A linear approach to visuo-inertial fusion for homography-based filtering and estimation

Alexandre Eudes and Pascal Morin

***Abstract*— A solution to visuo-inertial filtering and estimation based on homography, angular velocity, and specific acceleration measurements is proposed. This corresponds to the typical situation of a mono-camera/IMU sensor facing a (locally) planar environment. By lifting the estimation state space to a higher-dimensionnal space, we show that the problem can be formulated as a *linear* estimation problem. This allows for the application of classical estimation techniques, e.g., Kalman filtering. Based on this linear formulation, we also determine explicitly the motion conditions that ensure uniform observability of the system, and we propose a simple linear Luenberger-like observer. A validation of the proposed solution based on real data is presented.**

I. Introduction

For decades, vision has been a major topic of interest in robotics due to its importance in many applications such as visual-servoing and SLAM. Motivated by challenging problems in highly dynamic environments, recent years have seen an increased interest for visuo-inertial fusion, where inertial data provides complementary information, especially in term of bandwith (see, e.g., [7] for an introduction and [16], [14], [12], [19] for recent contributions). The litterature on the subject can be roughly splitted into two categories. In the first one, it is assumed that the cartesian pose can be recovered from vision, as in the case of a stereo-camera system. In the second one, typically associated with a mono-camera system, the cartesian pose is not directly available and the problem partly consists in estimating this pose. Feature-based estimation techniques are then often used, with features positions being estimated together with the pose and other parameters [16], [20]. The present paper also belongs to the second category but it does not rely on features positions estimation. More precisely, by assuming that the sensor is facing a locally planar environment, vision can provide the homography matrix that relates two images of the planar scene (see, e.g., [18] for more details). The problem addressed in this paper is to use IMU measurements (i.e., obtained from gyrometers and accelerometers) to obtain in real-time filtered values of the homography matrix, estimates of the sensor velocity, and estimates of the environment structure (i.e. the so-called Structure from Motion problem). Filtering of the homography matrix is instrumental in improving homography estimation vision algorithms, especially those based on direct methods (see [8] for more details). This is also needed to provide good inputs to homography-based visual-servoing algorithms [2], [13], [17], [4], [24], [9], [25]. Velocity estimates are typically needed in visual-servoing applications (see [11], [12] for related work). As for environment structure estimation, it can be used for both visual-servoing and SLAM.

Existing methods of visual inertial fusion for homography-based filtering and estimation make use of non-linear techniques like the EKF [10], [23], the UKF [10], or nonlinear geometric observers [19], [8]. Kalman-like approaches (EKF, UKF) can provide covariance estimates, but in the nonlinear framework stability can usually not be guaranteed and performance may be poor. Nonlinear geometric observers are of interest for their computational simplicity, large stability domains and explicit stability conditions, but to our knowledge no solution of this type has been proposed in order to incorporate accelerometer measurements. In particular, the solution proposed in [19] can only cope with gyrometers measurements. The solution proposed in this paper departs from those existing methods. It is based on the observation that the problem can be formulated as a *linear* estimation problem by lifting the estimation state space to a higher-dimensionnal space. This observation is as simple as powerful since it allows for the application of linear estimation techniques, e.g. the Kalman filter, which typically come with guarantees of stability in term of the system's uniform observability, and optimality of the filter. As a matter of fact, we derive in this paper conditions for the uniform observability of the system. Concurrently, the linear structure also allows us to derive a very simple Luenberger-like linear observer and to establish its stability assuming the system's uniform observability, as in the case of the Kalman filter.

The paper is organized as follows. Section II provides technical background. Section III presents the main results of the paper, which consist of the problem formulation as a linear estimation problem, the uniform observability analysis, and the design of a linear Luenberger-like observer. Section IV concerns the validation of the proposed observer. The paper ends with some concluding remarks and perspective for future work.

II. Technical background

A. Notation

- $S(x)$ is the antisymmetric matrix associated with the cross product by x, i.e. $S(x)y = x \times y$ with $\times$ the cross product.

The authors are with Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, and CNRS UMR 7222, Paris, France. Emails: eudes@isir.upmc.fr, morin@isir.upmc.fr. The authors have been supported by the "Chaire d'excellence en Robotique RTE-UPMC" and by the French ANR Project SCAR.

- I_n is the $n \times n$ identity matrix and $0_n \in \mathbb{R}^n$ is the null vector.
- $\mathrm{tr}(M)$ is the trace of the matrix M, $\det(M)$ its determinant, and $\|M\| := \sqrt{\mathrm{tr}(M^T M)}$ its norm.

B. Homographies

Consider two images $\mathbf{I}_A$ and $\mathbf{I}_B$ of a planar scene taken by a monocular camera. Each image $\mathbf{I}_*$ ($* \in \{A, B\}$) is taken from a specific pose of the camera and we denote by $\mathcal{F}_*(* \in \{A, B\})$ an associated camera frame with origin corresponding to the optical center of the camera and third basis vector aligned with the optical axis. Furthermore, we denote by d_* and n_* respectively the distance from the origin of $\mathcal{F}_*$ to the planar scene and the normal to the scene expressed in $\mathcal{F}_*$.

