Abstract—This paper extends the recently developed Model-Aided Visual-Inertial Fusion (MA-VIF) technique for quadrotor Micro Air Vehicles (MAV) to deal with wind disturbances. The wind effects are explicitly modelled in the quadrotor dynamic equations excluding the unobservable wind velocity component. This is achieved by a nonlinear observability of the dynamic system with wind effects. We show that using the developed model, the vehicle pose and two components of the wind velocity vector can be simultaneously estimated with a monocular camera and an inertial measurement unit. We also show that the MA-VIF is reasonably tolerant to wind disturbances, even without explicit modelling of wind effects and explain the reasons for this behaviour. Experimental results using a Vicon motion capture system are presented to demonstrate the effectiveness of the proposed method and validate our claims.

I. INTRODUCTION

Quadrotor Micro Air Vehicles (MAV) have attracted a considerable interest in the recent past due to their versatility and simple construction. Employing autonomous quadrotor MAVs to perform routine or hazardous tasks in urban and indoor environments is already a much discussed and researched topic. One of the key obstacles that still needs to be overcome to achieve such an objective is the problem of pose estimation. Though established solutions such as Inertial Navigation Systems (INS) combining Inertial Measurement Units (IMU) and GPS exist for outdoor operation, it is well known that these are not suitable for urban or indoor environments where the quality of GPS measurements degrade substantially. Other sensor modalities can be employed to replace GPS in an INS and this has been the focus of a recent string research usually categorized as “sensing in GPS-denied environments”. Examples include laser range finders [1], depth cameras [2] and RGB cameras [3]. The third option, namely a sensor suite consisting of an IMU and a monocular camera, has emerged as a promising alternative to INS in the MAV state estimation problem. Known as Visual-Inertial Fusion (VIF), the validity and advantageous of this approach have already been demonstrated by other researchers, both in general [4] and in specific application to quadrotor MAVs [5].

This paper analyses and extends a novel VIF formulation for quadrotor MAVs, recently introduced in [6]. Termed “Model-Aided Visual-Inertial Fusion” (MA-VIF), this approach employs quadrotor MAV dynamic equations in the state estimator design resulting in improved state estimation accuracy and consistency. The dynamic model used in [6] is only valid when the quadrotor is operating in an environment with no wind. While this is generally the case for spacious indoor environments, for most other indoor and outdoor urban environments, wind disturbances are significant and must be taken into account. As our main contribution, we show that the MA-VIF can be extended to incorporate the effect of wind on the quadrotor motion while simultaneously estimating those components of wind that affect the quadrotor MAV. We employ a rigorous theoretical analysis as well as real world experiments to illustrate the validity of this claim.

As a secondary contribution we present the reason why the MA-VIF approach as it is in [6] has the capability of coping with reasonable amount of wind disturbances, albeit with reduced estimation accuracy and over-confident estimates. Through a careful analysis of the process and measurement equations and experimental results, we show that when wind disturbances are not explicitly modelled, the MA-VIF tends to adjust the accelerometer biases in an attempt to reconcile the sensor measurements with its incomplete description of the vehicle dynamics.

II. RELATED WORK

Visual-inertial fusion, especially with a monocular camera has been a much researched topic in the recent past. A tutorial level introduction to visual-inertial integration along with a summary of applications can be found in [7]. Conventionally VIF algorithms are platform agnostic because they do not make use of platform specific dynamics equations in the filter design. VIF formulations presented in [4] and [5] are typical examples of such designs. With such a formulation, non-zero accelerations of the camera-IMU rig are a necessary condition for full observability [8], [4].

Martin and Salaun [9] were the first to suggest the possibility of employing the quadrotor dynamic model for improved state estimation. Since then several researchers have designed and evaluated IMU based, model-aided state estimators for quadrotors, including [10] and [11]. MA-VIF design presented in [6] extended this by incorporating a monocular camera into the state estimator. They showed
that, in contrast to the conventional VIF algorithms, this formulation is fully observable even when the quadrotor is travelling at a constant velocity. However, they made the assumption that the quadrotor is operating in an enclosed environment with no wind. Here we analyse the case when this assumption is violated.

