# A Novel Observer Design for Monocular Visual SLAM

Pierre Gintrand, Minh-Duc Hua, Tarek Hamel, Guillaume Varra

# A Novel Observer Design for Monocular Visual SLAM

**Pierre Gintrand** [*,**] **Minh-Duc Hua** [*] **Tarek Hamel** [*]
**Guillaume Varra** [**]

[*] *Laboratory I3S, CNRS, Université Côte d'Azur, Sophia Antipolis, France (e-mail: gintrand(hua,thamel)@i3s.unice.fr)*
[**] *Airbus Helicopters, Marignane, France (e-mail: pierre.gintrand(guillaume.varra)@airbus.com)*

**Abstract:**
A novel cascaded observer design exploiting source points' bearing measurements is proposed for the monocular visual SLAM problem. A distinguishing feature of the present work with respect to most existing visual SLAM algorithms is the decoupling of the camera's pose estimation (i.e., localization) from source points' position estimation (i.e., map building), leading to a straightforward architecture that can handle a vast number of source points efficiently. Furthermore, the persistence of excitation of the camera's translational motion together with the source points' configuration (specified in our prior works) is the key to achieving (local) exponential stability of the camera's pose estimation and, subsequently, overcoming the well-known depth ambiguity associated with the use of a monocular camera. This ingredient has paved the way for the proposed cascaded observer architecture, in which the main contributions concern the design and stability analysis of the three proposed observers for source points' position estimation. Convincing comparative simulation and experimental results are reported to support the proposed approach.

*Keywords:* Nonlinear Observers and Filter Design, Visual SLAM, Sensor Fusion, Navigation

## 1. INTRODUCTION

SLAM (Simultaneous Localization And Mapping) is one of the most fundamental capabilities of autonomous robots to navigate and perform tasks. SLAM algorithms allow for a robotic system to estimate its pose (i.e. position and orientation) and to build its surrounding map at the same time using onboard exteroceptive sensors that sometimes can be fused with inertial sensors such as IMU, Doppler velocity sensors, etc. Visual SLAM (or vSLAM) refers to SLAM for which cameras or other image sensors are used (Cadena et al. (2016)). In particular, monocular vSLAM involving a single camera is of substantial interest and remains a challenging research topic principally due to the unknown depth. It has been extensively studied by the robotics community with classical Extended Kalman Filter (EKF) techniques (Davison et al. (2007); Durrant-Whyte and Bailey (2006)) and modern graph-based optimization methods (Lu and Milios (1997); Grisetti et al. (2010)) encompassing academic and industry-standard solutions. However, both these approaches suffer from significant limitations. For instance, EKF algorithms experience consistency issues along with computational complexity for covariance update (Bailey et al. (2006)) whereas modern graph-based methods require even higher computational cost. More recently, SLAM and vSLAM have started to attract attention from the control systems community (Guerreiro et al. (2013); Barrau and Bonnabel (2017)). The bi-dependence of the map-pose estimation processes makes the SLAM problem highly nonlinear and challenging (Cadena et al. (2016)). The SLAM state space's nonlinearity is commonly resulted from the rigid body's orientation evolving in the compact Lie group $SO(3)$ (van Goor et al. (2019)). Bonnabel *et al.* have introduced the Invariant Extended Kalman Filter (IEKF) that turned out to be competitive for the SLAM problem (Barrau and Bonnabel (2015)). Van Goor *et al.* have developed a nonlinear equivariant filter (van Goor et al. (2020)) exploiting the structure of the SLAM state-space manifold presented in van Goor et al. (2019). In most SLAM and vSLAM works, the pose and the map are estimated at the same time via the design of a single estimator. By contrast, Johansen and Brekke (2016) have separated the camera's orientation estimate from the map-position estimation problem. From there, several cascaded observers depending on the nature of measurements (e.g. range and/or bearing) have been proposed (Bjørne et al. (2017)).

The present paper adopts a new cascaded structure in order to perform monocular vSLAM. First, the pose estimation is carried out independently using the relative pose observer developed in our prior works (Hua et al. (2020); Gintrand et al. (2022)). The mapping is achieved afterwards using one of the three proposed observers for landmarks position estimation. The proposed architecture presents apparent advantages: the pose estimation does not depend on the mapping (i.e. landmarks' position estimation) and each source point is located independently. This structure brings simplicity and robustness; the computational load is more manageable; a larger number of source points can be processed compared to standard

SLAM algorithms; and abnormalities can be detected more easily.

The paper is organized as follows. Mathematical notions used throughout the paper, uniform observability definition and recalls on the Riccati observer framework (Hamel and Samson (2017)) are provided in Section 2. Section 3 defines the system kinematics and introduces the pose SLAM problem. Section 4 recalls the relative pose observer presented in Hua et al. (2020) along with complements on the observability conditions presented by Gintrand et al. (2022), and it presents three different observers for the source point position estimation along with convergence and stability analysis. Section 5 provides a comparative study of the proposed observers in terms of performance and computational efficiency throughout a numerical simulation. Section 6 describes practical implementation to ensure continuity of the pose estimation when visible source points are not the same over time, followed by experimental validation using the Urban Complex Dataset (Jeong et al. (2019)) that provides IMU data, images, and wheel odometry data. Finally, concluding remarks are drawn in Section 7.

## 2. PRELIMINARY MATERIAL

### 2.1 Mathematical notations

The following notations will be used throughout the paper:

- The identity matrix and the null matrix of $\mathbb{R}^{n\times n}$ are respectively denoted by $I_n$ and $0_n$.
- The closed ball in $\mathbb{R}^n$ of radius $r$ is denoted as $\mathbb{B}^n_r$.
- $SO(n)$ denotes the special orthogonal group of order $n$.
- $x_\times$ is the skew-symmetric matrix associated with $x$, i.e. $\forall x, y \in \mathbb{R}^3\ x_\times y = x \times y$.
- $\pi_x := I_3 - xx^\top = -x_\times^2$, with $x \in \mathbb{S}^2$, is the orthogonal projection operator in $\mathbb{R}^3$ onto the two-dimensional vector subspace orthogonal to $x$.

