. O’Keeffe, Q. Qian, and D. Boyle, “Kinojgm: A framework for efficient and accurate quadrotor trajectory generation and tracking in dynamic environments,” in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 11036–11043, IEEE, 2022.