Estimating the direction and speed of wind can be of critical importance to MAVs operating in out-door environments. If an estimate of the platforms ground velocity and velocity with respect to the air is available, then calculating wind velocity is trivial. Langelaan et al. [12] made use of air speed measurements from a pitot tube and GPS based ground velocity to directly calculate the wind velocity for a fixed wing aircraft. Using simulations, they showed that main source of error in estimating wind velocity is the air speed sensor. Zachariah et al. [13] presented a wind speed estimator suitable for a generic MAV by making use of measurements from three orthogonal anemometers and a monocular camera. There again, only simulation results were used for validation. Moreover, anemometers are hardly suitable for quadrotor MAVs with limited payload capabilities and low flight speeds. Waslander and Wang [14] overcame this issue by employing the quadrotor dynamic model to predict the MAV speed with respect to air. Using simulations they showed that it is possible to estimate wind velocity by only using a GPS and an IMU. Neumann et al. [15] followed a similar approach but made use of an experimentally derived relationship between the quadrotor inclination angle and the air speed instead of a dynamic model. With the use of experiments performed in a wind tunnel, they showed that this method can produce reasonably accurate wind speed and direction estimates. The last two approaches are close in spirit to the work presented here. However, to the best of our knowledge, this is the first time the observability of such an approach has been analysed and the conditions for observability identified.

III. VISUAL-INERTIAL FUSION: METHODOLOGY

A. System Description

The platform under consideration is a quadrotor MAV affixed with an IMU (consisting of a triad of orthogonal accelerometers and gyroscopes) and a monocular camera. The body coordinate frame \( \{ B \} \) is assumed to be at the center of mass of the quadrotor MAV with its \( b_x - b_y \) plane aligned with the propeller plane. Without loss of generality, we assume that the camera is placed at the origin of \( \{ B \} \) aligned with the same. We assume that the accelerometers are placed at the center of mass of the quadrotor, as is the case with most commercially available platforms. We also define an earth fixed inertial coordinate frame \( \{ E \} \), which is defined by the position and heading of \( \{ B \} \) at the start of the VSLAM estimator but with its vertical axis aligned with gravity (see Fig. 1). Throughout the paper we use boldface letters to denote vectors and leading superscripts to denote coordinate frame in which the vector is expressed. A trailing subscript denotes individual components of the vector.

The states we wish to estimate are: the position of the origin of \( \{ B \} \) expressed in \( \{ E \} \), \( ^e p \); velocity of the origin of \( \{ B \} \) measured in \( \{ E \} \), \( ^b v \); scale of the VSLAM estimates \( \lambda \); orientation of \( \{ B \} \) with respect to \( \{ E \} \) expressed in Z-Y-X Euler angles, \( \Theta = [\phi, \theta, \psi] \), the body rotational rate \( \Omega \), accelerometer bias \( \beta_a \) and gyroscope bias \( \beta_g \). The translational and rotational dynamics of the quadrotor MAV are governed by:

\[
^e \dot{v} = \mathbf{g} + \frac{1}{m} \mathbf{F}_A \\
^b \dot{\Omega} = J^{-1} ( ^b M_A - ^b \Omega \times ^b J^b \Omega )
\]

where \( \mathbf{g} \) is the gravity vector, \( m \) is the mass of the quadrotor, \( \mathbf{F}_A \), \( M_A \) are respectively the sum of aerodynamic forces and moments acting on the quadrotor and \( J \) is the moment of inertia matrix for the vehicle.

As illustrated in Fig. 2 there are two separate, but coupled, estimators in the system that we are about to describe. First is the monocular SLAM estimator, termed VSLAM, similar to that detailed in [16]. We assume that, given the recent advances in VSLAM algorithms, it is reasonable to consider the VSLAM block as a sensor, which produces unbiased position (up-to-scale) and orientation measurements with an error covariance. This assumption allows us to model the VSLAM estimates as:

\[
\begin{align*}
\mathbf{h}_{vp} &= \lambda \mathbf{p} + \eta_p \\
\mathbf{h}_{vo} &= \Theta + \eta_o
\end{align*}
\]

where \( \eta_p, \eta_o \) are White Gaussian Noise (WGN) terms and \( E[\eta_p \eta_o^{\top}] = P_{rr} \) is the VSLAM pose covariance. \( (E[\cdot]) \) denotes the expectation operator.