### 2.2 Uniform observability

Consider the following linear time-varying system

$$\begin{cases} \dot{x} = \bar{A}(t)x + \bar{B}(t)u \\ y = \bar{C}(t)x \end{cases}$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^l$, $y \in \mathbb{R}^m$.

*Definition 1.* The pair $(\bar{A}(t), \bar{C}(t))$ is called *uniformly observable* if there exist $\delta, \mu > 0$ such that for all $t \geq 0$:

$$W(t,\delta) := \frac{1}{\delta}\int_t^{t+\delta} \Phi^\top(s,t)\bar{C}^\top(s)\bar{C}(s)\Phi(s,t)ds \geq \mu I_n \quad (1)$$

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

The matrix $W(t,\delta)$ is the so-called *observability Gramian* associated with the pair $(\bar{A}(t), \bar{C}(t))$.

### 2.3 Riccati observer

The observers proposed thereafter are designed using the Riccati observer framework developed by Hamel and Samson (2017). Here is a short recall.

Consider the following class of nonlinear systems with state $x := [x_1^\top, x_2^\top]^\top \in \mathbb{B}^{n_1}_r \times \mathbb{R}^{n_2}$, input $u := [u_1^\top, u_2^\top]^\top \in \mathbb{R}^{n_1+n_2}$, and output $y \in \mathbb{R}^m$:

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

with $A(t)$ a continuous matrix-valued function uniformly bounded with respect to (w.r.t.) $t$ of the form

$$A(t) = \begin{bmatrix} A_{11}(t) & 0_{n_1\times n_2} \\ A_{21}(t) & A_{22}(t) \end{bmatrix}$$

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

Apply the input $u = -Ky$, with $K = [K_1^\top, K_2^\top]^\top := PC^\top Q(t)$ and $P \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ a symmetric positive definite matrix, the solution to the Continuous Riccati Equation (CRE)

$$\dot{P} = AP + PA^\top - PC^\top Q(t)CP + V(t)$$

where $P(0)$ is positive definite, and $Q$ and $V$ are bounded continuous positive semi-definite. Then the Corollary 3.2 by Hamel and Samson (2017) shows that the equilibrium $x = 0$ is locally uniformly exponentially stable when $Q(t)$ and $V(t)$ are both larger than some positive matrix and the pair $(A^\star(t), C^\star(t)) := (A(t), C(0,t))$ is uniformly observable.

## 3. PROBLEM STATEMENT

Let $\mathcal{A}$ be the moving camera-fixed frame, and $\mathring{\mathcal{A}}$ be the reference configuration of this frame. The relative orientation of $\mathcal{A}$ w.r.t. $\mathring{\mathcal{A}}$ is denoted as $R \in SO(3)$ and the relative position of $\mathcal{A}$ w.r.t. $\mathring{\mathcal{A}}$, expressed in $\mathcal{A}$ (resp. in $\mathring{\mathcal{A}}$) is denoted as $\xi \in \mathbb{R}^3$ (resp. $\xi^\circ = R\xi$). The vector $\Omega \in \mathbb{R}^3$ denotes the angular velocity of $\mathcal{A}$ w.r.t. $\mathring{\mathcal{A}}$ expressed in $\mathcal{A}$ and the vector $v \in \mathbb{R}^3$ represents the linear velocity of the camera origin expressed in $\mathcal{A}$. Both velocities $\Omega$ and $v$ are assumed to be measured using gyrometers and a Doppler sensor, respectively.

The dynamics of the camera pose $(R, \xi)$ are given by

$$\begin{cases} \dot{R} = R\Omega_\times \\ \dot{\xi} = -\Omega_\times \xi + v \end{cases} \quad (3)$$

The problem under consideration focuses on the pose SLAM from a monocular camera observing $N$ fixed unknown source points using bearing measures. Let $p_i := \frac{P_i}{|P_i|} = \frac{p_i^p}{|p_i^p|} \in S^2$ (resp. $p_i^\circ := \frac{P_i^\circ}{|P_i^\circ|} = \frac{p_i^{\circ p}}{|p_i^{\circ p}|} \in S^2$) denote the bearing that one can get from the perspective outputs $p_i^p$ (resp. $p_i^{\circ p}$) commonly used in computer vision literature.

Using the relation $P_i = R^\top P_i^\circ - \xi$, one deduces for each source points $i$ the following *epipolar* constraint

$$p_i^{\circ\top} R\xi_\times p_i = 0 \quad (4)$$

## 4. OBSERVER DESIGN FOR THE POSE SLAM

This section presents novel cascaded observers for the pose SLAM problem. A Riccati observer is first designed for the relative pose problem, followed by N parallel observers to estimate the position of the source points.

### 4.1 Pose Observer design

Let $(\hat{R}, \hat{\xi})$ denote an estimate of the pose $(R, \xi)$. The proposed observer has the following form

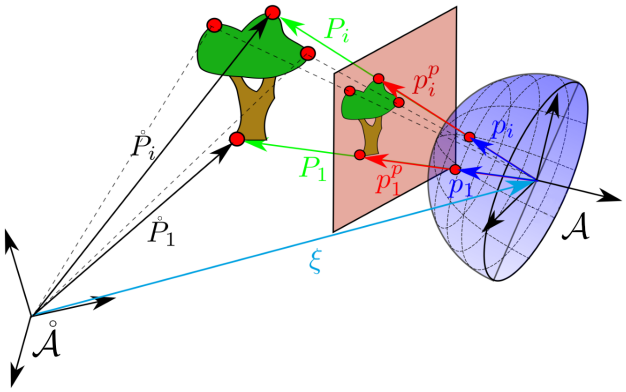


Fig. 1. Intuitive representation of inertial coordinates $P_i^\circ$, planar projective coordinates $p_i^p$ and spherical projective coordinates $p_i$ of the $i$th source point.

$$\begin{cases} \dot{\hat{R}} = \hat{R}\Omega_\times - \hat{R}\sigma_{R\times} \\ \dot{\hat{\xi}} = -\Omega_\times\hat{\xi} + v - \sigma_\xi \end{cases} \tag{5}$$

with initial conditions $\hat{R}(0) \in SO(3)$ and $\hat{\xi}(0) \in \mathbb{R}^3$, and with $\sigma_R$, $\sigma_\xi \in \mathbb{R}^3$ innovation terms to be designed.

