View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

Cooperative Visual-Inertial Sensor Fusion: Fundamental Equations and State Determination in Closed-Form

Agostino Martinelli, Alessandro Renzaglia, Alexander Oliva

To cite this version:

Agostino Martinelli, Alessandro Renzaglia, Alexander Oliva. Cooperative Visual-Inertial Sensor Fusion: Fundamental Equations and State Determination in Closed-Form. Autonomous Robots, Springer Verlag, In press, pp.1-19. 10.1007/s10514-019-09841-8 . hal-02013869

HAL Id: hal-02013869

https://hal.inria.fr/hal-02013869

Submitted on 11 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noname manuscript No.
(will be inserted by the editor)

Cooperative Visual-Inertial Sensor Fusion: Fundamental Equations and State Determination in Closed-Form

Agostino Martinelli, Alessandro Renzaglia and Alexander Oliva

Received: date / Accepted: date

Abstract This paper investigates the visual and inertial sensor fusion problem in the cooperative case and provides new theoretical and basic results. Specifically, the case of two agents is investigated. Each agent is equipped with inertial sensors (accelerometer and gyroscope) and with a monocular camera. By using the monocular camera, each agent can observe the other agent. No additional camera observations (e.g., of external point features in the environment) are considered. First, the entire observable state is analytically derived. This state contains the relative position between the two agents (which includes the absolute scale), the relative velocity, the three Euler angles that express the rotation between the two local frames and all the accelerometer and gyroscope biases. Then, the basic equations that describe this system are analytically obtained. The last part of the paper describes the use of these equations to obtain a closed-form solution that provides the observable state in terms of the visual and inertial measurements provided in a short time interval. This last contribution is the extension of the results presented in [18,31,32] to the cooperative case. The impact of the presence of the bias on the performance of this closed-form solution is also investigated and a simple and effective method to obtain the gyroscope bias is proposed. Extensive simulations clearly show that the proposed method is successful. It is worth noting that it is possible to automatically retrieve the absolute scale and simultaneously calibrate the gyroscopes not only without any prior knowledge (as in [18]), but also without external point features in the environment.

Keywords Visual-Inertial Sensor Fusion · Observability · Cooperative Sensor Fusion · Closed-Form Solution

A. Martinelli is with INRIA Rhône-Alpes, Grenoble, France, E-mail: agostino.martinelli@inria.fr
A. Renzaglia is with INRIA Grenoble Rhône-Alpes and INSA Lyon CITI Lab, France, E-mail: alessandro.renzaglia@inria.fr
A. Oliva is with INRIA Rhône-Alpes, Grenoble, France, E-mail: alexander.oliva@inria.fr

1 Introduction

When a team of mobile robots cooperates to fulfill a task, an optimal localization strategy must take advantage of relative observations (detection of other robots). This problem has been considered in the past by following different approaches and it is often referred as *Cooperative Localization*. In Cooperative Localization (CL), several communicating robots use relative measurements (such as distance, bearing and orientation between the robots) to jointly estimate their poses. This problem has been investigated for a long time and several approaches have been introduced in earlier works [6,10,19,29,40–42]. Then, a great effort has been devoted to decentralize the computation among the team members and, simultaneously, to minimize the communication among the robots without deteriorating the localization performance [3,14,20–22,25,27,43]. Specific cases of cooperative localization have been considered both in 2D and in 3D.

For instance, in the framework of Micro Aerial Vehicles (MAV), a critical issue is to limit the number of on-board sensors to reduce weight and power consumption. Several methods consider the use of bearing-only sensors [35,39,44,45] or only range measurements [46]. A common setup is otherwise 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.

The problem of fusing visual and inertial data for single robots has been extensively investigated in the past [2,4,11,17,24]. Recently, this sensor fusion problem has been successfully addressed by enforcing observability constraints [9,13], and by using optimization-based approaches [5,12,16,23,28,36,37]. These optimization 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. For this reason, a closed-form solution able to automatically determine the state without initialization has been introduced [18,31,32].

Visual and inertial sensors have also been used in a cooperative scenario to estimate the relative state [1] and for cooperative mapping [7]. However, in the cooperative case, a solution able to automatically determine the state without initialization (as in [18,31,32]) is still missing.

Any estimation approach, either filter based or optimization based, is built upon the fundamental equations that fully characterize the considered sensor fusion problem. These equations are the differential equations that describe the dynamics of the observable state together with the equations that express the observations in terms of this observable state. Hence, to successfully solve a given estimation problem, the first step to be accomplished is the determination of the observable state. Regarding the single-agent visual-inertial sensor fusion problem, this state has been analytically derived by many authors and it consists of the absolute scale, the speed expressed in the local frame and the absolute roll and pitch angles. This result even holds if only a single point feature is available in the environment.

In this paper we study the visual-inertial sensor fusion problem in the cooperative case. We investigate the extreme case where no point features are available. Additionally, we consider the critical case of only two agents. In other words, we are interested in investigating the minimal case. If we prove that the absolute scale is observable, we can conclude that it is observable in all the other cases. Each agent is equipped with an Inertial Measurement Unit (IMU) and a monocular camera. By using the monocular camera, each agent can observe the other agent. Note that, we do not assume that these camera observations contain metric information (due for instance to the known size of the observed agent). The two agents can operate far from each other and a single camera observation only consists of the bearing of the observed agent in the frame of the observer. In other words, each agent acts as a moving point feature with respect to the other agent.

The first questions we wish to answer are: *Is it possible to retrieve the absolute scale in these conditions? And the absolute roll and pitch angles?* More generally, we want to determine the entire observable state, i.e., all the physical quantities that it is possible to determine by only using the information contained in the sensor data (from the two cameras and the two IMUs) during a short time interval. In [33] we provided the answers to these questions in the case when the inertial measurements are unbiased. These results are provided in section 3. Then, in section 5, we provide a full answer even in presence of biased measurements (both the ones from the accelerometers and the ones from the gyroscopes) and we also obtain that it suffices that only one agent is equipped with a camera. In addition, it suffices that this camera is a linear camera, i.e., which only provides the azimuth of the other agent in its local frame.

Note that part of these questions have already been answered in [1]. However, the results here provided in sections 3-5 are more general for the following reasons:

- They account for the bias on all the inertial measurements (both on the accelerometers and the gyroscopes);
- As mentioned above, we also prove that the same observability properties hold when only one of the agents is equipped with a camera and that this single camera can even be a linear camera;
- In [1] it is proved that the relative state is observable while here it is also proved that no other states are observable (e.g., the absolute roll and pitch of each agent is unobservable);
- In [1] it is assumed that the camera directly provides the relative position (up to a scale) and the relative orientation. In our derivation we do not require the latter assumption. In other words, it suffices that the camera detects one single point on the observed agent (which represents the origin of its local frame). This does not require more restrictive assumptions (e.g., the two agents can operate very far from each other).

In section 4 we provide the basic equations that describe the cooperative visual-inertial sensor fusion problem. These equations are:

- The differential equations that describe the dynamics of the observable state expressed only in terms of the components of the observable state and the accelerations and the angular speeds (i.e., the quantities measured by the two IMUs);

– The equations that provide the analytic expression of the two camera observations in terms of the components of the observable state.

These are the fundamental equations that fully characterize the problem of fusing visual and inertial data in the cooperative case. These equations can then be used to build any method (e.g., filter-based or optimization-based) to carry out the state estimation. In [33] we used them to introduce an EKF-based estimation method. Note that an EKF-based estimation method was also introduced in [1]. In that work, the authors did not need the equations derived in section 4 because, as mentioned above, they used a more restrictive camera model, which assumes that the camera also provides the relative orientation (we relax this assumption). In this paper we use these fundamental equations to obtain a closed-form determination of the observable state in terms of the measurements delivered during a short time interval by the cameras and the IMUs that belong to the two agents. This solution is provided in section 6. This is precisely the extension of the closed-form solution in [31,32] to the cooperative case. For clarity sake, in section 6 we directly provide the solution by addressing the reader to the appendix B for its analytic derivation (and to [34] for further technical details). Then, the paper demonstrates the efficiency of this solution. A closed-form solution directly returns the state in terms of the measurements collected during a short time interval and, thus, does not require any initialization. We perform simulations with plausible MAV motions and synthetic noisy sensor data (section 7). This allows us to identify limitations of the solution and bring modifications to overcome them. In practice, we perform exactly the same investigation done in [18] for the case of a single agent. Specifically, we investigate the impact of biased inertial measurements. We show that a large bias on the accelerometer does not significantly worsen the performance (section 7.5). One major limitation is the impact of biased gyroscope measurements (section 7.6). In other words, the performance becomes very poor in presence of a bias on the gyroscopes of the two agents and, in practice, the overall method can only be successfully used with very precise - and expensive - gyroscopes. In section 8, we introduce a simple method that automatically estimates both these biases. By adding this new method for the bias estimation to the solution presented in section 6, we obtain results that are equivalent to the ones in absence of bias (section 8.1).

Note that the implementation of the closed-form solution requires that two MAVs observe one each other. This could seem restrictive since most MAVs do not have omni-directional cameras, but rather a front facing camera with a limited field of view. However, the big advantage of a closed form solution is that, if at a given time it fails (loss of visual contact or any other unmodeled event), this does not have any impact on its performance at successive times. In addition, it suffices that the two MAVs observe one each other not more than 10 times during a time period of not more than 4 seconds.

Finally, it is important to note that, even though in this paper we particularly focus on multi-MAV systems, this method is suitable for any kind of robots moving in 3D that operate in extreme conditions (e.g., GPS-denied environments, absence of point features, etc.) and need to recover the absolute scale in few seconds. The solution does not need initialization. Additionally, it is robust to the bias and automatically calibrates the gyroscopes.

2 The system

We consider two vehicles that move in a $3D$ environment. Each vehicle is equipped with an Inertial Measurement Unit (IMU), which consists of three orthogonal accelerometers and three orthogonal gyroscopes. Additionally, each vehicle is equipped with a monocular camera. We assume that, for each vehicle, all the sensors share the same frame. Without loss of generality, we define the vehicle local frame as this common frame. The accelerometer sensors perceive both the gravity and the inertial acceleration in the local frame. The gyroscopes provide the angular speed in the local frame. Finally, the monocular camera of each vehicle provides the bearing of the other vehicle in its local frame (see Fig. 1 for an illustration). Additionally, we assume that the z-axis of the global frame is aligned with the direction of the gravity.
We adopt the following notations:

– $r^1 = [r^1_x,\ r^1_y,\ r^1_z]$ and $r^2 = [r^2_x,\ r^2_y,\ r^2_z]$ are the positions of the two vehicles in the global frame;
– $v^1 = [v^1_x,\ v^1_y,\ v^1_z]$ and $v^2 = [v^2_x,\ v^2_y,\ v^2_z]$ are the velocities of the two vehicles in the global frame;
– $q^1 = q^1_t + q^1_x i + q^1_y j + q^1_z k$ and $q^2 = q^2_t + q^2_x i + q^2_y j + q^2_z k$ are the two unit quaternions that describe the rotations between the global and the two local frames, respectively[1].

