2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Interacting Multiple Model Navigation System for Quadrotor Micro Aerial Vehicles Subject to Rotor Drag

Mahmoud A.K. Gomaa, Oscar De Silva, George K.I. Mann, and Raymond G. Gosine

***Abstract*— This paper presents the design of an Interacting Multiple Model (IMM) filter for improved navigation performance of Micro Aerial Vehicles (MAVs). The paper considers a navigation system that incorporates rotor drag dynamics and proposes a strategy to overcome the sensitivity of the system to external wind disturbances. Two error state Kalman filters are incorporated in an IMM filtering framework. The first filter has a model that uses conventional Inertial Navigation System (INS) mechanization equations, while the second filter considers a dynamic model with rotor drag forces of the MAV. In order to support the two error state Kalman filters, the generic IMM algorithm [1] is modified for error state implementation, handle dissimilar state definitions, and adaptive switching during operation. Numerical simulations and experimental validation using the EuRoC dataset are conducted to evaluate the performance of the proposed IMM filter design for changing flight conditions and external wind disturbance scenarios.**

I. INTRODUCTION

Reported visual intertial navigation systems (VINS) designs primarily rely on either optimization-based approaches [2] or filtering-based approaches [3, 4]. Filtering-based approaches are preferred for computationally efficient implementation of VINS as compared in [5]. The number of camera poses stored in filtering-based VINS dictates the type of measurement model used for filter update. The number of camera poses appended to the state vector could be as small as two, as in the Epipolar constraint measurement model [4], or more [3]. Incorporating more camera poses improves the estimation accuracy but with an increase in the computational resource requirement of the filter.

For improved performance the process model of the filter should be modified to address the aerodynamic rotor drag forces of the micro-aerial vhicle (MAV) [6]. Recent studies show performance improvement of VINS when the drag force model is used in the navigation equations [4]. Performance improvement as a result of the drag force model is also reported in feedback controller design for aggressive flight of MAVs [7]. However, when there is external disturbance such as wind, ground effects, and modeling errors, or if the MAV has landed, the drag force model is invalid [8, 9]. Many research work in the literature assumes a disturbance-free environment which is only valid for limited number of applications [4, 9]. One approach to address periods with external disturbance is to adaptively switch to the conventional kinematic model, which is unaffected by the existence of wind. This means that there should be a multiple model filtering technique in place to address these changing conditions, i.e., transition between a conventional kinematic model VINS and a drag force model VINS to improve the estimation accuracy.

Literature shows successful application of multi-model techniques, specifically when it comes to the Interacting Multiple Model (IMM) algorithm [10, 11]. IMMs are typically applied to linear systems or system models with low order non-linearities. For instance, target tracking IMMs [12] and image tracking IMMs [13] deal with linear models in the multi-model filtering bank. Similarly, the ground vehicle models and steering geometry models used for IMM filters in [10, 14] exhibit stable dynamics with non-linearities which are significantly different from spatial geometric and projective non-linearities present in VINS filters [3]. Furthermore, IMM is generally applied to filters that share similar state vectors and measurement models, whereas the conventional VINS [15] and drag force VINS [4] filters have a set of conceptually different state and measurement models.

To address these drawbacks, this paper develops an error state implementation of the IMM algorithm incorporating geometrically consistent error definitions, which allow to address non-linearities of VINS filters in the IMM algorithm. The paper introduces an approach to handle dissimilar states and measurement models of the two filters in the filter bank. The performance of the proposed IMM-VINS filter is compared to the drag force VINS proposed in [4]. To the best of authors' knowledge, this paper performs the first reported application of IMM for VINS filters on MAVs.

II. ESTIMATOR DESIGN

A. System Description

Fig 1 defines the coordinate system of the proposed VINS on MAV. Frames $\{G\}$ is the Global frame, $\{B\}$ is attached to MAV center of gravity. The IMU of the MAV is located at frame $\{I\}$ and a forward-facing monocular camera is at frame $\{C\}$. For simplicity, assume that $\{I\}$ and $\{B\}$ are aligned, which will be relaxed when validating the filter for experimental data.

The nonlinear state-space model of the system is given as,

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \mathbf{n}_w) \quad , \quad \mathbf{y} = h(\mathbf{x}, \mathbf{n}_\nu) \tag{1}$$

where $\mathbf{x}$ = system state, $\mathbf{u}$ = input, $\mathbf{n}_w$ = process noise with zero-mean Gaussian, $\mathbf{y}$ = measurement , and $\mathbf{n}_\nu$ = measurements noise with zero-mean Gaussian. The state vector of the system has a dimension of 23 and is defined as follows.

$$\mathbf{x} = \begin{bmatrix} {}^G\mathbf{p}_B{}^T & {}^B\mathbf{q}_G{}^T & {}^B\mathbf{v}^T & \mathbf{b}_a{}^T & \mathbf{b}_g{}^T & {}^G\acute{\mathbf{p}}_B{}^T & {}^B\acute{\mathbf{q}}_G{}^T \end{bmatrix}^T$$

*This work was supported by the Natural Sciences and Engineering Research Council of Canada and Memorial University of Newfoundland.

Authors are with Memorial University of Newfoundland, Canada. makamel, oscar.desilva,gmann,rgosine@mun.ca

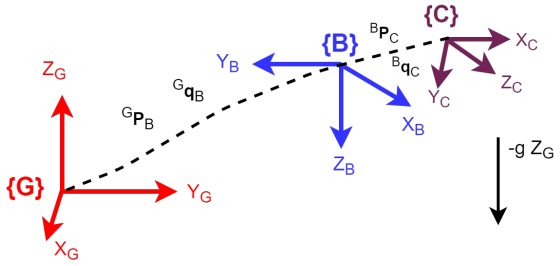

Fig. 1: Coordination systems related to VINS on MAV.

