Absolute scale velocity determination combining visual and inertial measurements for micro aerial vehicles

Jacques Kaiser and Agostino Martinelli

***Abstract*— Hi**

I. INTRODUCTION

Autonomous mobile robots navigating in unknow 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. Additionaly, 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 by an external initial state.

The initialization of a filter based method is critical. Due to non-linearity of the system, a poor initialization can result into converging towards local minima and providing faulty states with high confidence. Indeed, another shortcoming of filters is that they can silently fail.

In this work we demonstrate the efficiency of a recent closed-form solution introduced in [2][1] that 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. It is assumed that the camera is calibrated and the transformation between the IMU and the camera is known. This is a fair assumption for industrial drones to come pre-calibrated.

In this work, we have studied the recent closed-form solution proposed by [1] that performs visual-inertial sensor fusion without requiring an initialization. We implemented this method in order to test it with real terrain data. This allowed us to identify its bottlenecks and bring modifications to overcome them.

Specifically, we started by reformulating the equations for greater numerical stability. This lead to a major leap in the quality of the estimations.

We then investigated the impact of biased inertial measurements. Despite the case of biased accelerometer was originally studied in [1] we show that its low impact on the system makes it hard to estimate.

One major bottleneck of this method was 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 could only be successfully used with a very precise - and expensive - gyroscope. We then introduced a simple method that automatically estimates this bias.

However, this method requires a significant amount of data to provide correct estimates of the gyroscope bias (either many features or long time of integration). We were able to get rid of this limitation with a simple statistical trick by adding a regularization term to our cost function.

By adding this new method for the bias estimation to the original method we obtain results which are equivalent to the ones in absence of bias. Compared to the original method, the new method is now robust to the gyroscope bias, and also provides the gyroscope bias.

II. RELATED WORK

III. THE CLOSED-FORM SOLUTION

In this paper, we do not provide a new derivation of the closed-form solution. Instead, we consider the latest derivation proposed in [1]. Specifically, the author expresses the state of the MAV with respect to the visual and inertial measurements in Equation 6:

$$S_j = \lambda_1^i \mu_1^i - V t_j - G\frac{t_j^2}{2} - \lambda_j^i \mu_j^i \tag{6}$$

With:

- μ_j^i the normalized bearing of point feature i at time t_j in the initial local frame;

- λ_j^i the distance to the point feature i at time t_j;
- V the initial velocity in the initial local frame;
- G the initial gravity in the initial local frame;
- S_j the integration up to time t_j of the rotated linear acceleration data.

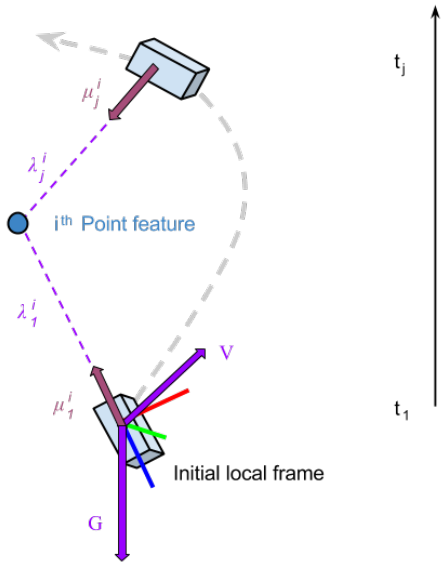

Fig. 1: Visual representation of Equation 6. The unknowns of the equation are colored in purple.

The unknowns of Equation 6 are the distances 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 camera observations and gyroscope measurements, and the vectors S_j are determined by accelerometer and gyroscope measurements.

As Equation 6 holds for each three dimensions of all point features $i = 1,...,N$ and each observation starting from the second one $j = 2,...n_i$, we therefore have a system consisting of $3(n_i-1)N$ equations in $6+Nn_i$ unknowns. Indeed, note that when the first observation occurs, at $t_j = 0$, Equation 6 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 [1], the author proceeded to one more step before expressing the underlying linear system. For an observation at time t_j, the equation of the first point feature $i = 1$ happening at time t_j is subtracted to all other point features $1 < i <= N$ at time t_j (Equation 7). This additional step has the effect to corrupt all measurements with the first measurement, hence worsening the performance of the closed-form solution. In this paper, we do not take to this additional step.