In the following, for each vector defined in the $3D$ space, the subscript q will be adopted to denote the corresponding imaginary quaternion. For instance, regarding

[1] A quaternion $q = q_t + q_x i + q_y j + q_z k$ is a unit quaternion if the product with its conjugate is 1, i.e.: $qq^* = q^*q = (q_t + q_x i + q_y j + q_z k)(q_t - q_x i - q_y j - q_z k) = (q_t)^2 + (q_x)^2 + (q_y)^2 + (q_z)^2 = 1$

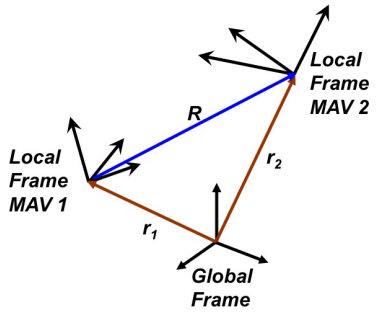


Fig. 1 The global frame and the two local frames (attached to the first and the second aerial vehicle, respectively). r_1 and r_2 are their position, expressed in the global frame. R is the relative position of the second vehicle with respect to the first vehicle, expressed in the local frame of the first vehicle.

the position of the first vehicle, we have: $r^1_q = 0 + r^1_x\, i + r^1_y\, j + r^1_z\, k$. Additionally, we denote by A^1, A^2, Ω^1 and Ω^2 the following physical quantities:

- $A^{rob} = [A^{rob}_x,\ A^{rob}_y,\ A^{rob}_z]$, $(rob = 1, 2)$, is the vehicle acceleration perceived by the IMU mounted on the first and the second vehicle (this includes both the inertial acceleration and gravity);
- $\Omega^{rob} = \left[\Omega^{rob}_x\ \Omega^{rob}_y\ \Omega^{rob}_z\right]$, $(rob = 1, 2)$, is the angular speed of the first and the second vehicle expressed in the respective local frame (and $\Omega^{rob}_q = 0 + \Omega^{rob}_x\, i + \Omega^{rob}_y\, j + \Omega^{rob}_z\, k$).

The dynamics of the first/second vehicle are:

$$\begin{cases} \dot{r}^{rob}_q = v^{rob}_q \\ \dot{v}^{rob}_q = q^{rob} A^{rob}_q (q^{rob})^* - gk \\ \dot{q}^{rob} = \dfrac{1}{2} q^{rob} \Omega^{rob}_q \end{cases} \tag{1}$$

where g is the magnitude of the gravity, $rob = 1, 2$, and k is the fourth fundamental quaternion unit ($k = 0 + 0\, i + 0\, j + 1\, k$).

The monocular camera on the first vehicle provides the position of the second vehicle in the local frame of the first vehicle, up to a scale. The position of the second vehicle in the local frame of the first vehicle is given by the three components of the following imaginary quaternion:

$$p^1_q = (q^1)^* (r^2_q - r^1_q)\, q^1\,. \tag{2}$$

Hence, the first camera provides the quaternion p^1_q up to a scale. For the observability analysis, it is convenient to use the ratios of its components:

$$h^1 \triangleq [h^1_u,\ h^1_v]^T = \left[\frac{[p^1_q]_x}{[p^1_q]_z},\ \frac{[p^1_q]_y}{[p^1_q]_z}\right]^T \tag{3}$$

where the subscripts x, y and z indicate respectively the i, j and k component of the corresponding quaternion. Similarly, the second camera provides:

$$h^2 \triangleq [h^2_u,\ h^2_v]^T = \left[\frac{[p^2_q]_x}{[p^2_q]_z},\ \frac{[p^2_q]_y}{[p^2_q]_z}\right]^T \tag{4}$$

where p^2_q is the imaginary quaternion whose three components are the position of the first vehicle in the local frame of the second, namely:

$$p^2_q = (q^2)^* (r^1_q - r^2_q)\, q^2\,. \tag{5}$$

Note that, using the ratios in (3) and (4) as observations can generate problems due to singularities and, when the camera measurements are used to estimate a state, it is more preferable to adopt different quantities (e.g., the two bearing angles, i.e., the azimuth and the zenith). For the observability analysis, this problem does not arise.

3 Observable state

The goal of this subsection is to obtain the entire observable state for the system defined in section 2. First of all, we characterize this system by the following state:

$$X = [(r^1)^T,\ (v^1)^T,\ q^1,\ (r^2)^T,\ (v^2)^T,\ q^2]^T\,. \tag{6}$$

The dimension of this state is equal to 20. Actually, the components of this state are not independent. Both q^1 and q^2 are unit quaternions. In other words, we have:

$$(q^1_t)^2 + (q^1_x)^2 + (q^1_y)^2 + (q^1_z)^2 = (q^2_t)^2 + (q^2_x)^2 + (q^2_y)^2 + (q^2_z)^2 = 1\,. \tag{7}$$

The dynamics of the state defined in (6) are given by (1). The observation functions are the four scalar functions h^1_u h^1_v h^2_u h^2_v given by equations (2-5). Additionally, we need to add the two observation functions that express the constraint that the two quaternions, q^1 and q^2, are unit quaternions. The two additional observations are:

$$h^{rob}_{const}(X) \triangleq (q^{rob}_t)^2 + (q^{rob}_x)^2 + (q^{rob}_y)^2 + (q^{rob}_z)^2 = 1 \tag{8}$$

with $rob = 1, 2$. We investigate the observability properties of this system. Since both the dynamics and the six observations are nonlinear with respect to the state, we use the observability rank condition in [8]. The dynamics are affine in the inputs, i.e., they have the expression

$$\dot{X} = f_0(X) + \sum_{i=1}^{12} f_i(X) u_i \tag{9}$$

where u_i are the system inputs, which are the quantities measured by the two IMUs. Specifically, we set:

- u_1, u_2, u_3 the three components of A^1;
- u_4, u_5, u_6 the three components of Ω^1;
- u_7, u_8, u_9 the three components of A^2;
- u_{10}, u_{11}, u_{12} the three components of Ω^2.

Then, by comparing (1) with (9) it is immediate to obtain the analytic expression of all the vector fields $f_0, f_1, \cdots f_{12}$; for instance, we have:

$$f_0 = [v_x^1, v_y^1, v_z^1, 0, 0, -g, 0_4, v_x^2, v_y^2, v_z^2, 0, 0, -g, 0_4]^T$$

$$f_1 = [0_3, (q_t^1)^2 + (q_x^1)^2 - (q_y^1)^2 - (q_z^1)^2, 2(q_t^1 q_z^1 + q_x^1 q_y^1),$$
$$2(q_x^1 q_z^1 - q_t^1 q_y^1), 0_{14}]^T$$

$$f_4 = \frac{1}{2}[0_6, -q_x^1, q_t^1, q_z^1, -q_y^1, 0_{10}]^T$$

where 0_n is the n-line zero vector.

For systems with the dynamics given in (9) the application of the observability rank condition can be automatically done by a recursive algorithm. In particular, this algorithm automatically returns the observable codistribution[2] by computing the Lie derivatives of all the system outputs along all the vector fields that characterize the dynamics. In the following, we provide a very simple description of the observability rank condition for systems with the dynamics given in (9), i.e., dynamics nonlinear in the state and affine in the inputs (for a detailed description the reader is addressed to the first chapter of [15]). In accordance with the observability rank condition, the observable codistribution provides all the observability properties. The dimension of this vector space (the observable coditribution) cannot exceed the dimension of the state X. If this dimension is equal to the dimension of the state X, this means that the entire state is observable (actually, weakly locally observable [8]). If this dimension is smaller than the dimension of the state X, the entire state is not observable and it is possible to detect the observable states by computing its Killing vectors in order to obtain the system symmetries [30]. The recursive algorithm that returns the observable codistribution, for systems with the dynamics given in (9), is the following:

Algorithm 1 *Observable codistribution Λ*

Set $\Lambda_0 = span\{\nabla h_u^1, \nabla h_v^1, \nabla h_u^2, \nabla h_v^2, \nabla h_{const}^1, \nabla h_{const}^2\}$
while $\Lambda_m \neq \Lambda_{m-1}$ ***do***
$\Lambda_m = \Lambda_{m-1} + \mathcal{L}_{f_0}\Lambda_{m-1} + \sum_{i=1}^{12} \mathcal{L}_{f_i}\Lambda_{m-1}$
end while

where Λ_m, with $m \geq 1$, is the codistribution at the m-th step and the symbol ∇ denotes the gradient with respect to the state X.
We remind the reader that the Lie derivative of a scalar function $h(X)$ along the vector field $f(X)$ is defined as follows:

$$\mathcal{L}_f h \triangleq \nabla h \cdot f$$

which is the product of the row vector ∇h with the column vector f. Hence, it is a scalar function. Additionally, by definition of Lie derivative of covectors, we have: $\mathcal{L}_f \nabla h = \nabla \mathcal{L}_f h$. Finally, given two vector spaces V_1 and V_2, we denoted by $V_1 + V_2$ their sum, i.e., the span of all the generators of both V_1 and V_2.

In [15] it is proved that algorithm 1 converges. In particular, it is proved that it has converged when $\Lambda_m = \Lambda_{m-1}$. An interesting consequence of this result is that the convergence is achieved in at most $n-1$ steps, where n is the dimension of the state (see lemmas 1.9.1, 1.9.2 and 1.9.6 in [15]).

We provide few insights to figure out how Algorithm 1 works and in particular how it can be implemented in practice. As we mentioned above, the observable codistribution is the span of line vectors. In practice, Algorithm 1 builds a matrix whose lines are these vectors (e.g., at the first step we include the six lines: $\nabla h_u^1, \nabla h_v^1, \nabla h_u^2, \nabla h_v^2, \nabla h_{const}^1, \nabla h_{const}^2$). At each subsequent step, we include a new set of lines and we compute the rank of the matrix (the new set of lines is obtained by computing the Lie derivatives of all the lines of the matrix along all the directions allowed by the dynamics, i.e., along all the vector fields $f_0, f_1, \cdots f_{12}$). The algorithm has converged when the rank of the matrix remains equal to the rank of the matrix at the previous step. Note that each line will be a symbolic function of the state. To compute the rank we use the symbolic tool of MATLAB. In particular, we use the functions "rank" and "null". The latter provide the killing vectors of the matrix which are precisely the symmetries of the system, once the algorithm has converged [30]. If the set of the killing vectors of the matrix only consists of the null vector, this means that the systems does not have any symmetry and the entire state is observable.

For the specific case, we obtain that the algorithm converges at the third step, i.e., the observable codistribution is the span of the differentials of the previous Lie derivatives up to the second order. In particular, its dimension is 11 and, a choice of eleven Lie derivatives is:

[2] The reader unfamiliar with the concept of *codistribution*, as it is used in [15], should not be afraid by the term *distribution* and the term *codistribution*. Very simply speaking (and this is enough to understand the theory of nonlinear observability) they are both vector spaces. Specifically, a distribution is the span of a set of column-vector functions. A codistribution is the span of a set of line-vector functions. Hence, both a distribution and a codistribution can be regarded as vector spaces that change by moving on the space of the states (X), namely, vector spaces that depend on X.

$\mathcal{L}^0 h^1_u$, $\mathcal{L}^0 h^1_v$, $\mathcal{L}^0 h^2_u$, $\mathcal{L}^0 h^2_v$, $\mathcal{L}^0 h^1_{const}$, $\mathcal{L}^0 h^2_{const}$, $\mathcal{L}^1_{f_0} h^1_u$, $\mathcal{L}^1_{f_0} h^1_v$, $\mathcal{L}^1_{f_0} h^2_u$, $\mathcal{L}^2_{f_0 f_0} h^1_u$, $\mathcal{L}^2_{f_0 f_1} h^1_u$ [3].