The second estimator which fuses VSLAM estimates and inertial measurements, termed VIF (denoting Vision-Inertial Fusion), is the focus of this paper. VIF makes use
of accelerometer and gyroscope measurements which are assumed to be corrupted by a fixed, but unknown bias and a zero mean WGN term. These measurements can be modelled as:

\[
\begin{align*}
\begin{bmatrix}
\dot{b}a \\
\dot{b}v \\
\dot{b}\omega_g \\
\dot{b}\Omega
\end{bmatrix} &= \begin{bmatrix}
\begin{bmatrix}
\dot{b}a \\
\dot{b}v - R^Tg + \beta_a + \eta_a \\
\dot{b}\omega_g - R^Tg + \beta_g + \eta_\theta
\end{bmatrix} \\
\begin{bmatrix}
\eta_a \\
\eta_\theta
\end{bmatrix}
\end{bmatrix}
\end{align*}
\] (4)

where \( R \) is the rotation matrix that transforms a vector from \( \{B\} \) to \( \{E\} \), \( b\dot{a} = \{b\dot{a}_x, b\dot{a}_y, b\dot{a}_z\} \) and \( b\dot{\omega}_g = \{b\dot{\omega}_{gx}, b\dot{\omega}_{gy}, b\dot{\omega}_{gz}\} \) are, respectively, the accelerometer and gyroscope measurements in \( \{B\} \) and \( \eta_a = \{\eta_{ax}, \eta_{ay}, \eta_{az}\} \) and \( \eta_\theta = \{\eta_{\theta x}, \eta_{\theta y}, \eta_{\theta z}\} \) are zero mean WGN terms.

For most robotic platforms, the relationship between the actual inputs and the resulting forces and moments is a complex function involving numerous physical constants that are, at the least, difficult to measure. Moreover, for most quadrotor MAVs, the inputs - in this case propeller rotational rates - themselves are not available as measurements. Conventionally, roboticists have circumvented the need for explicit dynamic models by using inertial sensor measurements to replace the dynamics in (1) by rigid body kinematics. A state estimator designed under this conventional paradigm would, thus, feature a process model similar to equation 5 below.

\[
\begin{align*}
\begin{bmatrix}
\dot{\eta}_p \\
\dot{\eta}_v \\
\dot{\theta}_a \\
\dot{\theta}_g
\end{bmatrix} &= \begin{bmatrix}
\begin{bmatrix}
\dot{\eta}_p \\
\dot{\eta}_v - R^Tg + R\dot{\beta}_a + g - \eta_a \\
\dot{\theta}_a \\
\dot{\theta}_g - \beta_g - \eta_\theta
\end{bmatrix} \\
\begin{bmatrix}
\eta_\lambda \\
\eta_\beta_a \\
\eta_\beta_g
\end{bmatrix}
\end{bmatrix}
\end{align*}
\] (5)

where the state \( \eta_p \) was included to account for the vision measurements and \( \eta_\lambda, \eta_\beta_a \) and \( \eta_\beta_g \) are WGN terms and \( \Xi \) is a matrix that converts body rotational rates to Euler rates.

However, if a simple mathematical model that can satisfactorily explain some of the vehicle dynamics is available, then it is possible to revisit this conventional estimator design with the hope of improving the estimation accuracy. For example, Koifman and Bar-Itzhack [17] illustrated how the aircraft dynamics of a fixed wing MAV can be used to aid a low grade INS. The following section details how a similar behaviour can be accomplished with a quadrotor MAV.

IV. MODEL-AIDED VISUAL-INERTIAL FUSION

A MA-VIF formulation attempts explicitly model some of the dynamic equations in (1). In the case of a quadrotor MAV, this is facilitated by the fact that a simple relationship can be derived to satisfactorily explain translational motion in the \( b_x \) and \( b_y \) directions, as detailed below.

A quadrotor MAV in-flight is affected by three forces: gravity, thrust and aerodynamic drag. Rotor drag caused by blade flapping is the only significant drag force at low translational velocities [9]. The amount of blade flapping and in-turn rotor drag is proportional to the MAV translational velocity with respect to the wind, in the propeller plane. Thus

\[
F_A = \begin{bmatrix}
F_T \beta_3 - \alpha \sum_{i=1}^{4} \omega_i \tilde{v}_\infty 
\end{bmatrix}
\] (6)

where \( F_T \) is the magnitude of the summation of propeller thrust, \( \omega_i \) is the rotational velocity of \( i^{\text{th}} \) propeller, \( \alpha \) is a positive coefficient known as rotor drag coefficient, \( \beta_3 \) is the unit vector along \( b_z \) and \( \tilde{v}_\infty \) is the projection of \( \epsilon \tilde{v}_\infty \) on the propeller plane. The free stream velocity, \( \epsilon \tilde{v}_\infty \) is defined as:

