Deterministic observer design for vision-aided inertial navigation

Tarek Hamel, Minh-Duc Hua, Claude Samson

Deterministic observer design for vision-aided inertial navigation

Tarek Hamel, Minh-Duc Hua and Claude Samson

Abstract— The purpose of this tutorial is to report results developed over the last three years for state estimation problems arising with unmanned mobile robots equipped with a monocular camera and a 3-axis gyrometer, complemented with either a velocity sensor, or a 3-axis accelerometer, or an optical flow sensor. Definition and characterization of uniform observability for linear time-varying systems, followed by an observer design framework exploiting first-order approximations of a class of nonlinear systems, are first recalled. The resulting Riccati observers are locally uniformly exponentially stable when associated uniform observability conditions are satisfied. This framework is subsequently applied, with detailed explanations, to a set of practical problems, namely i) classical PnP camera pose estimation using known source points and bearing measurements, ii) the adaptation of this problem to unknown source points by using epipolar constraints, iii) camera pose and velocity estimation using bearing and IMU measurements, and iv) camera velocity and depth estimation using optical flow and IMU measurements from the observation of a planar target. The observer solutions proposed for these last two problems are validated with experimental data.

I. INTRODUCTION

State estimation (position, velocity, orientation, etc.) is a central problem in the fields of robotics and autonomous systems. It consists in designing sensor fusion algorithms that reconcile various measurements through the use of a dynamical model to produce state estimates. When the state estimates are used in a control law, these algorithms are required to operate in real time, with high bandwidth and low latency. This puts severe constraints on the sensor fusion algorithms, from both theoretical and practical points of view. A variety of methods and sensory modalities have been proposed in the literature for state estimation of autonomous systems.

The most commonly used proprioceptive sensors are inertial sensors (e.g., accelerometer, gyrometer) and Doppler velocity sensors that have the advantage of producing measurements at a high rate with relatively low latency. However, they only provide relative motion measurements that are furthermore often impaired by significant drifts. Proprioceptive sensors are often complemented with Global Positioning Systems (GPS) to provide a global position measurement at a relatively high rate [7], [10], but with low precision and, more annoyingly, with limited applicability when the GPS signal is severely degraded in urban canyon environments, to the point of being inoperable in indoor environments.

T. Hamel is with I3S, Université Côte d'Azur, Institut Universitaire de France, CNRS, Sophia Antipolis, France, `thamel@i3s.unice.fr`

M-D. Hua is with I3S, UCA-CNRS, Sophia Antipolis, France `hua@i3s.unice.fr`.

C. Samson is with INRIA and I3S, UCA-CNRS, Sophia Antipolis, France, `claude.samson@inria.fr`

These sensors are also often complemented with onboard exteroceptive sensors, such as laser range finders, acoustic sensors and stereo cameras that, beyond the estimation of the system state, provide information about the surrounding environment. Early approaches for state estimation were based on Extended Kalman Filters (EKF) [3], unscented Kalman filters (UKF) [14] and particle filters [1]. However, these early solutions have limitations, mostly in terms of robustness. The nonlinearity of autonomous systems dynamics is much related to the inherent nonlinearity of their state space, well exemplified by the Lie groups of rotation $SO(2)$ and $SO(3)$. Exploiting the structure of Lie groups for state estimation is a fruitful approach that has motivated, during the last fifteen years, an increasing interest in the design of nonlinear observers [2], [5], [16] and invariant extended Kalman filters (IEKF) [4] for autonomous systems. All observers in this class take into account the specificity of sensory devices that translate the action of a Lie group on a measurement space. However, this entails a certain number of complications at the design level that need to be overcome. In particular, the estimation of the pose of a moving monocular camera on the basis of proprioceptive sensors measurements, complemented with bearing measurements of a set of source points identified in the camera image, the coordinates of which in the inertial frame are either known (the case of the classical Perspective-n-Point (PnP) problem), or unknown (the essential matrix estimation problem), yields challenging problems. A variety of methods dealing with these problems have been proposed. Let us cite, for instance, algebraic algorithms [6], [20] and iterative algorithms based on gradient search [9] for the PnP problem, and nonlinear optimization algorithms for essential matrix estimation [15]. There is also a rich literature addressing these problems with EKF algorithms, typically in the context of Simultaneous Localization And Mapping (SLAM) and Visual Odometry [13].

In this paper we show that the deterministic Riccati observer framework developed in [8], which exploit first-order approximations of a class of nonlinear systems, and may formally be viewed as a generalisation of the so-called multiplicative extended Kalman filter (MEKF) [18], [19], applies to a number of applications involving proprioceptive sensors and monocular vision.

The remainder of the paper is organized as follows. In Section II, after specifying the notation used throughout the paper, basic definitions related to uniform observability of a time-varying linear system –a property whose satisfaction is central to proving the exponential stability of the proposed observers– and a generic dynamic system verifying a set of

structural properties, complemented with a generic Riccati observer, are recalled. Section III shows how the proposed observer framework applies to the estimation of the camera pose when using bearing measurements of a set of source points whose coordinates in the inertial frame are either known (the classical PnP problem) or unknown (use of epipolar constraints). In Section IV, an adaption of the observer to the generalized PnP problem for which the linear velocity is unknown is proposed, and its performance is validated with experimental data. In Section V, another dynamic problem involving the estimation of the camera's attitude, of its linear velocity and of the distance between the camera and a planar target (depth), on the basis of optical flow and IMU measurements, is addressed. The observer's performance is also validated with experimental data. Concluding remarks are presented in Section VI.

II. PRELIMINARY MATERIAL

A. Notation

We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space, by $\{e_1, \ldots, e_n\}$ the canonical basis of $\mathbb{R}^n$ and by $|x|$ the Euclidean norm of the vector $x \in \mathbb{R}^n$. The set $\mathcal{B}_r^n := \{x \in \mathbb{R}^n : |x| \le r\}$ is the closed ball in $\mathbb{R}^n$ with radius r. The null matrix and the identity matrix of dimension $n \times n$ are denoted by 0_n and I_n, respectively. The null matrix of dimension $m \times n$ is denoted by $0_{m \times n}$.

The set $\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : |x| = 1\}$ is the n-dimensional sphere embedded in $\mathbb{R}^{n+1}$ with radius equal to one. For any $p \in \mathbb{S}^n$, the projection onto the tangent space of the unit n-dimensional sphere at the point p is denoted by $\Pi_p := (I_{n+1} - pp^\top)$. The Special Orthogonal group of order n is denoted by $\mathrm{SO}(n) := \{R \in \mathbb{R}^{n \times n} : \det(R) = 1,\, R^\top R = RR^\top = I_n\}$. For any $x, y \in \mathbb{R}^3$, $x_\times$ is the skew symmetric matrix associated with the cross product, i.e. $x_\times y = x \times y$.

