Simultaneous State Initialization and Gyroscope Bias Calibration in Visual Inertial aided Navigation

Jacques Kaiser[1], Agostino Martinelli[1], Flavio Fontana[2] and Davide Scaramuzza[2]

***Abstract*— State of the art approaches for visual-inertial sensor fusion use filter-based or optimization-based algorithms. Due to the nonlinearity of the system, a poor initialization can have a dramatic impact on the performance of these estimation methods. Recently, a closed-form solution providing such an initialization was derived in [?]. That solution determines the velocity (angular and linear) of a monocular camera in metric units by only using inertial measurements and image features acquired in a short time interval. In this paper, we study the impact of noisy sensors on the performance of this closed-form solution. We show that the gyroscope bias, not accounted for in [?], significantly affects the performance of the method. Therefore, we introduce a new method to automatically estimate this bias. Compared to the original method, the new approach now models the gyroscope bias and is robust to it. The performance of the proposed approach is successfully demonstrated on real data from a quadrotor MAV.**

I. INTRODUCTION

Autonomous mobile robots navigating in unknown environments have an intrinsic need to perform localization and mapping using only on-board sensors. Concerning Micro Aerial Vehicles (MAV), a critical issue is to limit the number of on-board sensors to reduce weight and power consumption. Therefore, a common setup is to combine a monocular camera with an inertial measurements unit (IMU). On top of being cheap, these sensors have very interesting complementarities. Additionally, they can operate in indoor environments, where Global Positioning System (GPS) signals are shadowed. An open question is how to optimally fuse the information provided by these sensors.

Currently, most sensor-fusion algorithms are either filter-based or iterative. That is, given a current state and measurements, they return an updated state. While working well in practice, these algorithms need to be provided with an initial state. The initialization of these methods is critical. Due to nonlinearities of the system, a poor initialization can result into converging towards local minima and providing faulty states with high confidence.

In this paper, we demonstrate the efficiency of a recent closed-form solution introduced in [**?**][**?**], which fuses visual and inertial data to obtain the structure of the environment at the global scale along with the attitude and the speed of the robot. By nature, a closed-form solution is deterministic and, thus, does not require any initialization.

The method introduced in [**?**][**?**] was only described in theory and demonstrated in simulations. In this paper, we implement this method in order to test it with real world data. This allow us to identify its limitations and bring modifications to overcome them. Specifically, we investigate the impact of biased inertial measurements. Although the case of biased accelerometer was originally studied in [**?**], here we show that a large bias on the accelerometer does not significantly worsen the performance. One major limitation of [**?**] is the impact of biased gyroscope measurements. In other words, the performance becomes very poor in presence of a bias on the gyroscope and, in practice, the overall method can only be successfully used with a very precise - and expensive - gyroscope. Here, we introduce a simple method that automatically estimates this bias. By adding this new method for the bias estimation to the original method [**?**], we obtain results that are equivalent to the ones in absence of bias. Compared to [**?**], the new method is now robust to the gyroscope bias and automatically calibrates the gyroscope.

II. RELATED WORK

The problem of fusing visual and inertial data has been extensively investigated in the past. However, most of the proposed methods require a state initialization. Because of the system nonlinearities, lack of precise initialization can irreparably damage the entire estimation process. In literature, this initialization is often guessed or assumed to be known [**?**][**?**][**?**][**?**][**?**]. Recently, this sensor fusion problem has been addressed by using

*This work was supported by the French National Research Agency ANR through the project VIMAD

[1] INRIA Rhone Alpes, Grenoble, France. Email: jacko.kaiser@gmail.com, agostino.martinelli@ieee.org

[2] Robotics and Perception Group, University of Zurich, Switzerland. Email: fforster,sdavideg@ifi.uzh.ch

optimization-based approaches [**?**][**?**][**?**][**?**][**?**][**?**]. These methods outperform filter-based algorithms in terms of accuracy due to their capability of relinearizing past states. On the other hand, the optimization process can be affected by the presence of local minima. We are therefore interested in a deterministic solution that analytically expresses the state in terms of the measurements provided by the sensors during a short time-interval.