\[
\epsilon \tilde{v}_\infty = \epsilon \tilde{v} - \epsilon \tilde{v}_w
\]

where \( \epsilon \tilde{v}_w \) is the velocity of the wind in \( \{E\} \). To derive the process equations for the MA-VIF we first combining 6 with the equation for \( \epsilon \dot{\tilde{v}} \) in (5) and transform the result to \( \{B\} \) frame:

\[
b\dot{v} \approx R^Tg - \begin{bmatrix}
k_1 b\tilde{v}_\infty \\
k_1 \beta_3 \tilde{v}_w
\end{bmatrix} + \begin{bmatrix}
k_1^{\text{FERT}} \\
k_1^{\text{AD}}
\end{bmatrix} T
\] (7)

where the second order terms resulting from the coordinate transformation are neglected and \( k_1 = \frac{\epsilon}{m} \sum_{i=1}^{4} \omega_i \). The validity of this equation for situations where there is no wind has already been analysed by several authors including [10]. There, experimental results show that \( k_1 \) is approximately a constant, the value of which for a specific quadrotor MAV can be approximated via equation (7) when ground truth state estimates are available. We make the same assumption throughout the rest of this paper. Also, by combining equations (7) and (4), we note that it is possible to replace \( F_T^m \) by \( b\dot{a}_z - \beta_3 \), similar to the conventional estimator design.

At this point it is possible to incorporate (7) into the state estimator process model (5) given that a dynamic model for wind velocity is available. However, given the complexities involved in developing and adopting such a model, here we make the assumption that the wind velocity is smoothly varying in \( \{E\} \). Though we employ this assumption for both the estimator design and analysis, it will be later shown through experiments that this proves to be adequate for coping with real wind patterns.

With the above assumptions, it is now possible to complete the process equation for the MA-VIF:
Note that we have used the gyroscope and \( b_z \) accelerometer measurements as system inputs. Also:

\[
\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},
\]

and \( \eta_v, \eta_w \) are zero mean WGN terms denoting uncertainties in process equations due to the assumptions made previously.

It is also possible to augment the accelerometer measurement equation by combining (4) with (7). This leads to:

\[
h_i = \begin{bmatrix} ba_x \\ ba_y \end{bmatrix} = -k_1 \Upsilon (b v - R^T e v_w) + \Upsilon \beta_a + \Upsilon \eta_a \tag{9}
\]

where:

\[
\Upsilon = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\]

Equations (8), (3) and (9) make up the process and measurement models for the MA-VIF.

V. OBSERVABILITY OF VISUAL-INERTIAL FUSION

Since both process and measurement equations of the MA-VIF are nonlinear, we resort to the ideas presented by Hermann and Krener [18] to analyse the observability of the MA-VIF formulation. They proposed a rank condition test for “locally weak observability” of a non-linear system: a system is locally weakly observable if it satisfies the observability condition generically [18]. A system described by a process equation of the form \( X = f_0(X) + \Sigma f_i(u_i), i = 1 \ldots l \), and a measurement equation of the form \( y = h(X) \) satisfies the observability rank condition if any one of the possible matrices whose rows are of the form

\[
\mathcal{O} = \{ \nabla L_{i_1 \ldots i_l} h_k(x) | i, j = 0, \ldots, l; k = 1, \ldots m; n \in \mathbb{N} \}
\]

is of full column rank. Here \( m \) is the number of measurements. Also, \( \nabla L_{i_1 \ldots i_l} h_k(x) \) is the gradient of the \( n \)th order Lie derivatives of \( h_k(x) \) with respect to \( f_i \) and can be calculated iteratively as:

\[
\nabla L_{i_1 \ldots i_l} h_k = \frac{\partial h_k}{\partial x} \nabla L_{i_1} h_{k-1} \frac{\partial f_i}{\partial x}
\]

\[
\nabla L_{i_1 \ldots i_l} h_k = \nabla L_{i_1} h_{k-1} \frac{\partial f_i}{\partial x} + \left[ \frac{\partial}{\partial x} \left( \nabla L_{i_1 \ldots i_l} h_{k-1} \right) \right] T \tag{10}
\]

where we have omitted function arguments for brevity.

Locally weak observability is a necessary condition but not sufficient for strict observability. However, following the approach of [4] and many others, we assume that for most mobile robotic navigational tasks, locally weak observability implies that the considered system contains enough information to perform state estimation. Experimental results presented in sections VII serve to validate the claims made based on this assumption.