The state includes the position with respect to the global frame ${}^{G}\mathbf{p}_{B}$, unit quaternion ${}^{B}\mathbf{q}_{G}$ corresponding to the MAV rotation from $\{G\}$ to $\{B\}$, MAV velocity ${}^{B}\mathbf{v}$ of $\{B\}$ relative to $\{G\}$ expressed in $\{B\}$, accelerometer bias $\mathbf{b}_a$, and gyroscope bias $\mathbf{b}_g$ expressed in $\{B\}$. A keyframe based approach is used to handle visual measurements similar to [2, 4]. As a result, the position and orientation of the MAV corresponding to the previous keyframe are also stored in the state vector, which are denoted using ${}^{G}\acute{\mathbf{p}}_{B}$ and ${}^{B}\acute{\mathbf{q}}_{G}$ respectively. The error state vector corresponds to the geometric difference between the true state vector $\mathbf{x}$ and the estimated state vector $\hat{\mathbf{x}}$. The error state is defined as, $\tilde{\mathbf{x}} = \mathbf{x} \ominus \hat{\mathbf{x}}$, where the inverse mapping $\ominus$ is used to capture geometrically consistent error terms. This inverse mapping is same as standard subtraction in case of position, velocity, and bias states, but is different for quaternions. The quaternion error state is computed as, $\tilde{\mathbf{q}} = \mathbf{q} * \hat{\mathbf{q}}^{-1} \xrightarrow{\text{Linearize}} \delta\mathbf{q} = \begin{pmatrix} 1 & \frac{1}{2}\delta\boldsymbol{\theta}^T \end{pmatrix}^T$, where $\delta\boldsymbol{\theta}$ is a small angle approximation of rotation and $*$ denotes the quaternion multiplication.

First order linearization of the continuous error state dynamics $\dot{\tilde{\mathbf{x}}}$ results in,

$$\dot{\tilde{\mathbf{x}}} \xrightarrow{\text{Linearize}} \delta\dot{\mathbf{x}} = \mathbf{F}\delta\mathbf{x} + \mathbf{G}_w\delta\mathbf{n}_w \tag{2}$$

where $\mathbf{F}$ and $\mathbf{G}_w$ are the process model and noise Jacobians.

B. Mathematical model of the conventional VINS (C-VINS)

The C-VINS process model corresponds to the INS mechanization equations given by,

$$\begin{gathered} {}^{G}\dot{\mathbf{p}}_{B} = R({}^{B}\mathbf{q}_{G})^{T}\,{}^{B}\mathbf{v} \quad , \quad {}^{B}\dot{\mathbf{q}}_{G} = -\frac{1}{2}\,{}^{B}\boldsymbol{\omega} * {}^{B}\mathbf{q}_{G} \\ {}^{B}\dot{\mathbf{v}} = R({}^{B}\mathbf{q}_{G}){}^{G}g\bar{\mathbf{e}}_{\mathbf{3}} + {}^{B}\mathbf{a} - {}^{B}\boldsymbol{\omega} \times {}^{B}\mathbf{v} \\ {}^{G}\dot{\mathbf{p}}_{B} = \mathbf{0}_{3\times1} \quad , \quad {}^{B}\dot{\mathbf{q}}_{G} = \mathbf{0}_{3\times1} \end{gathered} \tag{3}$$

where the platform angular velocity input ${}^{B}\boldsymbol{\omega}$ and acceleration input ${}^{B}\mathbf{a}$ are driven using measurements from an inertial measurement unit given by,

$$\begin{gathered} \boldsymbol{\omega}_m = ({}^{B}\boldsymbol{\omega} + \mathbf{b}_g + \mathbf{n}_g) \quad , \quad \mathbf{a}_m = ({}^{B}\mathbf{a} + \mathbf{b}_a + \mathbf{n}_a) \\ \dot{\mathbf{b}}_a = \mathbf{n}_{ba} \quad , \quad \dot{\mathbf{b}}_g = \mathbf{n}_{bg}. \end{gathered} \tag{4}$$

In (3,4), $R({}^{B}\mathbf{q}_{G})$ is the rotation matrix from frame $\{G\}$ to $\{B\}$, ${}^{G}g$ is the gravitational acceleration expressed in $\{G\}$, the standard basis $\bar{\mathbf{e}}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$, $\boldsymbol{\omega}_m$ is the gyroscope measurements vector, $\mathbf{a}_m$ is the accelerometer measurements vector, and $\mathbf{n}_a$, $\mathbf{n}_g$, $\mathbf{n}_{ba}$, $\mathbf{n}_{bg}$ are stochastic Gaussian noise variables of the accelerometer measurement, gyroscope measurement, accelerometer bias random walk, and gyroscope bias random walk, respectively. The system noise vector given in (2) is defined as $\mathbf{n}_w = \begin{bmatrix} \mathbf{n}_g^T & \mathbf{n}_a^T & \mathbf{n}_{ba}^T & \mathbf{n}_{bg}^T \end{bmatrix}^T$. The pose stored for the previous keyframe has zero dynamics since it does not change with time. The $\mathbf{F}$ and $\mathbf{G}_w$ matrices corresponding to this model can be found as,

$$\mathbf{F} = \begin{pmatrix} \mathbf{O}_3 & -R(\hat{\mathbf{q}})^T\left[{}^{B}\hat{\mathbf{v}}\right]_\times & R(\hat{\mathbf{q}})^T & \mathbf{O}_3 & \mathbf{O}_3 & \mathbf{O}_{3\times6} \\ \mathbf{O}_3 & -\left[\boldsymbol{\omega}\right]_\times & \mathbf{O}_3 & \mathbf{O}_3 & -\mathbf{I}_3 & \mathbf{O}_{3\times6} \\ \mathbf{O}_3 & \left[R(\hat{\mathbf{q}})\,{}^{G}g\bar{\mathbf{e}}_{\mathbf{3}}\right]_\times & -\left[\boldsymbol{\omega}\right]_\times & -\mathbf{I}_3 & -\left[{}^{B}\hat{\mathbf{v}}\right]_\times & \mathbf{O}_{3\times6} \\ \mathbf{O}_{12\times3} & \mathbf{O}_{12\times3} & \mathbf{O}_{12\times3} & \mathbf{O}_{12\times3} & \mathbf{O}_{12\times3} & \mathbf{O}_{12\times6} \end{pmatrix}$$