In computer vision, several deterministic solutions have been introduced. These techniques, known as *Structure from Motion*, can recover the relative rotation and translation up to an unknown scale factor between two camera poses [**?**]. Such methods are currently used in state-of-the-art visual navigation methods for MAVs to initialize maps [**?**][**?**][**?**]. However, the knowledge of the absolute scale, and, at least, of the absolute roll and pitch angles, is essential for many applications ranging from autonomous navigation in GPS-denied environments to 3D reconstruction and augmented reality. For these applications, it is crucial to take the inertial measurements into consideration to compute these values deterministically.

A procedure to quickly re-initialize a MAV after a failure was presented in [**?**]. However, this method requires an altimeter to initialize the scale.

Recently, a closed-form solution has been introduced in [**?**]. From integrating inertial and visual measurements over a short time-interval, this solution provides the absolute scale, roll and pitch angles, initial velocity, and distance to 3D features. Specifically, all the physical quantities are obtained by simply inverting a linear system. The solution of the linear system can be refined with a quadratic equation assuming the knowledge of the gravity magnitude. This closed-form was improved in [**?**] to work with unknown camera-IMU calibration; however, since in this case the problem cannot be solved by simply inverting a linear system, a method to determine the six parameters that characterize the camera-IMU transformation was proposed. As a result, this method is independent of external camera-IMU calibration, hence, suitable for power-on-and-go systems.

A more intuitive expression of this closed-form solution was derived in [**?**]. While being mathematically sound, this closed-form solution is not robust to noisy sensor data. For this reason, to the best of our knowledge, it has never been used in an actual application. In this paper, we perform an analysis to find out its limitations. We start by reminding the reader the basic equations that characterize this solution (section **??**). In section **??**, we show that this solution is resilient to the accelerometer bias but strongly affected by the gyroscope bias. We then introduce a simple method that automatically estimates the gyroscope bias (section **??**). By adding this new method for the bias estimation to the original method, we obtain results that are equivalent to the ones obtained in absence of bias. Compared to the original method, the new method is now robust to the gyroscope bias and also calibrates the gyroscope. In section **??**, we validate our new method against real world data from a flying quadrotor MAV to prove its robustness against noisy sensors during actual navigation. Finally, we provide the conclusions in section **??**.

III. CLOSED-FORM SOLUTION

In this section, we provide the basic equations that characterize the closed-form solution proposed in [**?**]. We also provide the main features of this solution[1].

Let us refer to a short interval of time (e.g., of the order of 3 seconds). We assume that during this interval of time the camera observes simultaneously N point-features and we denote by $t_1,\ t_2, \cdots, t_{n_i}$ the times of this interval at which the camera provides an image of these points. Without loss of generality, we can assume that $t_1 = 0$. The following equation holds (see [**?**] for its derivation):

$$S_j = \lambda_1^i \mu_1^i - V t_j - G\frac{t_j^2}{2} - \lambda_j^i \mu_j^i \tag{1}$$

with:

- μ_j^i the normalized bearing of point feature i at time t_j in the local frame at time t_1;
- λ_j^i the distance to the point feature i at time t_j;
- V the velocity in the local frame at time t_1;
- G the gravity in the local frame at time t_1;
- S_j the integration in the interval $[t_1,\ t_j]$ of the rotated linear acceleration data (i.e., the integration of the inertial measurements).

The local frame refers to a frame of reference common to the IMU and the camera. In a real application, we would work in the IMU frame and have some additional constant terms accounting for the camera-IMU transformation. We do not express these constant calibration terms explicitly here for clarity reasons.

The unknowns of Equation **??** are the scalars λ_j^i and the vectors V and G. Note that the knowledge of G is equivalent to the knowledge of the roll and pitch angles. The vectors μ_j^i are fully determined by visual and gyroscope measurements [2], and the vectors S_j are determined

[1]Note that in this paper we do not provide a new derivation of this solution for which the reader is addressed to [**?**], section 3.

[2]The gyroscope measurements in the interval $[t_1,\ t_j]$ are needed to express the bearing at time t_j in the frame at time t_1

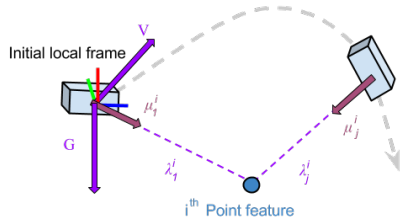

Fig. 1: Visual representation of Equation **??**. The unknowns of the equation are colored in purple.

by accelerometer and gyroscope measurements.