Let R denote the rotation matrix from $\mathcal{F}_B$ to $\mathcal{F}_A$ and $p \in \mathbb{R}^3$ the coordinate vector of the origin of $\mathcal{F}_B$ expressed in $\mathcal{F}_A$. The uncalibrated homography matrix is defined as (see, e.g., [18] for details):

$$G \propto KHK^{-1} \tag{1}$$

with

$$H = R^T - \frac{1}{d_A} R^T p n_A^T \tag{2}$$

K the calibration matrix of the camera, and where $\propto$ denotes equality up to a positive scalar factor. The matrix G maps pixel coordinates from $\mathbf{I}_A$ to $\mathbf{I}_B$. If the camera calibration matrix K is known, then the matrix H can be deduced from G up to a scale factor, i.e., $K^{-1}GK = \alpha H$. As a matter of fact, the scale factor α corresponds to the mean singular value of the matrix $K^{-1}GK$: $\alpha = \sigma_2(K^{-1}GK)$ (see, e.g., [18, Pg. 135]). Therefore, α can be computed together with the matrix H.

Another interesting matrix is

$$\bar{H} = \det(H)^{-\frac{1}{3}} H = \eta H \tag{3}$$

Indeed, $\det(\bar{H}) = 1$ so that $\bar{H}$ belongs to the Special Linear Group $SL(3)$.

C. Inertial data

Strapped down IMUs provide the angular velocity vector of the sensor (w.r.t. inertial frame) in body frame ω, and the so-called specific acceleration a_s. More precisely, assuming that $\mathcal{F}_A$ is an inertial frame and $\mathcal{F}_B$ the body frame[1], ω and a_s are defined by the following relations:

$$\dot{R} = RS(\omega)\,, \quad \ddot{p} = g_A + Ra_s \tag{4}$$

where g_A denotes the gravitational acceleration field expressed in $\mathcal{F}_A$. Note that a_s can also be defined by the relation

$$\dot{v} = -S(\omega)v + a_s + g_B \tag{5}$$

where $v = R^T \dot{p}$ denotes the velocity of $\mathcal{F}_B$ w.r.t. $\mathcal{F}_A$ expressed in $\mathcal{F}_B$, and $g_B = R^T g_A$ is the gravitational acceleration field expressed in $\mathcal{F}_B$.

[1]For simplicity it is assumed in this paper that the camera frame and IMU frame coincide. This means in practice that a preliminary calibration has been performed so as to bring visual and inertial measurements in the same frame (see, e.g., [21] for solutions to this latter problem).

D. Problem statement and motivation

Consider a visuo-inertial sensor consisting of an IMU rigidly attached to a monocular camera observing a planar scene. With the above notation, the objective is to fuse the homography measurements obtained from vision with inertial data. Two main types of solutions have been proposed, which differ by the estimation problem formulation. A first approach (see, e.g., [23]) essentially consists in defining the state to be estimated as $x = (R, p, v, n_A, d_A)$, with measurement $y = (H, \omega, a_s)$. Since H is a nonlinear function of x, the estimation problem is nonlinear and one must resort to nonlinear estimation techniques, like e.g. the EKF used in [23]. A second approach [19], [8], essentially consists in setting $x = (\bar{H}, M, n_A)$ where $\bar{H}$ is defined by (3) and $M = v\frac{n_A^T}{d_A}$, with measurement $y = (\bar{H}, \omega, a_s)$. The measurement then becomes a linear function of x, but the dynamics of x is nonlinear. The fact that $\bar{H}$ belongs to a Lie group has been exploited in [19], [8] to derive nonlinear observers with strong stability results, but under restrictive assumptions on the sensor motion that prevent the use of accelerometer measurements. We show in the next section that another choice of x leads to a linear form that much simplifies the estimation problem and yields strong estimation results.

III. Main results

Define the state to be estimated as

$$x = (H, M, n_s, Q) \tag{6}$$

with H defined by (2) and

$$M = vn_s^T, \quad n_s = \frac{n_A}{d_A}, \quad Q = g_B n_s^T \tag{7}$$

Then, using the fact that n_A and d_A are constant, one verifies from (4) and (5) the following relations:

$$\left\{ \begin{array}{rcl} \dot{H} &=& -S(\omega)H - M \\ \dot{M} &=& -S(\omega)M + Q + a_s n_s^T \\ \dot{n}_s &=& 0 \\ \dot{Q} &=& -S(\omega)Q \end{array} \right. \tag{8}$$

This expression calls for several remarks.

1) Since ω and a_s are known time-functions, the above system is a linear time-varying system in x. This is a strong asset for estimation purposes. Clearly, the price to pay is that the estimation state space is larger than what would be strictly needed for the estimation of H, v, n_A, d_A, and g_B. In other words, the estimation problem has been lifted to a higher-dimensional state space. The main point is that this lifting much simplifies the observer design and analysis.
2) Since n_A is a unit vector and the sign of one component of n_A is known (the camera is facing the scene), one can deduce from n_s both n_A and d_A. As a consequence, one can extract from H, M, n_s, and Q:

$$\begin{array}{l} n_A,\ d_A \\ v = d_A^2 M n_s = \frac{M n_s}{\|n_s\|^2} \\ g_B = d_A^2 Q n_s = \frac{Q n_s}{\|n_s\|^2} \end{array}$$

Recall also that R and p can be extracted from H and n_s using the standard homography decomposition.