Once we have obtained the observable codistribution, the next step is to obtain the observable state. This state has eleven components. Obviously, a possible choice would be the state that contains the previous eleven Lie derivatives. On the other hand, their expression is too complex and it is much more preferable to find an easier state, whose components have a clear physical meaning. By analytically computing the continuous symmetries of our system (i.e., the Killing vectors of the previous observable codistribution, [30]), we detect the following independent observable modes:

- The position of the second vehicle in the local frame of the first vehicle (three observable modes);
- The velocity of the second vehicle in the local frame of the first vehicle (three observable modes);
- The three Euler angles that characterize the rotation between the two local frames (three observable modes);
- Trivially, the norm of the two quaternions (two observable modes).

Therefore, we can fully characterize our system by a state whose components are the previous observable modes. It must be possible to express the dynamics of this state only in terms of its components and the twelve system inputs. Additionally, also the camera observations must be expressed only in terms of these nine components. This is actually trivial, since the first camera provides the first three components of this state, up to a scale. The second camera, provides the same unit vector rotated according to the previous three Euler angles. Regarding the dynamics, its derivation is a bit more complex. We provide all these analytic expressions in the next section.
We conclude this section with the following three important remarks.

- The absolute roll and pitch angles of each vehicle are not observable. This is a consequence of the fact that no feature in the environment has been considered. The observation consists only of the bearing angles of each vehicle in the local frame of the other vehicle. The presence of the gravity, which determines the observability of the absolute roll and pitch in the case of a single vehicle, acts in the same way on the two IMUs and its effect on the system observability vanishes, since it cannot be distinguished from the inertial acceleration.
- The choice of the above 11 independent Lie derivatives is not unique. In particular, it is possible to avoid the Lie derivatives of the functions that correspond to one of the two cameras (e.g., h^2_u and h^2_v). This means that we obtain the same observability properties when only one of the MAVs is equipped with a camera. In addition, it is also possible to avoid the Lie derivatives of the function h^1_v. This means that we obtain the same observability properties when only one MAV is equipped with a camera and this camera is a linear camera able to only provide the azimuth of the other MAV in its local frame. In section 5 we obtain that the same result holds even in presence of a bias on the inertial measurements (see also appendix A for computation details).
- In order to have 11 eleven independent Lie derivatives, at least one of them must be computed along a direction that corresponds to one of the axes of at least one of the two accelerometers (i.e., one direction among f_1, f_2, f_3, f_7, f_8 and f_9). Any selection that does not include at least one of them provides a codistribution whose dimension is smaller than 11. In particular, there will be a symmetry for this codistribution that corresponds to a scale invariance. This means that a necessary condition for the observability of the absolute scale is that the relative acceleration between the two MAVs does not vanish. Note that this same condition was found in [1] (the fact that in [1] the camera is assumed to directly provide the relative orientation does not impact the observability of the scale).

[3] Higher order Lie derivatives are recursively computed. For instance, for the second order Lie derivative $\mathcal{L}^2_{f_0 f_1} h$ we have $\mathcal{L}^2_{f_0 f_1} h = \nabla(\mathcal{L}_{f_0} h) \cdot f_1 = [\nabla(\nabla h \cdot f_0)] \cdot f_1$

4 Fundamental Equations

In accordance with the observability analysis carried out in the previous section, we characterize our system by the following state:

$$S = [R^T,\ V^T,\ q]^T \tag{10}$$

where:

- R is the position of the second vehicle in the local frame of the first vehicle;
- V is the velocity of the second vehicle in the frame of the first vehicle (note that this velocity is not simply the time derivative of R because of the rotations accomplished by the first local frame);
- q is the unit quaternion that describes the relative rotation between the two local frames.

In other words, the imaginary quaternions associated to R and V are:

$$R_q = (q^1)^*(r^2_q - r^1_q)q^1 \tag{11}$$

$$V_q = (q^1)^*(v_q^2 - v_q^1)q^1 \tag{12}$$

and

$$q = (q^1)^* q^2 \tag{13}$$

The fundamental equations of the cooperative visual-inertial sensor fusion problem are obtained by differentiating the previous three quantities with respect to time and by using (1) in order to express the dynamics in terms of the components of the state in (10) and the components of $A^1, A^2, \Omega^1, \Omega^2$. After some analytic computation, we obtain:

$$\begin{cases} \dot{R}_q = \frac{1}{2}(\Omega_q^1)^* R_q + \frac{1}{2} R_q \Omega_q^1 + V_q \\ \dot{V}_q = \frac{1}{2}(\Omega_q^1)^* V_q + \frac{1}{2} V_q \Omega_q^1 + q A_q^2 q^* - A_q^1 \\ \dot{q} = \frac{1}{2}(\Omega_q^1)^* q + \frac{1}{2} q \Omega_q^2 \end{cases} \tag{14}$$

As desired, the dynamics of the state is expressed only in terms of the components of the state and the system inputs (the angular speeds and the accelerations of both the vehicles). Finally, the camera observations can be immediately expressed in terms of the state in (10). The first camera provides the vector R up to a scale. Regarding the second camera, we first need the position of the first vehicle in the second local frame. The components of this position are the components of the following imaginary quaternion: $-q^* R_q q$. The second camera provides this position up to a scale.

In the last part of this section we provide the same equations, without using quaternions. We characterize our system by the two 3D vectors R and V, as before. Instead of the quaternion q, we use the matrix O that characterizes the rotation between the two local frames. From (14) it is immediate to obtain the dynamics of this state. They are:

$$\begin{cases} \dot{R} = \left[\Omega^1\right]_\times R + V \\ \dot{V} = \left[\Omega^1\right]_\times V + O A^2 - A^1 \\ \dot{O} = \left[\Omega^1\right]_\times^T O + O \left[\Omega^2\right]_\times \end{cases} \tag{15}$$

where $\left[\Omega^{rob}\right]_\times$, $rob = 1, 2$, are the skew-symmetric matrices associated to the vectors Ω^{rob}:

$$\left[\Omega^{rob}\right]_\times = \begin{bmatrix} 0 & \Omega_z^{rob} & -\Omega_y^{rob} \\ -\Omega_z^{rob} & 0 & -\Omega_x^{rob} \\ \Omega_y^{rob} & -\Omega_x^{rob} & 0 \end{bmatrix} \tag{16}$$

Finally, the two cameras provide the two vectors, R and $-O^T R$, up to a scale.

The cooperative visual-inertial sensor fusion problem is fully characterized by the dynamics equations given in (15) and the two observations given by R and $-O^T R$, up to a scale. These equations allow us to build any estimation strategy: filter-based, optimization-based or a closed-form solution, i.e. a solution that extends the solution given in [32] to the cooperative case.

5 Observable state in presence of bias

The goal of this section is to obtain the observable state when the inertial measurements are corrupted by the biases. Specifically, we introduce the following four vectors: B_A^1, B_A^2, B_Ω^1 and B_Ω^2. B_A^1 and B_A^2 are the biases on the accelerometers of the first and the second vehicle and B_Ω^1 and B_Ω^2 are the biases on the gyroscopes. Since the presence of the bias cannot improve the observability properties, we characterize our system by including in the observable state that holds in absence of bias (i.e., the state mentioned in the previous subsection), all the 12 components of the 4 bias vectors. If we prove that this state is observable, we can conclude that it is the entire observable state, i.e., any other physical quantity independent from its components is unobservable. Additionally, we will consider the case when only the first agent is equipped with a camera. Again, by proving that in these conditions the previous state is observable, we can conclude that the same observable state characterizes the case of two cameras.

Both the biases on the gyroscopes and on the accelerometers are time dependent. However, they change very slowly with time. In particular, they are modelled as random-walk processes driven by the zero-mean, white Gaussian noise $n_{B_\Omega}^1$, $n_{B_\Omega}^2$, $n_{B_A}^1$, $n_{B_A}^2$, respectively.

To characterize our system we define the extended state S_E by including the bias in the state (10):

$$S_E = [R^T,\ V^T,\ q,\ (B_\Omega^1)^T,\ (B_A^1)^T,\ (B_\Omega^2)^T,\ (B_A^2)^T]^T\,. \tag{17}$$

The dimension of this state is equal to 22. Actually, the components of this state are not independent, since q is a unit quaternion. In other words, we have:

$$(q_t)^2 + (q_x)^2 + (q_y)^2 + (q_z)^2 = 1\,. \tag{18}$$

The dynamics of the state defined in (17) are given by the following equations:

$$\begin{cases} \dot{R} &= \left[\Omega'^1\right]_\times R + V \\ \dot{V} &= \left[\Omega'^1\right]_\times V + O A'^2 - A'^1 \\ \dot{q} &= -\frac{1}{2}\Omega_q'^1 q + \frac{1}{2} q \Omega_q'^2 \\ \dot{B}_\Omega^1 = n_{B_\Omega}^1,\ \dot{B}_A^1 = n_{B_A}^1,\ \dot{B}_\Omega^2 = n_{B_\Omega}^2,\ \dot{B}_A^2 = n_{B_A}^2 \end{cases} \tag{19}$$

where:

- $\Omega'^1 = \Omega^1 + B^1_\Omega$, $A'^1 = A^1 + B^1_A$, $\Omega'^2 = \Omega^2 + B^2_\Omega$, $A'^2 = A^2 + B^2_A$.
- The matrix O, can be uniquely expressed in terms of the components of the quaternion q.
- Ω'^1_q is the imaginary quaternion associated with Ω'^1, i.e.,: $\Omega'^1_q = 0 + \Omega'^1_x i + \Omega'^1_y j + \Omega'^1_z k$. The same holds for Ω'^2_q

Note that, in the interval of few seconds, the time derivatives of the biases (last equation in (19)) can be set to zero. Since we will consider time intervals no longer than 4 seconds (and, as it will be shown, this will allow us to auto calibrate the inertial sensors with very high accuracy), we can assume that the biases are constant during the considered time interval (the same assumption is made in [18]).

The observation functions are the two scalar functions h_u h_v:

$$h \triangleq [h_u,\ h_v]^T = \left[\frac{R_x}{R_z},\ \frac{R_y}{R_z}\right]^T . \tag{20}$$

Additionally, we need to add the observation function that expresses the constraint that q is a unit quaternion. The additional observation is:

$$h_{const}(X) \triangleq (q_t)^2 + (q_x)^2 + (q_y)^2 + (q_z)^2 \tag{21}$$

The analytic derivation of this system observability is provided in appendix A. We summarize its result:

The system defined above is observable (i.e., the state in (17) is observable). This holds both in the case when both the MAVs are equipped with a camera and in the case when only one MAV is equipped with a camera. Additionally, the observable state remains the same even in the case when the camera is a linear camera, i.e., it only provides the azimuth of the other MAV in its local frame. Finally, as in the case without bias, a necessary condition for the observability of the absolute scale is that the relative acceleration between the two MAVs does not vanish.