$$\mathbf{G}_w = \begin{pmatrix} \mathbf{O}_3 & \mathbf{O}_3 & \mathbf{O}_{3\times6} \\ -\mathbf{I}_3 & \mathbf{O}_3 & \mathbf{O}_{3\times6} \\ -\left[{}^{B}\hat{\mathbf{v}}\right]_\times & -\mathbf{I}_3 & \mathbf{O}_{3\times6} \\ \mathbf{O}_{6\times3} & \mathbf{O}_{6\times3} & \mathbf{I}_6 \\ \mathbf{O}_{6\times3} & \mathbf{O}_{6\times3} & \mathbf{O}_6 \end{pmatrix}$$

where, $[\cdot]_\times$ is the skew symmetric matrix operator. For the sake of brevity, $R(\hat{\mathbf{q}})$ is the same as $R({}^{B}\hat{\mathbf{q}}_{G})$, $\mathbf{I}_i$ is an $i \times i$ identity matrix, $\mathbf{O}_i$ is an $i \times i$ zero matrix, $\mathbf{O}_{i\times j}$ is an $i \times j$ zero matrix, and $\boldsymbol{\omega} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_g$.

We consider visual measurements from one forward-facing monocular camera. Corresponding features $\mathbf{p}_i$ and $\acute{\mathbf{p}}_i$ between each pair of consecutive images are extracted using features detection and features matching techniques. These matched features are used to construct the visual measurement residual using the epipolar geometry constraints as follows,

$$\begin{gathered} \tilde{\mathbf{y}}_v = \tilde{h}(\mathbf{x}, \mathbf{p}_i, \acute{\mathbf{p}}_i) = (\mathbf{p}_i)^T\mathbf{K}^{-T}\mathbf{E}\mathbf{K}^{-1}\acute{\mathbf{p}}_i \\ \mathbf{E} = R({}^{C}\mathbf{q}_{B})R({}^{B}\mathbf{q}_{G})\left[{}^{G}\acute{\mathbf{p}}_{C} - {}^{G}\mathbf{p}_{C}\right]_\times R({}^{B}\acute{\mathbf{q}}_{G})^T R({}^{C}\mathbf{q}_{B})^T \end{gathered} \tag{5}$$

where $\mathbf{E}$ is the essential matrix, $R({}^{C}\mathbf{q}_{B})$ is the rotation matrix from frame $\{B\}$ to $\{C\}$, $\mathbf{K}$ is the camera intrinsic matrix, and ${}^{G}\mathbf{p}_{C}$ and ${}^{G}\acute{\mathbf{p}}_{C}$ are the position of $\{C\}$ with respect to $\{G\}$ of the current and previous poses, respectively.

The linearized error state visual measurement model is formulated as, $\delta\mathbf{y}_v = \mathbf{H}_v\delta\mathbf{x} + \mathbf{G}_v\delta\mathbf{n}_v$, where $\mathbf{H}_v$ and $\mathbf{G}_v$ are the Jacobian measurement and noise matrices of the visual measurement model given in (5). $\delta\mathbf{n}_v = \begin{pmatrix} \delta\mathbf{p}_i & \delta\acute{\mathbf{p}}_i \end{pmatrix}^T$ is the camera measurement noise vector in pixels where $\delta\mathbf{p}_i$ and $\delta\acute{\mathbf{p}}_i$ are the camera noise vectors in pixels for feature points in current and previous keyframes, respectively. Using the error state formulation, the matrices $\mathbf{H}_v$ and $\mathbf{G}_v$ are found as,

$$\mathbf{H}_v = \begin{pmatrix} \mathbf{p}_i^T\mathbf{A}R({}^{B}\hat{\mathbf{q}}_{G})[R({}^{B}\hat{\acute{\mathbf{q}}}_{G})^T\mathbf{B}\ \acute{\mathbf{p}}_i]_\times \\ \mathbf{H}_{v2} \\ \mathbf{O}_{9\times1} \\ -\mathbf{p}_i^T\mathbf{A}R({}^{B}\hat{\mathbf{q}}_{G})[R({}^{B}\hat{\acute{\mathbf{q}}}_{G})^T\mathbf{B}\ \acute{\mathbf{p}}_i]_\times \\ \mathbf{H}_{v7} \end{pmatrix}^T$$

$$\mathbf{G}_v = \begin{pmatrix} \acute{\mathbf{p}}_i^T\mathbf{B}^T R({}^{B}\hat{\acute{\mathbf{q}}}_{G})[\mathbf{C}]_\times^T & R({}^{B}\hat{\mathbf{q}}_{G})^T\mathbf{A}^T \\ \mathbf{p}_i^T\mathbf{A}R({}^{B}\hat{\mathbf{q}}_{G})[\mathbf{C}]_\times & R({}^{B}\hat{\acute{\mathbf{q}}}_{G})^T\mathbf{B} \end{pmatrix}^T$$

where $\mathbf{H}_{v2}$, $\mathbf{H}_{v7}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are defined as follows.

$$\begin{aligned} \mathbf{H}_{v2} = -\mathbf{p}_i^T\mathbf{A}R({}^{B}\hat{\mathbf{q}}_{G})[R({}^{B}\hat{\acute{\mathbf{q}}}_{G})^T\mathbf{B}\ \acute{\mathbf{p}}_i]_\times R({}^{B}\hat{\mathbf{q}}_{G})^T[{}^{B}\mathbf{p}_C]_\times - \\ \acute{\mathbf{p}}_i^T\mathbf{B}^T R({}^{B}\hat{\acute{\mathbf{q}}}_{G})[\mathbf{C}]_\times^T R({}^{B}\hat{\mathbf{q}}_{G})^T[\mathbf{A}^T\ \mathbf{p}_i]_\times \end{aligned}$$

$$\mathbf{H}_{v7} = \mathbf{p}_i^T \mathbf{A} R({}^B\hat{\mathbf{q}}_G)[R({}^B\hat{\mathbf{q}}_G)^T \ \mathbf{B} \ \acute{\mathbf{p}}_i]_\times R({}^B\hat{\mathbf{q}}_G)^T [{}^B\mathbf{p}_C]_\times - \mathbf{p}_i^T \mathbf{A} R({}^B\hat{\mathbf{q}}_G)[\mathbf{C}]_\times R({}^B\hat{\mathbf{q}}_G)^T [\mathbf{B} \ \acute{\mathbf{p}}_i]_\times$$

$$\mathbf{A} = \mathbf{K}^{-T} R({}^C\mathbf{q}_B) \quad , \quad \mathbf{B} = R({}^C\mathbf{q}_B)^T \mathbf{K}^{-1}$$

$$\mathbf{C} = {}^G\hat{\mathbf{p}}_B - {}^G\hat{\mathbf{p}}_B + \left(R({}^B\hat{\mathbf{q}}_G)^T - R({}^B\hat{\mathbf{q}}_G)^T\right) {}^B\mathbf{p}_C$$