The linear form of System (8) allows one to apply existing tools of linear estimation theory, starting with the Kalman Filter that ensures asymptotic convergence of the estimation errors to zero as soon as the system is *uniformly completely observable* (UCO) (See [15], [6], [3] for more details). Assuming that the state matrix is bounded, UCO is equivalent to the more classical notion of *uniform observability* (See, e.g., [5]). A first question is to determine conditions under which this latter property is satisfied for System (8). An answer to this question is provided in the following result.

Proposition *1:* Assume that

i) ω and a_s are continuous and bounded on $[0, +\infty)$, and their first, second, and third-order time-derivatives are well defined and bounded on $[0, +\infty)$;

ii) there exists two scalars $\delta, \sigma > 0$ such that

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\sigma} \|\dot{a}_s(\tau) + \omega(\tau) \times a_s(\tau)\| \, d\tau \tag{9}$$

Then, System (8) with measurement $y = H$ is uniformly observable.

The proof is given in Appendix.

The above proposition holds independently of the relation between ω, a_s and the derivatives of p and R. When ω and a_s are given by (4), however, one easily verifies that the observability condition (9) reduces to

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\sigma} \|\ddot{p}(\tau)\| \, d\tau \tag{10}$$

which essentially means that uniform observability is ensured provided that the linear acceleration is not constant.

As already mentionned, a natural observer candidate for System (8) is the Kalman filter, which ensures asymptotic (exponential) convergence of the estimation errors to zero under the above observability condition. We propose below another solution, in the form of a Luenberger-like observer. Although it is not expected that this observer performs better than the Kalman filter, it is computationally simpler.

Consider the following observer, which classically consists in a copy of the system's dynamics plus innovation terms:

$$\left\{ \begin{array}{rcl} \dot{\hat{H}} & = & -S(\omega)\hat{H} - \hat{M} + \alpha_H \\ \dot{\hat{M}} & = & -S(\omega)\hat{M} + \hat{Q} + a_s \hat{n}_s^T + \alpha_M \\ \dot{\hat{n}}_s & = & \alpha_n \\ \dot{\hat{Q}} & = & -S(\omega)\hat{Q} + \alpha_Q \end{array} \right. \tag{11}$$

where the α_* denote the innovations. Define also the estimation errors as

$$\tilde{H} = \hat{H} - H, \tilde{M} = \hat{M} - M, \tilde{n}_s = \hat{n}_s - n_s, \tilde{Q} = \hat{Q} - Q$$

so that the estimation error dynamics is given by:

$$\left\{ \begin{array}{rcl} \dot{\tilde{H}} & = & -S(\omega)\tilde{H} - \tilde{M} + \alpha_H \\ \dot{\tilde{M}} & = & -S(\omega)\tilde{M} + \tilde{Q} + a_s \tilde{n}_s^T + \alpha_M \\ \dot{\tilde{n}}_s & = & \alpha_n \\ \dot{\tilde{Q}} & = & -S(\omega)\tilde{Q} + \alpha_Q \end{array} \right. \tag{12}$$

Proposition *2:* Assume that a_s is constant and that ω and a_s satisfy Conditions *i)* and *ii)* of Proposition 1 for uniform observability. Define the innovation terms as:

$$\left\{ \begin{array}{rcl} \alpha_H & = & -k_1 \tilde{H} \\ \alpha_M & = & ((k_2 + k_4) I_3 + k_3 a_s a_s^T)\tilde{H} \\ \alpha_n & = & -k_3 \tilde{H}^T (S(\omega) - k_1 I_3) a_s \\ \alpha_Q & = & k_1 k_4 \tilde{H} \end{array} \right. \tag{13}$$

where $\{k_1, \cdots, k_4\}$ denotes any set of strictly positive constants. Then, the origin of System (12) is globally uniformly exponentially stable.

The proof is given in Appendix.

Proposition 2 provides a simple observer that yields the same asymptotic stability property as the Kalman filter, under the stronger assumption that a_s is constant. Asymptotic stability in the case a_s constant is sufficient to ensure good stability properties in practice, provided that a_s does not vary too rapidly. This will be illustrated in the experimental validation section. If necessary, it is possible to modify the above observer so as to better take into account possible variations of a_s.

Let us end this section with a last remark. There are benefits to maintaining the homography matrix estimate in the Special Linear Group $SL(3)$. In particular, this ensures that the estimate is consistent with the physical constraint that the camera remains on one side of the scene. The homography estimate as given by (11) does not a priori belong to $SL(3)$, and it cannot be projected onto $SL(3)$ via the scaling defined by (3) since one cannot guarantee that $\det(\hat{H}) \neq 0$. This limitation can be circumvented by coupling the observer (11) with the one proposed in [19]. More precisely, the output $\hat{M}$ of (11) can be used as input to the observer proposed in [19]. The latter then provides estimates $\hat{\bar{H}} \in SL(3)$ that benefit from an accurate estimation of M.