Define the attitude error $\tilde{R} := \hat{R}^\top R$ and the position error $\tilde{\xi} := \xi - \hat{\xi}$. From (3) and (5), one deduces

$$\dot{\tilde{R}} = -\Omega_\times\tilde{R} + \tilde{R}\Omega_\times + \sigma_{R\times}\tilde{R}$$

From the Rodrigues' formula, the first-order approximation of $\tilde{R}$ is given by

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

with $\tilde{\lambda} \in \mathbb{B}_2^3$ equal to twice the vector part of the quaternion associated with $\tilde{R}$. One deduces from the two previous relations that

$$\begin{bmatrix} \dot{\tilde{\lambda}} \\ \dot{\tilde{\xi}} \end{bmatrix} = \begin{bmatrix} -\Omega_\times & 0_3 \\ 0_3 & -\Omega_\times \end{bmatrix} \begin{bmatrix} \tilde{\lambda} \\ \tilde{\xi} \end{bmatrix} + \begin{bmatrix} \sigma_R + O(|\tilde{\lambda}||\sigma_R|) + O(|\tilde{\lambda}|^2) \\ \sigma_\xi \end{bmatrix} \tag{6}$$

By considering the epipolar constraint (4) and using the fact that $R = \hat{R}\tilde{R}$ and $\xi = \hat{\xi} + \tilde{\xi}$, one gets

$$\begin{aligned} 0 &= {p_i^\circ}^\top\hat{R}\tilde{R}(\hat{\xi} + \tilde{\xi})_\times p_i \\ &= {p_i^\circ}^\top\hat{R}\hat{\xi}_\times p_i + {p_i^\circ}^\top\hat{R}\tilde{\lambda}_\times\hat{\xi}_\times p_i + {p_i^\circ}^\top\hat{R}\tilde{\xi}_\times p_i + O(|\tilde{\lambda}||\tilde{\xi}|) + O(|\tilde{\lambda}|^2) \end{aligned}$$

By setting $y_i = {p_i^\circ}^\top\hat{R}\hat{\xi}_\times p_i$, one deduces

$$y_i = {p_i^\circ}^\top\hat{R}(\hat{\xi}_\times p_i)_\times\tilde{\lambda} + {p_i^\circ}^\top\hat{R}{p_i}_\times\tilde{\xi} + O(|\tilde{\lambda}||\tilde{\xi}|) + O(|\tilde{\lambda}|^2) \tag{7}$$

From (5)-(7), one obtains the compact form of the Riccati observer (2) with

$$\begin{cases} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} := \begin{bmatrix} \tilde{\lambda} \\ \tilde{\xi} \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} := \begin{bmatrix} \sigma_R \\ \sigma_\xi \end{bmatrix}, \\ A := \begin{bmatrix} -\Omega_\times & 0_3 \\ 0_3 & -\Omega_\times \end{bmatrix}, \; y := \begin{bmatrix} {p_1^\circ}^\top\hat{R}\hat{\xi}_\times p_1 \\ \vdots \\ {p_N^\circ}^\top\hat{R}\hat{\xi}_\times p_N \end{bmatrix}, \\ C_1 := \begin{bmatrix} {p_1^\circ}^\top\hat{R}(\hat{\xi}\times p_1)_\times \\ \vdots \\ {p_N^\circ}^\top\hat{R}(\hat{\xi}\times p_N)_\times \end{bmatrix}, \; C_2 := \begin{bmatrix} {p_1^\circ}^\top\hat{R}{p_1}_\times \\ \vdots \\ {p_N^\circ}^\top\hat{R}{p_N}_\times \end{bmatrix} \end{cases} \tag{8}$$

The innovation terms involved in (5) can then be deduced:

$$\sigma_R = -K_1 y, \; \sigma_\xi = -K_2 y \tag{9}$$

with $K_1$ and $K_2$ computed according to Subsection 2.3.

Comprehensive observability analysis of the above pose observer has been carried out in (Gintrand et al. (2022)). It is mainly based on the persistence of excitation (PE) of the translation motion. By defining the matrix

$$\Pi_i(t,\delta) := \frac{1}{\delta}\int_t^{t+\delta} \frac{\xi^\circ(s){\xi^\circ}^\top(s)}{|P_i^\circ - \xi^\circ(s)|^2} ds$$

and $\lambda_1^i(t,\delta)$, $\lambda_2^i(t,\delta)$, $\lambda_3^i(t,\delta)$ denoting the eigenvalues of $\Pi_i(t,\delta)$ in increasing order (i.e. $\lambda_1^i(t,\delta) \leq \lambda_2^i(t,\delta) \leq \lambda_3^i(t,\delta)$), one defines the PE as follows:

*Definition 2.* (Strong and weak persistent excitation). The camera translational motion is called *strongly persistently exciting* if there exist $\delta, \beta > 0$ such that for all $t \geq 0$: $\lambda_1^i(t,\delta) \geq \beta$. It is called *weakly persistently exciting*, if $\text{rank}(\Pi_i(t,\delta)) = 2$ and $\lambda_2^i(t,\delta) \geq \beta$.