C. Mathematical model of the drag-force VINS (DF-VINS)

The system dynamics of the DF-VINS process model is similar to the process model of C-VINS given in (3) except for the velocity dynamics which is given as,

$${}^B\dot{\mathbf{v}} = R({}^B\mathbf{q}_G){}^G g\bar{\mathbf{e}}_\mathbf{3} - \bar{D}_L {}^B\mathbf{v} + \mathbf{f}_{ip} + \mathbf{n}_m \tag{6}$$

where

$$\mathbf{f}_{ip} = \bar{T}_L \bar{\mathbf{e}}_3 - (\boldsymbol{\omega}_m - \mathbf{b}_g + \mathbf{n}_g) \times {}^B\mathbf{v}$$
$$\bar{T}_L = a_{m_z} - b_{az}, \ \bar{D}_L = diag(k_{1_x}, \ k_{1_y}, \ k_{1_z})$$

The random Gaussian noise in the MAV drag force model is denoted by $\mathbf{n}_m$, a_{mz} is the accelerometer measurement in z direction, and b_{az} is the accelerometer bias in z direction. The mass normalized thrust is denoted by $\bar{T}_L$. The mass normalized drag parameters matrix $\bar{D}_L$ is a diagonal matrix with elements $k_{1_x}, k_{1_y}, k_{1_z}$ and is crucial for estimation accuracy of the DF-VINS. The drag parameters in x-axis and y-axis can be estimated following a least squared optimization procedure while it is reasonable to assume that the drag parameter in z-axis to be zero similar to work in [7]. The system model noise vector given in (2) is defined as $\mathbf{n}_w = \begin{bmatrix}\mathbf{n}_g^T & \mathbf{n}_m^T & \mathbf{n}_{ba}^T & \mathbf{n}_{bg}^T\end{bmatrix}^T$. The filtering matrices $\mathbf{F}$ and $\mathbf{G}_w$ corresponding to the DF-VINS are similar to the matrices of C-VINS except for the 7^{th}, 8^{th}, and 9^{th} rows of $\mathbf{F}$ matrix ($\mathbf{F}(7:9,:)$), which were found to be,

$$\mathbf{F}(7:9,:) = \begin{pmatrix}\mathbf{O}_3 & [R(\hat{\mathbf{q}}) {}^G g\bar{\mathbf{e}}_\mathbf{3}]_\times & -\bar{D}_L - [\boldsymbol{\omega}]_\times & -\bar{\mathbf{e}}_\mathbf{3}\bar{\mathbf{e}}_\mathbf{3}^T & -[{}^B\hat{\mathbf{v}}]_\times & \mathbf{O}_{3\times 6}\end{pmatrix}$$

Two measurement models are considered for the DF-VINS model; the first one is the inertial measurement model that contains the accelerometer measurements along x and y directions,

$$\mathbf{h}_a = \Upsilon(-\bar{D}_L {}^B\mathbf{v} + \mathbf{b}_a + \mathbf{n}_a) \tag{7}$$

where $\Upsilon = \begin{bmatrix}\bar{\mathbf{e}}_1^T \ ; \ \bar{\mathbf{e}}_2^T\end{bmatrix}$ is a 2×3 matrix used to extract the first two rows, and $\bar{\mathbf{e}}_1 = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^T$ and $\bar{\mathbf{e}}_2 = \begin{bmatrix}0 & 1 & 0\end{bmatrix}^T$ are the first and second standard basis vectors. The linearized residual measurement model is defined by, $\delta\mathbf{y}_i = \mathbf{H}_i\delta\mathbf{x} + \mathbf{G}_i\delta\mathbf{n}_i$, where $\mathbf{H}_i$ and $\mathbf{G}_i$ are the measurement and noise Jacobian matrices of the inertial measurement model given in (7), and $\mathbf{n}_i$ is the accelerometer measurement noise vector along x-axis and y-axis and defined as, $\mathbf{n}_i = \begin{pmatrix}n_{ax} & n_{ay}\end{pmatrix}^T$. The filtering matrices $\mathbf{H}_i$ and $\mathbf{G}_i$ are defined as,

$$\mathbf{H}_i = \begin{pmatrix}\mathbf{O}_{1\times 6} & -k_{1_x}\bar{\mathbf{e}}_1^T & \bar{\mathbf{e}}_1^T & \mathbf{O}_{1\times 9} \\ \mathbf{O}_{1\times 6} & -k_{1_y}\bar{\mathbf{e}}_2^T & \bar{\mathbf{e}}_2^T & \mathbf{O}_{1\times 9}\end{pmatrix} \quad , \quad \mathbf{G}_i = \mathbf{I}_2$$

The second measurement model of the DF-VINS is same as the visual measurement model of C-VINS given in (5).