With f denoting a vector-valued function depending on the two variables x and y, and on the time variable t, we write $f = O(|x|^{k_1}|y|^{k_2})$ with $k_1 \ge 0$ and $k_2 \ge 0$ if $\forall t : |f(x, y, t)|/(|x|^{k_1}|y|^{k_2}) \le \gamma < \infty$ in the neighbourhood of $(x = 0, y = 0)$. If f depends only on x and t then we write $f(x, t) = O(|x|^k)$ if $\forall t : |f(x, t)|/|x|^k \le \gamma < \infty$ in the neighbourhood of $x = 0$.

B. Uniform observability of a linear time-varying system

Consider a Linear Time-Varying (LTV) system of the form:
$$\begin{cases} \dot{x} = \bar{A}(t)x + \bar{B}(t)u \\ y = \bar{C}(t)x \end{cases}$$
with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^l$, $y \in \mathbb{R}^m$, and $\bar{A}(t)$, $\bar{B}(t)$, $\bar{C}(t)$ denoting continuous matrix-valued functions with adequate dimensions. Let $Q(t)$ denote a continuous $m \times m$-dimensional matrix-valued function, positive definite for all $t \in \mathbb{R}^+$. The *Riccati observability Gramian* associated with the triplet $(\bar{A}, \bar{C}, Q)$ is the non-negative definite matrix-valued function defined by
$$W_Q^{\bar{A}, \bar{C}}(t, t + \delta) := \frac{1}{\delta} \int_t^{t+\delta} \Phi^\top(s, t)\bar{C}^\top(s)Q(s)\bar{C}(s)\Phi(s, t)ds$$

where $\Phi(s, t)$ is the transition matrix associated with $\bar{A}$, i.e. $\frac{d}{ds}\Phi(s, t) = \bar{A}(s)\Phi(s, t), \forall s \ge t,\ \Phi(t, t) = I_n$.

If $\bar{A}(t)$ and $\bar{C}(t)$ are bounded and if there exist $\delta > 0$ and $\epsilon > 0$ such that $W_{I_{n_2}}^{\bar{A}, \bar{C}}(t, t + \delta) > \epsilon I_n,\ \forall t \ge 0$, then we say that the pair $(\bar{A}(t), \bar{C}(t))$ is uniformly observable.

C. Riccati Observer for a Class of nonlinear Systems

In this section we define a class of nonlinear systems of interest, along with a concise description of the proposed observer design methodology. This material is taken from [8].

Every member of this class of systems lives in $\mathcal{B}_r^{n_1} \times \mathbb{R}^{n_2}$, and its state is denoted as $x = [x_1^\top, x_2^\top]^\top$, with $x_1 \in \mathcal{B}_r^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$. The output of this system, denoted as $y \in \mathbb{R}^m$, is assumed to be a function of x, $\hat{x}_2$, and t, with $\hat{x}_2$ denoting an estimate of x_2. This system evolves according to the following equations:

$$\begin{cases} \dot{x} = A(x_1, t)x + \begin{bmatrix} u_1 \\ u_2(t) \end{bmatrix} + O(|x_1|^2) + O(|x_1||u_1|) \\ y(x, \hat{x}_2, t) = C_1(x_1, \hat{x}_2, t)x_1 + C_2(x_1, \hat{x}_2, t)x_2 \\ \qquad + O(|x_1|^2) + O(|x_1||x_2 - \hat{x}_2|) + O(|x_2 - \hat{x}_2|^2) \end{cases} \tag{1}$$

where

- $A(x_1, t)$ is a continuous matrix-valued function uniformly bounded with respect to (w.r.t.) t and uniformly continuous w.r.t. x_1. Furthermore, it is of the form
$$A(x_1, t) = \begin{bmatrix} A_{11}(t) & 0_{n_1 \times n_2} \\ A_{21}(x_1, t) & A_{22}(t) \end{bmatrix}$$

- $C(x, \hat{x}_2, t) := [C_1(x, \hat{x}_2, t)^\top,\ C_2(x, \hat{x}_2, t)^\top]^\top \in \mathbb{R}^{m \times (n_1 + n_2)}$ is a continuous matrix-valued function uniformly bounded w.r.t. t and uniformly continuous w.r.t. x and $\hat{x}_2$.

It is also assumed that $u_2(t)$ is a bounded input and that the solutions to this system belong to a compact set $\mathcal{D}$ independently of the choice of the input u_1, whose role is to make x_1 converge to zero. By setting
$$u_1 = -K_1(y - C_2\hat{x}_2), \tag{2}$$
an estimate of x is $(\hat{x}_1^\top, \hat{x}_2^\top)^\top$, with $\hat{x}_1 \equiv 0$ and $\hat{x}_2$ calculated according to:
$$\begin{cases} \dot{\hat{x}}_2 = A_{22}\hat{x}_2 + u_2 + K_2(y - C_2\hat{x}_2) \\ \dot{P} = AP + PA^\top - PC^\top Q(t)CP + V(t) \end{cases} \tag{3}$$
with $\hat{x}_2(0) \in \mathbb{R}^{n_2}$, $P(0) \in \mathbb{R}^{(n_1 + n_2) \times (n_1 + n_2)}$ a symmetric positive definite matrix, Q and V bounded continuous symmetric positive semi-definite matrix-valued functions, and $K := [K_1^\top, K_2^\top]^\top$ given by
$$K = k(t)PC^\top Q \tag{4}$$
with $0.5 \le k(t) \le k_{max} < \infty$ and P the solution, obtained by numerical integration, to the continuous Riccati equation (CRE) specified in (3).

Let $\tilde{x} := ((x_1 - \hat{x}_1)^\top, (x_2 - \hat{x}_2)^\top)^\top$ denote the estimation error. Then, as proven in [8, Corollary 3.2], the equilibrium

$\tilde{x} = 0$ is locally uniformly exponentially stable when $Q(t)$ and $V(t)$ are both larger than some positive matrix and the pair $(A(0, t), C(0, x_2(t), t))$ is uniformly observable.

III. OBSERVER DESIGN FOR KINEMATIC SYSTEMS INVOLVING MONOCULAR VISION

In this section we focus on kinematic systems evolving in $\mathrm{SO}(3) \times \mathbb{R}^3$, with measurements performed in the camera frame, and we show how the equations of this type of system can fit the framework described in the previous section in order to derive observers of interest in practice. The problem at hand consists in providing an estimate of the camera pose given measurements of the camera linear and angular velocities (both assumed to be bounded), and by using either the inertial coordinates of a set of $n \geq 3$ static source points identified in the image, or epipolar constraints when these coordinates are not known.