If the observability matrix \( \mathcal{O} \) is rank deficient, then one or more of the states are not observable. Martinelli [19] introduced a method for identifying any unobservable modes using the concept of continuous symmetry.

**PROPERTY 1:** \( g(X) \) is an observable mode if and only if its gradient is orthogonal to the null space of \( \mathcal{O} \). This property can be expressed by a system of partial differential equations, one for each symmetry:

\[
\sum_{i=1}^{n} w_{si}(X) \frac{\partial g}{\partial X_i} = 0 \tag{10}
\]

where \( w_{si}(X) \) is the \( i \)th component of the symmetry \( w_s \).

A. MA-VIF Estimator

Observability of the MA-VIF formulation with no wind has been analysed previously in [6]. Here we proceed to extend that analysis to the case with non-zero wind. The system under consideration is fully characterised by (8), (3) and (9), with zero noise terms as they do not affect the observability. Note that \( \dim(X) = 19 \).

By considering the non-trivial Lie derivatives of the system up to second order, we can construct the following observability matrix for the MA-VIF:

\[
\mathcal{O}_1 = \begin{bmatrix} \nabla L^0 h_{v_p}(x) & \nabla L^1 h_{v_p}(x) & \nabla L^0 h_{v_i}(x) & \nabla L^1 h_{v_i}(x) & \ldots & \nabla L^1 h_{v_p}(x) & \nabla L^2 h_{v_p}(x) \end{bmatrix}^T
\]

\[
\begin{bmatrix}
\lambda I_3 & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\
0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 1} & I_3 & 0_{3 \times 3} & 0_{3 \times 3} \\
0_{2 \times 3} & -k_1 \Upsilon & 0_{2 \times 1} & D_1 & \Upsilon & -k_1 \Upsilon R^T \\
0_{4 \times 3} & \lambda R & R \Upsilon & D_2 & 0_{3 \times 3} & 0_{3 \times 3} \\
0_{4 \times 3} & 0_{3 \times 3} & 0_{3 \times 1} & 0_{3 \times 3} & 0_{3 \times 3} & \Xi \lambda R^T \\
0_{2 \times 3} & k_1^2 \Upsilon & 0_{2 \times 1} & D_3 & 0_{3 \times 2} & 0_{3 \times 3} & k_1^2 \Upsilon R^T \\
0_{4 \times 3} & -k_1 \Upsilon \Lambda & D_4 & D_5 & \lambda R^T & 0_{3 \times 3} & \lambda k_1 \Upsilon \Lambda R^T
\end{bmatrix}
\]

where \( I_3 \) is the \( 3 \times 3 \) identity matrix, \( 0_{m \times n} \) is an all zero matrix of size \( m \times n \), \( D_i (i = 1 \ldots 5) \) are functions of \( \lambda, \Theta, b v, \beta_a, e v_w \), with:

\[
D_4 = g - k_1 \Upsilon R^T + k_1 \Upsilon \Lambda R^T e v_w + \Upsilon R \beta_a
\]

It can be shown that all other Lie derivatives of the considered system are either linearly dependent on the rows of \( \mathcal{O}_1 \) or are trivially zero.

**Lemma 1:** The matrix \( \mathcal{O}_1 \) is rank deficient. It has a column rank of 18 when at least one of \( b v_z \) and \( b v_y \) is non-zero.

**Proof:** We proceed using block Gaussian elimination to recover the rank of \( \mathcal{O}_1 \). The notation \( \mathcal{O}_1(p, q) \) is used to represent the \( (p, q) \)th block, and \( \mathcal{O}_1(p, :) \) represent the \( p \)th row block of \( \mathcal{O}_1 \).

