



Homography-based Riccati observer design for camera pose estimation

Tarek Bouazza, Tarek Hamel, Minh Duc Hua, Robert Mahony

► To cite this version:

Tarek Bouazza, Tarek Hamel, Minh Duc Hua, Robert Mahony. Homography-based Riccati observer design for camera pose estimation. 61st IEEE Conference on Decision and Control, 2022, Cancun, Mexico. ⟨hal-03851200⟩

HAL Id: hal-03851200

<https://hal.science/hal-03851200v1>

Submitted on 14 Nov 2022

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

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



HAL Authorization

Homography-based Riccati observer design for camera pose estimation

Tarek Bouazza, Tarek Hamel, Minh Duc Hua, Robert Mahony

Abstract—This paper introduces a novel approach for estimating the relative pose of a mobile robot equipped with an onboard IMU, a velocity sensor complemented with a monocular camera observing a planar scene. The proposed solution relies on the design of a deterministic Riccati observer that exploits the first-order approximations of a class of nonlinear systems. It uses the point-feature correspondences of a sequence of images and exploits the homography constraint to derive the system’s measurement equation. The observability analysis, which highlights the uniform observability condition under which local exponential stability is guaranteed, is performed. Moreover, an extension of the observer to depth estimation is provided. Finally, the proposed observer solution is validated through simulation and experimental results.

I. INTRODUCTION

State estimation (position, velocity, orientation, etc.) is a central problem in robotics and autonomous systems. It involves designing sensor fusion algorithms that combine various measurements using a dynamical model to produce state estimates. The most commonly used sensors for navigation problems are proprioceptive sensors, such as inertial measurement units, which are often complemented by on-board exteroceptive sensors, such as monocular cameras or stereovision systems which (besides estimating the state of the system) provide valuable information on the surrounding environment.

In computer vision, homography is a projective transformation that provides pointwise mapping between two images of the same planar scene, it encodes the camera relative pose (i.e., position and orientation), the distance between the camera and the scene, and its normal vector into a single matrix. Homographies have been widely used in robotic navigation applications as a vision primitive, most notably for visual odometry [15] and visual servoing [12].

Homography estimation is a highly developed topic in the classical computer vision literature. Classical algorithms for homography estimation consist in computing the homography on a frame-by-frame basis by solving algebraic constraints related to image feature correspondences (points, lines, conics, contours, etc.) [1], [10], [9]. These methods, however, are not suited for robotic applications as they do not exploit temporal correlation and as a result cannot improve (or filter) homography over time. Recent work by the authors on homography estimation in [11], [4], [7] involves nonlinear observer design based on the $SL(3)$ group structure of the set of homographies allowing to exploit velocity information.

T. Bouazza, T. Hamel and M.-D. Hua are with I3S, Université Côte d’Azur, CNRS, Sophia Antipolis, France, bouazza(t.hamel, hua)@i3s.unice.fr.

R. Mahony is with the Australian Centre for Robotic Vision, Australian National University, Robert.Mahony@anu.edu.au.

The process of estimating the camera pose and scene parameters given an estimate of the homography matrix is known as homography decomposition. Traditional homography decomposition approaches, such as Faugeras-SVD based [2] and Zhang-SVD based [16] algorithms use singular value decomposition to obtain numerical solutions. Malis and Vargas [13] solved the decomposition problem analytically. However, all the mentioned methods fall into the algebraic category, which focuses on solving the homography decomposition problem on a frame-by-frame basis but not on filtering the measurement noise. Manerikar *et al.*, [14] used the deterministic Riccati observer developed by Hamel and Samson [5] to compute pose using homography estimates as inputs. This design, however, depends entirely on the homography estimate obtained in the prior step.

The present paper introduces a novel direct approach to solving the homography-based pose estimation problem as an alternative to the classical homography estimation-decomposition process. Such that to obtain an estimate of the camera motion, the planar homography constraint between the two views of the scene is used directly to construct the system measurement equation needed for the Riccati observer design [3]. The design of the proposed observer builds on previous work of the authors [6] on pose estimation using epipolar constraints, itself an extension of their previous work on the PnP pose estimation problem, but departs from it in the modeling of the system and the output measurements. The output measurement equation used in [6] is slightly modified in this work to meet the requirements of the Riccati observer framework [3] when using the planar homography constraint.

For the sake of clarity, let us consider a single moving camera observing a stationary plane scene. The task of the proposed observer is to provide an estimate of the relative position and attitude of the camera with respect to a chosen reference frame, as well as the normal direction between the reference and the planar scene, given a temporal sequence of images and measurements of the angular and linear velocities of the camera (both assumed to be bounded), provided by an IMU and a linear velocity sensor. For this initial design, no assumption on the reference frame is made, the inertial frame is considered unknown and therefore the accelerometer measurements are not taken into account.