Let $\{\mathcal{I}\}$ denote an inertial (fixed) frame of reference and let $\{\mathcal{B}\}$ denote a frame of reference attached to the camera. Let $R \in \mathrm{SO}(3)$ denote the orientation of the camera frame $\{\mathcal{B}\}$ w.r.t. the inertial frame $\{\mathcal{I}\}$. Let $\mathring{\xi} \in \mathbb{R}^3$ denote the position of the camera in the inertial frame $\{\mathcal{I}\}$ and define $\xi := R^\top \mathring{\xi} \in \mathbb{R}^3$, its expression in $\{\mathcal{B}\}$ w.r.t. $\{\mathcal{I}\}$ expressed in $\{\mathcal{B}\}$. Then the equations of motion of the camera satisfy

$$\begin{cases} \dot{R} = R\Omega_\times \\ \dot{\xi} = -\Omega_\times \xi + V \end{cases} \quad (5)$$

where $V \in \mathbb{R}^3$ and $\Omega \in \mathbb{R}^3$ denote the linear and the angular velocity of the camera expressed in the frame $\{\mathcal{B}\}$, respectively. Assume that V and Ω are measured by means of a linear velocity sensor (e.g., a Doppler sensor) and an Inertial Measurement Unit (IMU), respectively.

The inertial coordinates of the source points are denoted by $\mathring{P}_i$, and we assume that their calibrated projective coordinates p_i^{p} onto the camera plane expressed w.r.t. the frame $\{\mathcal{B}\}$ are measured (see Figure 1). Denoting by P_i the coordinates of the ith source point w.r.t. the frame $\{\mathcal{B}\}$ expressed in $\{\mathcal{B}\}$, and by z_i its third component, one verifies that $p_i^{\mathrm{p}} = P_i / z_i$.
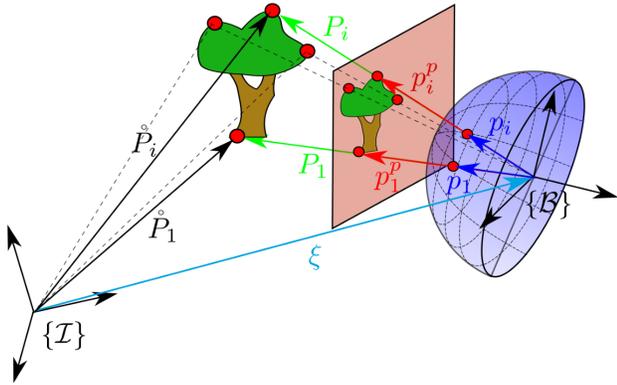

Fig. 1. Intuitive representation of inertial coordinates $\mathring{P}_i$, planar projective coordinates p_i^p and spherical projective coordinates p_i of the ith source point.

Rather than perspective outputs, typically used in computer vision algorithms, we prefer using the direction (bearing) vectors:

$$p_i := \frac{p_i^{\mathrm{p}}}{|p_i^{\mathrm{p}}|} = \frac{R^\top(\mathring{P}_i - \mathring{\xi})}{|\mathring{P}_i - \mathring{\xi}|} = \frac{R^\top \mathring{P}_i - \xi}{|R^\top \mathring{P}_i - \xi|} \in \mathbb{S}^2 \quad (6)$$

obtained by projecting the source points onto a virtual spherical image. These outputs can always be calculated from the image data. They differ from perspective outputs only by a scaling factor, and using them contributes to the simplicity of the observer design methodology exposed in the present paper.

A. Observer equations and model adaptation

The estimate of R is denoted by $\hat{R} \in \mathrm{SO}(3)$, and the estimate of ξ (resp. $\mathring{\xi}$) is denoted by $\hat{\xi} \in \mathbb{R}^3$ (resp. $\hat{\mathring{\xi}} \in \mathbb{R}^3$). The proposed observer has the following form:

$$\begin{cases} \dot{\hat{R}} = \hat{R}\Omega_\times - \hat{R}\sigma_{R\times} \\ \dot{\hat{\xi}} = -\Omega_\times \hat{\xi} + V - \sigma_\xi \end{cases} \quad (7)$$

with initial conditions $(\hat{R}(0), \hat{\xi}(0)) \in \mathrm{SO}(3) \times \mathbb{R}^3$, and $\sigma_R, \sigma_\xi \in \mathbb{R}^3$ the innovation terms to be determined thereafter. In practice, it is advantageous to calculate the unit quaternion $(\hat{\lambda}_0 \in \mathbb{R}, \hat{\lambda}_v \in \mathbb{R}^3)$ associated with $\hat{R}$, and then translate this quaternion into a rotation matrix. The corresponding equation, which then replaces the first equation in (7), is:

$$\begin{aligned} \frac{d}{dt}(\hat{\lambda}_0, \hat{\lambda}_v) &= 0.5\,(\hat{\lambda}_0, \hat{\lambda}_v) \circ (0, \Omega - \sigma_R) \\ &+ k\,(1 - \hat{\lambda}_0^2 - |\hat{\lambda}_v|^2)\,(\hat{\lambda}_0, \hat{\lambda}_v) \end{aligned} \quad (8)$$

with $\circ$ denoting the quaternion product, and k a positive gain ensuring the numerical stability of the quaternion's norm (equal to one).

Define the attitude error matrix

$$\tilde{R} := \hat{R}^\top R \quad (9)$$

One could also use $\tilde{R} := R\hat{R}^\top$ as an alternate definition of this error. From (5) and (7) one verifies that the attitude error kinematics are given by

$$\dot{\tilde{R}} = -\Omega_\times \tilde{R} + \tilde{R}\Omega_\times + \sigma_{R\times}\tilde{R} \quad (10)$$

The following step involves developing a first-order approximation of (10). From the Rodrigues' formula, the first-order approximation of $\tilde{R}$ is given by

$$\tilde{R} = I_3 + \tilde{\lambda}_\times + O(|\tilde{\lambda}|^2) \quad (11)$$

with $\tilde{\lambda} \in \mathcal{B}_2^3$ equal to twice the vector part of the quaternion associated with the attitude error matrix $\tilde{R}$, and whose convergence to zero implies the convergence of $\hat{R}$ to R. One then deduces from (10), (11), and the identity $a_\times b_\times - b_\times a_\times = (a \times b)_\times$, $\forall a, b \in \mathbb{R}^3$, that

$$\dot{\tilde{\lambda}} = -\Omega_\times \tilde{\lambda} + \sigma_R + O(|\tilde{\lambda}|^2) + O(|\tilde{\lambda}||\sigma_R|) \quad (12)$$

From (12) and the second equation of (5), and by setting $x = [x_1^\top, x_2^\top]^\top := [\tilde\lambda^\top, \xi^\top]^\top$, $u_1 := \sigma_R$ and $u_2(t) := V$, one obtains the system equation in (1) with

$$A(x_1, t) = A(0, t) := \begin{bmatrix} -\Omega_\times & 0_3 \\ 0_3 & -\Omega_\times \end{bmatrix}$$

Note that this latter matrix valued-function is independent of x_1 for this particular case. Using $R\hat R^\top$, instead of $\hat R^\top R$ for the attitude error, and using the corresponding error quaternion, would have led to the same result but with a different matrix A and a different input u_1.