D. Error state Kalman filtering

In this work, the filter propagation and update follow an error state filtering formulation as presented in *Algorithm 1- The general continuous-discrete EKF* of [16]. The DF-VINS filter has, (1) inertial measurement update at IMU rate, e.g., 200 Hz for the EuRoC dataset, and (2) visual measurement update that triggers at image acquisition rate, e.g., 20 Hz for the EuRoC dataset. The C-VINS filter only uses the visual measurement update. The visual update is only executed at each new keyframe registered when there is sufficient feature disparity between two images. Use of keyframes avoids inconsistent corrections of the filter when the MAV is stationary because when the camera is stationary, the residual $\tilde{\mathbf{y}}_v$ is a very small value for any given essential matrix $\hat{\mathbf{E}}$. As a result, visual measurements do not contain any useful information about the estimated state vector $\hat{\mathbf{x}}$. Therefore, turning the visual update off minimizes the risk of inconsistent updates leading to divergence. The feature disparity f_d between two corresponding sets of features is calculated as given in (8), where n_f is the number of matched features between the two images. After the measurement correction step is implemented, the state vector and covariance matrix are augmented to update the previous pose information similar to standard state augmentation procedure used in VINS filters [15]. VINS filters have a four-dimensional unobservable space spanning the position state and rotations about the gravity axis. Observability consistency rules are enforced for both filters following the approach in [17] in order to preserve the unobservable space during filter execution.

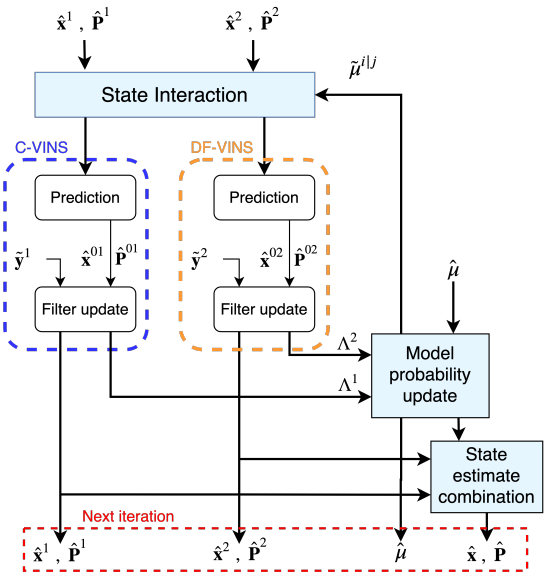

Fig. 2: Structure of IMM algorithm for 2 filters.

$$f_d = \frac{1}{n_f}\sum_{i=1}^{n_f} \|\mathbf{p}_i - \acute{\mathbf{p}}_i\| \tag{8}$$

E. Interacting Multiple Model VINS (IMM-VINS)

An IMM algorithm is proposed to combine estimated state $\hat{\mathbf{x}}^1$ from C-VINS (filter 1) and state $\hat{\mathbf{x}}^2$ from DF-VINS (filter 2) to find a combined estimate $\hat{\mathbf{x}}$ corresponding to the two models. The structure of the IMM algorithm is shown in Fig. 2, where $\hat{\mathbf{x}}^{01}$ and $\hat{\mathbf{x}}^{02}$ are the mixed states, $\hat{\mathbf{P}}^{01}$ and $\hat{\mathbf{P}}^{02}$ are

the mixed covariances which are calculated following a state interaction step. The mixed estimates $\hat{\mathbf{x}}^{01}$ and $\hat{\mathbf{x}}^{02}$ are used in filtering algorithms C-VINS and DF-VINS, respectively. The state interaction step makes use of the model probabilities $\hat{\mu}$ and probability for switching from one model to another $\tilde{\mu}^{i|j}$ which is calculated using likelihoods of the two filter models Λ^1 and Λ^2 at each iteration. The state estimates of the two filters are combined to find the combined estimates for state vector $\hat{\mathbf{x}}$ and covariance $\hat{P}$.

Algorithm 1. Error state IMM algorithm

1: Initialize estimated state vector $\hat{\mathbf{x}}^i$, covariance matrix $\mathbf{P}^i$, and initial probabilities $\hat{\mu}^i$ for each filter.

2: Compute the mixed estimated states $\hat{\mathbf{x}}^{0j}$ and covariance $\mathbf{P}^{0j}$ as,

$$\hat{\mathbf{x}}^{0j} = \sum_{i=1}^{N} \hat{\mathbf{x}}^i \tilde{\mu}^{i|j}$$

$$\mathbf{P}^{0j} = \textstyle\sum_{i=1}^{N} \tilde{\mu}^{i|j} \left[\mathbf{P}^i + (\hat{\mathbf{x}}^i \ominus \hat{\mathbf{x}}^{0j})(\hat{\mathbf{x}}^i \ominus \hat{\mathbf{x}}^{0j})^T\right]$$

$$\tilde{\mu}^{i|j} = \frac{1}{\bar{\psi}^j}\, \rho^{ij} \hat{\mu}^i \;,\; \bar{\psi}^j = \sum_{i=1}^{N} \rho^{ij} \hat{\mu}^i$$

3: Propagate and update estimated states and covariance for each filter model.

4: Compute the likelihood Λ^j and estimated probability for each filter using the innovations $\tilde{\mathbf{y}}^j$ and the innovations covariance matrix $\mathbf{S}^j$,

$$\Lambda^j = \frac{1}{\sqrt{|2\pi \mathbf{S}^j|}} e^{\left[-0.5(\tilde{\mathbf{y}}^j)^T (\mathbf{S}^j)^{-1} (\tilde{\mathbf{y}}^j)\right]}$$

$$\tilde{\mathbf{y}}^j = \mathbf{y} \ominus \hat{\mathbf{y}}^j \quad , \quad \mathbf{S}^j = \mathbf{H}^j \mathbf{P}^{oj} (\mathbf{H}^j)^T + \mathbf{R}$$

$$\hat{\mu}^i = \frac{1}{c} \Lambda^i \bar{\psi}^i \;,\; c = \sum_{i=1}^{N} \Lambda^i \bar{\psi}^i$$

5: Combine both estimated states and covariances based on estimated probabilities,

$$\hat{\mathbf{x}} = arg \min_{\hat{\mathbf{x}} \in M} \sum_{i=1}^{N} \vartheta(\hat{\mathbf{x}}^i, \hat{\mu}^i)$$

$$\mathbf{P} = \textstyle\sum_{i=1}^{N} \hat{\mu}^i \left[\mathbf{P}^i + (\hat{\mathbf{x}}^i \ominus \hat{\mathbf{x}})(\hat{\mathbf{x}}^i \ominus \hat{\mathbf{x}})^T\right]$$

