2013 IEEE International Conference on Robotics and Automation (ICRA)
Karlsruhe, Germany, May 6-10, 2013

Visual-Inertial Fusion for Quadrotor Micro Air Vehicles with Improved Scale Observability*

Dinuka Abeywardena, Zhan Wang, Sarath Kodagoda and Gamini Dissanayake

Abstract— **This paper presents a novel algorithm for fusing monocular vision and inertial information for quadrotor Micro Air Vehicles by incorporating the unique dynamic characteristics of that platform into the state estimation process. The dynamics of a quadrotor is unique in that a dual axis accelerometer mounted parallel to the propeller plane provides measurements that are directly proportional to vehicle velocities in that plane. By exploiting these dynamic characteristics, we show that all vehicle states, including the absolute scale, become observable in all motion patterns. This distinguishes our method with other visual-inertial fusion methods, which either assume zero accelerometer bias, or require "sufficiently exciting" motion, such as non-zero acceleration, to ensure observability of the scale. The advantages of our method over existing visual-inertial fusion algorithms are proved through a theoretical analysis using Lie Derivatives and verified by extensive simulations and experiments.**

I. INTRODUCTION

The quadrotor Micro Aerial Vehicle (MAV) is a simple robotic platform to construct. In its basic form, it is no more than two counter rotating propeller pairs attached symmetrically to a rigid cross-like frame, along with the means to control the speed of each propeller individually. This symmetric design has enabled the quadrotor to become a simple yet powerful vertical take-off and landing aerial platform that is popular among the robotics community.

MAVs are - by design - limited in their payload capacity and with that limitation, obtaining accurate and fast state estimates becomes a challenge. Inertial Measurement Units (IMU) have been extensively used for - fast but relatively less accurate in the long run - attitude estimation of MAVs [1]. Absolute positioning systems such as GPS are desirable but not available in many environments. Recent advances in sensing technology have resulted in a string of so-called "sensing in GPS-denied environments" in which three main sensing modalities are evaluated for their suitability as sensors in MAV state estimators. All three modalities, namely laser range finders, depth cameras and RGB cameras, have already been evaluated in the context of quadrotor state estimation (see for example [2], [3] and [4]).

None of the three modalities discussed above provide a perfect solution for the state estimation problem. In many cases, a combination of sensors have to be used to design a MAV state estimator that works in practice. This paper focuses on one promising possibility, namely the fusion of visual information from a monocular camera and inertial information from a triad of body mounted accelerometers for the purpose of state estimation of a quadrotor MAV. From the control theory perspective, the purpose of visual-inertial fusion is to derive both fast and accurate state estimates that would eventually drive a suitable controller for the quadrotor. Moreover, with such an integration, producing true metric scale translational estimates becomes a possibility, which in the case of pure vision based estimators is not possible.

The contribution of this paper is twofold. First, a novel algorithm for visual-inertial integration based on the dynamic model of the quadrotor MAV is presented. Second, with the help of an observability analysis, we compare the proposed algorithm to a "generic" form of visual-inertial fusion such as in [5] and show the advantages of our algorithm. Almost all visual-inertial fusion algorithms to date, including [5], [6], [7] and [8], fall into this generic category in the sense that they do not incorporate any platform specific equations into the fusion algorithms. We show, using both theoretical derivation and extensive simulations, that our design has superior properties compared to these generic forms. Specifically, it is shown that with the proposed algorithm, all states of interest including the absolute scale remain observable for all motion patterns of the MAV and therefore can be robustly estimated. This is in contrast to the restrictions imposed by the generic visual-inertial integrations such as [5], [6] and [7], where the sensor platform has to undergo "sufficiently exciting" motion to ensure observability of scale. As will be detailed in the following sections, the advantages in our method stem from the incorporation of the special dynamic properties of the quadrotor, as reported in our previous work [9], into the estimator design.

The remainder of this paper is organised as follows. Section II presents a brief overview of state-of-the-art visual-inertial fusion algorithms and the observability of such systems. In Section III, our visual-inertial integration algorithm is presented. For comparison, Section IV introduces a slightly extended version of the generic visual-inertial approach utilised in [5]. In Section V, an observability analysis of both systems are presented and the advantages of the novel design are highlighted. Simulation and experimental results are presented respectively in Section VI and VII to verify the merits and validity of our method.