Once there, it remains to determine the output y and the matrix-valued function C, depending on the problem at hand, i.e. either the PnP problem with known source points coordinates in the inertial frame, or the case of unknown source points' coordinates and the use of epipolar constraints.

B. System output associated with the PnP problem

In this case the inertial coordinates $\mathring P_i$ of the source points are known. From (6), and by definition of the projector Π_{p_i}, one has $\Pi_{p_i} p_i = 0$, $\forall i = \{1, \ldots, n\}$. It follows that

$$
\begin{aligned}
0 &= -|\mathring P_i - \mathring\xi| \Pi_{p_i} p_i \\
&= \Pi_{p_i} \xi - \Pi_{p_i} R^\top \mathring P_i \\
&= \Pi_{p_i} \xi - \Pi_{p_i} \tilde R^\top \hat R^\top \mathring P_i \\
&= \Pi_{p_i} \xi - \Pi_{p_i} (I_3 - \tilde\lambda_\times) \hat R^\top \mathring P_i + O(|\tilde\lambda|^2)) \\
&= \Pi_{p_i} \xi - \Pi_{p_i} \hat R^\top \mathring P_i - \Pi_{p_i} (\hat R^\top \mathring P_i)_\times \tilde\lambda + O(|\tilde\lambda|^2)
\end{aligned}
$$

By setting $y_i := \Pi_{p_i} \hat R^\top \mathring P_i$ one obtains

$$y_i = \Pi_{p_i} \xi - \Pi_{p_i} (\hat R^\top \mathring P_i)_\times \tilde\lambda + O(|\tilde\lambda|^2) \quad (13)$$

for $i = 1, \ldots, n$. Now, by defining the system output as $y := [y_1^\top, \ldots, y_n^\top]^\top$, one easily verifies that the output equation in (1) is satisfied with

$$C_1 = \begin{bmatrix} -\Pi_{p_1} (\hat R^\top \mathring P_1)_\times \\ \vdots \\ -\Pi_{p_n} (\hat R^\top \mathring P_n)_\times \end{bmatrix}, \quad C_2 = \begin{bmatrix} \Pi_{p_1} \\ \vdots \\ \Pi_{p_n} \end{bmatrix}$$

From there, one deduces from (3) and (7) the innovation terms $\sigma_R = u_1 = -K_1(y - C_2\hat\xi)$ and $\sigma_\xi = -K_2(y - C_2\hat\xi)$.

Uniform observability of $(A(0, t), C(0, x_2(t), t))$ then guarantees the (local) exponential stability of the equilibrium $(\tilde R, \xi - \hat\xi) = (I_3, 0)$. This observability issue is not treated here due to space limitation, but one may refer to [8] where it is addressed in relation to the number and possible singular locations of source points, combined with body motion and the measurement of the camera linear velocity, either in the body-fixed frame or in the inertial frame.

C. System output when using Epipolar Constraints in the case of Unknown Source Points

In the case where the camera pose w.r.t. to the inertial frame is unknown, alike the inertial coordinates of the source points, it is not possible to recover the camera pose from the sole bearing measurements of the source points. However, if the inertial frame is taken as the *initial* frame of the camera (i.e. at time t=0), then the camera pose ($R = I_3, \xi = 0$) is initially known, by definition of this particular frame. This entails that, $\forall i = \{1, \ldots, n\}$, the (constant) unit vector $\mathring p_i := \frac{\mathring P_i}{|\mathring P_i|}$, which is initially equal to p_i, is measured. It is then possible to estimate future camera poses w.r.t. this frame by using source points bearing measurements (that do not depend on the chosen inertial frame). To this aim, one can exploit the fact that $P_i = R^\top(\mathring P_i - \mathring\xi) = R^\top \mathring P_i - \xi$, from which the ith epipolar constraint:

$$\mathring p_i^\top R\xi_\times p_i = 0, \quad (i = 1, \ldots, n) \quad (14)$$

is easily deduced. This constraint is typically presented in the form $\mathring p_i^\top E p_i = 0$, where $E := Rt_\times$ (with $t := \xi/|\xi|$) is the essential matrix involved in the constraint (14).

From (14), and using (11), one deduces that

$$
\begin{aligned}
0 &= -\mathring p_i^\top \hat R \tilde R \xi_\times p_i \\
&= -\mathring p_i^\top \hat R (I_3 + \tilde\lambda_\times) \xi_\times p_i + O(|\tilde\lambda|^2) \\
&= \mathring p_i^\top \hat R(\xi \times p_i)_\times \tilde\lambda + \mathring p_i^\top \hat R p_{i\times} \xi + O(|\tilde\lambda|^2) \\
&= \mathring p_i^\top \hat R(\hat\xi \times p_i)_\times \tilde\lambda + \mathring p_i^\top \hat R p_{i\times} \xi + O(|\tilde\lambda|^2) + O(|\tilde\xi||\tilde\lambda|)
\end{aligned}
$$

Therefore, by defining y as the vector in $\mathbb R^n$ with zero entries, the output equation in (1) is satisfied with

$$C_1 = \begin{bmatrix} \mathring p_1^\top \hat R(\hat\xi \times p_1)_\times \\ \vdots \\ \mathring p_n^\top \hat R(\hat\xi \times p_n)_\times \end{bmatrix}, \quad C_2 = \begin{bmatrix} \mathring p_1^\top \hat R p_{1\times} \\ \vdots \\ \mathring p_n^\top \hat R p_{n\times} \end{bmatrix}$$

As in the PnP case, one deduces that the innovation terms are $\sigma_R = u_1 = -K_1(y - C_2\hat\xi)$ and $\sigma_\xi = -K_2(y - C_2\hat\xi)$.

The issue of uniform observability of $(A(0, t), C(0, x_2(t), t))$, whose verification ensures the local exponential stability of the observer so obtained, is addressed in [12]. In particular, uniform observability is proven in the case of three non-aligned source points provided that the camera's translational motion is sufficiently exciting.