The IMM algorithm is summarized in Algorithm 1, where N is the number of filter models, ρ^{ij} is the ij element of the Markov transition probability matrix and represents the switching from model i to model j, the matrix elements are assigned such that $\sum_j^N \rho^{ij} = 1$, and $\bar{\psi}^i$ is a normalization vector used to normalize the model probability. Compared with the generic IMM algorithm [1, 18], the one proposed in the paper has several key modifications. Error state and measurement residual definitions are used in the state interaction, model probability update, and state combination steps. This makes the algorithm applicable for the error state Kalman filter VINS formulations presented in this work. Furthermore, a generalized state averaging is performed in the state estimation combination step, where ϑ is the averaging function. Function ϑ corresponds to the usual vector averaging in case of position, velocity, and biases states, while optimal quaternion averaging [19] is used for averaging the quaternion states as given in (9). The averaging function effectively minimizes the weighted sum of the squared lengths of the error quaternions $\tilde{\mathbf{q}}$.

$$\hat{\mathbf{q}} = \pm \frac{\left[\left(\hat{\mu}^1 - \hat{\mu}^2 + z\right) \hat{\mathbf{q}}_1 + 2\hat{\mu}^2 \left(\hat{\mathbf{q}}_1^T \hat{\mathbf{q}}_2\right) \hat{\mathbf{q}}_2\right]}{\left\|\left(\hat{\mu}^1 - \hat{\mu}^2 + z\right) \hat{\mathbf{q}}_1 + 2\hat{\mu}^2 \left(\hat{\mathbf{q}}_1^T \hat{\mathbf{q}}_2\right) \hat{\mathbf{q}}_2\right\|} \tag{9}$$

$$\text{where, } z \triangleq \sqrt{\left(\hat{\mu}^1 - \hat{\mu}^2\right)^2 + 4\hat{\mu}^1 \hat{\mu}^2 \left(\hat{\mathbf{q}}_1^T \hat{\mathbf{q}}_2\right)^2}$$

III. Results

A. Numerical validation

A MATLAB simulator is implemented, and compared its performance with the stand-alone estimators (C-VINS and DF-VINS). The simulated arena included 495 feature points uniformly distributed on a cylinder with a radius of $6m$ and a height of $2m$, as shown in Fig. 3. The MAV was simulated using the kinematic model given in (3) with the inputs of the platform $[{}^B\mathbf{a}, {}^B\boldsymbol{\omega}]$. The inputs were designed such that the MAV follows a circular trajectory of radius $4m$ completing two laps with additional excitation along the z-axis to result in a wavelike trajectory as shown in Fig. 3. The input acceleration and angular speeds were designed to adhere to differential flatness constraints related to the drag force model [7]. This implicitly enforces the dynamic constraints related to model (6) during the simulation as long as there is no external disturbance acting on the system. In order to verify the switching capability of the proposed IMM estimator in the presence of external disturbance, $1.76m/s$ wind has been added in the second lap for a short duration, as shown in Fig. 3. The wind disturbance was incorporated into the velocity differential equation in the MAV model as an external force acting on the MAV. The IMU and camera measurements of the MAV were simulated at rates of $100Hz$ and $10Hz$. The noise covariance $\mathbf{Q}$ for C-VINS is set as $Diag(1.1e-3\ \mathbf{I}_3;\ 1.3e-2\ \mathbf{I}_3;\ 1.8e-2\ \mathbf{I}_3;\ 1.7e-4\ \mathbf{I}_3)$ and for DF-VINS is set as $Diag(1.1e-3\ \mathbf{I}_3;\ 1.8e-4\ \mathbf{I}_3;\ 1.8e-2\ \mathbf{I}_3;\ 1.7e-4\ \mathbf{I}_3)$. The standard deviation of camera measurement is set as 1 pixel. The IMM algorithm was implemented with the transition probability matrix and initial model probability vector selected as follows,

$$\rho = \begin{bmatrix} 0.96 & 0.04 \\ 0.04 & 0.96 \end{bmatrix} \;,\; \hat{\mu}^i = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \tag{10}$$

Fig. 3 illustrates the actual and IMM-estimated trajectories of the MAV. Fig. 4 illustrates the position and orientation estimation accuracy of the stand-alone filters and the IMM-VINS. As seen in Fig. 4, DF-VINS filter exhibits improved performance than the C-VINS when there is no external wind disturbance acting on the system; on the other hand, the C-VINS filter is not significantly affected in the presence of the wind disturbance as it relies on a more robust kinematic system model. The IMM estimator dynamically approximates the model probability, as shown in Fig. 5 and generates a combined estimate of the state, which is more accurate than the standalone filters. Moreover, it can also

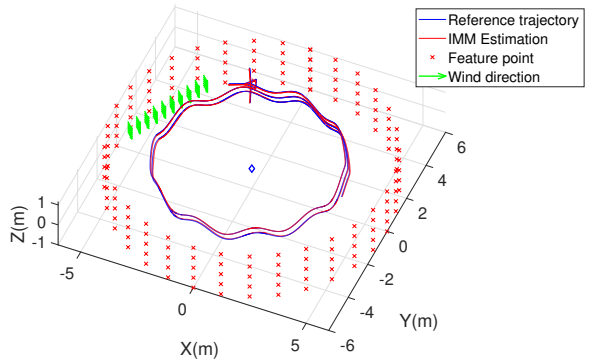

Fig. 3: 3D view of the simulator arena

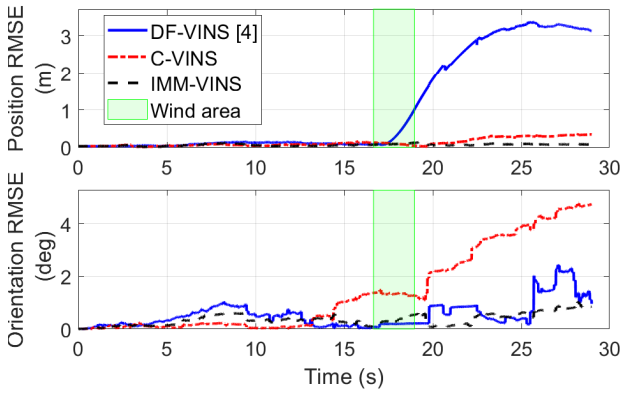

Fig. 4: RMSE of the position and orientation of the IMM and the two stand-alone filters in the presence of wind

handle the high drift of the DF-VINS during the external disturbance period by switching to the correct model based on model probability calculations.