IV. Experimental validations

In this section we validate the linear observer presented in the previous sections and evaluate its performance. This algorithm is used in conjunction with an ESM homography computer vision algorithm [2]. Recall that ESM is an intensity-based registration method. It relies on an efficient optimization algorithm to estimate transformation parameters (here homography parameters) that best align a reference image with a current one. For each image, the observer is used to make a prediction of the raw homography that is used as initialization of the ESM algorithm. The visual method provides two results: an homography estimation and the correlation score between the current image and the reference

image. If the correlation score is good enough (> 0.85) the estimate is considered as "good" and is used as measurement in the observer.

Three issues are investigated: *i)* tracking quality and ability of the filter to follow the pattern in the presence of fast dynamics; *ii)* ability of the filter to interpolate the homography between two frames and provide estimation of the homography at higher rate; *iii)* quality of the other estimates (normal, velocity, gravity, scale factor).

A. Experimental setup

We use a sensor consisting of a xSens MTiG IMU working at a frequency of 200 [Hz] and an AVT Stingray 125B camera that provides 40 images of 800×600 [pixel] resolution per second. The camera and the IMU are synchronized. The camera uses wide-angle lenses (focal 1.28 [mm]). The target is placed over a surface parallel to the ground and is printed out on a 376×282 [mm] sheet of paper to serve as a reference for the visual system. The reference image is 300×225 [pixel]. So the distance d_A can be determined as 0.2115[m]. The processed video sequence presented in the accompanying video is 1321 frames long and presents high velocity motion (rotations up to 5[rad/s], translations, scaling change) and occlusions. In particular, a complete occlusion of the pattern occurs little after $t = 10$[s]. If the pattern is lost, we reset the algorithm with the ground-truth homography (see below). The sequence is subsampled at 20Hz to obtain a more challenging sequence. Four images of the sequence are presented on Figure 1.

Unfortunately, we do not have an external system (e.g., Vicon) to provide ground truth. We compare the output of the observer with "pseudo ground-truth" values obtained as follows. Since the pattern geometry and size and the camera calibration parameters are known, we are able to compute the ground-truth normal n_A and distance d_A. A ground-truth homography for each frame of the sequence is computed thanks to a global estimation of the homography by SIFT followed by the ESM algorithm. Using the homography decomposition (see [18, Pg. 136] for the decomposition algorithm) and the ground-truth normal and distance mentionned above, we get ground-truth rotation R and position p. From the ground-truth rotation and a ground-truth g_A obtained from the accelerometers, we deduce a ground-truth g_B. Finally, an approximation $\hat{v}_{gt}$ of the ground-truth velocity is obtained applying a derivative filter to the ground-truth position:

$$\begin{cases} \dot{\hat{p}}_{gt} = \hat{v}_{gt} - k_5(\hat{p}_{gt} - \dfrac{p}{d_A}) \\ \dot{\hat{v}}_{gt} = -k_6(\hat{p}_{gt} - \dfrac{p}{d_A}) \end{cases} \tag{14}$$

with $k_5 = 27$ and $k_6 = 225$. We are aware that v_{gt} may significantly differ from v, so that the comparison of $\hat{v}$ with v_{gt} must be interpreted with caution.

The observer gains have been chosen as $k_1 = 25$, $k_2 = 75$, $k_3 = 0.4$, $k_4 = 38.5$. Following the notation of the description available at *http://esm.gforge.inria.fr/ESM.html*, the ESM algorithm is used with the following parameter values: $prec = 2$,$iter = 50$.

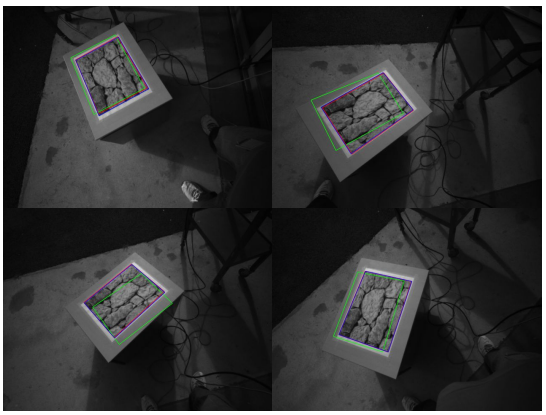

Fig. 1. Four images of the sequence at 20[Hz]: pattern position at previous frame (green), vision estimate (blue), and prediction of the filterIMU (red).

B. Performance evaluation

First, it should be noted that the ESM estimator alone is able to correctly track the pattern on only 72% of the sequence. Thanks to the observer, we are able to track the pattern on 97% of the sequence (a loss of the pattern occurs at $t =$10[s] because of the complete occlusion).

We now discuss the estimation results for each state component.