This paper is organized as follows. In Section II some preliminary notation and definitions of observability, the deterministic Riccati observer framework and planar homography are recalled. Section III highlights the proposed Riccati observer design and explains the derivation of the required measurement equation from the homography constraint. In

addition, a sufficient condition to ensure uniform observability, under which local exponential stability is guaranteed, was presented, along with an extension of the observer to depth estimation. Section IV describes simulation results and outlines experiments conducted to validate the theoretical results. Concluding remarks are presented in Section V.

II. PRELIMINARY MATERIAL

A. Notation

We denote by \mathbb{R}^n the n -dimensional Euclidean space, by $\{e_1, \dots, e_n\}$ the canonical basis of \mathbb{R}^n , by $|x|$ the Euclidean norm of the vector $x \in \mathbb{R}^n$, and by $\|A\| = \max_{|x|=1} |Ax|$ the induced matrix norm of $A \in \mathbb{R}^{n \times n}$. The set $\mathbb{B}_r^n := \{x \in \mathbb{R}^n : |x| \leq r\}$ is the closed ball in \mathbb{R}^n of radius r . The null matrix and identity matrix of dimension $n \times n$ are denoted by 0_n and I_n , respectively. the null matrix of dimension $n \times m$ is denoted $0_{n \times m}$.

$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : |x| = 1\}$ is the n -dimensional sphere embedded in \mathbb{R}^{n+1} with radius equal to one.

x_{\times} is the skew symmetric matrix associated to the cross product $x_{\times}y = x \times y$, $\forall x, y \in \mathbb{R}^3$.

$\Pi_p := (I_{n+1} - pp^\top)$ is the projection onto the tangent space of the unit n -dimensional sphere at the point $p \in \mathbb{S}^n$.

$\text{SL}(n) := \{H \in \mathbb{R}^{n \times n} : \det(H) = 1\}$ is the Special Linear group of order n .

$\text{SO}(n) := \{R \in \mathbb{R}^{n \times n} : \det(R) = 1, RR^\top = R^\top R = I_n\}$ is the Special Orthogonal group of order n .

B. Uniform observability of a linear time-varying system

We consider a generic linear time-varying (LTV) system of the form

$$\begin{cases} \dot{\mathbf{x}} = A(t)\mathbf{x} + B(t)\mathbf{u} \\ \mathbf{y} = C(t)\mathbf{x} \end{cases} \quad (1)$$

with $\mathbf{x} \in \mathbb{R}^n$ the system state vector, $\mathbf{u} \in \mathbb{R}^s$ the system input vector, and $\mathbf{y} \in \mathbb{R}^m$ the system output vector and $A(t)$, $B(t)$, $C(t)$ denoting continuous matrix-valued functions with adequate dimensions.

Let $S(t)$ 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 (A, C, S) is the nonnegative semidefinite matrix-valued function defined by

$$W(t, t+\delta)_S^{A,C} = \frac{1}{\delta} \int_t^{t+\delta} \Phi^\top(s, t) C^\top(s) S(s) C(s) \Phi(s, t) ds \quad (2)$$

where $\Phi(s, t)$ is the transition matrix associated with $A(t)$, such that

$$\begin{cases} \frac{d}{dt} \Phi(s, t) = A(t) \Phi(s, t), \quad \forall s \geq t \\ \Phi(t, t) = I_n \end{cases} \quad (3)$$

If $A(t)$ and $C(t)$ are bounded and if there exists $\delta > 0$ and $\epsilon > 0$ such that $W(t, t+\delta)_S^{A,C} \geq \epsilon I_n$ for all $t \geq 0$, then we say that the pair (A, C) is uniformly observable.

C. Riccati Observer for a Class of nonlinear Systems

The presented observer is based on a modified version of the Riccati design framework (presented in [14]). Consider a class of nonlinear systems, whose state $\mathbf{x} = [x_1^\top \ x_2^\top]^\top$ with $x_1 \in \mathbb{B}_r^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ (with $n_1 + n_2 = n$), that evolves according to the following equations:

$$\begin{cases} \dot{\mathbf{x}} = A(t)\mathbf{x} + \mathbf{u} + O(|\mathbf{x}|^2) + O(|\mathbf{x}||\mathbf{u}|) \\ \mathbf{y} = C(\mathbf{x}, t)\mathbf{x} + O(|\mathbf{x}|^2) \end{cases} \quad (4)$$