The performance of the C-VINS, DF-VINS, and IMM-VINS filters for three different trajectories are given in Table I. Trajectory 1 does not include any wind disturbance denoted by vector $\mathbf{V}_w$, while trajectory 2 and 3 include wind disturbances of $\mathbf{V}_w = \begin{bmatrix} 0.83 & -0.83 & 0 \end{bmatrix}^T m/s$ and $\mathbf{V}_w = \begin{bmatrix} 1.76 & -1.76 & 0 \end{bmatrix}^T m/s$, respectively. The improvements of the IMM over C-VINS and DF-VINS are given in Table II for the three trajectories. The IMM filter has 26% less drift of its pose when compared with the DF-VINS filter.

B. Experimental validation

Experimental validation is performed using the EuRoC Vicon dataset. The helicopter used to collect the data was equipped with a Visual-Inertial sensor unit, having (a) MEMS IMU operating at 200 Hz (b) two monocular cameras operating at 20 Hz. Only *cam0* measurement is used in this

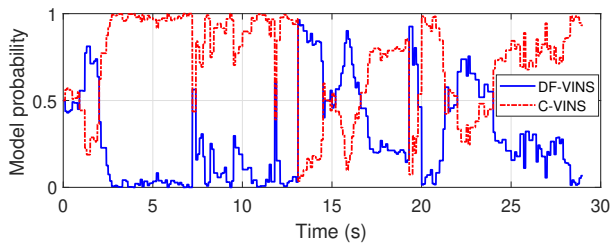

Fig. 5: Models probability of IMM filters

TABLE I: Estimation accuracy (RMSE) of position (m) and orientation (deg) in the simulation results

	C-VINS		DF-VINS [4]		IMM	
	Pos.	Orien.	Pos.	Orien.	Pos.	Orien.
Traj. 1	0.126	1.42	0.087	0.32	**0.064**	**0.31**
Traj. 2	0.128	1.46	0.279	0.48	**0.065**	**0.39**
Traj. 3	0.124	1.43	1.05	0.54	**0.062**	**0.36**

TABLE II: Performance comparison percentage (%) of the improvement in RMSE in the simulation results

	DF-VINS / C-VINS		IMM / C-VINS		IMM / DF-VINS	
	Pos.	Orien.	Pos.	Orien.	Pos.	Orien.
Traj. 1	30.95	77.46	49.21	78.17	26.44	3.13
Traj. 2	-117.97	67.12	49.22	73.29	76.70	18.75
Traj. 3	-746.77	62.24	50	74.83	94.1	33.33

paper with feature tracker front-end data imported to MATLAB from the VINS mono ROS package [20]. The Vicon room 2 datasets are used for this experimental validation of the IMM filter. *V1_01 easy* has a trajectory length of 58.6 m and duration of 144 s. The average linear and angular velocity are 0.41 m/s and 0.28 rad/s, respectively. *V1_02 medium* has a trajectory length of 75.9 m and duration of 83.5 s. The average linear and angular velocity are 0.91 m/s and 0.56 rad/s, respectively. A nonlinear least-squared optimization was used to estimate the drag parameters $\bar{D}_L$ using the ground truth data and the IMU measurements of the V1_02 dataset. The optimal values were found as, $k_{1x} = 0.2$, $k_{1y} = 0.2$, and $k_{1z} = 0.0$.

Since the IMU frame $\{I\}$ of EuRoC dataset is not co-incident with the Center Of Gravity (COG) located at the body frame $\{B\}$ of the MAV, all applied forces including thrust force and drag force is transformed to the body frame $\{B\}$ in order for the drag-force model to work properly. The accelerometer bias of the DF-VINS is updated to include the thrust force and is expressed in $\{B\}$. The accelerometer bias of the C-VINS filter is also expressed in $\{B\}$. As a consequence of these changes, the accelerometer biases of the two filters are dissimilar however, both are expressed in the same frame. This issue is addressed following the unbiased approach for dissimilar state IMM proposed in [21]. In addition, an improved IMM algorithm proposed in [22] is used to adaptively updated the transition probability matrix, which enables fast switching between the filters to accommodate the changing flight conditions of the datasets.

$$\rho = \begin{bmatrix} 0.9999997 & 0.0000003 \\ 0.0001 & 0.9999 \end{bmatrix}, \quad \hat{\mu}^i = \begin{bmatrix} 0.99999997001 \\ 0.00000002999 \end{bmatrix}$$

The performance of the DF-VINS, C-VINS, and IMM filters are illustrated in Fig. 6 and Table III. The IMM has improved performance with minimum position RMSE while the IMM exhibits intuitive dynamic selection of the valid model suitable for the flight condition. The IMM filter might have slightly different performance than the stand-alone filters at some regions because the two filters in the IMM filtering bank exhibit different performance than the stand-alone filters due to the state interaction (mixing) given

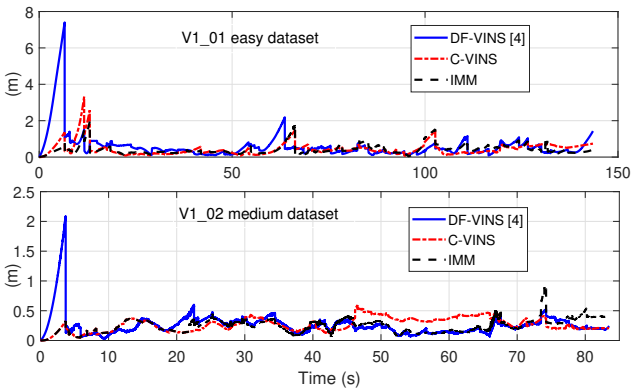

Fig. 6: RMSE of the position given by IMM and other filters.