IV. OBSERVER DESIGN FOR POSE AND LINEAR VELOCITY ESTIMATION FROM BEARING AND IMU MEASUREMENTS

In this section, an extension of the PnP problem, addressing the combined estimation of the camera's pose and of the camera's linear velocity, is presented. The material is taken from [17]. This extension is motivated by a large number of applications where sensors measuring this velocity directly are not available. It also contributes to demonstrate the applicability of the proposed observer design methodology for a broad range of estimation problems encountered in practice.

The so-called *specific acceleration* $a_\mathcal{B} \in \mathbb R^3$ of the camera, i.e. the difference between the camera's linear acceleration and the gravity acceleration, both expressed in the body-fixed frame, can be measured by a 3-axis accelerometer. By definition (and assuming that the accelerometer is fixed to the camera):

$$a_\mathcal{B} = \dot V + \Omega_\times V - gR^\top e_g \quad (15)$$

with g denoting the gravity constant and $e_g \in \mathbb{S}^2$ denoting the gravity direction expressed in the inertial frame $\{\mathcal{I}\}$. In practice, the estimates of both gravity constant and gravity direction may be slightly erroneous. This leads to a bias between the gravity acceleration expressed in the inertial frame $a_g := g e_g$ and its estimate a_g^m. From experience, a good estimate of this bias significantly improves the quality of the camera pose and velocity observer.

Let $b_g := R^\top (a_g - a_g^m)$ denote the gravity bias expressed in the body-fixed frame $\{\mathcal{B}\}$. The time-derivative of this vector satisfies the relation

$$\dot{b}_g = -\Omega_\times b_g \qquad (16)$$

so that, in view of (15)

$$\dot{V} = -\Omega_\times V + R^\top a_g^m + b_g + a_{\mathcal{B}}$$

Using the first-order approximation (11) of $\tilde{R} \,(= \hat{R}^\top R)$

$$
\begin{aligned}
\dot{V} &= -\Omega_\times V + \tilde{R}^\top \hat{R}^\top a_g^m + b_g + a_{\mathcal{B}} \\
&= -\Omega_\times V + (I_3 - \tilde{\lambda}_\times)\hat{R}^\top a_g^m + b_g + a_{\mathcal{B}} + O(|\tilde{\lambda}|^2) \ (17) \\
&= -\Omega_\times V + \bar{b}_g + \hat{R}^\top a_g^m + a_{\mathcal{B}} + O(|\tilde{\lambda}|^2)
\end{aligned}
$$

with

$$\bar{b}_g := b_g + (\hat{R}^\top a_g^m)_\times \tilde{\lambda}$$

Using (12) and the first equation of (7) in the calculation of the time-derivative of the new bias $\bar{b}_g$, one verifies that

$$\dot{\bar{b}}_g = -\Omega_\times \bar{b}_g + (\hat{R}^\top a_g^m)_\times \sigma_R + O(|\tilde{\lambda}|^2) + O(|\tilde{\lambda}||\sigma_R|) \ (18)$$

In view of (5), (17) and (18), the proposed observer takes the following form:

$$
\begin{cases}
\dot{\hat{R}} = \hat{R}\Omega_\times - \hat{R}\sigma_{R\times} \\
\dot{\hat{\xi}} = -\Omega_\times \hat{\xi} + \hat{V} - \sigma_\xi \\
\dot{\hat{V}} = -\Omega_\times \hat{V} + \hat{\bar{b}}_g + \hat{R}^\top a_g^m + a_{\mathcal{B}} - \sigma_V \\
\dot{\hat{\bar{b}}}_g = -\Omega_\times \hat{\bar{b}}_g + (\hat{R}^\top a_g^m)_\times \sigma_R - \sigma_b
\end{cases}
$$

with initial conditions $(\hat{R}(0), \hat{\xi}(0), \hat{V}(0), \hat{\bar{b}}_g(0)) \in SO(3) \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$. As for the innovation terms $\sigma_R, \sigma_\xi, \sigma_V, \sigma_b \in \mathbb{R}^3$, they are again derived from the observer design framework.

The bearing measurements of n source points whose inertial coordinates are known yield the output equations (13), as for the PnP problem addressed in subsection III-B.

Define the state variable $x := [x_1^\top, x_2^\top]^\top$, with $x_1 := \tilde{\lambda} \in \mathcal{B}_2^3$ and $x_2 := [\xi^\top, V^\top, \bar{b}_g^\top]^\top \in \mathbb{R}^9$, and the output $y := [y_1^\top, \ldots, y_n^\top]^\top \in \mathbb{R}^{3n}$, with $y_i := \Pi_{p_i}\hat{R}^\top \mathring{P}_i$. Then, from (12), (13), (17), (18) and the second equation of (5), one obtains the system equations (1) with

$$
\begin{cases}
A_{11} = -\Omega_\times, \quad A_{21} = 0_{3\times9}, \qquad u_1 = \sigma_R \\
A_{22} = \begin{bmatrix} -\Omega_\times & I_3 & 0_3 \\ 0_3 & -\Omega_\times & I_3 \\ 0_3 & 0_3 & -\Omega_\times \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0_{3\times1} \\ \hat{R}^\top a_g^m + a_{\mathcal{B}} \\ (\hat{R}^\top a_g^m)_\times \sigma_R \end{bmatrix} \\
C_1 = \begin{bmatrix} -\Pi_{p_1}(\hat{R}^\top \mathring{P}_1)_\times \\ \vdots \\ -\Pi_{p_n}(\hat{R}^\top \mathring{P}_n)_\times \end{bmatrix}, \quad C_2 = \begin{bmatrix} \Pi_{p_1} & 0_3 & 0_3 \\ \vdots & \vdots & \vdots \\ \Pi_{p_n} & 0_3 & 0_3 \end{bmatrix}
\end{cases}
$$

The innovation terms, derived from (3) and (4), are

$$\sigma_R = u_1 = -K_1(y - C_2\hat{x}_2)$$
$$[\sigma_\xi^\top, \sigma_V^\top, \sigma_b^\top]^\top = -K_2(y - C_2\hat{x}_2)$$

Establishing the uniform observability of the matrix pair $(A(0,t), C(0, x_2(t), t))$, which in turn guarantees the (local) uniform exponential stability of the observer, is not a simple task. It is out of scope of the present paper. However, by analogy with the PnP problem, explicit conditions of non-uniform observability are established in [17].