6 Closed-form solution

In this section we provide a closed-form solution that allows us to determine R, V and O by only using the measurements provided by the visual and the inertial sensors during a short time interval. In this section, we only provide the solution. The analytic derivation is provided in appendix B and more details about this derivation are available in [34]. Additionally, for brevity sake, we only deal with the case when only the first MAV is equipped with a camera. The case when both the MAVs are equipped with a camera is very similar (both for the analytic derivation and the solution) and can be found in [34]. Note that this solution is obtained by assuming noiseless and unbiased measurements. Hence, it is exact only in the noiseless and unbiased case. On the other hand, the impact of the bias on its performance will be evaluated in the next section. As we will see, it is precisely the strong sensitivity on the bias that will allow us to determine the bias itself (as in [18]).

Let us consider a given time interval $(t_A,\ t_B)$. Let us denote by R_A, V_A and O_A, the values of R, V and O at time t_A. Our goal is to obtain R_A, V_A and O_A in closed-form, only in terms of the measurements provided during the considered time interval. Note that, the length of the considered time interval (i.e., $t_B - t_A$) is very small (4 seconds). We assume that, during our time interval, the camera performs n observations at the times t_j, $(j = 1, \cdots, n)$, with $t_1 = t_A$ and $t_n = t_B$.

Let us denote by $M^1(t)$ and $M^2(t)$ the orthonormal matrices that characterize the rotations made by the first and the second MAV, respectively, between t_A and $t \in (t_A,\ t_B)$. $M^1(t)$ and $M^2(t)$ can be computed by integrating the following first order differential equations:

$$\dot{M}_1 = \left[\Omega^1\right]^T_\times M_1 \quad \dot{M}_2 = \left[\Omega^2\right]^T_\times M_2 \tag{22}$$

with initial conditions: $M_1(t_A) = M_2(t_A) = I_3$, ($I_3$ is the 3×3 identity matrix) and $\left[\Omega^1\right]_\times$ and $\left[\Omega^2\right]_\times$ are the matrices defined in (16). Note that, since $t_B - t_A$ does not exceed 4 seconds, these two matrices can be obtained with very high accuracy by using the measurements from the gyroscopes delivered in the considered time interval. In particular, the drift due to the noise in the gyroscope measurements is negligible. Regarding the bias, in Section 8 we will show that it can be removed.

Let us introduce the following two vector quantities:

$$\beta^1(t) = \int_{t_A}^{t}\int_{t_A}^{\tau} M^1(\tau')A^1(\tau')d\tau' d\tau \tag{23}$$

$$\beta^2(t) = \int_{t_A}^{t}\int_{t_A}^{\tau} M^2(\tau')A^2(\tau')d\tau' d\tau$$

Note that these vectors are computed by only using the IMU measurements delivered in the interval (t_A, t). In particular, the matrices $M^{1,2}(\tau')$ are obtained by integrating the differential equations in (22) in the interval (t_A, τ'), and this only requires the gyroscope measurements in this interval.

Now we are ready to provide the extension of the closed-form solution in [32] to the cooperative case. We

obtain the components of R_A, V_A and O_A by simply solving the linear system:

$$\Xi x = b \tag{24}$$

where:

- Ξ is a matrix with dimension $3n \times (15+n)$ given in (27) (top of next page), with:
  - 0_{33} the 3×3 zero matrix, 0_3 the zero 3×1 vector.
  - $\mu_1, \cdots, \mu_n$ the unit vectors provided by the camera (i.e., the directions of the second MAV in the frame of the first MAV at times $t_1, \cdots, t_n$) rotated by pre-multiplying them by the matrix $M^1(t_j)^T$.
  - $\Delta_j \equiv t_j - t_1 = t_j - t_A$ $(j = 2, \cdots, n)$.
  - β^2_{xj}, β^2_{yj} and β^2_{zj} $(j = 2, \cdots, n)$ the three components of the vector $\beta^2(t_j)$.
- x is the vector that contains all the unknowns, i.e.:

$$x \equiv [R_A^T, V_A^T, O_{A_{11}}, O_{A_{21}}, O_{A_{31}}, O_{A_{12}}, O_{A_{22}}, O_{A_{32}}, O_{A_{13}}, O_{A_{23}}, O_{A_{33}}, \lambda_1, \cdots, \lambda_n]^T \tag{25}$$

  where $\lambda_1, \cdots, \lambda_n$, are the distances between the two MAVs at the times $t_1, \cdots, t_n$.
- b is a vector with dimension $3n$:

$$b \equiv [{\beta_1^1}^T, {\beta_2^1}^T, \cdots, {\beta_j^1}^T, \cdots, {\beta_n^1}^T]^T \tag{26}$$

  where $\beta_j^1 = \beta^1(t_j)$, with $j = 1, \cdots, n$.

In the case when both the MAVs are equipped with a camera and the observations are synchronized (i.e., both the cameras return the direction of the other MAV at the same times $t_1, \cdots, t_n$), the solution is given always by solving the linear system in (24). The new expressions of Ξ, x and b are available in [34].

We conclude this section by remarking that all the components of the matrix Ξ and the vector b depend only on the measurements from the IMUs and the camera delivered during the time interval $(t_A,\ t_B)$. As a result, the solution is able to obtain the entire observable state without any prior knowledge (e.g., initialization). In particular, it provides the state as a simple expression of the measurements delivered during the time interval $(t_A,\ t_B)$.

It is also worth remarking that the communication needed between the two MAVs is very limited. Specifically, if MAV 2 performs the implementation, MAV 1 must provide the quantities β^1 in (23) and the unit vectors μ previously defined. The crucial advantage of this solution is that such data exchange is not required at the high frequency of the inertial sensors: only the β^1 and μ at the times of the camera images used are necessary (over a period of 3-4 seconds less than ten images are sufficient).

Finally, note that we include in x all the entries of the matrix O_A. This means that, by obtaining the state through the inversion of the linear system in (24), we are considering independent the entries of O_A. The fact that the matrix O_A is orthonormal, means that we are ignoring 6 quadratic equations, i.e., the equations that express the fact that each column of the matrix is a unit vector (3 equations) and that the three columns are orthogonal one each other (3 equations). We are currently working on this important issue. We need to define a new state that only includes independent components. Then, by using the results obtained in this section (and in appendix B) it is possible to obtain a new equations system, which will be different from the one in (24). In particular, we already found that, by a suitable choice of the new state and some analytic compuation, instead of a linear system, the new equations system will include three polynomial equations of second degree and several linear equations.

7 Limitations of the Closed-Form Solution

The goal of this section is to find out the limitations of the solution provided in section 6 when it is adopted in a real scenario. In particular, special attention will be devoted to the case of a MAV equipped with low-cost camera and IMU sensors. For this reason, this section evaluates the impact of the following sources of error on the performance:

1. Varying noise on the camera and inertial measurements (section 7.2);
2. Erroneous camera extrinsic calibration, i.e., imperfect knowledge of the transformation between the camera and the IMU frame (section 7.3);
3. Erroneous synchronization between the two cameras (section 7.4);
4. Bias on the accelerometers (section 7.5);
5. Bias on the gyroscopes (section 7.6).

7.1 Simulation setup

We simulate two MAVs that execute random trajectories. Specifically, the trajectories are simulated as follows. Each trial lasts 4 s. The first MAV starts at the origin and the second MAV starts at a random position, normally distributed, centered at the origin, and with covariance matrix 1 $m^2 I_3$. The initial velocities are randomly generated. Specifically, their values are normally distributed, with zero mean, and covariance matrix 1 $(m/s)^2 I_3$. Finally, the initial orientations are characterized by the roll, pitch and yaw angles. These

$$\Xi = \left[\begin{array}{c|c|c|c|c|c|c|c|c|c|c} I_3 & 0_{33} & 0_{33} & 0_{33} & 0_{33} & -\mu_1 & 0_3 & \cdots & \cdots & \cdots & 0_3 \\ I_3 & \Delta_2 I_3 & \beta^2_{x2} I_3 & \beta^2_{y2} I_3 & \beta^2_{z2} I_3 & 0_3 & -\mu_2 & 0_3 & \cdots & \cdots & 0_3 \\ I_3 & \Delta_3 I_3 & \beta^2_{x3} I_3 & \beta^2_{y3} I_3 & \beta^2_{z3} I_3 & 0_3 & 0_3 & -\mu_3 & 0_3 & \cdots & 0_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ I_3 & \Delta_j I_3 & \beta^2_{xj} I_3 & \beta^2_{yj} I_3 & \beta^2_{zj} I_3 & 0_3 & \cdots & 0_3 & -\mu_j & 0_3 & 0_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ I_3 & \Delta_{n-1} I_3 & \beta^2_{x\,n-1} I_3 & \beta^2_{y\,n-1} I_3 & \beta^2_{z\,n-1} I_3 & 0_3 & \cdots & \cdots & \cdots & 0_3 & -\mu_{n-1} \\ I_3 & \Delta_n I_3 & \beta^2_{xn} I_3 & \beta^2_{yn} I_3 & \beta^2_{zn} I_3 & 0_3 & \cdots & \cdots & \cdots & 0_3 & -\mu_n \end{array}\right] \quad (27)$$

are also randomly generated, with zero mean and covariance matrix $(50\ deg)^2 I_3$.

The angular speeds, i.e. Ω^1 and Ω^2, are Gaussian. Specifically, their values at each step of $0.1s$ follow a zero-mean Gaussian distribution with covariance matrix equal to $(30deg)^2 I_3$, where I_3 is the identity 3×3 matrix. At each time step, the two MAV inertial accelerations are generated as random vectors with zero-mean Gaussian distribution. In particular, the covariance matrix of this distribution is set equal to $(1ms^{-2})^2 I_3$.

The MAVs are equipped with inertial sensors able to measure at the frequency of $0.5kHz$ the acceleration (the sum of the gravity and the inertial acceleration) and the angular speed. These measurements are affected by errors. Specifically, each measurement is generated by adding to the true value a random error that follows a Gaussian distribution. The mean value of this error is zero. The standard deviation will be denoted by σ_{Accel} for the accelerometer and σ_{Gyro} for the gyroscope (these values will be specified for each result). Regarding the camera measurements, they are generated at a lower frequency. Specifically, the measurements are generated at $5Hz$. Also these measurements are affected by errors. Specifically, each measurement is generated by adding to the true value a random error that follows a zero-mean Gaussian distribution, with standard deviation σ_{Cam}.

To evaluate the performance, we define the metrics as follows. The error on the absolute scale is defined as the relative error averaged over all the estimated distances between the MAVs at the times of the camera measurements $(t_1, \cdots, t_n)$, i.e.:

$$Err_{scale} \triangleq \frac{1}{n} \sum_{i=1}^{n} \frac{|\lambda_i^{est} - \lambda_i^{true}|}{\lambda_i^{true}}$$

For the speed, the error is defined as

$$Err_V \triangleq \frac{\|V_A^{est} - V_A^{true}\|}{\|V_A^{true}\|}$$

Finally, for the relative orientation, the error is computed by averaging on the roll pitch and yaw, that define the relative rotation between the two local frames.