A is a continuous matrix-valued function uniformly bounded with respect to t . Furthermore, it is of the form

$$A(t) = \begin{bmatrix} A_{1,1}(t) & 0_{n_1 \times n_2} \\ A_{2,1}(t) & A_{2,2}(t) \end{bmatrix} \in \mathbb{R}^{n \times n}$$

and C is a continuous matrix-valued function uniformly bounded with respect to t and uniformly continuous with respect to \mathbf{x} .

$$C(\mathbf{x}, t) := [C_1(\mathbf{x}, t)^\top \ C_2(\mathbf{x}, t)^\top]^\top \in \mathbb{R}^{m \times n}$$

Assume that $D(t)$ is a bounded continuous symmetric positive semi-definite matrix-valued function and $S(t)$ a bounded continuous symmetric positive definite matrix-valued function and that both $D(t)$ and $S(t)$ are bounded below. Applying the input

$$\mathbf{u} = -PC^\top(t)D\mathbf{y} \quad (5)$$

with $P \in \mathbb{R}^{n \times n}$ a symmetric positive definite matrix solution to the continuous Riccati equation (CRE)

$$\dot{P} = AP + PA^\top - PC^\top D(t)CP + S(t), P(0) = P_0 > 0 \quad (6)$$

then if the pair $(A^*(t), C^*(t))$ with $A^*(t) = A(t)$ and $C^*(t) := C(0, t)$ are uniformly observable, the equilibrium $\mathbf{x} = 0$ is locally uniformly exponentially stable [5, Theorem 3.1 and Corollary 3.2].

D. Homography definition

Consider a moving camera observing a textured planar scene. Let $\{\mathcal{A}\}$ and $\{\mathcal{B}\}$ be the reference frame and the current frame respectively. Let $\xi \in \mathbb{R}^3$ denote the position of frame $\{\mathcal{B}\}$ with respect to frame $\{\mathcal{A}\}$ expressed in frame $\{\mathcal{A}\}$. The orientation of frame $\{\mathcal{B}\}$ with respect to frame $\{\mathcal{A}\}$ is represented by a rotation matrix $R \in \text{SO}(3)$.

Let \hat{d} (resp. d) and $\hat{\eta}$ (resp. $\eta \in \mathbb{S}^2$) denote the distance from the origin of $\{\mathcal{A}\}$ (resp. $\{\mathcal{B}\}$) to the planar scene and the normal vector pointing to the scene expressed in $\{\mathcal{A}\}$ (resp. $\{\mathcal{B}\}$), one easily verifies that

$$\begin{cases} \eta = R^\top \hat{\eta} \\ d = \hat{d} - \hat{\eta}^\top \xi \end{cases} \quad (7)$$

Let $i = 1, \dots, N$ source points belonging to the planar scene. The coordinate vectors $\hat{P}_i \in \{\mathcal{A}\}$ and $P_i \in \{\mathcal{B}\}$ of the same i -th point on the scene are related by

$$P_i = R^\top (\hat{P}_i + \xi) \quad (8)$$

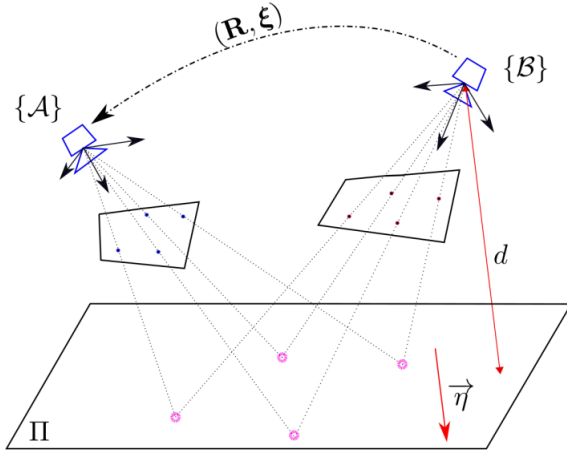


Fig. 1. Euclidean homography

Since the considered points belong to the observed planar scene

$$\Pi := \{\forall P \in \mathbb{R}^3 : \eta^\top P - d = 0\} = \{\forall \dot{P} \in \mathbb{R}^3 : \dot{\eta}^\top \dot{P} - \dot{d} = 0\} \quad (9)$$

and using the fact that $\frac{\dot{\eta}^\top \dot{P}_i}{\dot{d}} = 1$ along with (8), one gets:

$$P_i = R^\top (I_3 - \frac{\xi \dot{\eta}^\top}{\dot{d}}) \dot{P}_i = H^{-1} \dot{P}_i \quad (10)$$

where $H^{-1} = R^\top (I_3 - \frac{\xi \dot{\eta}^\top}{\dot{d}})$ (with $\bar{\xi} := \frac{\xi}{\dot{d}}$), is the inverse of the Euclidean homography matrix H that maps Euclidean coordinates of the scene's points from $\{\mathcal{B}\}$ to $\{\mathcal{A}\}$ and given by

$$H := R + \frac{\xi \eta^\top}{d} \quad (11)$$

Since a homography matrix H is only defined up to scale then any homography matrix is associated with a unique matrix $\bar{H} \in \text{SL}(3)$ by re-scaling

$$\bar{H} = \frac{1}{\det(H)^{\frac{1}{3}}} H \quad (12)$$

such that $\det(\bar{H}) = 1$.