*This work is supported by the Centre for Intelligent Mechatronic Systems (CIMS), University of Technology, Sydney.

Dinuka Abeywardena, Zhan Wang, Sarath Kodagoda and Gamini Dissanayake are with Centre for Intelligent Mechatronic Systems, Faculty of Engineering and IT, University of Technology, Sydney, Australia {dinuka.abeywardena, zhan.wang, sarath.kodagoda, gamini.dissanayake} at uts.edu.au

II. Related Work

An interesting application of visual-inertial fusion was reported in [5] where a quadrotor MAV equipped with an IMU and a downward facing camera was shown to exhibit robust estimator-controller structure stemming from visual-inertial integration. Details of a similar system which uses an additional air pressure sensor can be found in [8]. In both cases, the visual information was derived by processing the images from a monocular camera with a Simultaneous Localization and Mapping (SLAM) algorithm. The SLAM algorithm itself was considered as a black box that produces up-to-scale position and velocity estimates which are later integrated with measurements from an accelerometer and/or an air pressure sensor.

It is well known that when visual information is obtained from a visual SLAM (VSLAM) algorithm using a single camera, the scene structure and the translational motion of the camera can only be estimated up to an unknown scale. Resolving the scale ambiguity in bearing-only SLAM is one major motivation for visual-inertial integration. Observability analysis of bearing-only SLAM is quite common in literature. Analysis on the non-linear observability of bearing-only SLAM is reported in [10] and [11] by using differential geometric characterisation of observability and the Lie-derivative based observability matrices originally introduced in [12]. However, rarely has the observability been analysed theoretically for visual-inertial fusion systems. The work in [7] focuses mainly on the possibility of automatic inter-sensor calibration in a visual-inertial fusion setup and concludes that such a system is observable under two conditions. Firstly, the initial camera orientation has to be "locked down" or in other words, known perfectly. Secondly, the observability is satisfied only when the sensor platform undergoes non-zero accelerations. This approach is extended in [6] by deriving closed form solutions to attitude, speed, absolute scale and bias of a visual-inertial fusion setup. The first condition proposed in [7] is relaxed and as a result, the yaw angle becomes unobservable.

III. Estimator Design

A. System Description

As illustrated in Fig. 1 there are two separate - but linked - estimators in the system that we are about to describe. First is the monocular SLAM estimator termed VSLAM (See for example [13]). For the purpose of this paper, we make the observation that it is reasonable to consider VSLAM algorithm itself as a sensor, producing up-to-a-scale position, velocity measurements and absolute orientation measurements with an error covariance. The second estimator which fuses VSLAM estimates and inertial measurements, termed VIF (denoting Vision-Inertial Fusion), is the focus of this paper and will be discussed in detail below.

The system under consideration is a quadrotor MAV affixed with an IMU (consisting of a triad of orthogonal accelerometers) and a monocular camera. Without loss of generality, we assume that the IMU is located at the centre

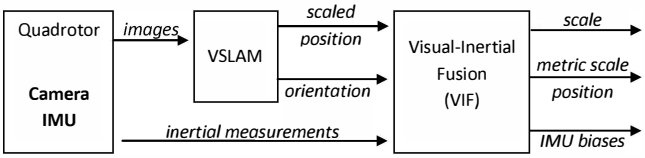


Fig. 1: Visual-inertial fusion for a quadrotor

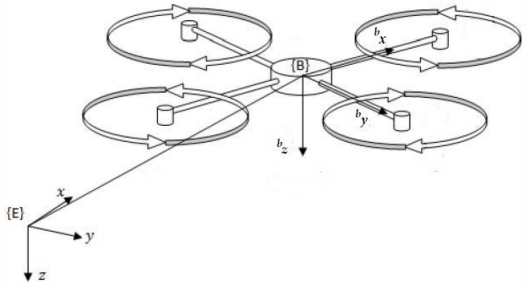


Fig. 2: Quadrotor and the coordinate systems

of gravity of the quadrotor and that the camera and IMU coordinate systems are aligned. We name this coordinate system the body coordinates $\{B\}$ and also define an earth fixed inertial coordinate system $\{E\}$, which is defined by the position and orientation of $\{B\}$ at the starting position of the VSLAM estimator (see Fig. 2). Throughout the paper we use boldface letters to denote vectors and a leading superscript to denote coordinate frame in which the vector is expressed. A trailing subscript denotes individual components of the vector.