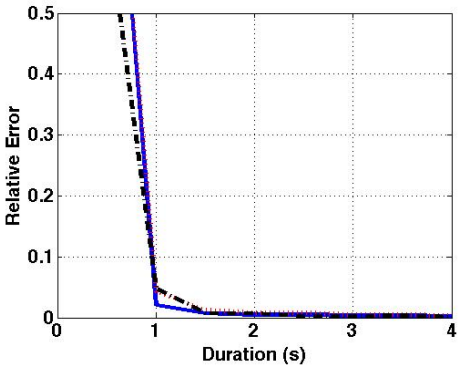


Fig. 2 Relative error of the closed-form solution in determining the absolute scale (solid blue), the relative speed (dotted red) and the relative orientation (dashed black). The two agents observe one each other over a variable duration of integration. $\sigma_{Accel} = 0.03\ ms^{-2}$ and $\sigma_{Gyro} = 0.1\ deg\ s^{-1}$. All the values are averaged on 1000 trials.

In the next subsections, we will present the results obtained with the closed-form solution provided in Section 6 on the simulated data. In section 8, we introduce a simple method to autocalibrate the bias.

7.2 Performance with varying sensor noise

In Fig. 2, we show the performance of the Closed-Form solution in estimating absolute scale, relative speed and relative orientation. The performance is given as a function of the duration of the time interval $(t_B - t_A)$. In this case, the sensor noise is set as follows: $\sigma_{Accel} = 0.03\ ms^{-2}$, $\sigma_{Gyro} = 0.1\ deg\ s^{-1}$ and $\sigma_{Cam} = 1\ deg$. All the values are averaged over 1000 trials. From the results, it is clear how the evaluations improve as we increase the duration of the integration time.

Fig. 3 displays the relative error for the same quantities showed in Fig. 2 but for a variable noise on the inertial measurements. Specifically, $\sigma_{Accel} = (s \cdot 0.03)\ ms^{-2}$, $\sigma_{Gyro} = (s \cdot 0.1)\ deg\ s^{-1}$ and $\sigma_{Cam} = (s \cdot 0.5)\ deg$. In this case, the two agents observe one each other over 3 seconds. The general behavior remains the same. Note that the noise is very large (standard sensors are charac-

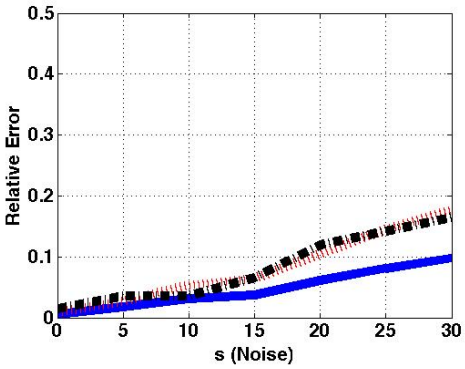


Fig. 3 As in Fig. 2 but for a variable noise on the inertial measurements ($\sigma_{Accel} = (s \cdot 0.03)\ ms^{-2}$, $\sigma_{Gyro} = (s \cdot 0.1)\ deg\ s^{-1}$ and $\sigma_{Cam} = (s \cdot .5)\ deg$). The two agents observe one each other over 3 seconds.

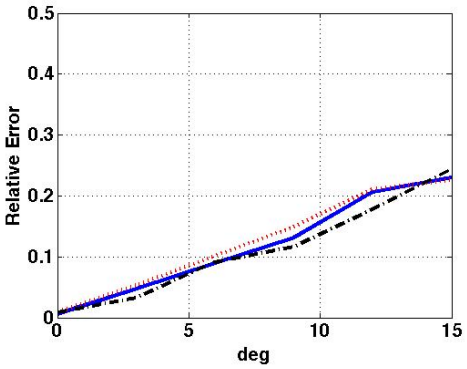


Fig. 4 As in Fig. 2 but for a variable angle between the camera and the IMU frame (of both MAVs). The two agents observe one each other over 3 seconds.

terized by $s \simeq 1$). The performance remains very good also for very large noise.

7.3 Performance with imperfect camera extrinsic calibration

Figures 4 and 5 display the relative error for the same quantities showed in Fig. 3 but for a variable error in the camera extrinsic calibration. Specifically, Fig. 4 displays the results when the camera and the IMU frames are not perfectly aligned (this holds for both MAVs). The x-y planes of the two frames make a variable angle and the performance is provided when this angle is in the range $(0,\ 15)\ deg$. We remark that this source of error has the same effect on the scale, on the speed and on the relative orientation. A misalignment of 6 deg produces a 10% relative error.

Fig. 5 displays the results when the origin of the camera frame does not coincide with the origin of the IMU frame. In particular, the origin of the former has coordinates $\rho[1,\ 1,\ 1]/\sqrt{3}$ and the performance is provided when ρ is in the range $(0,\ 0.1)\ m$. We remark that this source of error does not impact the performance on the relative speed and the relative orientation.

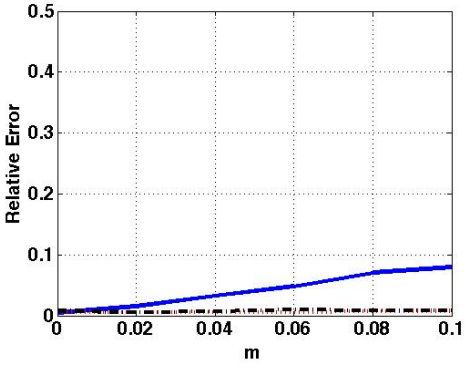


Fig. 5 As in Fig. 2 but for a variable position of the origin of the camera frame in the IMU frame (of both MAVs). The two agents observe one each other over 3 seconds.

7.4 Performance with imperfect synchronization between the two cameras

Fig. 6 displays the relative error for the same quantities showed in Fig. 3 but for a variable synchronization error between the two cameras. Specifically, the measurements of the second camera are generated with a delay of Δt seconds. The performance remains good (relative error smaller than 8%) for $\Delta t \leq 0.02\ s$.

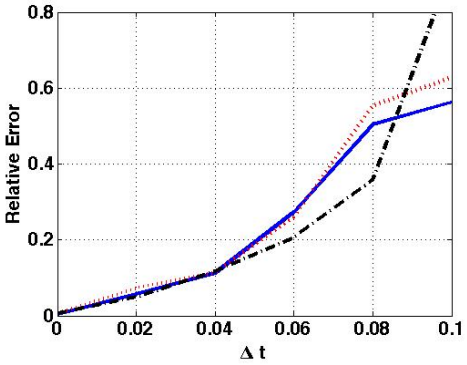


Fig. 6 As in Fig. 2 but for a variable synchronization error (Δt, in seconds). The two agents observe one each other over 3 seconds.

7.5 Impact of accelerometer bias on the performance

In order to visualize the impact of the accelerometer bias on the performance, we corrupt the accelerometer measurements by a bias (Fig. 7). Despite a high accelerometer bias, the closed-form solution still provides good results. Note that, even in the case of a bias with magnitude $0.1ms^{-1}$ (black dashed line in Fig. 7), the error attains its minimum after $1.5s$ and it is less than 3% for the scale and less than 10% for the relative speed (note that, the larger error on the speed is due to its smaller absolute value).

7.6 Impact of gyroscope bias on the performance

To visualize the impact of the gyroscope bias on the performance, we corrupt the gyroscope measurements by an artificial bias (Fig. 8). As seen in Fig. 8, 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.

8 Estimating the Gyroscope Bias (B^1_Ω and B^2_Ω)

In this section, we propose an optimization approach to estimate the gyroscope bias using the closed-form solution.

Let us consider a given experiment, i.e., a set of inertial and camera measurements obtained during a given time interval $(t_A,\ t_B)$. As shown in section 6 (and in the appendix B), these measurements provide all the ingredients to compute the matrix Ξ and the vector b. By solving the linear system in (24) we compute the vector x. Finally, we compute the residual $\|\Xi x - b\|^2$. We can repeat this procedure by changing the measurements provided by the two gyroscopes and by leaving all the other measurements unaltered. In particular, we subtract from the gyroscope measurements a fixed quantity (i.e., which is constant on the considered time interval). In other words, for each $t \in (t_A,\ t_B)$, we replace $\Omega^1(t)$ and $\Omega^2(t)$ with $\tilde{\Omega}^1(t) \triangleq \Omega^1(t) - B^1_\Omega$ and $\tilde{\Omega}^2(t) \triangleq \Omega^2(t) - B^2_\Omega$, respectively. Then we compute the new matrix Ξ and the new vector b. We solve the new linear system in (24) and we compute the new vector x. Finally, we compute the residual $\|\Xi x - b\|^2$.

In accordance with the above procedure, for a given experiment, we can regard $\|\Xi x - b\|^2$ as a function of the two vectors: B^1_Ω and B^2_Ω. We introduce the following function:

$$Cost(B) = \|\Xi x - b\|^2 \tag{28}$$

with:

- B is a vector with six components, which are the components of the bias of the first and the second gyroscope, i.e.,: $B = [B^1_\Omega,\ B^2_\Omega]$.
- Ξ and b are computed by removing from the measurements provided by the two gyroscopes, the corresponding components of B.

By minimizing this cost function, we recover the gyroscope bias B and the vector x. Note that the minimization is carried out over the six components of B, i.e., the bias of the two gyroscopes. Since this minimization requires an initialization and the cost function is non-convex, the optimization process can be stuck in local minima. However, by running extensive simulations we found that the cost function is convex around the true value of the bias. In addition, even if it is true that the bias can significantly increase with time, it increases quite slowly. By continuously estimating its value, and by initializing the minimization of the cost function with the last estimate of the bias, we always remain in the region where the cost function is convex.

8.1 Performance Overall Evaluation

This section analyzes the performance of the closed-form solution completed with the bias estimator introduced in section 8. The setup is the one described in section 7.1. Also in this case, the results are averaged on 1000 trials. We consider the same five values of the bias of the gyroscopes considered in Fig. 8. Finally, we set the magnitude of the accelerometer bias equal to zero (Fig. 9) and equal to $0.1ms^{-2}$ (Fig. 10). Fig. 9 shows a performance comparable to the one exhibited in Fig. 2. Fig. 10 shows a performance even better than the one exhibited in Fig. 7. This demonstrates that the effect of the bias has been fully compensated.