If the camera is calibrated (the intrinsic parameters of the camera are known) one can write

$$\dot{p}_{img} = K \dot{P}, \quad p_{img} = K P \quad (13)$$

where $K \in \mathbb{R}^{3 \times 3}$ is the camera calibration matrix and $\dot{p}_{img} \in \{\mathcal{A}\}$ (resp. $p_{img} \in \{\mathcal{B}\}$) is the image of the point \dot{P} (resp. P) when the camera is aligned with frame $\{\mathcal{A}\}$ (resp. $\{\mathcal{B}\}$) and can be written in the form $(u, v, 1)^\top$ using the homogeneous coordinate representation for that 2D image point.

The image homography matrix that maps pixel coordinates from the current frame to the reference frame is given by

$$H_{im} := K H K^{-1} = K (R + \frac{\xi \eta^\top}{d}) K^{-1} \quad (14)$$

Rather than source points, normalized coordinates (direction vectors) can be used

$$\begin{cases} \dot{p}_i := \frac{\dot{P}_i}{|\dot{P}_i|} = \frac{K^{-1} \dot{p}_{img}}{|K^{-1} \dot{p}_{img}|} \in \mathbb{S}^2 \\ p_i := \frac{P_i}{|P_i|} = \frac{K^{-1} p_{img}}{|K^{-1} p_{img}|} \in \mathbb{S}^2 \end{cases} \quad (15)$$

From equations (10) and (15), the projected points satisfy

$$p_i := \frac{H^{-1} \dot{p}_i}{|H^{-1} \dot{p}_i|} \quad (16)$$

known as the planar homography constraint.

III. HOMOGRAPHY-BASED RICCATI OBSERVER DESIGN

A. Observer design

Let $\zeta = R^\top \xi$ denote the position of the current camera frame $\{\mathcal{B}\}$ with respect to reference camera frame $\{\mathcal{A}\}$ expressed in frame $\{\mathcal{B}\}$, and let $\Omega \in \mathbb{R}^3$ and $V \in \mathbb{R}^3$ denote the camera angular velocity and linear velocity of the current frame $\{\mathcal{B}\}$ w.r.t the reference frame $\{\mathcal{A}\}$, expressed in $\{\mathcal{B}\}$. The equations of motion are

$$\begin{cases} \dot{R} = R \Omega_\times \\ \dot{\zeta} = -\Omega_\times \zeta + V \end{cases}$$

To avoid using minimal parameterization techniques to parameterize the unit normal vector $\dot{\eta} \in \mathbb{S}^2$, we over-parameterize it by introducing an auxiliary rotation matrix $Q \in \text{SO}(3)$ (see [8]) such that

$$\dot{\eta} := Q^\top e_3 \quad (17)$$

Since $\frac{d}{dt} \dot{\eta} = 0$, the following system is obtained

$$\begin{cases} \dot{Q} = 0_{3 \times 3} \\ \dot{R} = R \Omega_\times \\ \dot{\zeta} = \dot{\zeta} = -\Omega_\times \zeta + \frac{V}{d} \end{cases} \quad (18)$$

Let $\hat{Q} \in \text{SO}(3)$, $\hat{R} \in \text{SO}(3)$, $\hat{\zeta} \in \mathbb{R}^3$ denote the estimates of Q , R , ζ , respectively. Such that the normal vector estimate is given by

$$\hat{\eta} := \hat{Q}^\top e_3 \quad (19)$$

In the case where \dot{d} is known, the proposed observer takes the following form

$$\begin{cases} \dot{\hat{Q}} = -\hat{Q} \sigma_Q \times \\ \dot{\hat{R}} = \hat{R} \Omega_\times + \sigma_R \times \hat{R} \\ \dot{\hat{\zeta}} = -\Omega_\times \hat{\zeta} + \frac{V}{d} - \sigma_\zeta \end{cases} \quad (20)$$

with initial conditions $(Q(0), R(0), \zeta(0)) \in \text{SO}(3) \times \text{SO}(3) \times \mathbb{R}^3$, and $\sigma_Q, \sigma_R, \sigma_\zeta \in \mathbb{R}^3$ the innovation terms to be determined thereafter.