A. Experimental validation

In this subsection the performance of the proposed observer, when using a time-synchronised camera-IMU system, is tested experimentally. Details about the experimental setup, the observer implementation, and tuning measures are given in [17]. In particular, point-feature detection and matching operations are performed with the ArUco library included in OpenCV.

The proposed observer is compared to the popular algebraic algorithm *solvePnP()* in OpenCV, and to the ground-truth data obtained from an Optitrack motion capture system MOCAP.

Figure 2 shows the time evolution of the estimated camera position obtained with the proposed Riccati observer (red), with *solvePnP()* (yellow) and with ground truth data (blue). One observes that the position estimates obtained with the proposed Riccati observer and the *solvePnP()* algorithm are very similar. These estimates are close to the ground-truth values except for the third component, when the distance between the camera and the target increases significantly during the time-interval (85 sec, 140 sec). We attribute this degradation of the estimates to the low resolution of the camera in relation to the ArUco points extraction algorithm.

On the other hand, one observes from Figure 3 showing the time evolution of the estimated and ground-truth attitudes (represented by roll, pitch and yaw Euler angles) that the proposed observer outperforms the *solvePnP()* algorithm, especially when the camera's pose changes quickly. More precisely, when the camera's motion combines a rapid displacement along the third axis (cf. third subplot of Figure 2) and a fast yaw rotation (cf. third subplot of Figure 3) during the time-interval (145 sec, 185 sec), the figure shows that the *solvePnP()* algorithm switches between two possible pitch angle solutions.

V. OBSERVER DESIGN FOR LINEAR VELOCITY, GRAVITY DIRECTION, AND DEPTH ESTIMATION OF A CAMERA OBSERVING A PLANAR TARGET FROM OPTICAL FLOW AND IMU MEASUREMENTS

This classical problem has been revisited in [11] by adopting the proposed observer design framework. A monocular camera and an IMU provide the measurements needed for the estimation process. More precisely, it is assumed that the camera images and IMU are perfectly time synchronised so that one can extract from consecutive images the optical flow $\phi := V/d$, with d the distance between the camera
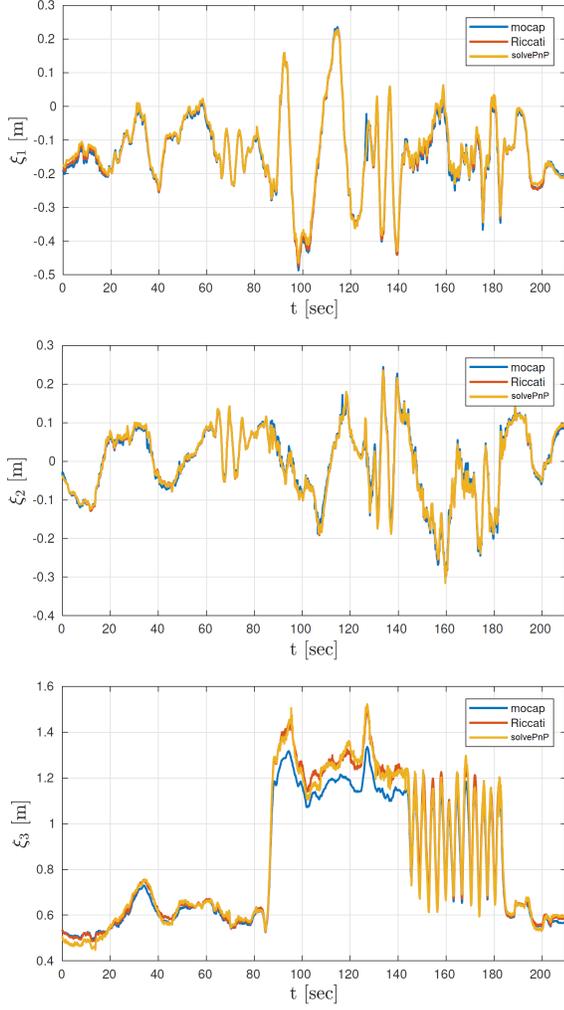
Fig. 2. Camera positions given by the proposed Riccati observer, *solvePnP()* algorithm and MOCAP ground truth.

Fig. 3. Camera attitudes (Euler angles) given by the proposed Riccati observer, *solvePnP()* algorithm and MOCAP ground truth.

and the observed planar scene, and also the flow divergence $\phi^\perp := -\dot{d}/d$ (see, e.g., [11] for more explanations). It is further assumed that the measurements of the angular velocity Ω and of the specific acceleration $a_\mathcal{B}$ provided by the IMU are available at a fast sampling rate.

By denoting $s := 1/d$, the equations of motion involved in the observer design are

$$\begin{cases} \dot{R} = R\Omega_\times \\ \dot{V} = -\Omega_\times V + gR^\top e_3 + a_\mathcal{B} \\ \dot{s} = \phi^\perp s \end{cases} \quad (19)$$

In view of the above equations, the proposed observer has the following form

$$\begin{cases} \dot{\hat{R}} = \hat{R}\Omega_\times - \sigma_{R\times}\hat{R} \\ \dot{\hat{V}} = -\Omega_\times \hat{V} + g\hat{R}^\top e_3 + a_\mathcal{B} - \sigma_V \\ \dot{\hat{s}} = \phi^\perp \hat{s} - \sigma_s \end{cases}$$

with initial conditions $(\hat{R}(0), \hat{V}(0), \hat{s}(0)) \in SO(3) \times \mathbb{R}^3 \times \mathbb{R}^3$, and with $\sigma_R, \sigma_V, \sigma_s \in \mathbb{R}^3$ the innovation terms to be derived from the observer design framework, using the measurement of ϕ.
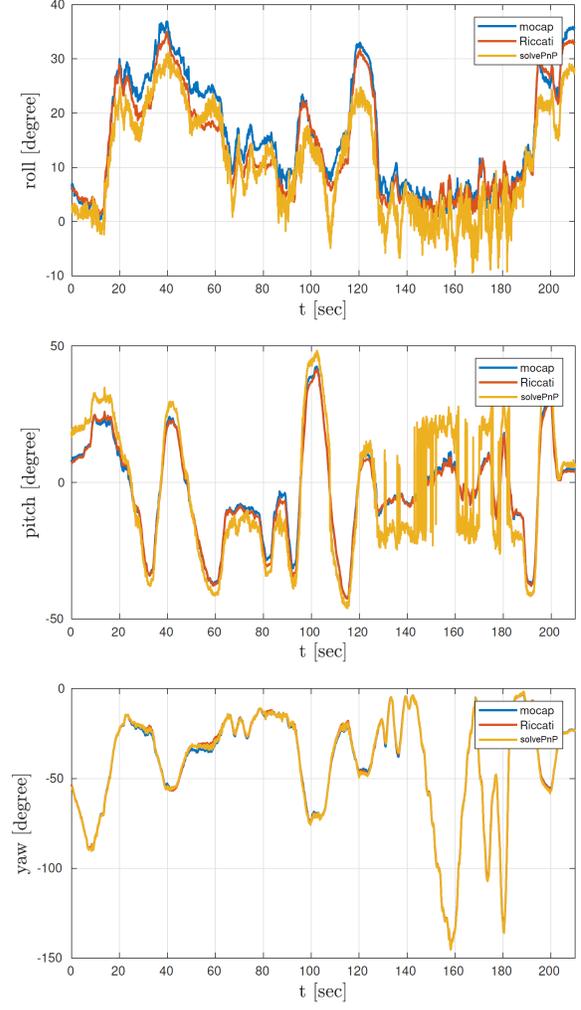
The next step consists in transforming the considered system into the form (1). First, the following attitude error is considered (instead of (9))