Figure 2 shows the results of the homography estimation. The blue curve corresponds to the matrix norm of the difference between the predicted homography (by the observer) and the ground-truth at current fame. In order to show the benefit brought by the observer, we have also displayed in green the matrix norm of the difference between the ground-truth at previous frame and the ground-truth at current frame. This clearly shows the capacity of the observer to provide accurate predictions.

Figure 3 shows the results of the scale factor estimation (i.e., d_A). More precisely, the blue curve corresponds to the ratio of $\hat{d}_A$ (as deduced from $\hat{n}_s$) over the ground-truth d_A. One can notice that the scale estimation is very accurate (ultimate error is about 0.5%).

Figure 4 presents the angular error between estimate and ground-truth for the normal n_A (top) and the gravity vector g_B (bottom). The figure exhibits the convergence of the normal estimate to the ground-truth (the final error is less than one degree). The direction of the gravity is also correctly estimated. The estimation seems less accurate and more noisy than for the normal but let us remark that the ground-truth is obtained from the visual processing and it is not completely accurate either. A better evaluation would require a better ground-truth.

Figure 5 shows the velocity components for both the estimation $\hat{v}$ (blue) and the ground truth v_{gt} (red). The trends are similar but as in the previous case, the reader must be aware that v_{gt} is not a very reliable ground-truth. One can observe that the estimation error on the scale factor at the begin of the sequence explains the larger difference between

ground-truth and estimate. As time goes by, the two curves get closer.

Let us finally remark that these performances are obtained despite the fact that the assumption of constant specific acceleration (upon which the observer stability was established) is violated.

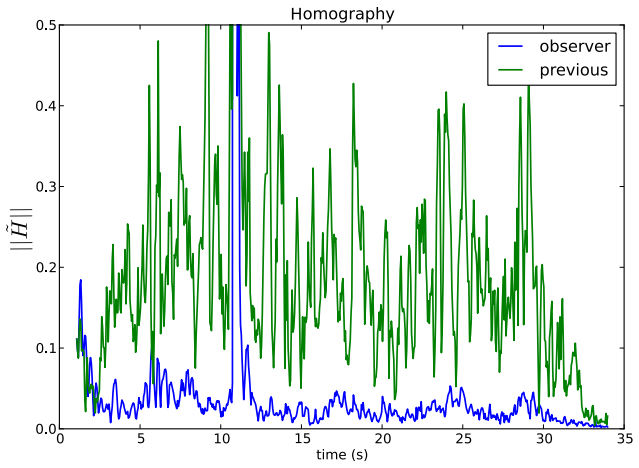

Fig. 2. Homography: $||\tilde{H}||$ (blue), error between homography at current and previous frames (green)

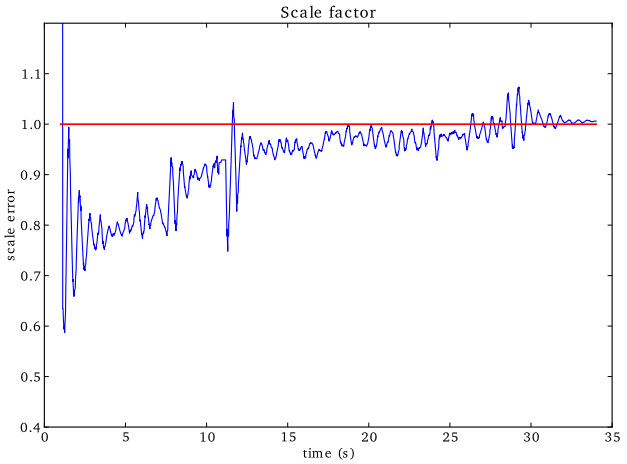

Fig. 3. Scale factor: ratio $\hat{d}_A/d_A$

Conclusion

We have shown that the problem of visuo-inertial estimation based on homography measurements can be formulated as a linear estimation problem by lifting the state space to a higher-dimensional space. This allows for the direct application of linear estimation theory. Based on this formulation, a uniform observability analysis has been derived and a Luenberger-like linear observer has been proposed together with its stability analysis in term of the system's observability. Experimental validation of the proposed solutions have shown the capacity of the observer to provide good-quality estimates of homography, linear velocity, and

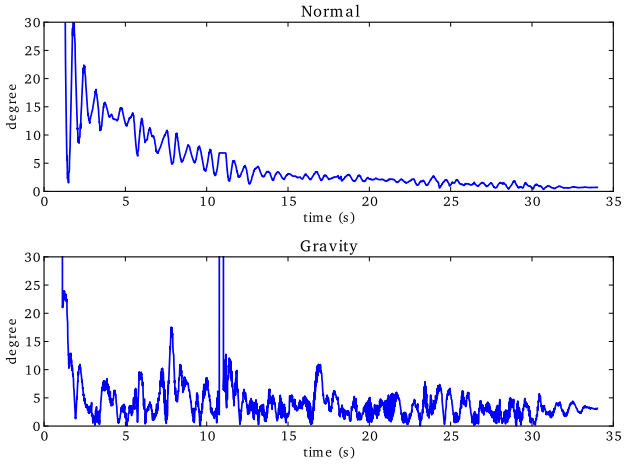

Fig. 4. Normal n_A and gravity g_B: errors in degrees between estimation and ground-truth