B. Process Model

The state vector $\boldsymbol{X}$ consists of position of the origin of $\{B\}$ expressed in $\{E\}$ ${}^e\boldsymbol{p}$, velocity of the origin of $\{B\}$ measured in $\{E\}$, but expressed in $\{B\}$ ${}^b\boldsymbol{v}$, scale of the VSLAM estimates λ, orientation of $\{B\}$ with respect to $\{E\}$ expressed in Z-Y-X Euler angles $\boldsymbol{\Theta} = \{\phi, \theta, \psi\}$ and accelerometer bias $\boldsymbol{\beta_a}$. The evolution of states are defined by:

$$\left.\begin{aligned} {}^e\dot{\boldsymbol{p}} &= R\,{}^b\boldsymbol{v} \\ {}^b\dot{\boldsymbol{v}} &= R^T\boldsymbol{g} - \begin{bmatrix} k_1\,{}^bv_x + \eta_{vx} \\ k_1\,{}^bv_y + \eta_{vx} \\ (\beta_{az} - a_z) + \eta_{az} \end{bmatrix} \\ \dot{\lambda} &= \eta_\lambda \\ \dot{\boldsymbol{\Theta}} &= \boldsymbol{\eta}_\Theta \\ \dot{\boldsymbol{\beta}}_{\boldsymbol{a}} &= \boldsymbol{\eta}_{\beta a} \end{aligned}\right\} \quad (1)$$

where k_1 is a positive constant, $\boldsymbol{g}$ is gravity vector in $\{E\}$, R is the rotation matrix that transforms a vector from $\{B\}$ to $\{E\}$ and $\boldsymbol{a} = \{a_x, a_y, a_z\}$ are the accelerometer measurements. $\eta_{vx}, \eta_{vy}, \eta_{az}, \eta_\lambda, \boldsymbol{\eta}_{\beta a}$ and $\boldsymbol{\eta}_\Theta$ are zero mean White Gaussian Noise (WGN) terms denoting uncertainties

in process equations. The accelerometer measurement along ${}^b z$ axis is considered as a control input to the system. Second equation of (1) which describes the relationship between the orientation and the translational motion is unique to a quadrotor MAV. As will be evident later, the advantages of the novel VIF estimator presented in this paper stem from this unique relationship. More details on this equation along with experimental validation can be found in [9].

C. Measurement Model

Three types of measurements are used by the VIF, namely the position (up-to-scale), orientation estimates from the VSLAM algorithm and the measurements from the ${}^b x$ and ${}^b y$ accelerometers. Each measurement is assumed to be corrupted by zero mean WGN. The measurement equations are:

$$\boldsymbol{h}_{vp} = \lambda\, {}^e\boldsymbol{p} + \boldsymbol{\eta}_p \tag{2}$$

$$\boldsymbol{h}_{vo} = \boldsymbol{\Theta} + \boldsymbol{\eta_o} \tag{3}$$

$$\boldsymbol{h}_i = \begin{bmatrix} a_x \\ a_y \end{bmatrix} = \begin{bmatrix} -k_1\, {}^b v_x + \beta_{ax} + \eta_{ax} \\ -k_1\, {}^b v_y + \beta_{ay} + \eta_{ay} \end{bmatrix}. \tag{4}$$

Again equation (4), which describes what body mounted ${}^b x$ and ${}^b y$ accelerometers would measure, is unique to a quadrotor MAV and can be easily derived from equation (1).

D. Estimator Structure

An Extended Kalman Filter (EKF) is proposed to estimate $\boldsymbol{X}$ considering the non-linearities in both process and measurement equations. Structure of the proposed estimator follows the standard EKF format (see for example [14]) and we refrain from detailing it here to conserve space.

IV. Generic Visual-Inertial Fusion

For comparison purposes, we include here one of the generic methods of visual-inertial integration. In the following, two extensions are made to the method presented in [5]. First, their method only considers the component of motion along one axis of $\{E\}$, whereas we extend the approach to all three axes. Second, their algorithm does not include accelerometer bias as a state even though low cost accelerometers exhibit a considerable bias drift. Realising this, we resort to online bias estimation by including it as a state.