9 Conclusion

In this paper, we studied the problem of cooperative visual inertial sensor fusion. Specifically, the case of two agents was investigated. Each agent was equipped with inertial sensors (accelerometer and gyroscope) and with a monocular camera. By using the monocular camera, each agent can observe the other agent. Specifically, the camera only returns the position (up to a scale) of the observed agent in its local frame. No additional camera observations (e.g., of external point features in the environment or of known pattern on the observed agent able to directly provide the relative orientation between

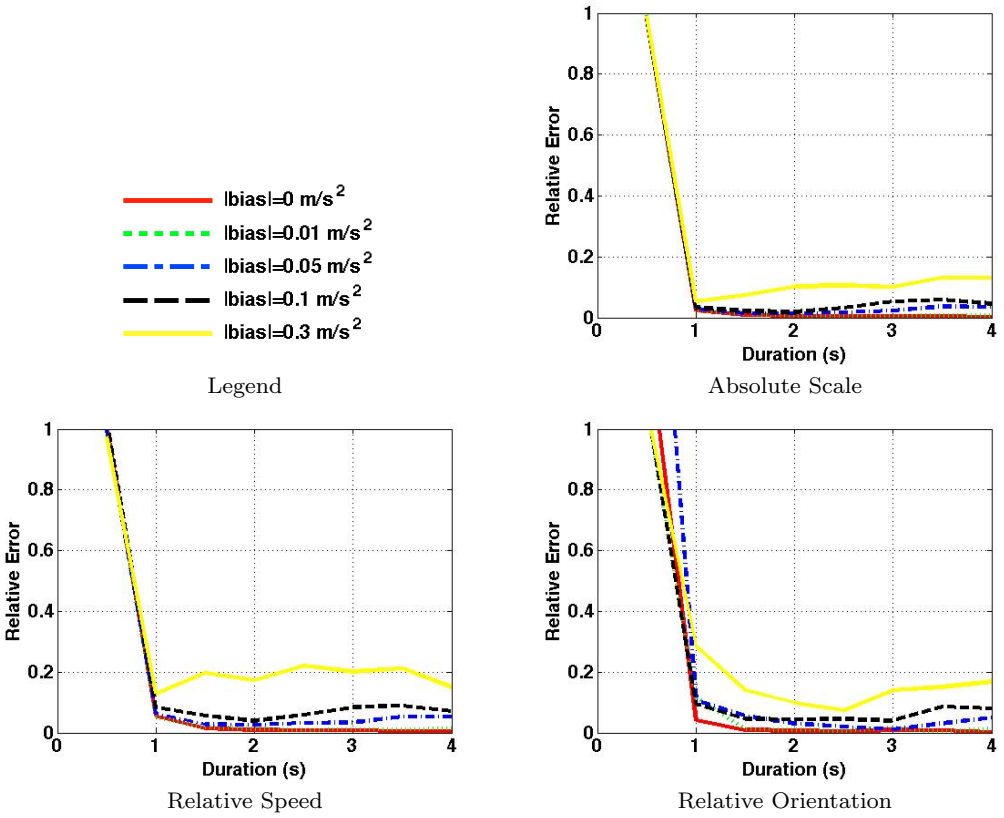


Fig. 7 Impact of the accelerometer bias on the performance of the closed-form solution. The two MAVs observe one each other over a variable duration of integration.

the agents, as in [1]) were considered. All the inertial sensors were assumed to be affected by a bias. First, the entire observable state was analytically derived. To this regard, we proved that the entire observable state consists of the following independent physical quantities:

- The position of one of the agents in the local frame of the other agent (this means that the absolute scale is observable).
- The relative speed between the two agents expressed in the local frame of one of them.
- The three Euler angles that characterize the rotation between the two local frames attached to the two agents.
- All the bias that affect the inertial measurements (both the accelerometers and the gyroscopes).

It is interesting to remark that this result holds even in the case when only one of the two agents is equipped with a camera and, very surprisingly, even when this camera is a linear camera, i.e., it only provides the azimuth of the other agent in its local frame.

Then, the paper provided a closed-form solution, able to determine the observable state by only using visual and inertial measurements delivered in a short time interval (4 seconds). This solution extended the solution in [31,32] to the cooperative case. It is remarkable that it is possible to retrieve the absolute scale even when no point features are available in the environment.

Following the analysis conducted in [18], the paper focused on investigating all the limitations that characterize this solution when used in a real scenario. Specifically, the impact of the presence of the bias on the performance of this closed-form solution was investigated. As in the case of a single agent, this performance is significantly sensitive to the presence of a bias on the gyroscope, while the presence of a bias on the accelerometer is less important. A simple and effective method to obtain the gyroscope bias was proposed. Extensive simulations clearly showed that the proposed method is successful. It is fascinating that it is possible to automatically retrieve the absolute scale and simultaneously calibrate the gyroscopes not only without any prior knowledge (as in [18] for a single agent), but also without external point features in the environment.

Future works will be focused on additional theoretical investigation. This will include the study of the case of more than two agents. This study requires to address

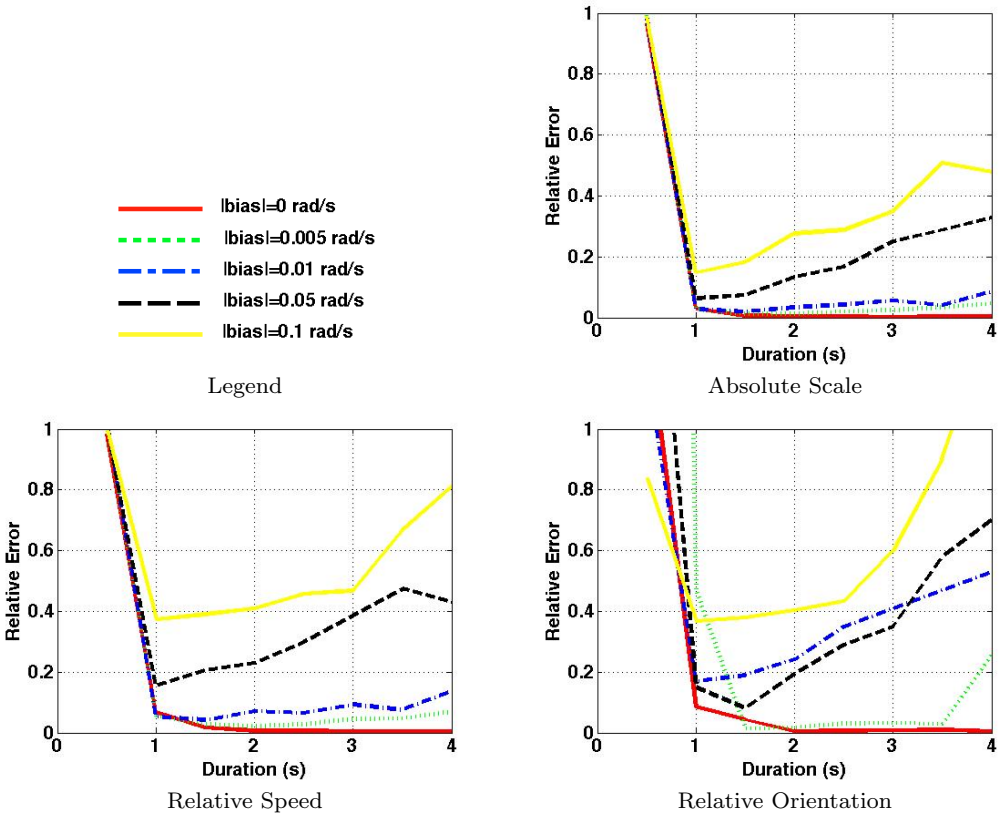


Fig. 8 Impact of the gyroscope bias on the performance of the closed-form solution. The two MAVs observe one each other over a variable duration of integration.

several important issues. In the case of two agents, to obtain the linear system, we had to consider a state whose components are not independent (e.g., all the nine entries of the rotation matrix are included in the state). In the case of more than two agents, obtaining a linear system by minimizing the redundancy in the state, is a first issue to be investigated. Note that, as mentioned at the end of section 6, our objective to be achieved firstly in the case of two agents, is to consider a state whose components are independent and obtaining the equation system that characterizes the problem. This equation system will be a Polynomial Equation System (PES), instead of a linear system. So far, we already found a partial solution to this problem. The PES consists of three polynomial equations in three unknowns and several linear equations (this provides up to eight solutions in the minimal case). The analysis of this PES, that fully characterizes the problem, provides all the theoretical features of the problem. This analysis is currently under our investigation and will be the extension of the analysis provided in [32] for the case of a single agent (in that case the PES consists of a single polynomial equation of second degree and several linear equations). In the case of more than two agents, we expect that the PES becomes much more complex and this issue certainly deserves to be investigated. From a practical point view, there are many issues to be considered in the case of more than two agents. The visual constraint due to the limited camera field of view becomes more important. In particular, the new issue to be investigated is how the performance changes by varying the number of agents that can be seen from each agent. In addition, the cameras synchronization is harder to be realized in the case of many agents. Finally, the problem of communication delays and how robust is the solution vs communication troubles becomes certainly more relevant.

A Observability with bias

We analytically obtain the observability properties of the system defined by the state in (17), the dynamics in (19) (where the last equation has been replaced by $\dot{B}^1_\Omega = \dot{B}^1_A = \dot{B}^2_\Omega = \dot{B}^2_A = 0$) and the three observations in (20) and (21). Since both the dynamics and the observations are nonlinear with respect to the state, we use the observability rank condition in [8]. The dynamics are affine in the inputs, i.e., they have the expression given in (9). Specifically, we set:

– u_1, u_2, u_3 the three components of A^1;

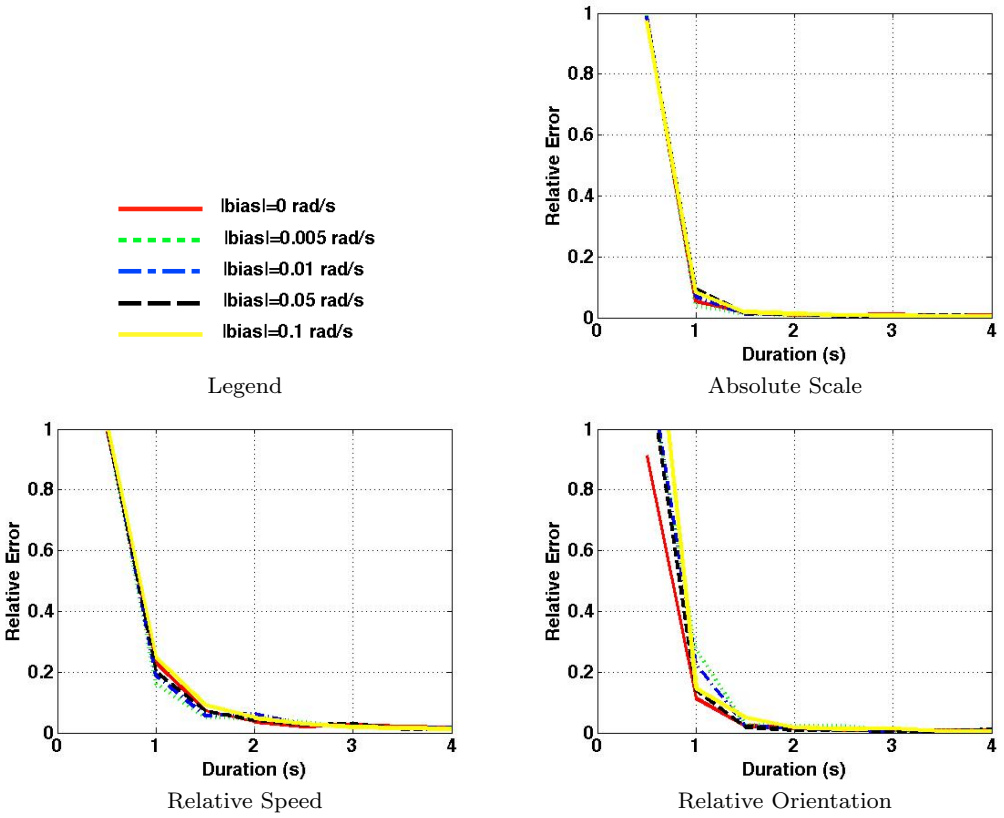


Fig. 9 Impact of the gyroscope bias on the performance of the closed-form solution completed with the bias estimator. The accelerometers are unbiased. The two MAVs observe one each other over a variable duration of integration.

- u_4, u_5, u_6 the three components of Ω^1;
- u_7, u_8, u_9 the three components of A^2;
- u_{10}, u_{11}, u_{12} the three components of Ω^2.