Equation (**??**) provides three scalar equations for each point feature $i = 1,...,N$ and each frame starting from the second one $j = 2,...n_i$. We therefore have a linear system consisting of $3(n_i-1)N$ equations in $6+Nn_i$ unknowns. Indeed, note that, when the first frame is taken at $t_1 = 0$, Equation (**??**) is always satisfied; thus does not provide information. We can write our system using matrix formulation. Solving the system is equivalent to inverting a matrix of $3(n_i-1)N$ rows and $6+Nn_i$ columns.

In [**?**], the author proceeded to one more step before expressing the underlying linear system. For a given frame j, the equation of the first point feature $i = 1$ is subtracted from all other point feature equations $1 < i <= N$ (Equation (7)). This additional step, very useful to detect system singularities, has the effect to corrupt all measurements with the first measurement, hence worsening the performance of the closed-form solution. Therefore, in this paper we discard this additional step.

The linear system in Equation (**??**) can be written in the following compact form:

$$\Xi X = S. \tag{2}$$

Matrix Ξ and vector S are fully determined by the measurements, while X is the unknown vector. We have:

$$S \equiv [S_2^T,...,S_2^T,S_3^T,...,S_3^T,...,S_{n_i}^T,...,S_{n_i}^T]^T$$
$$X \equiv [G^T,V^T,\lambda_1^1,...,\lambda_1^N,...,\lambda_{n_i}^1,...,\lambda_{n_i}^N]^T$$
$$\Xi \equiv$$

$$\left[\begin{array}{c|c|c|c|c|c|c|c|c|c|c}
T_2 & S_2 & \mu_1^1 & 0_3 & 0_3 & -\mu_2^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\
T_2 & S_2 & 0_3 & \mu_1^2 & 0_3 & 0_3 & -\mu_2^2 & 0_3 & 0_3 & 0_3 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
T_2 & S_2 & 0_3 & 0_3 & \mu_1^N & 0_3 & 0_3 & -\mu_2^N & 0_3 & 0_3 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
T_{n_i} & S_{n_i} & \mu_1^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & -\mu_{n_i}^1 & 0_3 & 0_3 \\
T_{n_i} & S_{n_i} & 0_3 & \mu_1^2 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & -\mu_{n_i}^2 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
T_{n_i} & S_{n_i} & 0_3 & 0_3 & \mu_1^N & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & -\mu_{n_i}^N
\end{array}\right],$$

where $T_j \equiv -\frac{t_j^2}{2} I_3$, $S_j \equiv -t_j I_3$ and I_3 is the identity 3×3 matrix, 0_3 is the 3×1 zero matrix. Note that matrix Ξ and vector S are slightly different from the ones proposed in [**?**]. This is due to the additional step that, as we explained in the previous paragraph, we discarded for numerical stability reasons (see [**?**] section 3 for further details).

The sensor information is completely contained in the above linear system. Additionally, in [**?**], the author added a quadratic equation assuming the gravitational acceleration is a priori known. Let us denote the gravitational magnitude by g. We have the extra constraint $|G| = g$. We can express this constraint in matrix formulation:

$$|\Pi X|^2 = g^2, \tag{3}$$

with $\Pi \equiv [I_3,0_3,...,0_3]$. We can therefore recover the initial velocity, the roll and pitch angles, and the distances to the point features by finding the vector X that satisfies (**??**) and (**??**).

In the next sections, we will evaluate the performance of this method on real-world data. This will allow us to identify its weaknesses and bring modifications to overcome them.

IV. LIMITATIONS OF [**?**]

The goal of this section is to find out the limitations of the solution proposed in [**?**]. This is obtained by running the closed-form solution on a combination of both real and synthetic data. In particular, we will add an artificial bias to the real measurements delivered by the inertial sensors (both the accelerometers and the gyroscopes). This will allow us to evaluate the impact of the bias on the performance.

A. Considered data

The setup is similar to the one we use in the experiments on real data described in Section **??**. Specifically, the ground truth pose of the MAV is obtained with a motion-capture system. However, unlike Section **??**, we simulate the camera measurements by creating artificial 3D-landmarks, and re-projecting these landmarks into the camera views (μ_j^i in Equation (**??**)). Simulating the camera measurements allows us to compare our estimations of the distances to the features (λ_j^i in Equation (**??**)) with the ground truth. We measure our error on the absolute scale by computing the mean error over all estimated distances to point features λ_j^i.

Inertial data were recorded from a quadrotor MAV while executing a circular trajectory of about 1m radius (see Fig. **??**).