The following error variables are defined

$$\tilde{Q} = Q \hat{Q}^\top, \quad \tilde{R} = \hat{R} R^\top, \quad \tilde{\zeta} = \zeta - \hat{\zeta}$$

Then the objective of observer design consists in stabilizing $(\tilde{Q} e_3, \tilde{R}, \tilde{\zeta})$ about $(e_3, I_3, 0)$.

Let $\mathbf{q} = [q_0 \ \bar{q}^\top]^\top$ be the unit quaternion associated with a rotation matrix $R \in \text{SO}(3)$, the expression of R as a function of \mathbf{q} is given by Rodrigues' formula

$$R(\mathbf{q}) = I_3 + 2\bar{q}_\times(q_0 I_3 + \bar{q}_\times)$$

From this, one can write

$$R = I_3 + \lambda_\times + O(|\lambda_\times|^2), \text{ with } \lambda \triangleq 2\text{sign}(q_0)\bar{q} \in \mathcal{B}_2^3$$

One deduces the following first-order approximations

$$\begin{aligned}\tilde{R} &= I_3 + \lambda_{\tilde{R}\times} + O(|\lambda_{\tilde{R}}|^2) \\ \tilde{Q} &= I_3 + \lambda_{\tilde{Q}\times} + O(|\lambda_{\tilde{Q}}|^2)\end{aligned}\quad (21)$$

From (21) and the first two equations of (18) and (20) one derives the error dynamics of \tilde{R} and \tilde{Q}

$$\begin{aligned}\dot{\tilde{R}} &= \sigma_{R\times} \tilde{R} \\ \dot{\tilde{Q}} &= \tilde{Q} \sigma_{Q\times}\end{aligned}$$

One then deduces that in first-order approximations

$$\begin{aligned}\dot{\lambda}_{\tilde{R}} &= \sigma_R + O(|\lambda_{\tilde{R}}||\sigma_R|) \\ \dot{\lambda}_{\tilde{Q}} &= \sigma_Q + O(|\lambda_{\tilde{Q}}||\sigma_Q|)\end{aligned}\quad (22)$$

B. System output with Homography Constraints

The measurement equation is obtained through the planar homography constraint (16), written as follows

$$\frac{\hat{H}p_i}{|\hat{H}p_i|} = \frac{\tilde{H}\tilde{p}_i}{|\tilde{H}\tilde{p}_i|}\quad (23)$$

where \hat{H} is the homography estimate and $\tilde{H} = \hat{H}H^{-1}$ the homography error. By expanding the expression of \tilde{H} and multiplying both sides by $\Pi_{\tilde{p}_i}$, expression (23) writes

$$\Pi_{\tilde{p}_i} \frac{\hat{H}p_i}{|\hat{H}p_i|} = \Pi_{\tilde{p}_i} (I_3 + \frac{\hat{R}\hat{\zeta}\hat{\eta}^\top}{1 - \hat{\eta}^\top \hat{R}\hat{\zeta}})(\tilde{R} - \hat{R}\hat{\zeta}\hat{\eta}^\top)\tilde{p}_i$$

Using first-order approximations (21) yields

$$\Pi_{\tilde{p}_i} \frac{\hat{H}p_i}{|\hat{H}p_i|} \approx \Pi_{\tilde{p}_i} \hat{\Gamma} (\lambda_{\tilde{R}\times} - \hat{R}\hat{\zeta}e_3^\top \tilde{Q} - \hat{R}\hat{\zeta}e_3^\top \lambda_{\tilde{Q}\times} \tilde{Q}) \tilde{p}_i$$

with $\hat{\Gamma} = I_3 + \frac{\hat{R}\hat{\zeta}\hat{\eta}^\top}{1 - \hat{\eta}^\top \hat{R}\hat{\zeta}} = \hat{H}\hat{R}^\top$, one then deduces

$$\begin{aligned}\Pi_{\tilde{p}_i} \frac{\hat{H}p_i}{|\hat{H}p_i|} &\approx -(\hat{Q}\tilde{p}_i)_3 \Pi_{\tilde{p}_i} \hat{\Gamma} \hat{R}\hat{\zeta} - (\hat{Q}\tilde{p}_i)_2 \Pi_{\tilde{p}_i} \hat{\Gamma} \hat{R}\hat{\zeta} \lambda_{\tilde{Q},1} + \\ &\quad (\hat{Q}\tilde{p}_i)_1 \Pi_{\tilde{p}_i} \hat{\Gamma} \hat{R}\hat{\zeta} \lambda_{\tilde{Q},2} - \Pi_{\tilde{p}_i} \hat{\Gamma} \tilde{p}_i \times \lambda_{\tilde{R}}\end{aligned}$$