The states of interest for scale estimation are ${}^e\boldsymbol{p}$, ${}^e\boldsymbol{v}$, λ and β_a. The differential equations governing the states are:

$$\left.\begin{aligned} {}^e\dot{\boldsymbol{p}} &= {}^e\boldsymbol{v} \\ {}^e\dot{\boldsymbol{v}} &= {}^e\boldsymbol{a} - \boldsymbol{\beta}_a + \boldsymbol{\eta}_a \\ \dot{\lambda} &= \eta_\lambda \\ \dot{\boldsymbol{\beta}}_a &= \boldsymbol{\eta}_{\beta a} \end{aligned}\right\} \tag{5}$$

where ${}^e\boldsymbol{a}$ is the "gravity compensated" accelerometer measurement in inertial coordinate frame $\{E\}$, which can be derived through

$${}^e\boldsymbol{a} = R\boldsymbol{a} - \begin{bmatrix} 0 & 0 & g \end{bmatrix}$$

by constructing R from the VSLAM orientation estimates. In equation (5), ${}^e\boldsymbol{a}$ is considered as a control input. The only measurement is the up-to-a-scale position from the VSLAM algorithm

$$\boldsymbol{h}_v = \lambda\, {}^e\boldsymbol{p} + \boldsymbol{\eta}_p. \tag{6}$$

Similar to the previous section, an EKF is proposed for state estimation, details of which are omitted.

V. Observability of Visual-Inertial fusion algorithms

One of the main objectives of the estimators presented in Sections III and IV is to produce metric scale position and velocity estimates by fusing the information contained in both VSLAM estimates and inertial measurements. Given that no scale information is present in the bearing-only VSLAM estimates and the presence of biases in inertial sensors, the possibility of achieving this objective is not apparent. Therefore, observability of the states need to be analysed before the designed estimators are put to test.

Since both the considered systems are non-linear, we resort to the methods presented by Hermann and Krener [12] to analyse their observability. They proposed a rank condition test for what they termed "local weak observability" of a non-linear system. Specifically, it is stated that *a system is locally weakly observable if and only if it satisfies the observability rank condition* [12]. For a system described by a process equation of the form $f_0(\boldsymbol{X}) + \Sigma f_i(u_i)$ and measurement equation $h(\boldsymbol{X})$ to satisfy the observability rank condition, any possible observability matrices whose rows are of the form

$$\mathbf{O} = \{\nabla L^l_{f_i \ldots f_j} h_k(x) | i, j = 0, ..., l; k = 1, ...m; l \in \mathbb{N}\}$$

should be of full column rank. Here m is the number of measurements and $\nabla L^k_f h(\mathbf{x})$ is the gradient of the k^{th} order Lie derivatives of $h(\mathbf{x})$ with respect to f and can be calculated iteratively as:

$$dL^0_f h = \frac{\partial h}{\partial \mathbf{x}}$$

$$dL^k_{f_i} h = dL^{k-1}_{f_i} h \frac{\partial f_i}{\partial \mathbf{x}} + \left[\frac{\partial dL^{k-1}_{f_i} h^T}{\partial \mathbf{x}} f_i\right]^T$$

where we have omitted function arguments for brevity.

Obviously, in order to estimate the states, the locally weak observability of the considered system is necessary but not sufficient. However, following the approach of [15] and many others, this paper assumes that for most of mobile robotic navigational tasks, local weak observability implies that the considered system contains enough information to perform state estimation. On the other hand, simulation results presented in Section VI serve to validate the claims made in this paper based on this assumption.

To simplify the ensuing analysis, we restrict our attention to one dimensional motion of the quadrotor. We begin with aligned $\{B\}$ and $\{E\}$ coordinate frames and analyse the situation when only the roll angle ϕ is changed causing the quadrotor to move in the direction of ${}^e y$.