Then, by comparing (19) with (9) it is immediate to obtain the analytic expression of all the vector fields $f_0, f_1, \cdots, f_{12}$; for instance, we have:

$$f_4 = [0, -R_z, R_y, 0, -V_z, V_y, q_x/2, -q_t/2, q_z/2, -q_y/2, 0_{12}]^T$$

$$f_7 = [0, 0, 0, q_t^2 + q_x^2 - q_y^2 - q_z^2, 2q_tq_z + 2q_xq_y, 2q_xq_z - 2q_tq_y, 0, 0, 0, 0, 0_{12}]^T$$

For systems with the dynamics given in (9) the application of the observability rank condition can be automatically done by a recursive algorithm (Algorithm 1). In particular, this algorithm automatically returns the observable codistribution by computing the Lie derivatives of all the system outputs along all the vector fields that characterize the dynamics (see Chapter 1 of [15]). For the specific case, we obtain that the algorithm converges at the fourth step, i.e., the observable codistribution is the span of the differentials of the previous Lie derivatives up to third order. In particular, its dimension is 22 meaning that all the state components are observable.

A choice of 22 Lie derivatives is:

$$\mathcal{L}^0 h_u,\ \mathcal{L}^0 h_{const},\ \mathcal{L}^1_{f_0} h_u,\ \mathcal{L}^1_{f_4} h_u,\ \mathcal{L}^2_{f_0 f_0} h_u,\ \mathcal{L}^2_{f_0 f_1} h_u,$$
$$\mathcal{L}^2_{f_0 f_7} h_u,\ \mathcal{L}^2_{f_0 f_8} h_u,\ \mathcal{L}^2_{f_0 f_4} h_u,\ \mathcal{L}^2_{f_0 f_5},\ \mathcal{L}^2_{f_0 f_6} h_u,\ \mathcal{L}^3_{f_0 f_0 f_0} h_u,$$
$$\mathcal{L}^3_{f_0 f_0 f_1} h_u,\ \mathcal{L}^3_{f_0 f_0 f_2} h_u,\ \mathcal{L}^3_{f_0 f_0 f_7} h_u,\ \mathcal{L}^3_{f_0 f_0 f_4} h_u,\ \mathcal{L}^3_{f_0 f_0 f_5} h_u,$$
$$\mathcal{L}^3_{f_0 f_0 f_{10}} h_u,\ \mathcal{L}^3_{f_0 f_7 f_0} h_u,\ \mathcal{L}^3_{f_0 f_8 f_0} h_u,\ \mathcal{L}^3_{f_0 f_0 f_0} h_v,\ \mathcal{L}^3_{f_0 f_0 f_5} h_v\,.$$

Note that the choice of these 22 independent Lie derivatives is not unique. In particular, it is possible to avoid the Lie derivatives of the functions h_v. Specifically, in the previous choice, only the last two Lie derivatives are Lie derivatives of the function h_v. It is possible to avoid these two functions. On the other hand, in this case we need to include fourth order Lie derivatives of h_u. For instance, we can replace the last two functions with $\mathcal{L}^4_{f_0 f_0 f_0 f_0} h_u$, $\mathcal{L}^4_{f_0 f_0 f_0 f_5} h_u$. This means that we obtain the same observability properties when the first agent is equipped with a linear camera able to only provide the azimuth of the second agent in its local frame. Finally, as in the unbiased case, a necessary condition to have 22 independent Lie derivatives is that at least one of them must be computed along a direction that corresponds to one of the axes of at least one of the two accelerometers (note that in the above selection we have Lie derivatives computed along f_1, f_2, f_7 and f_8). This means that a necessary condition for the observability of the absolute scale is that the relative acceleration between the two MAVs does not vanish.

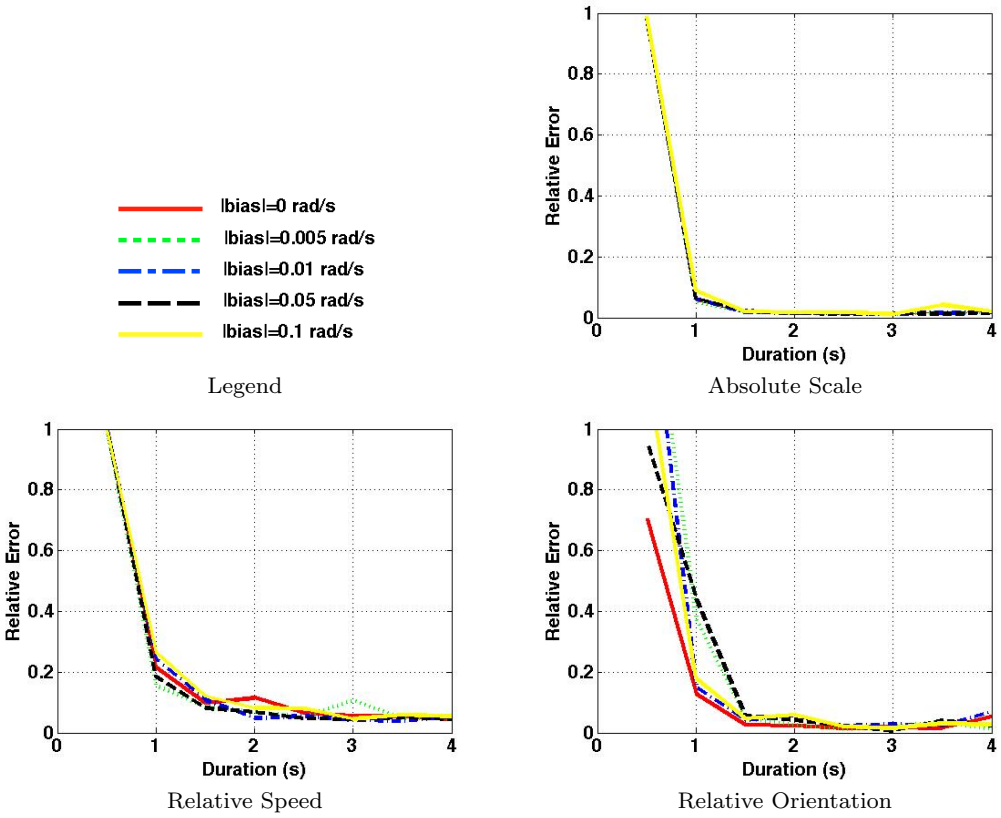


Fig. 10 As in Fig. 9 but the magnitude of the accelerometer bias is set to $0.1ms^{-2}$ for both the MAVs.

B Analytic derivation of the closed-form solution

In this appendix we provide the analytic steps to obtain the linear system given in section 6 (equation (24)). For the brevity sake, we only consider the case of a single camera. Specifically, we assume that only the first vehicle is equipped with a camera. The derivation in the case of two synchronized cameras follows the same schema and is available in [34].

We start our derivation by introducing a new local frame for each vehicle. Each new frame is defined as follows. It shares the same origin with the original local frame. Additionally, it does not rotate and its orientation coincides with the one of the original frame at the time t_A. From now on, we will refer to this frame as to the *new* frame. Additionally, we will refer to the original local frame, namely the one defined at the beginning of section 2, as to the *original* frame.
Let us introduce the following notation:

- $\mathcal{V}_1$ and $\mathcal{V}_2$ denote the first and the second vehicle;
- ξ is the position of $\mathcal{V}_2$ in the new local frame of $\mathcal{V}_1$;
- η is the relative velocity of $\mathcal{V}_2$ with respect to $\mathcal{V}_1$, expressed in the new local frame of $\mathcal{V}_1$;

By construction we have:

$$\xi_A \equiv \xi(t_A) = R_A \qquad \eta_A \equiv \eta(t_A) = V_A \tag{29}$$

From (15) we have the following dynamics in the new coordinates:

$$\begin{cases} \dot{\xi} &= \eta \\ \dot{\eta} &= O_A \mathcal{A}^2 - \mathcal{A}^1 \\ \dot{O}_A &= 0 \end{cases} \tag{30}$$

where:

- $\mathcal{A}^1$ is the acceleration (gravitational and inertial) of $\mathcal{V}_1$ expressed in the first new local frame (i.e., $\mathcal{A}^1 = M^1 A^1$);
- similarly, $\mathcal{A}^2 = M^2 A^2$.

Let us introduce the following notation:

- w^1, w^2 and w^3 are the three columns of the matrix O_A, i.e., $O_A = \begin{bmatrix} w^1 & w^2 & w^3 \end{bmatrix}$;
- $\alpha^1(t) = [\alpha^1_x(t),\ \alpha^1_y(t),\ \alpha^1_z(t)]^T = \int_{t_A}^t \mathcal{A}^1(\tau) d\tau$;
- $\alpha^2(t) = [\alpha^2_x(t),\ \alpha^2_y(t),\ \alpha^2_z(t)]^T = \int_{t_A}^t \mathcal{A}^2(\tau) d\tau$.

Note that the quantities $\beta^1(t)$ and $\beta^2(t)$ defined in section 6 are $\beta^1(t) = \int_{t_A}^t \alpha^1(\tau) d\tau$ and $\beta^2(t) = \int_{t_A}^t \alpha^2(\tau) d\tau$.

Let us integrate the second equation in (30) between t_A and a given $t \in [t_A, t_B]$. We obtain:

$$\eta(t) = \eta_A + w^1 \alpha^2_x(t) + w^2 \alpha^2_y(t) + w^3 \alpha^2_z(t) - \alpha^1(t) \tag{31}$$

and by substituting in the first equation in (30) and integrating again, we obtain:

$$\xi(t) = \xi_A + \eta_A (t - t_A) + w^1 \beta^2_x(t) + w^2 \beta^2_y(t) + w^3 \beta^2_z(t) - \beta^1(t) \tag{32}$$

Note that this equation provides $\xi(t)$ as a linear expression of 15 unknowns, which are the components of the 5 vectors: ξ_A, η_A, w^1, w^2 and w^3. In the following, we build a linear system in these unknowns together with the unknown distances when the camera performs the measurements.

The camera (on $\mathcal{V}_1$) provides the vector $R(t) = M^1(t)\xi(t)$, up to a scale. We denote by $\lambda(t)$ this scale (this is the distance between $\mathcal{V}_1$ and $\mathcal{V}_2$ at the time t). We have $\xi(t) = \lambda(t)\mu(t)$, where $\mu(t)$ is the unit vector with the same direction of $\xi(t)$. Note that our sensors (specifically, the camera together with the gyroscope on $\mathcal{V}_1$) provide precisely the unit vector $\mu(t)$: the camera provides the unit vector along $R(t)$; then, to obtain $\mu(t)$ it suffices to pre multiply this unit vector by $[M^1(t)]^T$.

We remind the reader that the camera performs n observations at the times t_j, $(j = 1, \cdots, n)$, with $t_1 = t_A$ and $t_n = t_B$. For notation brevity, for a given time dependent quantity (e.g., $\lambda(t)$), we will denote its value a the time t_j by the subscript j (e.g., $\lambda_j = \lambda(t_j)$). In this notation, equation (32) becomes:

$$\lambda_j \mu_j = \xi_A + \eta_A(t_j - t_A) + w^1\beta^2_{xj} + w^2\beta^2_{yj} + w^3\beta^2_{zj} - \beta^1_j \quad (33)$$

This is a linear equation in $15 + n$ unknowns. The unknowns are:

- The distances $\lambda_1, \cdots, \lambda_n$.
- The three components of ξ_A.
- The three components of η_A.
- The components of the vectors w_1, w_2 and w_3, i.e., the nine entries of the matrix O_A.