By setting the measurement as $y_i = \Pi_{\tilde{p}_i} \frac{\hat{H}p_i}{|\hat{H}p_i|}$, one finds

$$\begin{aligned}y_i &= -(\hat{Q}\tilde{p}_i)_2 \Pi_{\tilde{p}_i} \hat{H}\hat{\zeta} \lambda_{\tilde{Q},1} + (\hat{Q}\tilde{p}_i)_1 \Pi_{\tilde{p}_i} \hat{H}\hat{\zeta} \lambda_{\tilde{Q},2} - \\ &\quad \Pi_{\tilde{p}_i} \hat{\Gamma} \tilde{p}_i \times \lambda_{\tilde{R}} - (\hat{Q}\tilde{p}_i)_3 \Pi_{\tilde{p}_i} \hat{H}\hat{\zeta}\end{aligned}\quad (24)$$

Let the system output be $\mathbf{y} := [y_1^\top \ \dots \ y_N^\top]^\top$, it has the form of the measurement equation of (4), where

$$\left\{ \begin{aligned} \mathbf{x} &:= \begin{bmatrix} \lambda_{\tilde{Q},1} \\ \lambda_{\tilde{Q},2} \\ \lambda_{\tilde{R}} \\ \tilde{\zeta} \end{bmatrix}, \quad \mathbf{u} := \begin{bmatrix} \sigma_{Q,1} \\ \sigma_{Q,2} \\ \sigma_R \\ \sigma_\zeta \end{bmatrix}, \\ A &:= \left[\begin{array}{ccc|c} 0 & 0 & 0_{1 \times 3} & 0_{1 \times 3} \\ 0 & 0 & 0_{1 \times 3} & 0_{1 \times 3} \\ \hline 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 3} & -\Omega_\times \end{array} \right], \\ C_1 &:= \begin{bmatrix} -(\hat{Q}\tilde{p}_1)_2 \Pi_{\tilde{p}_1} \hat{H}\hat{\zeta} & (\hat{Q}\tilde{p}_1)_1 \Pi_{\tilde{p}_1} \hat{H}\hat{\zeta} & -\Pi_{\tilde{p}_1} \hat{\Gamma} \tilde{p}_1 \times \\ \vdots & \vdots & \vdots \\ -(\hat{Q}\tilde{p}_N)_2 \Pi_{\tilde{p}_N} \hat{H}\hat{\zeta} & (\hat{Q}\tilde{p}_N)_1 \Pi_{\tilde{p}_N} \hat{H}\hat{\zeta} & -\Pi_{\tilde{p}_N} \hat{\Gamma} \tilde{p}_N \times \end{bmatrix} \\ C_2 &:= \begin{bmatrix} -(\hat{Q}\tilde{p}_1)_3 \Pi_{\tilde{p}_1} \hat{H} \\ \vdots \\ -(\hat{Q}\tilde{p}_N)_3 \Pi_{\tilde{p}_N} \hat{H} \end{bmatrix} \end{aligned} \right. \quad (25)$$

From there, the expressions of the innovation terms are computed according to (5), where the matrices S and D (involved in (6)) are chosen larger than some constant positive matrix.

Note that only the first two components of the innovation term σ_Q (i.e. $\sigma_{Q,1}$ and $\sigma_{Q,2}$) are involved in the design process of the Riccati observer. Its last component can therefore be set to zero for the sake of simplicity, such that $\sigma_Q = [\sigma_{Q,1}, \sigma_{Q,2}, 0]^\top$.

Remark 1. Note that another possible setting of this problem can be considered, assuming that the measured linear velocity $v \in \mathbb{R}^3$ is expressed in the reference frame, which amounts to simpler dynamics

$$\begin{cases} \dot{R} = R\Omega_\times \\ \dot{\zeta} = \frac{v}{d} \end{cases}$$

This case, although impractical, is interesting to investigate from a theoretical standpoint as it leads to a simplified version of (25) with $A = 0_{8 \times 8}$.