TABLE III: Estimation accuracy in experimental validation

	C-VINS		DF-VINS [4]		IMM	
	Pos.	Orien.	Pos.	Orien.	Pos.	Orien.
V1_01	0.44	2.35	0.59	2.73	**0.40**	2.36
V1_02	0.27	0.35	0.26	0.35	**0.24**	0.35

in Algorithm 1, where both filters are mixed prior to state update at the beginning of each iteration. Then the updated states are combined to provide the IMM estimated states after state update based on the model probability of each filter. Therefore, the two filters in the IMM bank will be having slightly different performance, but with similar behavior. Fig. 7 depicts the model probability of IMM filters and shows how the IMM switched to the C-VINS filter while the MAV is stationary in the beginning. At slow speeds, the C-VINS model is used by the IMM as seen in the V1_01 dataset, while it switched to the DF-VINS in most of the trajectory in the V1_02 dataset due to the validity of the drag force model during fast and aggressive maneuvering. The multimedia attachment demonstrates these effects along with the estimation performance of the proposed approach.

IV. CONCLUSIONS

This work proposes the IMM-VINS filter to address the drawbacks of the stand-alone drag force VINS filters used for MAV state estimation. In the proposed design, the stand-alone filters in the IMM bank are meant to serve different navigation capabilities (accuracy, stability, robustness) and different flight conditions (aggressive, hover, landing). The IMM allows synergistic combination of these capabilities of the filters during flights to generate improved performance over the stand-alone versions, as shown in this work.

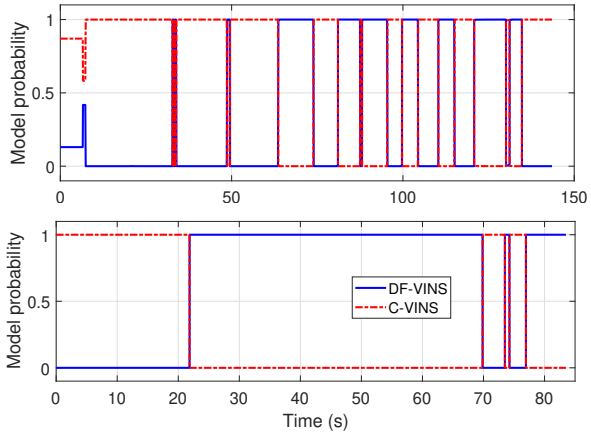

Fig. 7: Variation of model probabilities in the IMM filter. V1_01 dataset (upper) and V1_02 dataset (lower)

REFERENCES

[1] A. F. Genovese, "Interacting multiple model algorithm for accurate state estimation of maneuvering targets," *Johns Hopkins APL Technical Digest*, 2001.

[2] S. Leutenegger and et al., "Keyframe-based visual–inertial odometry using nonlinear optimization," *The Int. J. of Robotics Research*, 2015.

[3] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *2007 IEEE ICRA*, 2007.

[4] D. Abeywardena, S. Huang, B. Barnes, G. Dissanayake, and S. Kodagoda, "Fast, on-board, model-aided visual-inertial odometry system for quadrotor micro aerial vehicles," in *2016 IEEE ICRA*, 2016.

[5] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *2018 IEEE ICRA*, 2018.

[6] X. Zhang, X. Li, K. Wang, and Y. Lu, "A survey of modelling and identification of quadrotor robot," *Abstract and Applied Analysis*, vol. 2014, p. 1–16, 2014.

[7] M. Faessler, A. Franchi, and D. Scaramuzza, "Differential flatness of quadrotor dynamics subject to rotor drag for accurate tracking of high-speed trajectories," *IEEE Robotics and Automation Letters*, 3(2) 2018.

[8] 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. on Intelligent Robots and Systems*, Sep. 2014.

[9] A. Antonini, "Pre-integrated dynamics factors and a dynamical agile visual-inertial dataset for uav perception," Master's thesis, Massachusetts Institute of Technology, MA, USA, 2018.

[10] P. J. Glavine, O. De Silva, G. Mann, and R. Gosine, "Gps integrated inertial navigation system using interactive multiple model extended kalman filtering," in *IEEE MERCon*, 2018.

[11] Li Duan, X. han Huang, Bin Luo, and Qi-yuan Li, "Target tracking with interactive multiple model in geodetic coordinate system for naval ships cooperative engagement," in *11th Int. Conf. on Information Fusion*, 2008.

[12] C. H. Cho, S. Y. Chong, and T. L. Song, "Imm filtering for vehicle tracking in cluttered environments with glint noise," in *Int. Conf. on Control, Automation and Information Sciences*, Oct 2017.

[13] P. Tissainayagam and D. Suter, "Visual tracking with automatic motion model switching," *Pattern Recognition*, 34(3) 2001.

[14] K. Jo, K. Chu, K. Lee, and M. Sunwoo, "Integration of multiple vehicle models with an imm filter for vehicle localization," in *IEEE Intelligent Vehicles Symposium*, June 2010.

[15] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. Jose Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robot. and Autom. Lett.*, vol. PP, 11 2017.

[16] K. Wu and et al, "An invariant-ekf vins algorithm for improving consistency," in *IEEE IROS*, Sep. 2017.

[17] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Camera-imu-based localization: Observability analysis and consistency improvement," *The Int. J. of Robotics Research*, 33(1) 2014.

[18] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: a survey," *IEEE Trans. on Aerospace and Electronic Systems*, 34(1) 1998.

[19] F. L. Markley, Y. Cheng, J. L. Crassidis, and Y. Oshman, "Averaging quaternions," *J. of Guidance, Control, and Dynamics*, 30(4), 2007.

[20] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. on Robotics*, 34(4), 2018.

[21] K. Granström, P. Willett, and Y. Bar-Shalom, "Systematic approach to imm mixing for unequal dimension states," *IEEE Trans. on Aerospace and Electronic Systems*, 51(4), 2015.

[22] B. Li, F. Pang, C. Liang, X. Chen, and Y. Liu, "Improved interactive multiple model filter for maneuvering target tracking," in *33rd Chinese Control Conference*, 2014.