First, given that \( \lambda > 0 \), we divide row block one by \( \lambda \). Also since block \( \mathcal{O}_1(2, 4) \) is an identity matrix, we eliminate all non-zero entries in \( \mathcal{O}_1(:, 4) \). Rotational transformation matrix \( \Upsilon \) is orthogonal and thus is always full rank. This allows us to multiply \( \mathcal{O}_1(4, :) \) by \( (\Upsilon R)^{-1} \), \( \Xi \) transforms the body rotational rates to Euler rotational rates and is also full rank. Thus we can multiply \( \mathcal{O}_1(5, :) \) by \( \Xi^{-1} \). Next, we multiply \( \mathcal{O}_1(7, :) \) by \( \Xi^{-1} \) and add the result to \( \mathcal{O}_1(6, :) \)
to obtain:
\[
\begin{bmatrix}
 I_3 & 0_{3\times3} & \epsilon p & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{2\times3} & -k_1 \Upsilon & 0_{2\times3} & 0_{2\times3} & \Upsilon & 0_{3\times3} & -k_1 \Upsilon R^T \\
 0_{3\times3} & I_3 & k_1 \Upsilon & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{2\times3} & 0_{2\times3} & D_7 & 0_{2\times3} & 0_{2\times3} & 0_{3\times3} & 0_{2\times3} \\
 0_{3\times3} & -\lambda k_1 R A & D_4 & 0_{3\times3} & \lambda R \Gamma & 0_{3\times3} & \lambda k_1 R A R^T \\
\end{bmatrix}
\]
where
\[
D_7 = \frac{k_1}{\lambda} \Upsilon R^T (g - k_1 R A \v b + k_1 R A R^T \epsilon v_w + R \Gamma \beta_a) = \frac{k_1}{\lambda} \Upsilon (R^T g - k_1 (b \v - R^T \epsilon v_w)) = \frac{k_1}{\lambda} \Upsilon b \v = \begin{bmatrix} b \v_x & b \v_y \end{bmatrix}^T
\]
If either one of \( b \v_x \) and \( b \v_y \) is non-zero, then at least one element of the \( 2 \times 1 \) vector \( D_7 \) is non-zero and we can proceed to eliminate all remaining non-zero elements of \( \mathcal{O}_1(4,:) \). This results in \( \mathcal{O}_1(4,:) \) being all zero except in column block two. Thus we can then proceed to eliminate all remaining non-zero elements in \( \mathcal{O}_1(2,:) \) to obtain:
\[
\begin{bmatrix}
 I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{2\times3} & 0_{2\times3} & 0_{2\times3} & 0_{2\times3} & \Upsilon & 0_{3\times3} & -k_1 \Upsilon R^T \\
 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & I_3 & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & \lambda R \Gamma & 0_{3\times3} & \lambda k_1 R A R^T \\
\end{bmatrix}
\]
and finally moving the last row of \( \mathcal{O}_2(2,:) \) to \( \mathcal{O}_2(1,:) \) to immediately after \( \mathcal{O}_2(1,:) \) we obtain:
\[
\begin{bmatrix}
 I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\
 0_{3\times3} & 0_{3\times3} & k_1 \Upsilon R^T \\
\end{bmatrix}
\]
which has a column rank of 5 for any given \( R \). Thus returning to \( \mathcal{O}_1 \) we see that its column rank is 18.

B. Unobservaable Modes

The above analysis proves that the dimension of the null space of \( \mathcal{O}_1 \) is 1. It is easy to show that
\[
n_1 = \begin{bmatrix}
 0_{3\times1}; & 0_{3\times1}; & 0; & 0_{3\times1}; & 0_{3\times1}; & 0_{3\times1}; & R \\
 0 & 0 & 1 & \end{bmatrix}
\]
is a basis for the said null space.

It is possible to show that \( g(X) = b \v_{wz} = \Gamma_3 b \v_w = \Gamma_3 R^T \epsilon v_w \) is not orthogonal to \( n_1 \), where \( \Gamma_3 \) is the third column of \( \Gamma \). Thus, making use of property 1, we can conclude that the unobservable dimension corresponding to \( n_1 \) is the \( b_z \) component of wind. We can also conclude that all other states including the position, velocity and orientation of the quadrotor remain observable, despite wind disturbances.

VI. Estimator Design

Considering the non-linearities in the process and measurement equations, we propose an Extended Kalman Filter (EKF) based estimator for the MA-VIF. To proceed with the estimator design, we need to remove the unobservable mode, namely the \( b_z \) component of wind velocity from the state vector. This is not straightforward however, as the assumption about slow varying wind velocity in \( \{E\} \) meant that wind velocity had to be expressed in \( \{E\} \) in the state vector.