Note that equation (33) is a vector equations, providing 3 scalar equations. Since this holds for each $j = 1, \cdots, n$, we obtain a linear system of $3n$ equations in $15 + n$ unknowns. This is precisely the linear system given in (24) with the vector x given in (25), the matrix A given in (27) and the vector b given in (26).

Acknowledgment

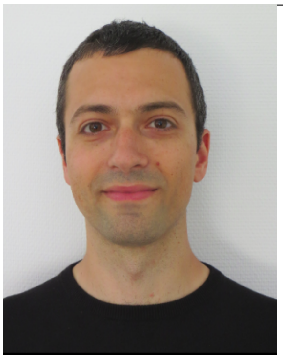
This work was supported by the French National Research Agency ANR 2014 through the project VIMAD.

References

1. Achtelik, M., Weiss, S., Chli, M., Dellaert, F., & Siegwart, R. (2011). Collaborative stereo. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2242-2248.
2. L. Armesto, J. Tornero and M. Vincze, "Fast ego-motion estimation with multi-rate fusion of inertial and vision," The International Journal of Robotics Research (IJRR), vol. 26, no. 6, pp. 577-589, 2007.
3. L. C. Carrillo-Arce, E. D. Nerurkar, J. L. Gordillo and S. I. Roumeliotis, "Decentralized multi-robot cooperative localization using covariance intersection," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1412-1417, 2013.
4. C. Forster, M. Pizzoli and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in IEEE International Conference on Robotics and Automation (ICRA), 2014.
5. C. Forster, L. Carlone, F. Dellaert and D. Scaramuzza, "IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation," in Robotics: Science and Systems (RSS), 2015.
6. R. Grabowski, L.E. Navarro-Serment, C.J.J. Paredis, P.K. Khosla, "Heterogeneous Teams of Modular Robots for Mapping and Exploration," Autonomous Robots, Vol. 8, no. 3, pp. 293–308, 2000.
7. H. X. Guo, K. Sartipi, R. C. DuToit, G. A. Georgiou, R. Li, J. OLeary, E. D. Nerurkar, J. A. Hesch and S. I. Roumeliotis, "Large-Scale Cooperative 3D Visual-Inertial Mapping in a Manhattan World," in IEEE International Conference on Robotics and Automation (ICRA), 2016.
8. R. Hermann and A.J. Krener, "Nonlinear Controllability and Observability," IEEE Transaction On Automatic Control, vol. 22, no. 5, pp. 728–740, 1977.
9. J. Hesch, D. Kottas, S. Bowman and S. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," IEEE Transactions on Robotics, vol. 30, no. 1, pp. 158-176, 2014.
10. A. Howard, M. Mataric and G. Sukhatme, "Localization for Mobile Robot Teams Using Maximum Likelihood Estimation," in International Conference on Intelligent Robots and Systems (IROS), 2002.
11. G. P. Huang, A. I. Mourikis and S. I. Roumeliotis, "On the complexity and consistency of UKF-based SLAM," in International Conference on Robotics and Automation (ICRA), 2009.
12. G. P. Huang, A. Mourikis, S. Roumeliotis et al., "An observability-constrained sliding window filter for slam," in International Conference on Intelligent Robots and Systems (IROS), 2011.
13. G. Huang, M. Kaess and J. Leonard, "Towards consistent visual-inertial navigation," in International Conference on Robotics and Automation (ICRA), 2015.
14. V. Indelman, P. Gurfil, E. Rivlin and H. Rotstein, "Graph-based distributed cooperative navigation for a general multi-robot measurement model," The International Journal of Robotics Research, vol. 31, no. 9, pp. 1057–1080, 2012.
15. A. Isidori, "Nonlinear Control Systems," 3rd ed., Springer Verlag, London, 1995.
16. V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," Robotics and Autonomous Systems, pp. 721-738, 2013.
17. E. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," The International Journal of Robotics Research, vol. 30, no. 4, pp. 407–430, 2011.
18. J. Kaiser, A. Martinelli, F. Fontana and D. Scaramuzza, "Simultaneous State Initialization and Gyroscope Bias Calibration in Visual Inertial Aided Navigation," IEEE Robotics and Automation Letters, vol. 2, no. 1, 2017.
19. K. Kato, H. Ishiguro and M. Barth, "Identifying and Localizing Robots in a Multi-Robot System Environment," in International Conference on Intelligent Robots and Systems (IROS), 1999.
20. S. Kia, S. Rounds, and S. Martinez, "Cooperative localization for mobile agents: a recursive decentralized algorithm based on Kalman filter decoupling," IEEE Control Systems Magazine, vol. 36, no. 2, pp. 86-101, 2016.
21. B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in International Conference on Robotics and Automation (ICRA), 2010.

22. K. Y. Leung, T. Barfoot and H. Liu, "Decentralized localization of sparsely-communicating robot networks: A centralized equivalent approach," IEEE Transactions on Robotics, vol. 26, no. 1, pp. 62-77, 2010.
23. S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige and R. Siegwart, "Keyframe-based visual-inertial odometry using nonlinear optimization," International Journal of Robotics Research (IJRR), 2014.
24. M. Li and A. I. Mourikis, "High-precision, consistent EKF based visual-inertial odometry," The International Journal of Robotics Research (IJRR), vol. 32, no. 6, pp. 690-711, 2013.
25. H. Li and F. Nashashibi, "Cooperative multi-vehicle localization using split covariance intersection filter," IEEE Intelligent Transportation Systems Magazine, vol. 5, no. 2, pp. 33-44, 2013.
26. T. Liu and S. Shen, "Spline-Based Initialization of Monocular Visual-Inertial State Estimators at High Altitude," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017.
27. L. Luft, T. Schubert, S. I. Roumeliotis and W. Burgard, "Recursive decentralized collaborative localization for sparsely communicating robots," in Robotics: Science and Systems (RSS), 2016.
28. T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," IEEE Transactions on Robotics, vol. 28, no. 1, pp. 61-76, 2012.
29. A. Martinelli, F. Pont and R. Siegwart, "Multi-Robot Localization Using Relative Observations," in International Conference on Robotics and (ICRA), 2005.
30. A. Martinelli, "State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration," IEEE Transactions on Robotics, vol. 27, no. 2, pp. 239–255, 2011
31. A. Martinelli, "Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," IEEE Transactions on Robotics, vol. 28, no. 1, pp. 44–60, 2012.
32. A. Martinelli, "Closed-form solution of visual-inertial structure from motion," International Journal of Computer Vision (IJCV), vol. 106, no. 2, pp. 138–152, 2014.
33. A. Martinelli and A. Renzaglia, "Cooperative Visual-Inertial Sensor Fusion: Fundamental Equations," in IEEE International Symposium on Multi-Robot and Multi-Agent Systems (MRS), 2017.
34. A.Martinelli, "Closed-form Solution to Cooperative Visual-Inertial Structure from Motion," arXiv:1802.08515 [cs.RO].
35. G. Michieletto, A. Cenedese and A. Franchi, "Bearing rigidity theory in SE (3)," in IEEE Conference on Decision and Control (CDC), 2016.
36. A. Mourikis, S. Roumeliotis, et al., "A multi-state constraint Kalman filter for vision-aided inertial navigation," in International Conference on Robotics and Automation (ICRA), 2007.
37. A. Mourikis, S. Roumeliotis, et al., "A dual-layer estimator architecture for long-term localization," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2008.
38. J. Mustaniemi, J. Kannala, S. Sarkka, J. Matas and J. Heikkila, "Inertial-Based Scale Estimation for Structure from Motion on Mobile Devices," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017.
39. G. Philippe, I. Rekleitis, and M. Latulippe, "I see you, you see me: Cooperative localization through bearing-only mutually observing robots," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012.
40. I. Rekleitis, G. Dudek and E. Milios, "Multi-robot cooperative localization: a study of trade-offs between efficiency and accuracy," in International Conference on Intelligent Robots and Systems (IROS), 2002.
41. S.I. Roumeliotis and G.A. Bekey, "Distributed Multi-robot Localization," IEEE Transaction On Robotics And Automation, vol. 18, no.5, pp. 781–795, 2002.
42. J.R. Spletzer and C.J. Taylor, "A Bounded Uncertainty Approach to Multi-Robot Localization," in International Conference on Intelligent Robots and Systems (IROS), 2003.
43. N. Trawny, S. I. Roumeliotis, and G. B. Giannakis, "Cooperative multi-robot localization under communication constraints," in IEEE International Conference on Robotics and Automation (ICRA), 2009.
44. R. Tron, L. Carlone, F. Dellaert and K. Daniilidis, "Rigid Components Identification and Rigidity Enforcement in Bearing-Only Localization using the Graph Cycle Basis," in American Control Conference (ACC), 2015.
45. D. Zelazo, P.R. Giordano and A. Franchi, "Bearing-only formation control using an SE(2) rigidity theory," IEEE Conference on Decision and Control (CDC), 2015.
46. X.S. Zhou and S.I. Roumeliotis, "Robot-to-robot relative pose estimation from range measurements," IEEE Transactions on Robotics, vol. 24, no. 6, pp. 1379–1393, 2008.

Agostino Martinelli received the M.Sc. degree in theoretical physics from the University of Rome Tor Vergata, Rome, Italy, in 1994 and the Ph.D. degree in astrophysics from the University of Rome La Sapienza, in 1999. While working toward the Ph.D. degree, he spent one year at the University of Wales, Cardiff, U.K., and one year with the Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy. His research focused on the chemical and dynamical evolution in elliptical galaxies, in quasars, and in the intergalactic medium. After receiving the Ph.D. degree, his interests moved to the problem of autonomous navigation. He was with the University of Rome Tor Vergata for two years, and, in 2002, he moved to the Autonomous Systems Laboratory, Ecole Polytechnique Federale de Lausanne, Switzerland, as a Senior Researcher. Since September 2006, he has been a Researcher with the Institut National de Recherche en Informatique et en Automatique (INRIA), Rhone Alpes, Grenoble, France. His current research focuses on three main topics: (i) Visual-Inertial Structure from Motion, (ii) Non linear observability, (iii) Overdamped Brownian motion and the Fokker-Plank equation without detailed balance.



Alessandro Renzaglia received his M.S. degree in Physics from the University of Rome La Sapienza, Italy, in 2007 and his Ph.D. degree in Computer Science from the University of Grenoble, France, in 2012. He was Postdoctoral Researcher from 2012 to 2014 with the Computer Science & Engineer Department at the University of Minnesota, Minneapolis, USA, and from 2014 to 2016 with the Laboratory for Analysis and Architecture of Systems (LAAS), Toulouse, France. Since 2017 he joined the INRIA Grenoble-Rhône Alpes, France. His research interests include multi-robot systems, path planning and optimization.

Alexander Oliva received his M.S. degree in Computer and Automation Engineering from Universita' degli studi dell'Aquila, Italy, in 2015. After his experience in industry as Embedded Software Engineer, he joins in 2017 the INRIA Grenoble-Rhone Alpes, France, as R&D Engineer. Since October 2018 he is a Ph.D. student in the field of Vision-force control of manipulators at the INRIA Rennes-Bretagne Atlantique. His research interests include Autonomous navigation, sensor fusion, state estimation and sensor-based automatic control.