The linear system in Equation 6 can be written in the following compact format:

$$\Xi X = S \tag{9}$$

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

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

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

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 x 3 matrix; 0_{33} is the 3 x 3 zero matrix. Note that the matrix Ξ and the vector S are slightly different from the one proposed in [1]. This is due to the additional step we did not take for numerical stability reasons.

The sensor information is completely contained in the above linear system. Additionally, in [1], 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:

$$|\Pi X|^2 = g^2 \tag{10}$$

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 which satisfies 9 and 10.

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

IV. PERFORMANCE BOTTLENECKS

A. Test setup

The MAV performs a motion while being tracked with an optical Vicon system. We can therefore compare our estimations with the ground truth. We define the relative error as the euclidian distance between the estimation and the ground truth, normalized by the groud truth. We measure our error on the absolute scale by computing the mean error over all estimated distances to point features λ_j^i.

To identify the performance bottlenecks, we used IMU data obtained from terrain acquisitions while we simulated the point feature observations. This separation allowed us to know the ground truth for the distance to the point features and also better understand the weaknesses of our method.

We represent this setup in Figure **??**.

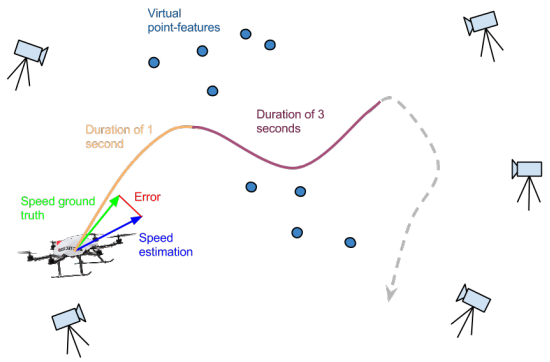

Fig. 2: Test setup for identifying the performance bottlenecks. The drone is equipped with an IMU, and the visual measurements are simulated. It performs a motion while being tracked by a Vicon system.

In general, we use one camera observation every 0.3 seconds, even if the camera provides significantly more frames. Indeed, we can discard most of the camera observations. If two observations are too close to each other, then the additional equations do not bring much information to our system. Reducing the number of considered frames reduces the size of the matrices, thus speeds up the computations.

As an example, over a time interval of 3 seconds, we obtain 11 distinct frames. When observing 7 features, it yields a system of $3 \times 10 \times 7 = 210$ equations and $6 + 7 \times 10 = 76$ unknowns.

The method we use to solve the overconstrained linear system $\Xi X = S$ is a singular value decomposition (SVD) since it yields numerically robust solutions.

In this section, we will start by presenting the results obtained with the original closed-form solution on terrain IMU data. Our goal is to identify its performance bottlenecks and introduce modifications to overcome them.

B. Original closed-form solution performance

The original closed-form solution described in Equation 9 will be used as a basis for our work. Moreover, we can also use the knowledge of the gravity magnitude in order to refine our results with Equation 10. In this case, we are minimizing a linear objective function with a quadratic constraint. in Figure **??**, we represent the quality of the evaluations with and without this additional constraint.

Note how the evaluations get better as we increase the integration duration. Indeed, our equations come from an extended triangulation [2]. Therefore, it requires a significant difference in the measurements over time to robustly estimate the state. Also note that the gravity is robustly estimated, whereas the estimations of the speed and the distance to the features is more erroneous.

Since the gravity is well estimated with the standard closed-form, it comes without surprise that constraining its magnitude does not impact the results much. That is why the estimations with and without the refinement are so similar.

With around 45% of error on the speed and the distance to the features after 4 seconds of integration, the original closed-form solution does not perform very well on terrain data. In the following sections, we will introduce modifications to improve over these results.

V. ESTIMATING THE GYROSCOPE BIAS

VI. RESULTS

VII. CONCLUSION

REFERENCES

[1] Agostino Martinelli. Closed-form solution of visual-inertial structure from motion. *International Journal of Computer Vision*, 106(2):138–152, 2014.

[2] Agostino Martinelli and Roland Siegwart. Vision and IMU Data Fusion: Closed-Form Determination of the Absolute Scale, Speed, and Attitude. *Handbook of Intelligent Vehicles*, 28(1):1335–1354, 2012.