*Proposition 3.* (see Gintrand et al. (2022)). Define 4 degenerated source points configurations as follow:

(1) The number of observed source points $n \leq 2$.
(2) The observed source points are aligned ($n \geq 3$).
(3) There are three non-aligned source points ($n = 3$) and the origin of the reference frame $\mathring{\mathcal{A}}$ belongs to the *danger cylinder* [1] generated by the 3 source points.
(4) There are at least four non-aligned source points ($n \geq 4$) located on a *horopter curve* [2] whose origin coincides with the origin of $\mathring{\mathcal{A}}$.

One verifies the 3 following statements:

(1) If one of the above conditions is fulfilled, then $(A^*, C^*)$ cannot be uniformly observable.
(2) If the camera motion is strongly PE and if none of the above conditions is satisfied, then $(A^*, C^*)$ is uniformly observable.
(3) If the camera motion is only weakly PE, then $(A^*, C^*)$ is uniformly observable if none of the above conditions is granted and if there are at least 3 bearings not orthogonal to $\ker(\Pi_i(t,\delta))$.

When the uniform observability is granted, one ensures that the equilibrium $(I_3, 0)$ of relative pose error $(\tilde{R}, \tilde{\xi})$ is locally exponentially stable.

### *4.2 Landmark position observers*

We consider now the second fundamental task of the pose SLAM problem which consists in estimating the source points' location in the reference frame $P_i^\circ$, $i = \{1, \ldots, N\}$.

*Definition 4.* The bearing $\bar{p}_i := Rp_i$ is called PE if there exist $\varepsilon, \delta > 0$ such that for all $t > 0$

$$\frac{1}{\delta}\int_t^{t+\delta} \pi_{\bar{p}_i(s)} ds \geq \varepsilon I_3 \tag{10}$$

Using the fact that $\dot{P}_i^\circ = 0$ and the output measure which is either explicitly given by $y_i = \pi_{\bar{p}_i}\xi^\circ$ or implicitly by $\pi_{\bar{p}_i}(\xi^\circ - P_i^\circ) = 0$, one verifies that the left-hand side term of (10) is nothing more than the Gramian.

*Lemma 5.* If the camera motion is either strongly PE or weakly PE such that $\bar{p}_i$ is uniformly non-collinear with $\xi^\circ$

[1] The danger cylinder is a circular cylinder generated by the circle passing through three source points whose axis is orthogonal to the plane containing the source points.
[2] A horopter curve is the intersection of a circular cylinder and an elliptic cone.

(i.e. there exists $\epsilon > 0$ such that $|\pi_{\bar{p}_i}\xi^\circ| \geq \epsilon$), then $\bar{p}_i$ is PE.

**Proof.** The critical situation consists in having a weakly PE linear camera motion. That is a linear motion that does not contain the origin of frame $\mathring{\mathcal{A}}$ and for which it is straightforward to verify that $\bar{p}_i$ is PE (according to Def. 4) if the source point $P_i^\circ$ is not in the direction of motion.

*Theorem 6.* Consider the system dynamics (3) and the relative pose observer (5) along with the innovation terms given by (9). Consider also the following source point observer:

$$\dot{\hat{P}}_i^\circ = k\pi_{\hat{R}p_i}(\hat{\xi}^\circ - \hat{P}_i^\circ), \text{ with } i = 1, \dots, N \tag{11}$$

with $k > 0$, $\hat{P}^\circ(0) \in \mathbb{R}^3$ and $\hat{\xi}^\circ := \hat{R}\hat{\xi}$. Assume that all source points $P_i^\circ \in \mathring{\mathcal{A}}$ are static, and none of Proposition 3's conditions is fulfilled. Assume also that the condition of Lemma 5 holds. Then, one ensures that

(1) $\hat{R}p_i$ is PE;
(2) the equilibrium $(\tilde{R}, \tilde{\xi}, \tilde{P}_i^\circ, \cdots, \tilde{P}_N^\circ) = (I_3, 0, 0, \cdots, 0)$ of the corresponding cascaded error system, with $\tilde{P}_i^\circ := P_i^\circ - \hat{P}_i^\circ$, is locally exponentially stable.

**Proof.** Using the fact that $(\tilde{R}, \tilde{\xi}) = (I_3, 0)$ is uniformly locally exponentially stable, one ensures that $(\bar{R}, \tilde{\xi}^\circ) := (\hat{R}\tilde{R}^\top\hat{R}^\top, \hat{R}\xi - \xi^\circ) = (I_3, 0)$ is also uniformly locally exponentially stable.

Now, since $\bar{p}_i$ is PE and using the fact that $\pi_{\hat{R}p_i} = \pi_{\bar{R}\bar{p}_i}$ along with the fact that $\bar{R} = I_3 - \bar{\lambda}_\times + O(|\bar{\lambda}|^2)$ with $\bar{\lambda} = \hat{R}\tilde{\lambda}$ one verifies

$$\frac{1}{\delta}\int_t^{t+\delta} \pi_{\hat{R}p_i(s)} ds = \frac{1}{\delta}\int_t^{t+\delta} (\pi_{\bar{p}_i(s)} - [\bar{\lambda}_\times(s), \pi_{\bar{p}_i(s)}] + O(|\bar{\lambda}|^2) ds \tag{12}$$

with $[\bar{\lambda}_\times, \pi_{\bar{p}_i}] = \bar{\lambda}_\times\pi_{\bar{p}_i} - \pi_{\bar{p}_i}\bar{\lambda}_\times$. Since $\bar{\lambda}$ is uniformly converging to zero exponentially, by choosing $t$ sufficiently large one ensures that $\exists \epsilon > 0$ such that $|\bar{\lambda}| < \frac{\epsilon}{4}$ and hence $\frac{1}{\delta}\int_t^{t+\delta} \pi_{\hat{R}p_i(s)} ds \geq \frac{\epsilon}{2} I_3$ which implies that $\hat{R}p_i$ is also PE. Then by direct application of (Le Bras et al., 2017, Lemma 5) one concludes that $\hat{P}_i^\circ$ is bounded. Combining (11) with the fact that $\pi_{\hat{R}p_i} = \pi_{\bar{R}\bar{p}_i} = \bar{R}\pi_{\bar{p}_i}\bar{R}^\top$ along with $\pi_{\bar{p}_i}(\xi^\circ - P_i^\circ) = 0$ and $\dot{P}_i^\circ = 0$, one verifies