$$\tilde{R} := R\hat{R}^\top$$

whose time-derivative is given by (instead of (10))

$$\dot{\tilde{R}} = \tilde{R}\sigma_{R\times}$$

By defining $\tilde{\lambda}$ as twice the vector part of the unit quaternion associated with the attitude error matrix $\tilde{R}$, one verifies that

$$\begin{aligned} \tilde{R} &= I_3 + \tilde{\lambda}_\times + O(|\tilde{\lambda}|^2) \\ \dot{\tilde{\lambda}} &= \sigma_R + O(|\tilde{\lambda}||\sigma_R|) \end{aligned} \quad (20)$$

Rewriting the dynamics of V in (19) using the first-order approximation of $\tilde{R}$, one gets

$$\begin{aligned} \dot{V} &= -\Omega_\times V + g\hat{R}^\top \tilde{R}^\top e_3 + a_\mathcal{B} \\ &= -\Omega_\times V + g\hat{R}^\top (I_3 - \tilde{\lambda}_\times)e_3 + a_\mathcal{B} + O(|\tilde{\lambda}|^2) \\ &= -\Omega_\times V + g\hat{R}^\top e_{3\times}\tilde{\lambda} + g\hat{R}^\top e_3 + a_\mathcal{B} + O(|\tilde{\lambda}|^2) \quad (21) \\ &= -\Omega_\times V + g\hat{R}^\top e_2\tilde{\lambda}_1 - g\hat{R}^\top e_1\tilde{\lambda}_2 \\ &\quad + g\hat{R}^\top e_3 + a_\mathcal{B} + O(|\tilde{\lambda}|^2) \end{aligned}$$

Concerning the measurement of $\phi\ (=\ \frac{V}{d}\ =\ Vs)$, one has

$$\phi = ((V - \hat{V}) + \hat{V})((s - \hat{s}) + \hat{s})$$
$$= \hat{V}s + (V - \hat{V})\hat{s} + O(|V - \hat{V}||s - \hat{s}|)$$
$$= \hat{V}s + (\hat{s}I_3)V - \hat{V}\hat{s} + O(|V - \hat{V}||s - \hat{s}|)$$

By defining the system output as $y := \phi + \hat{V}\hat{s}$, the previous equality rewrites as

$$y = \hat{V}s + (\hat{s}I_3)V + O(|V - \hat{V}||s - \hat{s}|) \qquad (22)$$

Let us set $x := [x_1^\top, x_2^\top]^\top$, with $x_1 := [\tilde{\lambda}_1, \tilde{\lambda}_2]^\top \in \mathcal{B}_2^2$ and $x_2 := [s, V^\top]^\top \in \mathbb{R}^4$. In view of (21), (22), the third equation of (19) and the second equation of (20), one obtains the system equation (1) with

$$\begin{cases} A_{11} = 0_2, \ A_{21} = \begin{bmatrix} 0 & 0 \\ g\hat{R}^\top e_2 & -g\hat{R}^\top e_1 \end{bmatrix}, \ u_1 = \sigma_R \\[2mm] A_{22} = \begin{bmatrix} \phi^\perp & 0_{1\times 3} \\ 0_{3\times 1} & -\Omega_\times \end{bmatrix}, \ u_2 = \begin{bmatrix} 0 \\ g\hat{R}^\top e_3 + a_\mathcal{B} \end{bmatrix} \\[2mm] C_1 = \begin{bmatrix} 0_{3\times 1} & 0_{3\times 1} \end{bmatrix}, \ C_2 = \begin{bmatrix} \hat{V} & \hat{s}I_3 \end{bmatrix} \end{cases}$$

The innovation terms, derived from (3) and (4), are

$$\sigma_R = u_1 = -K_1(y - C_2\hat{x}_2)$$
$$[\sigma_s, \sigma_V^\top]^\top = -K_2(y - C_2\hat{x}_2)$$

A persistent excitation condition, involving only the camera's translational motion, whose satisfaction ensures uniform observability of the matrix pair $(A(0,t), C(0, x_2(t), t))$ and thus (local) uniform exponential stability of the observer, is given in [11].

A. Experimental validation

Experimental validations of the proposed observer has been carried out using a time-synchronised camera-IMU system. Details about the experimental setup, the observer implementation and tuning measures are reported in [11]. Ground-truth data are obtained from an Optitrack motion capture system. The experiment is performed with the camera looking downward to observe an horizontal textured planar pattern on the ground. A demo video of this experiment is available at: `https://youtu.be/R09oTjr4s40`

The time evolutions of the estimated ground-truth attitudes, linear velocities and depths are shown in Figures 4–6. During the first 60 seconds, depth, gravity direction (i.e. roll and pitch Euler angles) and linear velocity estimates approach the corresponding ground-truth values closely, due to the satisfaction of condition of persistent excitation. By contrast, the camera being kept motionless during the time interval (60 sec, 69 sec), the condition of persistent excitation is not satisfied and one can observe that the depth estimate slightly drifts away from the ground-truth value. At the same time, both gravity direction and linear velocity estimates remain close to the corresponding ground-truth values. Once the condition of persistent excitation is again satisfied on the time interval (69 sec, 89 sec), as a consequence of the camera
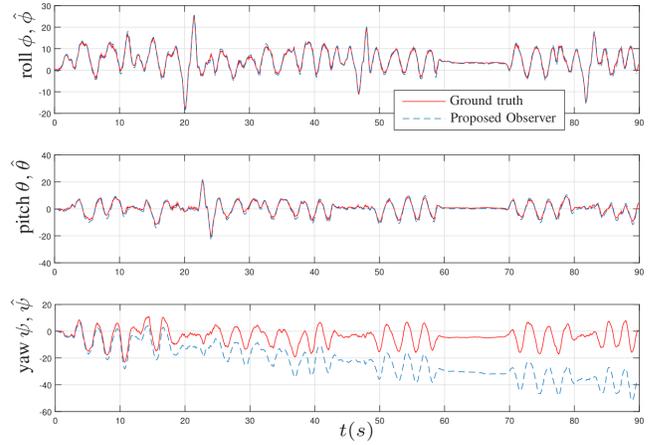