A. Novel Estimator

After simplifying equations (1) - (4) and removing the random noise terms (as they do not affect the observability) the following can be derived

$$\left.\begin{aligned} {}^{e}\dot{p}_y &= \cos(\phi)\,{}^{b}v_y \\ {}^{b}\dot{v}_y &= g\sin(\phi) - k_1\,{}^{b}v_y \\ \dot{\lambda} &= 0 \\ \dot{\phi} &= 0 \\ \dot{\beta_{ay}} &= 0 \end{aligned}\right\} \tag{7}$$

$$\left.\begin{aligned} h_1 &= \lambda\,{}^{e}p_x \\ h_2 &= \phi \\ h_3 &= a_x = -k_1\,{}^{b}v_x + \beta_{ay} \end{aligned}\right\}. \tag{8}$$

To employ the Lie-derivative based rank condition test, we need to express process equations in the input linear form. Since equation (7) is already in this form, we define a function of states $\mathbf{x} = \{{}^{e}p_y,\ {}^{b}v_y, \lambda, \phi, \beta_{ay}\}$ that makes this form explicit

$$f(\mathbf{x}) = \begin{bmatrix}\cos(\phi)\,{}^{b}v_y & -k\,{}^{b}v_y + g\sin(\phi) & 0 & 0 & 0\end{bmatrix}^T.$$

By considering the gradient of up to 2^{nd} order Lie derivatives, we have found the following to make up a full rank observability matrix

$$\begin{aligned}\mathbf{O_{s1}} &= \begin{bmatrix}\nabla L_f^0 h_1(x) & \nabla L_f^0 h_3(x) & \nabla L_f^1 h_1(x)\ldots \\ \nabla L_f^1 h_3(x) & \nabla L_f^2 h_1(x)\end{bmatrix}^T \\ &= \begin{bmatrix}\lambda & 0 & {}^{e}p_y & 0 & 0 \\ 0 & -k_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & \lambda\cos(\phi) & {}^{b}v_y\cos(\phi) & 0 & \lambda\,{}^{b}v_y\sin(\phi) \\ 0 & k_1^2 & 0 & 0 & gk_1\cos(\phi)\end{bmatrix}.\end{aligned}$$

It can be seen that $\boldsymbol{O_{s1}}$ is full rank when ${}^{b}v_y$ is non-zero, and thus we conclude that system is locally weakly observable when the quadrotor is in motion.

B. Generic Estimator

By simplifying the of process and measurement equations (5) - (6) and neglecting noise terms, we have

$$\begin{aligned} f(\mathbf{x},\mathbf{u}) &= f_0(\mathbf{x}) + f_1(\mathbf{u}) \qquad (9)\\ &= \begin{bmatrix}{}^{e}v_y & -\beta_{ay} & 0 & 0\end{bmatrix}^T + \begin{bmatrix}0 & {}^{e}a_y & 0 & 0\end{bmatrix}^T \\ h(\mathbf{x}) &= \lambda\,{}^{e}p_y \qquad (10)\end{aligned}$$

where $\boldsymbol{x} = \{{}^{e}p_y, {}^{e}v_y, \lambda, \beta_{ay}\}$ are the states and $u = \{a_y\}$ is the control input.

For the system described by equations (9) and (10), $\mathbf{O_{s2}}$ is the gradient of all non-trivial Lie derivatives

$$\begin{aligned}\mathbf{O_{s2}} &= \begin{bmatrix}\nabla L_f^0 h(x) & \nabla L_{f_\bullet}^1 h(x) & \nabla L_{f_\bullet}^2 h(x) & \nabla L_{f_1 f_\bullet}^2 h(x)\end{bmatrix}^T \\ &= \begin{bmatrix}\lambda & 0 & {}^{e}p_y & 0 \\ 0 & \lambda & {}^{e}v_y & 0 \\ 0 & 0 & -\beta_a & -\lambda \\ 0 & 0 & {}^{e}a_y & 0\end{bmatrix}.\end{aligned}$$

All higher order Lie derivatives not included in $\mathbf{O_{s2}}$ are identically zero. It can be seen that $\mathbf{O_{s2}}$ is of full rank when ${}^{e}a_y$, which is the accelerometer measurement, is non-zero. Thus we conclude that the system is locally weakly observable under this condition.

C. Discussion

Having proven that both considered systems are observable, this section is devoted to discuss some important properties arising from the observability analysis. We begin this discussion by pointing out the fact that the generic estimator is capable of observing the absolute scale of the VSLAM algorithm even in the presence of accelerometer bias. This finding indicates that it is possible to augment the visual-inertial fusion algorithms reported in [5] and [8] with online bias estimation to improve accuracy.