$$\begin{aligned}
\dot{\tilde{P}}_i^\circ &= -k\bar{R}\pi_{\bar{p}_i}\bar{R}^\top(\hat{\xi}^\circ - \hat{P}_i^\circ) \\
&= -k(I_3 - \bar{\lambda}_\times)\pi_{\bar{p}_i}(I_3 + \bar{\lambda}_\times)(\hat{\xi}^\circ - \hat{P}_i^\circ) + O(|\bar{\lambda}|^2) \\
&= -k\pi_{\bar{p}_i}(\hat{\xi}^\circ - \hat{P}_i^\circ) + k[\bar{\lambda}_\times, \pi_{\bar{p}_i}](\hat{\xi}^\circ - \hat{P}_i^\circ) + O(|\bar{\lambda}|^2) \\
&= k\pi_{\bar{p}_i}\tilde{P}_i^\circ - k\pi_{\bar{p}_i}\tilde{\xi}^\circ + k[\bar{\lambda}_\times, \pi_{\bar{p}_i}](\hat{\xi}^\circ - \hat{P}_i^\circ) + O(|\bar{\lambda}|^2)
\end{aligned}$$

Now, using the fact that $(\bar{\lambda}, \tilde{\xi}^\circ) = (0, 0)$ is exponentially stable while $\hat{\xi}^\circ$ and $\hat{P}_i^\circ$ are uniformly bounded, direct application of (Le Bras et al., 2017, Proposition 1) shows that $\tilde{P}_i^\circ := P_i^\circ - \hat{P}_i^\circ$ converges exponentially to zero. This allows one to conclude the proof.

The interest of this result lies in the extreme simplicity of the observer design. However, it is not the best option in terms of performance since the PE property of $\bar{p}_i$ is only exploited in the proof. We hence propose two alternative options that may significantly improve the source point estimation performance as shown by simulation results.

*Proposition 7.* Consider the statement of Theorem 6 along with the following observer instead of (11):

$$\dot{\hat{P}}_i^\circ = K_i\pi_{\hat{R}p_i}(\hat{\xi}^\circ - \hat{P}_i^\circ), \text{ with } i = 1, \dots, N \tag{13}$$

with $\hat{P}_i^\circ(0) \in \mathbb{R}^3$, $K_i := M_i\pi_{\hat{R}p_i}Q$, and $M_i$ solution to the Riccati equation $\dot{M}_i = -M_i\pi_{\hat{R}p_i}Q\pi_{\hat{R}p_i}M_i + V$, $M_i(0)$, $Q$ and $V$ positive definite matrices. Then, one ensures that

(1) $M_i(t)$ is a positive definite matrix, with bounded condition numbers;
(2) the equilibrium $(\tilde{R}, \tilde{\xi}, \tilde{P}_i^\circ, \cdots, \tilde{P}_N^\circ) = (I_3, 0, 0, \cdots, 0)$ of the corresponding cascaded error system, with $\tilde{P}_i^\circ := P_i^\circ - \hat{P}_i^\circ$, is locally exponentially stable.

**Proof.** The proof of Property 1) is a direct application of (Hamel and Samson, 2016, Lemma 2.1) using the fact that $\hat{R}p_i$ is also PE. Now, analogously to the proof of Theorem 6, one has:

$$\dot{\tilde{P}}_i^\circ = -K_i\pi_{\bar{p}_i}\tilde{P}_i^\circ + K_i\pi_{\bar{p}_i}\tilde{\xi}^\circ - K_i[\bar{\lambda}_\times, \pi_{\bar{p}_i}](\hat{\xi}^\circ - \hat{P}_i^\circ) + O(|\bar{\lambda}|^2)$$

Consider the following storage function $\mathcal{L}_i = \frac{1}{2}\tilde{P}_i^{\circ\top}M_i^{-1}\tilde{P}_i^\circ$. One verifies that

$$\dot{\mathcal{L}}_i = -\frac{1}{2}\tilde{P}_i^{\circ\top}(\pi_{\bar{p}_i}Q\pi_{\bar{p}_i} + M_i^{-1}VM_i^{-1})\tilde{P}_i^\circ + \tilde{P}_i^{\circ\top}\varepsilon(\bar{\lambda}, \tilde{\xi}^\circ)$$

where $\varepsilon(\bar{\lambda}, \tilde{\xi}^\circ)$ is a bounded exponentially vanishing term. From there, one concludes that $\tilde{P}_i^\circ$ locally exponentially converges to zero. This allows one to conclude the proof.

The second suggestion here proposed directly exploits the measured source point Gramian (12) in the observer design.

*Proposition 8.* Consider the statement of Theorem 6 along with the following observer instead of (11):

$$\dot{\hat{P}}_i^\circ(t) = -k\hat{P}_i^\circ(t) + \frac{k}{T}W_i(t)^{-1}\int_{t-T}^t \pi_{\hat{R}p_i(s)}\hat{\xi}^\circ(s)ds$$

with $i \in \{1, \dots, N\}$, $k > 0$, $\hat{P}^\circ(0) \in \mathbb{R}^3$, and $W_i(t) := \frac{1}{T}\int_{t-T}^t \pi_{\hat{R}p_i(s)}ds$ the measured Gramian. Then, the equilibrium $(\tilde{R}, \tilde{\xi}, \tilde{P}_i^\circ, \cdots, \tilde{P}_N^\circ) = (I_3, 0, 0, \cdots, 0)$ of the corresponding cascaded error system, with $\tilde{P}_i^\circ := P_i^\circ - \hat{P}_i^\circ$, is locally exponentially stable.