We begin by analysing the process noise matrix \( Q \) required for the EKF design. Under the assumption that all noise terms are uncorrelated with each other, \( Q \) will be a diagonal matrix. Considering the sub-matrix \( Q_{w} \) of \( Q \) that corresponds to wind velocity states:

\[
e \dot{v}_w = \eta_w \\
Q_w = E[\eta_w \eta_w^T] = \sigma_w^2 I_3
\]

where, \( \eta_w \in \mathcal{N}(0_{3\times1}, \sigma_w^2 I_3) \). Analysing the same covariance sub-matrix when wind velocity is expressed in \( \{B\} \):

\[
\dot{b} \v_w = R^T \epsilon \dot{v}_w = R^T \eta_w \\
\dot{\tilde{Q}}_w = \frac{\partial b \v_w}{\partial \eta_w} Q_w (\frac{\partial b \v_w}{\partial \eta_w})^T = \sigma_w^2 R^T I_3 R = \sigma_w^2 I_3.
\]

This shows that, for the purpose of the estimator design, the assumption about smoothly varying wind velocity can be made either in \( \{E\} \) or \( \{B\} \) with equivalent results. Having expressed wind velocity in \( \{B\} \), it is now straightforward

![Vicon motion capture arena](Image 313x646 to 434x737)

Fig. 3: (a) AR Drone II attached with four reflective markers and body coordinate frame, (b) Top view of the motion capture environment with dashed arrows indicating approximate direction of wind.
to remove the unobservable component $b\dot{v}_{wz}$ from the state vector. The resulting process and measurement equations are then:

$$
\begin{align*}
\begin{bmatrix}
\dot{p}
\
\dot{q}
\
\dot{\lambda}
\
\dot{\omega}
\end{bmatrix}
&= \\
R^T g - k_1 \Lambda(bv - b\dot{v}_{w}) - \Gamma \beta_a + \eta_v \\
&= -k_2 \Phi - \eta_\Phi
\end{align*}
\end{equation}

$$

(13)

$$

h_i = \begin{bmatrix} b_{ax} & b_{ay} \end{bmatrix} = -k_1 \Upsilon(bv - b\dot{v}_{w}) + \Upsilon \beta_a + \Upsilon \eta_a
$$

(14)

where $\dot{v}_{w} = [b\dot{v}_{wx} b\dot{v}_{wy}]^T$. VSLAM measurement equations remain the same. Note that we do not make the assumption that the $bz$ component of wind velocity is zero. The removal of the $bz$ component of wind velocity from the state vector is made possible by the fact that when posed in this formulation, the process equation for $b\dot{v}$ is independent of that component.

An observability analysis similar to that carried out in section V can be employed to show that the considered system is now locally-weakly observable, thus allowing us to proceed with EKF design in the standard manner as detailed in [20]. We omit the details for brevity.

Additionally, note that it is possible to rewrite (14) by combining $\beta_a$ and $b\dot{v}_w$ as they both have the same process dynamics.

$$
\begin{align*}
\begin{bmatrix}
b_{ax}
\end{bmatrix}
&= -k_1 \Upsilon(bv - b\dot{v}_{w}) + \Upsilon \beta_a + \Upsilon \eta_a
\end{align*}
\end{equation}

(15)

where $\beta_a = \beta_a + k_1 b\dot{v}_w$. This implies that it is not possible to distinguish between the effect of wind and bias by only analysing the accelerometer measurements. Wind velocity and accelerometer biases are only made observable as the same replacement cannot be done in the process dynamics for $b\dot{v}_{win}(13)$, where $\beta_{ax}$ and $\beta_{ay}$ do not appear. However, in practise this implies that when the wind effects are not explicitly modelled, as is the case in [6], it is possible for a state estimator to account for the different between true and predicted measurements by adjusting the accelerometer biases. Experimental results presented in the next section shows that this is indeed the case.

VII. Experiments

A. Experimental Setup

The quadrotor platform used for the experiments was the Parrot AR Drone II with a total flight weight of 450g (see Fig. 3a). The AR Drone II is equipped with two cameras facing forward and downward respectively, as well as a triad of accelerometers and gyroscopes. Accelerometer and gyroscope measurements captured at 200Hz, were timestamped on-board the MAV and wirelessly transmitted to a ground station computer. We collected several flight data sets by manually piloting the AR Drone II in an environment with approximate dimensions of $4 \times 4 \times 3m$. A Vicon motion capture system was used to produce real-time estimates for the vehicle states at 120Hz. An industrial grade fan situated at one side of the Vicon capture area was used to generate wind predominately along the x axis of {E} frame (see Fig. 3b). A hand held anemometer was used to obtain a rough measurement of the wind speed which at the center of the arena was about 2.5m/s.