Secondly, we observe that the observability of the generic estimator depends on the quadrotor maintaining a non-zero acceleration. This is similar to the results reported in [6] and [7]. This imposes a rather restrictive condition on the flight pattern of the quadrotor and can not be satisfied during most manoeuvres. As a result, though the generic estimator is theoretically observable, its estimates will be poorly conditioned for most part of the flight. More importantly for the proposed method, it can be easily shown that $\boldsymbol{O_{s1}}$ remains full rank even when the quadrotor is traversing at a constant velocity. This means that the state estimates of this design will converge for a much wider range of flight envelops.

VI. Simulation Results

To analyse the performance of the novel estimator presented in Section III and also to verify the validity of the conclusions made in Section V for the general case of three dimensional motion, a variety of simulations were performed.

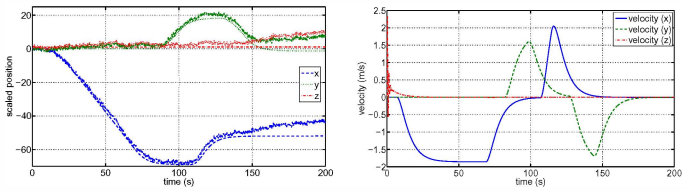
A. Simulation Environment

Dynamics of the quadrotor MAV were modelled according to the equations presented in [16]. Apart from the MAV dynamics presented therein, we also made use of four separate PID controllers - pitch, roll, yaw and thrust controllers - to control the quadrotor motion.

One of the inputs to both novel and generic VIF estimators is the up-to-scale position estimates from a suitable VSLAM algorithm. The implementation of a full fledged VSLAM algorithm would unnecessarily complicate the simulator. Thus we resorted to a simplification in which we scaled the true position and added suitable noise terms to simulate the SLAM position estimates. The scaled position estimates of the VSLAM algorithm were derived as:

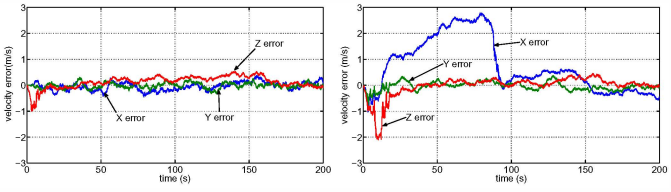
$$\begin{aligned}\dot{\tilde{\boldsymbol{p}}} &= {}^{e}\boldsymbol{v} + \boldsymbol{\eta_v} \\ \boldsymbol{h_{vp}} &= \lambda\tilde{\boldsymbol{p}} + \boldsymbol{\eta_p}\end{aligned}$$

where $\boldsymbol{\eta_v} \sim \mathcal{N}(0, 0.5)$ and $\boldsymbol{\eta_p} \sim \mathcal{N}(0, 1)$. The addition of a random walk type drift in SLAM position estimates was to better simulate real world SLAM estimates. Fig. 3 shows the true position and velocity of the quadrotor MAV in one



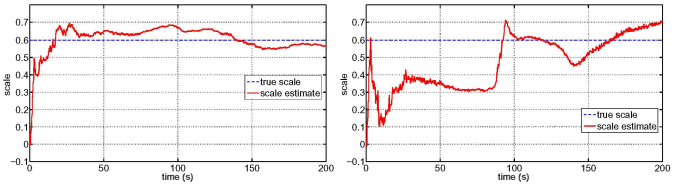
(a) (b)

Fig. 3: True states from simulation along with simulated VSLAM estimates. (a) True position with (noisy) VSLAM position estimates, (b) True velocity



(a) (b)

Fig. 4: Velocity estimation error (a) novel_VIF, (b) generic_VIF.



(a) (b)

Fig. 5: Absolutes scale estimations (a) novel_VIF, (b) generic_VIF

simulation run along with the simulated VSLAM estimates. Accelerometer measurements were assumed to be corrupted by an unknown constant bias and zero mean WGN with a standard deviation of $0.5ms^{-2}$. The VSLAM estimates were produced at a rate of 10Hz and inertial measurements at a rate of 200Hz.