**Proof.** Direct application of Theorem 6 shows that the measured Gramian is a well-defined positive definite matrix, with bounded condition numbers. From there, and using the fact that $\pi_{\hat{R}\bar{p}_i}\bar{R}(\xi^\circ - \bar{R}P_i^\circ) = 0$, one shows that:

$$\begin{aligned}
\dot{\tilde{P}}_i^\circ &= k\hat{P}_i^\circ(t) - \frac{k}{T}W_i(t)^{-1}\int_{t-T}^t \pi_{\hat{R}p_i(s)}\hat{\xi}^\circ(s)ds \\
&= \frac{k}{T}W_i^{-1}\int_{t-T}^t \pi_{\hat{R}p_i(s)}(\hat{P}_i^\circ - \hat{\xi}^\circ(s) + \bar{R}(\xi^\circ(s) - P_i^\circ))ds \\
&= -\frac{k}{T}W_i^{-1}\int_{t-T}^t \pi_{\hat{R}p_i(s)}(\tilde{P}_i^\circ(t))ds + \frac{k}{T}W_i^{-1}\varepsilon(\bar{\lambda}, \tilde{\xi}^\circ, t) \\
&= -k\tilde{P}_i^\circ(t) + \frac{k}{T}W_i^{-1}\varepsilon(\bar{\lambda}, \tilde{\xi}^\circ, t)
\end{aligned}$$

One verifies that $\varepsilon(\bar{\lambda}, \tilde{\xi}^\circ, t) = \int_{t-T}^t \pi_{\hat{R}p_i(s)}(\tilde{\xi}^\circ(s) - (\bar{\lambda}_\times - O(|\bar{\lambda}|^2))(\xi^\circ(s) - P_i^\circ))ds$ is a bounded exponentially vanishing term and, consequently, $\tilde{P}_i^\circ$ converges exponentially to zero. This allows one to conclude the proof.

Note that above Gramian-based point source observer is computationally more efficient compared to the Riccati-based point source observer proposed in Proposition 7. It

is also more reliable to detect observable and unobservable source points from the direct computation of the measured Gramian.

## 5. SIMULATIONS

In this section, the three source point position observers proposed in Section 4.2 are compared through numerical simulations. The camera performs the periodic planar motion defined by $\mathring{\xi}(t) = [8\sin(\pi t/4),\ 12\sin(\pi t/3),\ 0]^\top$. The camera's angular velocity is set to $\Omega(t) = \frac{\pi}{180}[5\cos(t),\ 10\cos(2t),\ 45\cos(2t)]^\top$.

We consider 5 source points spread in the 3D space: $P_1^\circ = [-6,-3,-3]^\top$, $P_2^\circ = [0,-2.5,0]^\top$, $P_3^\circ = [3,-3,-4]^\top$, $P_4^\circ = [-2,-5,-2]^\top$ and $P_5^\circ = [-2,-4,-5]^\top$ in order to ensure the observability. The parameters of the source point observers are described in Table 1. Simulations have been performed using Matlab. For the two first observers, *ode45* solver has been used, whereas, for the third one, *dde23* solver has been chosen to manage the propagation of variables with a constant delay. Figure 2 shows the estimation error for the landmark $P_1^\circ$ using the three observers. It points out that the Riccati-based and Gramian-based observers outperform the constant gain observer in terms of convergence rate. In contrast, the Riccati-based observer is much more demanding in computation. The comparison was performed using sampled versions of the observers at $0.001s$ and by taking the average over 20 iterations. The constant gain observer is less computationally demanding since it requires $8.8s$ to compute the relative pose and the five landmarks position of a $50s$ simulation. The Riccati-based observer takes $13.2s$, and the Gramian-based observer $12.3s$.

| | |
|---|---|
| Theo. 6 observer | $k = 1$ |
| Prop. 7 observer | $M(0) = 100I_3,\ V = 0.1I_3,\ Q = 100I_5$ |
| Prop. 8 observer | $k = 10,\ T = 0.2s$ |

Table 1. Parameters of the observers

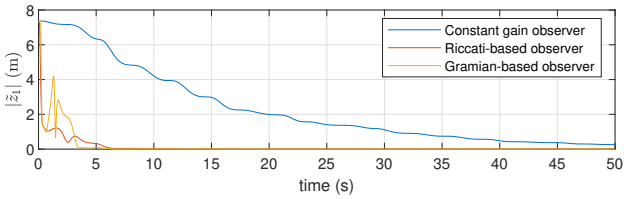


Fig. 2. Estimation error of the landmark position $P_1^\circ$ versus time (s) for the three observers

## 6. EXPERIMENTAL RESULTS

### 6.1 Handling landmarks appearance and disappearance

The relative pose observer (5) along with (9) supposes that every single bearing is available for all time $t$. However, in practice, source points are typically visible only for a short period $[t_k, t_{k+1}]$, and hence the measure equation (8) should be modified accordingly. Define $S_k$ the set of source points available from $t_k$ to $t_{k+1}$, $R_k := R(t_k)$, $\xi_k := \xi(t_k)$, $\xi_{kt} := \xi(t) - R^\top R_k \xi_k$ and $p_{ik} := p_i(t_k)$. Let us express the epipolar constraint verified by each points of the set $S_k$ between time $t_k$ and $t \in [t_k, t_{k+1}]$

$$
\begin{aligned}
0 &= p_{ik}^\top (R_k^\top R)(\xi_{kt})_\times p_i \\
&= p_{ik}^\top R_k^\top R \xi_\times p_i - p_{ik}^\top \xi_{k\times} R_k^\top R p_i \\
&= p_{ik}^\top R_k^\top \hat{R}\hat{\xi}_\times p_i + p_{ik}^\top R_k^\top \hat{R}\tilde{\xi}_\times p_i + p_{ik}^\top R_k^\top \hat{R}\tilde{\lambda}_\times \hat{\xi}_\times p_i \\
&\quad - p_{ik}^\top \xi_{k\times} R_k^\top \hat{R} p_i - p_{ik}^\top \xi_{k\times} R_k^\top \hat{R}\tilde{\lambda}_\times p_i + O(|\tilde{\lambda}|^2) + O(|\tilde{\lambda}||\tilde{\xi}^\circ|)
\end{aligned}
$$