C. Observability analysis

This section provides sufficient conditions that ensure the exponential stability of the observer's origin error. The equilibrium $(\tilde{Q}e_3, \tilde{R}, \tilde{\zeta}) = (e_3, I_3, 0)$ is locally exponentially stable, provided that the pair $(A^*(t), C^*(t))$ with $A^*(t) := A(t)$ and $C^*(t) := C(0, t)$ is uniformly observable, the observability Gramian is as follows

$$W(t, t+\delta)_{I_n}^{A^*, C^*} = \frac{1}{\delta} \int_t^{t+\delta} \Phi^{*\top}(\tau, t) C^{*\top}(\tau) C^*(\tau) \Phi^*(\tau, t) d\tau \quad (26)$$

with

$$C^* = \begin{bmatrix} (Qp_1)_2 \Pi_{\tilde{p}_1} H\bar{\zeta} & -(Qp_1)_1 \Pi_{\tilde{p}_1} H\bar{\zeta} & \Pi_{\tilde{p}_1} \Gamma \tilde{p}_1 \times & (Qp_1)_3 \Pi_{\tilde{p}_1} H \\ \vdots & \vdots & \vdots & \vdots \\ (Qp_N)_2 \Pi_{\tilde{p}_N} H\bar{\zeta} & -(Qp_N)_1 \Pi_{\tilde{p}_N} H\bar{\zeta} & \Pi_{\tilde{p}_N} \Gamma \tilde{p}_N \times & (Qp_N)_3 \Pi_{\tilde{p}_N} H \end{bmatrix}$$

and

$$\Phi^*(\tau, t) = \begin{bmatrix} I_5 & 0_{3 \times 3} \\ 0_{3 \times 5} & R(\tau)^\top R(t) \end{bmatrix}$$

In order to guarantee that the Gramian (26) is always positive, it suffices to prove that $C^*(t)$ is of full rank. The approach taken is inspired by the previous work in [4].

One has from the expression of \tilde{H} that in first-order approximation $\tilde{H} \approx I_3 + \Delta$, with

$$\Delta = \hat{\Gamma}(\lambda_{\tilde{R} \times} - \hat{R} \tilde{\zeta} e_3^\top \hat{Q} - \hat{R} \tilde{\zeta} e_3^\top \lambda_{\tilde{Q} \times} \hat{Q}) \quad (27)$$

Let us consider the equation

$$\Pi_{\tilde{p}_i} \tilde{H} \tilde{p}_i = 0, \quad \forall i \in \{1 \dots N\} \quad (28)$$

which can be developed as follows:

$$\begin{aligned} \Pi_{\tilde{p}_i} \Delta \tilde{p}_i &= 0, \quad \forall i \in \{1 \dots N\} \\ \Leftrightarrow \Pi_{\tilde{p}_i} \hat{\Gamma}(\lambda_{\tilde{R} \times} - \hat{R} \tilde{\zeta} e_3^\top \hat{Q} - \hat{R} \tilde{\zeta} e_3^\top \lambda_{\tilde{Q} \times} \hat{Q}) \tilde{p}_i &= 0 \\ \Leftrightarrow -(\hat{Q} \tilde{p}_i)_2 \Pi_{\tilde{p}_i} \hat{H} \tilde{\zeta} \lambda_{\tilde{Q},1} + (\hat{Q} \tilde{p}_i)_1 \Pi_{\tilde{p}_i} \hat{H} \tilde{\zeta} \lambda_{\tilde{Q},2} - \\ \Pi_{\tilde{p}_i} \hat{\Gamma} \tilde{p}_i \times \lambda_{\tilde{R}} - (\hat{Q} \tilde{p}_i)_3 \Pi_{\tilde{p}_i} \hat{H} \tilde{\zeta} &= 0 \end{aligned}$$

From there, the expression of C in (25) can be identified, and one finds by evaluating the latter around the equilibrium $\mathbf{x} = 0$, the following system

$$\Pi_{\tilde{p}_i} [-(Q \tilde{p}_i)_2 \Pi_{\tilde{p}_i} H \tilde{\zeta} (Q \tilde{p}_i)_1 \Pi_{\tilde{p}_i} H \tilde{\zeta} - \Pi_{\tilde{p}_i} \Gamma \tilde{p}_i \times - (Q \tilde{p}_i)_3 \Pi_{\tilde{p}_i} H] \mathbf{x} = 0 \quad (29)$$

Finally, from (29) the following equivalence is deduced

$$C^* \mathbf{x} = 0 \iff \Pi_{\tilde{p}_i} \tilde{H} \tilde{p}_i = 0, \quad \forall i \in \{1 \dots N\} \quad (30)$$

It is straightforward that if $\tilde{H} = I_3$ then $\mathbf{x} = 0$. This being said, to prove that C^* is of full rank it is sufficient to show that $\tilde{H} = I_3$ around the equilibrium. This amounts to equation (28), which furthermore implies the existence of some non-null constants such that

$$\tilde{H} \tilde{p}_i = \lambda_i \tilde{p}_i \quad (31)$$

Definition 1 (Consistent set). A set \mathcal{M}_N of $N \geq 4$ vector directions $\tilde{p}_i \in \mathbb{S}^2$ ($i = \{1 \dots N\}$) is called consistent if it contains a subset $\mathcal{M}_4 \subset \mathcal{M}_N$ of 4 constant vector directions such that all its vector triplets are linearly independent