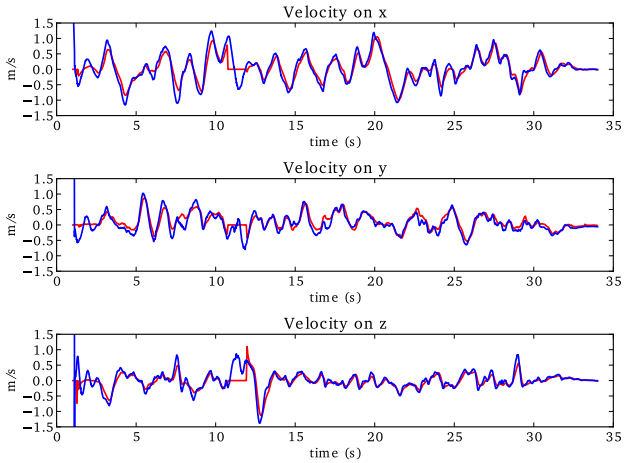

Fig. 5. Velocity : $\hat{v}$ (blue), v_{gt} (red)

environment structure parameters. Possible extensions of this work include, e.g., coupling the proposed solution with "geometric observers" that better preserve the structure of the problem, addressing possible IMU biaises, or applying the proposed solution to UAVs.

Appendix

Proof of Proposition 1

The proof is based on known relations between uniform observability and properties of the observability matrix (see, e.g. [5, Ch. 5]). It follows the same lines as [3], [22, App. A.4]. Recall that for a time-varying linear system

$$\left\{ \begin{array}{rcl} \dot{x} & = & A(t)x, \\ y & = & C(t)x \end{array} \right. \tag{15}$$

with $x \in \mathbb{R}^n$, the observability matrix is

$$\mathcal{O}(t) = \begin{pmatrix} N_0(t) \\ N_1(t) \\ \vdots \end{pmatrix}$$

with $N_0 = C$ and $N_{k+1} = N_k A + \dot{N}_k$ for $k = 1, \cdots$. We recall the following result, which is a particular case of [22, Lemma 3.1].

Lemma *1:* Consider the linear time-varying system (15) and assume that

1) C is constant,
2) A is continuous and bounded on $[0, +\infty)$,
3) The k-th order derivative of A is well defined and bounded on $[0, +\infty)$ up to order $k = K \geq 0$,
4) There exist an $m \times n$ matrix P, composed of row vectors of $N_0, \cdots, N_K$, and two scalars $\delta, \sigma > 0$ such that

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\sigma} \det(P(\tau)^T P(\tau))\, d\tau \quad (16)$$

Then, System (15) is uniformly observable.

We apply the above lemma to System (8), where columns vectors of H, M, and Q, are stacked vertically so as to obtain a vector $x \in \mathbb{R}^{9+9+3+9}$. Assumption 1 is clearly satisfied since $y = H$. Assumptions 2 and 3 with $K = 3$ are also satisfied by Assumption *i)* of Proposition 1. There remains to show that Assumption 4 above is satisfied. We simply set

$$P(\tau) = \begin{pmatrix} N_0(\tau) \\ N_1(\tau) \\ N_2(\tau) \\ N_3(\tau) \end{pmatrix}$$

One verifies that P has the following block structure:

$$P = \begin{pmatrix} I_9 & 0 & 0 \\ * & -I_9 & 0 \\ * & * & P_r \end{pmatrix} \quad (17)$$

where

$$P_r = \begin{pmatrix} -P_{33} & -I_9 \\ -P_{43} & P_{44} \end{pmatrix}$$

and the matrices P_{ij} are given by

$$P_{33} = \begin{pmatrix} a_s e_1^T \\ a_s e_2^T \\ a_s e_3^T \end{pmatrix}, \quad P_{43} = \begin{pmatrix} (\dot{a}_s - 2\omega \times a_s) e_1^T \\ (\dot{a}_s - 2\omega \times a_s) e_2^T \\ (\dot{a}_s - 2\omega \times a_s) e_3^T \end{pmatrix}$$

and

$$P_{44} = 3 \begin{pmatrix} S(\omega) & 0 & 0 \\ 0 & S(\omega) & 0 \\ 0 & 0 & S(\omega) \end{pmatrix}$$

with e_i the canonical vectors of $\mathbb{R}^3$. In the above expressions the $*$ denote matrices the expression of which is unimportant. Due to the block structure of P (cf. (17)), one can show that (16) is equivalent to

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\sigma} \det(P_r(\tau)^T P_r(\tau))\, d\tau \quad (18)$$

Furthermore, we remark that

$$P_r = \begin{pmatrix} 0 & I_9 \\ -(P_{43} + P_{44} P_{33}) & -P_{44} \end{pmatrix} \begin{pmatrix} I_3 & 0 \\ -P_{33} & -I_9 \end{pmatrix}$$

so that (18) can be further reduced to

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\sigma} \det(P_{rr}(\tau)^T P_{rr}(\tau))\, d\tau \quad (19)$$

with

$$P_{rr} = P_{43} + P_{44} P_{33} = \begin{pmatrix} (\dot{a}_s + \omega \times a_s) e_1^T \\ (\dot{a}_s + \omega \times a_s) e_2^T \\ (\dot{a}_s + \omega \times a_s) e_3^T \end{pmatrix}$$

Condition (19) follows from (9), after noting that $P_{rr}^T P_{rr} = \|\dot{a}_s + \omega \times a_s\|^2 I_3$.