The purpose of the experiments presented here is to verify the ability of the MA-VIF to produce accurate vehicle pose estimates amidst wind disturbances, given the measurements from a suitable VSLAM algorithm and an IMU. For this reason, we chose to simulate the VSLAM algorithm using the Vicon motion estimates. This was achieved by first subsampling the Vicon position and orientation estimates at 10Hz and then scaling and adding suitable levels of noise according to equation (3) with $\lambda = 1/5$, $\eta_\beta \in \mathcal{N}(0, (1/5)^2)$ (1 m standard deviation in metric scale) and $\eta_v \in \mathcal{N}(0, 4)$ (2 degrees standard deviation). The resulting position and orientation measurements are illustrated in Fig. 4.

B. Results

We performed two different experiments with each gathered data set. First the IMU and simulated VSLAM measurements were processed using a MA-VIF implementation designed for the case of no wind, originally presented in [6]. For ease of reference, we term this the MA-VIF-NW. Second, the same set of data were processed by the MA-VIF design presented here. For both experiments, position, velocity and scale estimates are presented in Fig. 5, 6 and 7 respectively.

First, we see that the MA-VIF is capable of quickly recovering the scale $\lambda$ thus producing accurate position and velocity estimates in metric scale. The scale estimate closely
matches the true value. It can also be seen that the scale estimate begins to deviate from the true value when the both $\hat{v}_x$, $\hat{v}_y$ are approximately zero. This agrees with the theoretical observability results derived in section V. The wind velocity estimates for the MA-VIF are presented in Fig. 8. Note how the wind is predominantly in the $x$ direction as is expected. Also note how when the quadrotor MAV moves away from the origin along $y$ axis, both the $x$ and $y$ components of wind estimates rapidly decrease (for example 290s - 300s and 320s - 340s). This indicates that the wind estimates are qualitatively accurate.

Focusing on the estimates of MA-VIF-NW, we see that the scale estimate takes longer to converge and as a result produce substantial errors in position estimates. Even in the long run, the MA-VIF scale estimate does not converge to the true value, creating inconsistencies in position estimates. Interestingly, we see that, even though the position and velocity estimates of MA-VIF-NW are less accurate than the MA-VIF, the reduction in accuracy is not substantial. However, upon a more thorough inspection, one can see that the MA-VIF-NW estimates are much less consistent than those of the MA-VIF. Velocity estimation errors of both estimators presented in Fig. 9 illustrates this fact clearly. This is as expected, since the MA-VIF-NW makes use of an incorrect model to describe the evolution of body frame velocity, thus resulting in over-confident and less accurate estimates.

The accelerometer bias estimates of both estimators are presented in Fig. 10. Note the similarity in MA-VIF-NW bias estimates for the $b_x$, $b_y$ axes and the wind velocity estimates for the same axes for the MA-VIF. In fact, as expected, it can be seen that the $b_x$, $b_y$ accelerometer biases of MA-VIF-NW agrees closely with the values predicted for $\beta_a$ by (15). We remark that this is also the reason why MA-VIF-NW estimates are not substantially erroneous, despite using an insufficient process model.

VIII. CONCLUSION

This paper extended the model-aided visual-inertial fusion for quadrotor micro air vehicles state estimation by considering the effects of wind disturbances. The translational dynamics of the quadrotor depend on the wind velocity and a state estimate aided by the same dynamics would needs to take the effect of wind into account for accurate state estimation. By explicitly modelling the said effects, we showed that it is possible to simultaneously estimate both the vehicle pose and two components of the wind velocity vector, only using a monocular camera and an inertial measurement unit. A nonlinear observability analysis and experimental results using a Vicon motion capture system were presented to validate these claims. Continuing further along these ideas, we wish to augment the MA-VIF with a propeller thrust model to improve the estimation accuracy in the vertical direction and also to estimate the full wind velocity vector.

Fig. 5: Position estimates with ground truth of MA-VIF (a) - x, (b) - y (c) - z and MA-VIF-NW (d) - (f).

Fig. 6: Velocity estimates with ground truth of MA-VIF (a) - (c) and MA-VIF-NW (d) - (f).
Fig. 7: Scale estimate of the MA-VIF (a) and MA-VIF-NW (b).

Fig. 8: Wind velocity estimates of the MA-VIF.

Fig. 9: Velocity estimation error with 3σ bounds of MA-VIF (a) - (c) and MA-VIF-NW (d) - (f).

Fig. 10: Accelerometer bias estimate of the MA-VIF (a) and MA-VIF-NW (b).

REFERENCES