Fig. 4. Estimated and ground-truth attitudes represented by roll, pitch and yaw Euler angles (deg) versus time (s)
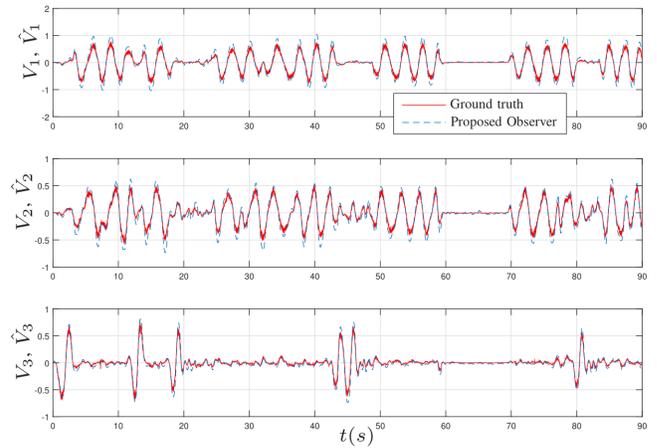

Fig. 5. Estimated and ground-truth linear velocity components in body-fixed frame (m/s) versus time (s)
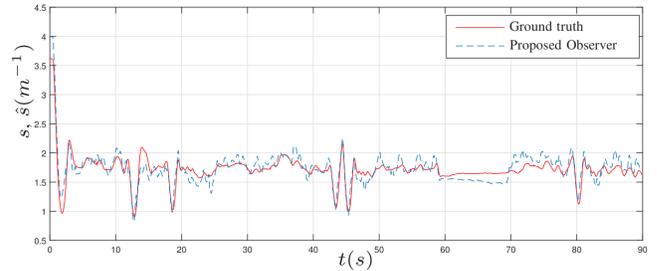

Fig. 6. Estimated and ground-truth depth inverse (m^{-1}) versus time (s)

motion, the depth estimate rejoins the ground-truth value. From Figure 4, the yaw angle estimate drifts away from the ground-truth value. This is expected, and not a shortcoming of the estimation of the state x. Indeed, the yaw angle error corresponds to the third component of $\tilde{\lambda}$ which is not part of the state. Its estimate, obtained via simple integration of the gyros, is for this reason imperfect. To summarize, the reported experiment shows that (practical) convergence of the gravity direction and linear velocity estimates to the ground-truth values is always achieved, whereas the convergence of the depth estimate to the corresponding ground-truth value is obtained only when the condition of persistent excitation

is satisfied.

VI. Conclusions

After briefly recalling a deterministic Riccati observer design framework proposed in [8], which exploits first-order approximations of a generic class of nonlinear systems, and formally generalises the multiplicative extended Kalman filter, the present paper addresses different estimation problems involving proprioceptive sensors and monocular vision to illustrate the applicability of the proposed observer design methodology for a range of estimation problems encountered in practice. The efficiency of the proposed observers for the last two problems are illustrated by means of experiments implementing a time-synchronised camera-IMU system. The proposed methodology is clearly applicable to many other practical problems, such as visual odometry (VO) and visual SLAM (VSLAM). This tutorial also points out the importance of verifying the satisfaction of adequate uniform observability conditions in order to ascertain the good conditioning and good performance of the obtained observers. In this respect, further work to simplify this verification and render it easier will be valuable.

References

[1] N. Gordon A. Doucet, N. de Freitas. *An Introduction to Sequential Monte Carlo Methods*. Springer, 2001.

[2] N. Aghannan and P. Rouchon. An intrinsic observer for a class of lagrangian systems. *IEEE Transactions on Automatic Control*, 48(6):936–945, June 2003.

[3] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[4] A. Barrau and S. Bonnabel. The invariant extended kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4):1797–1812, April 2017.

[5] S. Bonnabel, P. Martin, and P. Rouchon. Non-linear symmetry-preserving observers on lie groups. *IEEE Transactions on Automatic Control*, 54(7):1709–1713, July 2009.

[6] M. Dhome, M. Richetin, J.-T. Lapresté, and G. Rives. The inverse perspective problem from a single view for polyhedra location. *Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition*, pages 61–66, 1988.

[7] J. Farrell. *Aided Navigation: GPS with High Rate Sensors*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 2008.

[8] T. Hamel and C. Samson. Riccati observers for the nonstationary pnp problem. *IEEE Transactions on Automatic Control*, 63(3):726–741, 2017.

[9] R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim. Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1426–1446, Nov 1989.

[10] M.-D. Hua. Attitude estimation for accelerated vehicles using gps/ins measurements. *Control Engineering Practice*, 18(7):723 – 732, 2010. Special Issue on Aerial Robotics.

[11] M.-D. Hua, N. Manerikar, T. Hamel, and C. Samson. Attitude, linear velocity and depth estimation of a camera observing a planar target using continuous homography and inertial data. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1429–1435, 2018.

[12] M.-D. Hua, S. De Marco, T. Hamel, and R. W. Beard. Relative pose estimation from bearing measurements of three unknown source points. In *IEEE Conference on Decision and Control (CDC)*, 2020. submitted.

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

[14] S. J. Julier and J. K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In *Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068, pages 182 – 193. International Society for Optics and Photonics, SPIE, 1997.

[15] Y. Ma, J. Košecká, and S. Sastry. Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision*, 44(3):219–249, 2001.

[16] R. Mahony, J. Trumpf, and T. Hamel. Observers for kinematic systems with symmetry. *IFAC Proceedings Volumes*, 46(23):617 – 633, 2013. 9th IFAC Symposium on Nonlinear Control Systems.

[17] S. De Marco, M.-D. Hua, T. Hamel, and C. Samson. Position, velocity, attitude and accelerometer-bias estimation from imu and bearing measurements. In *European Control Conference (ECC)*, 2020.

[18] F. Markley. Attitude error representations for kalman filtering. *J. Guid. Control Dyn.*, 26(2):311–317, 2003.

[19] P. Martin and E. Salaun. Generalized multiplicative extended Kalman filter for aided attitude and heading reference system. In *AIAA Guidance, Navigation, and Control Conference*, pages 2010–8300, 2010.

[20] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004.