Proof of Proposition 2

Define two new variables $\bar{n}_s, \bar{Q}$ as:

$$\begin{cases} \bar{n}_s &= \tilde{n}_s + k_3 \tilde{H}^T a_s \\ \bar{Q} &= \tilde{Q} + k_4 \tilde{H} \end{cases}$$

and note that the mapping $(\tilde{H}, \tilde{M}, \tilde{n}_s, \tilde{Q}) \longmapsto (\tilde{H}, \tilde{M}, \bar{n}_s, \bar{Q})$ is a diffeomorphism. Using the assumption that a_s is constant, one verifies from (12) that

$$\begin{cases} \dot{\tilde{H}} &= -S(\omega)\tilde{H} - \tilde{M} + \alpha_H \\ \dot{\tilde{M}} &= -S(\omega)\tilde{M} + \bar{Q} + a_s \bar{n}_s^T + \bar{\alpha}_M \\ \dot{\tilde{n}}_s &= -k_3 \tilde{M}^T a_s + \bar{\alpha}_n \\ \dot{\tilde{Q}} &= -S(\omega)\bar{Q} - k_4 \tilde{M} + \bar{\alpha}_Q \end{cases} \quad (20)$$

with

$$\begin{cases} \bar{\alpha}_M &= \alpha_M - k_4 \tilde{H} - k_3 a_s a_s^T \tilde{H} \\ \bar{\alpha}_N &= \alpha_N + k_3 (\tilde{H}^T S(\omega) + \alpha_H^T) a_s \\ \bar{\alpha}_Q &= \alpha_Q + k_4 \alpha_H \end{cases} \quad (21)$$

Consider the candidate Lyapunov function

$$L = \frac{1}{2} \left(c_1 \|\tilde{H}\|^2 + c_2 \|\tilde{M}\|^2 + c_3 \|\bar{n}_s\|^2 + c_4 \|\bar{Q}\|^2 \right)$$

where the c_i's are positive constant specified latter. The time derivative of L is given by

$$\dot{L} = c_1 \operatorname{tr}(\tilde{H}^T \dot{\tilde{H}}) + c_2 \operatorname{tr}(\tilde{M}^T \dot{\tilde{M}}) + c_3 \bar{n}_s^T \dot{\tilde{n}}_s + c_4 \operatorname{tr}(\bar{Q}^T \dot{\bar{Q}})$$

Using the expressions of $\dot{\tilde{H}}, \dot{\tilde{M}}, \dot{\tilde{n}}_s, \dot{\bar{Q}}$ in (20), and the fact that for any square matrix P and any vectors $z_1, z_2 \in \mathbb{R}^n$, $\operatorname{tr}(P) = \operatorname{tr}(P^T)$ and $\operatorname{tr}(z_1 z_2^T) = z_1^T z_2$, one obtains:

$$\begin{aligned} \dot{L} \;=\; & c_1 \operatorname{tr}(-\tilde{H}^T \tilde{M} + \tilde{H}^T \alpha_H) + c_2 \operatorname{tr}(\tilde{M}^T \bar{\alpha}_M) \\ & + (c_2 - c_3 k_3) \bar{n}_s^T \tilde{M}^T a_s + (c_2 - c_4 k_4) \operatorname{tr}(\tilde{M}^T \bar{Q}) \\ & + c_3 \bar{n}_s^T \bar{\alpha}_n + c_4 \operatorname{tr}(\bar{Q}^T \bar{\alpha}_Q) \end{aligned} \quad (22)$$

Let us set

$$\begin{cases} \alpha_H = -k_1 \tilde{H} \\ \bar{\alpha}_M = k_2 \tilde{H} \\ \bar{\alpha}_N = 0 \\ \bar{\alpha}_Q = 0 \end{cases} \qquad \begin{cases} c_2 = \frac{c_1}{k_2} \\ c_3 = \frac{c_2}{k_3} \\ c_4 = \frac{c_2}{k_4} \end{cases}$$

and note, in view of (21), that this definition of $\alpha_H, \bar{\alpha}_M, \bar{\alpha}_N$ and $\bar{\alpha}_Q$ is equivalent to (13). One then deduces from (22) that

$$\dot{L} = -c_1 k_1 \operatorname{tr}(\tilde{H}^T \tilde{H}) = -k_1 c_1 \|\tilde{H}\|^2 \leq 0$$

This shows that the origin of System (12) is stable. Furthermore, it follows from [1, Th. 5] that the origin of System (12) is uniformly exponentially stable provided that the pair $(A(t), C)$ is uniformly observable, with $A(t)$ the state matrix of System (12)-(13), and C the constant matrix defined by $\tilde{H} = C\tilde{x}$ with $\tilde{x} = (\tilde{H}, \tilde{M}, \tilde{n}_s, \tilde{Q})$, i.e., provided that System (12)-(13) is uniformly observable with $\tilde{H}$ as measurement. Since the innovations only depend on $\tilde{H}$, this is equivalent to the uniform observability of the original system (8) with H as measurement. The latter property follows by application of Proposition 1.