In this process, we assume that the relative estimated pose at time $t_k$ is sufficiently accurate to approximate $(\hat{R}_k, \hat{\xi}_k) \approx (R_k, \xi_k)$. Then, vector $y$ and matrix $C$ in (8) are adjusted as follows:

$$
y = \begin{bmatrix} p_{1k}^\top (\hat{R}_k^\top \hat{R}\hat{\xi}_\times - \hat{\xi}_{k\times}\hat{R}_k^\top \hat{R}) p_1 \\ \vdots \\ p_{Nk}^\top (\hat{R}_k^\top \hat{R}\hat{\xi}_\times - \hat{\xi}_{k\times}\hat{R}_k^\top \hat{R}) p_N \end{bmatrix}
$$

$$
C = \begin{bmatrix} p_{1k}^\top (\hat{R}_k^\top \hat{R}(\hat{\xi}_\times p_1)_\times - \hat{\xi}_{k\times}\hat{R}_k^\top \hat{R} p_{1\times}) & p_{1k}^\top \hat{R}_k^\top \hat{R} p_{1\times} \\ \vdots & \vdots \\ p_{Nk}^\top (\hat{R}_k^\top \hat{R}(\hat{\xi}_\times p_N)_\times - \hat{\xi}_{k\times}\hat{R}_k^\top \hat{R} p_{N\times}) & p_{Nk}^\top \hat{R}_k^\top \hat{R} p_{N\times} \end{bmatrix}
$$

to account for points of the set $S_k$ and ensure continuity of the relative pose estimation with limited drift as long as the uniform observability conditions are granted.

### 6.2 Urban Complex Dataset

The Urban Complex Dataset (Jeong et al. (2019)), which provides the required data to test the proposed observer design methodology, is used. It provides the linear velocity of the vehicle through wheel encoders at a frequency of 100Hz together with the angular velocity at 200Hz. The images are captured with a global shutter camera at 10Hz. The ground truth of the dataset, which has been computed using a pose-graph SLAM algorithm, is provided.

The main drawback of using this dataset is that the car moves among objects that are not necessarily stationary (e.g., other moving cars and pedestrians) while our observer design assumes that all visible points to be fixed.To overcome this issue, only the image's upper part, typically related to static objects such as trees, buildings, and traffic lights, is considered. Figure 3 shows the original and truncated images.

(a) Original image

(b) Truncated image with $N = 213$ features drawn in red used by the observer

Fig. 3. Original image captured by the camera and the truncated image

The features are detected with the well-known SURF algorithm, and tracked using Kanade-Lucas-Tomasi (KLT)

algorithm. Initially, 300 features are identified in the image; when less than 150 features remain after the tracking, new ones are detected and tracked again. The constant bias of the gyroscope is computed by averaging the gyros measures when the IMU is at rest. The prediction loop runs at 100Hz, whereas the estimation loop operates at 10Hz. The following parameters have been chosen for the relative pose observer $Q = 10^{-5} I_N$, $V = \mathrm{diag}(I_3, 0.01 I_3)$ and $P_0 = 1000 I_6$ while the Gramian-based observer has been chosen to perform the environment mapping. This latter ignores unobservable points, that is, all points with an ill-conditioning measured Gramian.

Figure 4 shows the vehicle's trajectory during the beginning of the sequence #28 of the Urban Complex Dataset. The estimated and ground truth position of the car versus time is represented in Figure 5, and Figure 6 shows the heading estimation error.

Figure 7 compares the proposed observer against the Lidar-based algorithm proposed by Cho et al. (2020) and LeGO-LOAM (Shan and Englot (2018)). The pose estimation errors have been collected for multiple segments of 5 different lengths (1100m, 2200m, 3300m, 4400m and 5500m) and statistic data are displayed through a box plot. Although our observer deviates over time due to measurement noise, it outperformed the two Lidar-based odometer algorithms for both pose and heading estimation.

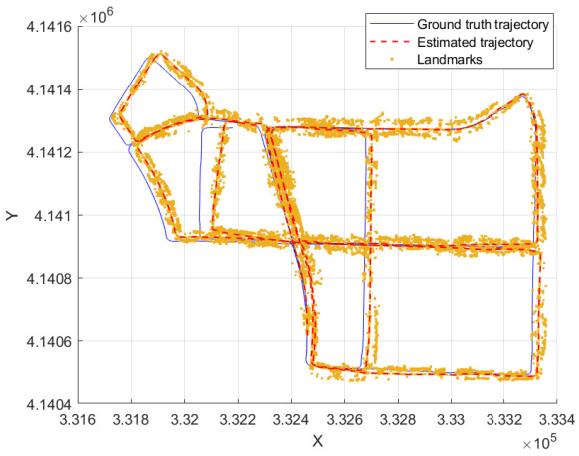


Fig. 4. Estimated and ground truth trajectory of the vehicle and landmarks estimated positions using the sequence #28 of the Urban Complex Dataset

## 7. CONCLUSIONS

In this paper, a novel observer design for the monocular VSLAM problem has been presented. In the first step, the pose estimation is carried out by a nonlinear observer (Hua et al. (2020)), which does not require any prior knowledge about the landmarks' position. Furthermore, the conditions under which the pose is not uniformly observable have been characterized in terms of the camera's translational motion and the number and position of the source points. In the second step, three observers are proposed to map the environment, and their convergence conditions have been specified. We show that the Gramian-based observer is particularly interesting. It is computationally efficient and directly allows the detection of the observable source points from the unobservable ones.