B. Simulation Results

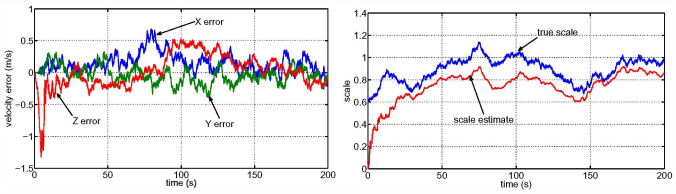
For the first set of simulations, the true value of λ was fixed at 0.6. To conserve space and also for clarity, we only present the velocity and scale estimates of the two estimators. All noise terms in the estimators were set to match the true noise levels of the simulator. For ease of notation, we use *novel_VIF* to denote the novel estimator presented in Section III and *generic_VIF* to represent the generic estimator presented in Section IV.

Fig. 4 presents the velocity estimation error of the novel_VIF and the generic_VIF. Fig. 5 compares the accuracy of the scale estimate of the novel_VIF and the generic_VIF. As expected, both estimators are capable of producing metric scale velocity estimates. However, the accuracy of the novel_VIF easily exceeds the other. Note the growth in generic_VIF estimation errors during the time period 20s - 70s as shown in Fig. 4b, where the velocity in all three dimensions are nearly constant. This increase in error can be accounted to the errors in scale estimate during the same period. As is evident from Fig. 5b the scale estimate of the generic_VIF shoots toward the true scale initially, due to the quadrotor undergoing vertical acceleration during takeoff. However, during the constant velocity period that follows after takeoff, the effects of scale unobservability can be seen clearly. In contrast, during the same period, the scale in novel_VIF remains observable and its estimates remain accurate. The scale estimate of the generic_VIF becomes sufficiently accurate, thus reducing the error in velocity estimates, only at around 90s when the quadrotor decelerates. These simulation results prove that the insights obtained from the theoretical observability analysis in Section V remain valid for the more complicated three dimensional motion.

To further analyse the performance of the novel_VIF, another set of simulations were performed with a varying true scale, where the true value of λ was produced by adding a random walk type noise to a constant value. The scale estimate and velocity estimation error of novel_VIF from this simulation are presented in Fig. 6. Fig. 6a shows that even with considerable variation in true scale, the velocity estimation error remains well bounded. Also, Fig. 6b clearly demonstrates that the scale of the novel_VIF remains observable during all phases of the flight. This points to a major advantage of the novel_VIF in real world scenarios, where the scale of VSLAM algorithms does tend to drift slowly with time.

VII. Experimental Results

Multiple experiments were performed with the use of a real world flight data set obtained by manoeuvring a quadrotor MAV in a vicon motion capture environment approximately of size $4 \times 4 \times 3(m)$. The quadrotor platform used was the Parrot AR Drone I with a total flight weight of 420g (see Fig. 7). The AR Drone was equipped with one front facing and one downward facing cameras, along with a triad of accelerometers and gyroscopes. The images from the front facing camera with a 93^0 wide-angle lens have a resolution of 640×480 pixels. These images, captured at approximately 10Hz along with accelerometer and

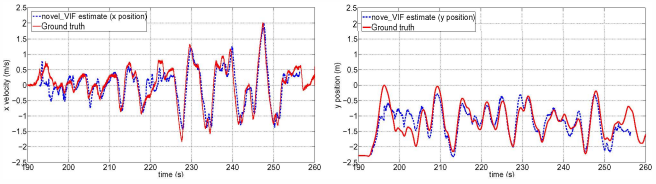


(a) (b)

Fig. 6: Velocity estimation error and scale estimate of the novel_VIF under varying true scale. (a) error in velocity estimates. (b) scale estimate

(a) (b)

Fig. 7: (a) AR Drone I, (b) Vicon motion capture environment, inset - one of the vicon IR cameras



(a) (b)

Fig. 8: Position estimates of the novel_VIF algorithm along with ground truth

gyroscope measurements (at 200Hz) were timestamped onboard the MAV and wirelessly transmitted to a ground station computer. The images were then processed using an open-source implementation [17] of EKF based monocular SLAM algorithm using inverse-depth parametrisation to derive the VSLAM pose estimates. That code was modified to accommodate SIFT feature detection and matching to overcome the difficulties faced by rather fast camera movements. All estimation tasks were performed off-board and off-line. Vicon real-time state estimates at 120Hz were stored in a different computer and were treated as ground truth in evaluating the VIF estimates.