REFERENCES

[1] B.D.O. Anderson and J.B. Moore. New results in linear system stability. *SIAM Journal on Control*, 7(3):398–414, 1969.

[2] S. Benhimane and E. Malis. Homography-based 2d visual tracking and servoing. *International Journal of Computer Vision*, 2007.

[3] P.J. Bristeau, N. Petit, and L. Praly. Design of a navigation filter by analysis of local observability. In *IEEE Conf. on Decision and Control*, pages 1298–1305, 2010.

[4] T.F. Gonçalves, J.R. Azinheira, and P. Rives. Homography-based visual servoing of an aircraft for automatic approach and landing. In *IEEE Conf. on Robotics and Automation*, pages 9–14, 2010.

[5] C.-T. Chen. *Linear system theory and design*. Oxford University Press, 1984.

[6] G. Besançon. *Nonlinear observers and applications*, volume 363 of *Lecture Notes in Control and Information Sciences*, chapter An overview on observer tools for nonlinear systems, pages 2–33. Springer, 2007.

[7] P. Corke, J. Lobo, and J. Dias. An introduction to inertial and visual sensing. *International Journal of Robotics Research*, 26:519–535, 2007.

[8] A. Eudes, P. Morin, R. Mahony, and T. Hamel. Visuao-inertial fusion for homography-based filtering and estimation. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 5186–5192, 2013.

[9] R. Garcia, J. Batle, X. Cufi, and J. Amat. Positioning an underwater vehicle through image mosaicking. In *IEEE Conf. on Robotics and Automation*, pages 2779–2784, 2001.

[10] P. Gemeiner, P. Einramhof, and M. Vincze. Simultaneous motion and structure estimation by fusion of inertial and vision data. *The International Journal of Robotics Research*, 26:591–605, 2007.

[11] V. Grabe, H.H. Bülthoff, and P. Robuffo Giordano. On-board velocity estimation and closed-loop control of a quadrotor uav based on optical-flow. In *IEEE Conf. on Robotics and Automation*, 2012.

[12] V. Grabe, H.H. Bülthoff, and P. Robuffo Giordano. A comparison of scale estimation schemes for a quadrotor uav based on optical-flow and imu measurements. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 5193–5200, 2013.

[13] J.Chen, W.E. Dixon, D.M. Dawson, and M. McIntyre. Homography-based visual servo tracking control of a wheeled mobile robot. *IEEE Trans. on Robotics*, 22:407–416, 2006.

[14] E.S. Jones and S. Soatto. Visual?inertial navigation, mapping and localization: A scalable real-time causal approach. *International Journal of Robotics Research*, 30(4):407–430, 2011.

[15] R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, pages 95–108, 1961.

[16] J. Kelly and G.S. Sukhatme. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *International Journal of Robotics Research*, 30(1):56–79, 2011.

[17] G. López-Nicolás, N.R. Gans, S. Bhattacharya, C. Sagüés, J.J. Guerrero, and S. Hutchinson. Homography-based control scheme for mobile robots with nonholonomic and field-of-view constraints. *IEEE Trans. on Systems, Man, and Cybernetics: Part B*, 10:1115–1127, 2010.

[18] Y. Ma, S. Soatto, J. Kosecka, and S.S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[19] R. Mahony, T. Hamel, P. Morin, and E. Malis. Nonlinear complementary filters on the special linear group. *International Journal of Control*, 85:1557–1573, 2012.

[20] G. Panahandeh and M. Jansson. Vision-aided inertial navigation using planar terrain features. In *First Int. Conference on Robot, Vision and Signal Processing*, pages 287–291, 2011.

[21] G. Scandaroli, P. Morin, and G. Silveira. A nonlinear observer approach for concurrent estimation of pose, imu bias and camera-to-imu rotation. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3335–3341, 2011.

[22] G.G. Scandaroli. *Visuo-inertial data fusion for pose estimation and self-calibration*. PhD thesis, Université de Nice-Sophia Antipolis, Available at http://tel.archives-ouvertes.fr/tel-00861858, 2013.

[23] F. Servant, P. Houlier, and E. Marchand. Improving monocular plane-based slam with inertial measures. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3810–3815, 2010.

[24] D. Suter, T. Hamel, and R. Mahony. Visual servo control using homography estimation for the stabilization of an x4-flyer. In *IEEE Conference on Decision and Control*, pages 2872–2877, 2002.

[25] S. van der Zwaan and J. Santos-Victor. Real-time vision-based station keeping for underwater robots. In *OCEANS, 2001. MTS/IEEE Conference and Exhibition*, pages 1058–1065, 2001.