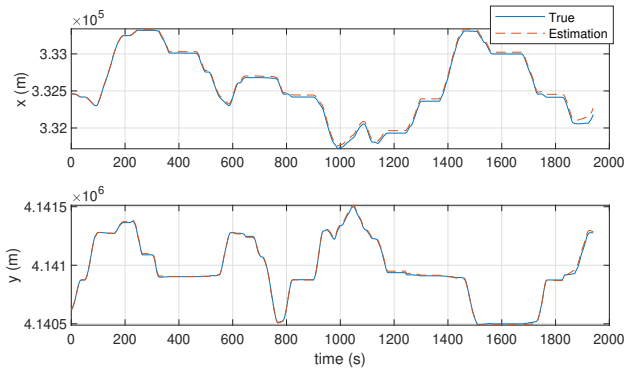


Fig. 5. Real and estimated positions (expressed in the reference frame $\mathring{\mathcal{A}}$) versus time (s)

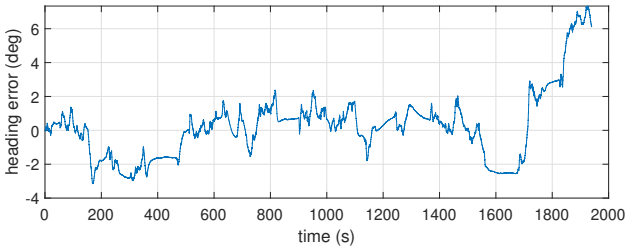


Fig. 6. Heading estimation error (deg) versus time (s)

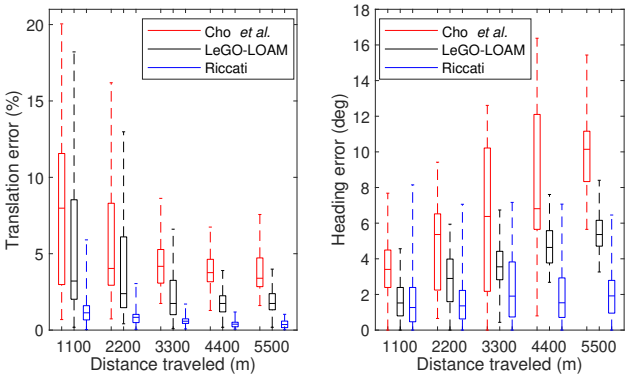


Fig. 7. Relative estimation error of 3 odometry algorithms for the sequence #28 of the Urban Complex Dataset. The box plot represents statistical data from the collection of error for all the sub-trajectory: the minimum and maximum (whiskers), the 1st and 3rd quartile (box) and the median (center line)

*Acknowledgements :* This work was supported by Airbus Helicopters and by the French ANR Astrid CONGRE (ANR-18-ASTR-0006).

## REFERENCES

Bailey, T., Nieto, J., Guivant, J., Stevens, M., and Nebot, E. (2006). Consistency of the EKF-SLAM Algorithm. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3562–3568.

Barrau, A. and Bonnabel, S. (2017). The Invariant Extended Kalman Filter as a Stable Observer. *IEEE Transactions on Automatic Control*, 62(4), 1797–1812.

Barrau, A. and Bonnabel, S. (2015). An EKF-SLAM algorithm with consistency properties.

Bjørne, E., Brekke, E.F., and Johansen, T.A. (2017). Cascade attitude observer for the SLAM filtering problem.

In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, 945–952.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., and Leonard, J.J. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6), 1309–1332.

Cho, Y., Kim, G., and Kim, A. (2020). Unsupervised Geometry-Aware Deep LiDAR Odometry. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2145–2152.

Davison, A.J., Reid, I.D., Molton, N.D., and Stasse, O. (2007). MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1052–1067.

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine*, 13(2), 99–110.

Gintrand, P., Hua, M.D., Hamel, T., and Varra, G. (2022). Relative pose estimation from bearing measurements of three unknown source points. In *2022 51th IEEE Conference on Decision and Control (CDC)*.

Grisetti, G., Kümmerle, R., Stachniss, C., and Burgard, W. (2010). A Tutorial on Graph-Based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4), 31–43.

Guerreiro, B.J.N., Batista, P., Silvestre, C., and Oliveira, P. (2013). Globally Asymptotically Stable Sensor-Based Simultaneous Localization and Mapping. *IEEE Transactions on Robotics*, 29(6), 1380–1395.

Hamel, T. and Samson, C. (2017). Riccati Observers for the Non-Stationary PnP Problem. *IEEE Transactions on Automatic Control*, 63(3), 726–741.

Hamel, T. and Samson, C. (2016). Riccati observers for position and velocity bias estimation from direction measurements. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2047–2053. doi: 10.1109/CDC.2016.7798565.

Hua, M.D., De Marco, S., Hamel, T., and Beard, R.W. (2020). Relative pose estimation from bearing measurements of three unknown source points. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 4176–4181.

Jeong, J., Cho, Y., Shin, Y.S., Roh., H., and Kim, A. (2019). Complex Urban Dataset with Multi-level Sensors from Highly Diverse Urban Environments. *International Journal of Robotics Research*, 38(6), 642–657.

Johansen, T.A. and Brekke, E. (2016). Globally exponentially stable Kalman filtering for SLAM with AHRS. In *2016 19th International Conference on Information Fusion (FUSION)*, 909–916.

Le Bras, F., Hamel, T., Mahony, R., and Samson, C. (2017). *Observers for Position Estimation Using Bearing and Biased Velocity Information*, 3–23. Springer International Publishing, Cham.

Lu, F. and Milios, E.E. (1997). Globally Consistent Range Scan Alignment for Environment Mapping. *Autonomous Robots*, 4, 333–349.

Shan, T. and Englot, B. (2018). LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4758–4765.

van Goor, P., Hamel, T., and Mahony, R. (2020). Equivariant Filter (EqF).

van Goor, P., Mahony, R., Hamel, T., and Trumpf, J. (2019). A Geometric Observer Design for Visual Localisation and Mapping. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2543–2549.