Due to the small size of the vicon motion capture volume, it was not possible for the quadrotor MAV to maintain a constant velocity for a satisfactory time period. This prevented an experimental validation of the observability properties of the two VIF estimators. However, to illustrate the validity of the novel_VIF estimator, the x and y position estimates along with the ground truth are presented in Fig. 8. This figure clearly demonstrates that the novel_VIF is capable of producing metric scale translational estimates.

VIII. Conclusions

This paper presented a novel method to integrate visual and inertial measurements to derive metric scale estimates for a quadrotor MAV. The proposed method makes use of the unique dynamic characteristics of the quadrotor to the benefit of the estimation task, unlike generic visual-inertial integration methods. Through an observability analysis and simulations, it was shown that the novel method outperforms existing methods amidst various flight conditions. In particular, it was shown that the scale remains observable even when the "sufficiently exciting" motion condition is not fulfilled, resulting in more accurate estimates.

In concluding the paper, we emphasize that the design presented here is suitable only for quadrotor MAVs, as it depends on the dynamics of the quadrotor. However, it is possible to extend our method to exploit the dynamics of other robotic vehicles to achieve similar results, and this is the focus of our current research.

References

[1] D. Kingston and R. W. Beard, "Real-time attitude and position estimation for small uavs using low-cost sensors," *AIAA 3rd Unmanned Unlimited Technical Conference*, 2004.

[2] A. Bachrach, S. Prentice, R. He, and N. Roy, "Range - robust autonomous navigation in gps-denied environments," *Journal of Field Robotics*, 2011.

[3] A. Bachrach, S. Prentice, R. He, P. Henry, A. S. Huang, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Estimation, planning and mapping for autonomous flight using an rgb-d camera in gps-denied environments," *International Journal of Robotics Research*, to appear.

[4] M. Bryson and S. Sukkarieh, "Building a robust implementation of bearing-only inertial slam for a uav," in *Journal of Field Robotics, Special issue on SLAM in the field*, vol. 24, no. 1-2, 2007, pp. 113–143.

[5] G. Ntzi, S. Weiss, D. Scaramuzza, and R. Siegwart, "Fusion of imu and vision for absolute scale estimation in monocular slam," *Journal of Intelligent and Robotic Systems*, vol. 61, pp. 287–299, 2011.

[6] A. Martinelli, "Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *Robotics, IEEE Transactions on*, vol. 28, no. 1, pp. 44 –60, feb. 2012.

[7] J. Kelly and G. Sukhatme, "Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration," in *Proc. IEEE International Symposium on Computational Intelligence in Robotics and Automation*, dec. 2009, pp. 360 –368.

[8] M. Achtelik, M. Achtelik, S. Weiss, and R. Siegwart, "Onboard imu and monocular vision based control for mavs in unknown in- and outdoor environments," in *Proc. IEEE International Conference on Robotics and Automation*, may 2011, pp. 3056 –3063.

[9] D. Abeywardena, S. Kodagoda, R. Munasinghe, and G. Dissanayake, "A virtual odometer for a quadrotor micro aerial vehicle," *Australasian Conference on Robotics and Automation*, December 2011.

[10] K. W. Lee, W. Wijesoma, and I. Javier, "On the observability and observability analysis of slam," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, oct. 2006, pp. 3569 –3574.

[11] L. Perera and E. Nettleton, "On the nonlinear observability and the information form of the slam problem," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, oct. 2009, pp. 2061 –2068.

[12] R. Hermann and A. Krener, "Nonlinear controllability and observability," *Automatic Control, IEEE Transactions on*, vol. 22, no. 5, pp. 728 – 740, oct 1977.

[13] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. IEEE International Conference on Computer Vision*, 2003, pp. 1403 –1410.

[14] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using Matlab*. Wiley-Interscience, 2001.

[15] A. Martinelli, "State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration," *Robotics, IEEE Transactions on*, vol. 27, no. 2, pp. 239 –255, april 2011.

[16] P. Martin and E. Salaun, "The true role of accelerometer feedback in quadrotor control," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, may 2010, pp. 1623 –1629.

[17] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel, "1-point ransac for extended kalman filtering: Application to real-time structure from motion and visual odometry," *Journal of Field Robotics*, vol. 27, no. 5, pp. 609–631, 2010.