Definition 1 implies that if the set \mathcal{M}_N is consistent then, for all $\tilde{p}_i \in \mathcal{M}_4$ there exist a unique set of three non vanishing scalars $b_j \neq 0$ ($j \neq i$) such that

$$\tilde{p}_i = \frac{y_i}{|y_i|} \text{ where } y_i = \sum_{j=1(j \neq i)}^4 b_j \tilde{p}_j \quad (32)$$

The relation (31) indicates that all λ_i are eigenvalues of \tilde{H} and all $\tilde{p}_i \in \mathbb{S}^2$ are the associated eigenvectors of \tilde{H} . Note that the fact that $\lambda_i \neq 0$ can be easily verified. For instance, if $\lambda_i = 0$, then $\tilde{p}_i = \lambda_i \tilde{H}^{-1} \tilde{p}_i = 0$ which contradicts the fact that $\tilde{p}_i \in \mathbb{S}^2$.

Consider the consistent set $\mathcal{M}_N = \{\tilde{p}_1 \dots \tilde{p}_N\}$, it follows at the limit (and without loss of generality) that $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3)$

are three independent vectors and therefore they represent three non collinear eigenvectors of \tilde{H} associated with the eigenvalues λ_i for $i = \{1, 2, 3\}$ such that $\tilde{H} \tilde{p}_i = \lambda_i \tilde{p}_i$.

Exploiting again the consistency of the set \mathcal{M}_N , it follows that there exists a constant direction \tilde{p}_k from the set $\{\tilde{p}_4 \dots \tilde{p}_N\}$ such that:

$$\tilde{p}_k = \frac{y_k}{|y_k|} \text{ where } y_k = \sum_{i=1}^3 b_i \tilde{p}_i, \quad b_i \in \mathbb{R}/\{0\}, \quad i = \{1, 2, 3\} \quad (33)$$

Since \tilde{p}_k can be seen as a forth eigenvector for \tilde{H} associated to the eigenvalue $\lambda_k = \pm |\tilde{H} \tilde{p}_k|$, this yields

$$\begin{aligned} \lambda_k \tilde{p}_k &= \tilde{H} \tilde{p}_k = \frac{1}{|y_k|} \sum_{i=1}^3 b_i \tilde{H} \tilde{p}_i \\ \lambda_k \sum_{i=1}^3 b_i \tilde{p}_i &= \sum_{i=1}^3 b_i \lambda_i \tilde{p}_i \end{aligned}$$

Exploiting the fact that the measured directions form a consistent set, it follows that $b_i \neq 0$, $i = \{1, 2, 3\}$ and using the fact that $\det(\tilde{H}) = 1$, a straightforward identification shows that $\lambda_k = \lambda_1 = \lambda_2 = \lambda_3 = 1$ and thus $\tilde{H} = I_3$.

It follows from (30) that with a minimum of four direction vectors \tilde{p}_k , $k = \{1, 2, 3, 4\}$ forming a consistent set \mathcal{M}_4 the pair (A^*, C^*) is uniformly observable, and the equilibrium $(\tilde{Q} e_3, \tilde{R}, \tilde{\zeta}) = (e_3, I_3, 0)$ is locally exponentially stable.

Remark 2. A special case occurs when the current and reference camera positions are overlapped i.e., $\tilde{\zeta} = 0$, the homography simplifies to the rotation matrix $H = R$, meaning that the Lie group $\text{SL}(3)$ is degenerated into $\text{SO}(3)$.

In this case, the planar homography constraint (16) becomes $\tilde{p}_i = R^\top \tilde{p}_i$, and the error reduces to $\tilde{H} = \tilde{R}$ which leads to the following measurement equation

$$0 = \Pi_{\tilde{p}_i} \tilde{p}_i \times \lambda_R, \quad \forall i \in \{1 \dots N\} \quad (34)$$

such that at the equilibrium $\tilde{R} = I_3$ one finds

$$C^*(t) = \begin{bmatrix} \tilde{p}_1 \times \\ \vdots \\ \tilde{p}_N \times \end{bmatrix} \quad (35)$$

$C^* \mathbf{x} = 0$ yields

$$\tilde{p}_i \times \lambda_R = 0, \quad \forall i \in \{1 \dots N\} \quad (36)$$

For $N = 2$ equation (36) implies the existence of $\lambda_1, \lambda_2 \in \mathbb{R}^*$, such that $\lambda_R = \lambda_1 \tilde{p}_1 = \lambda_2 \tilde{p}_2$, since the points are chosen so that \tilde{p}_j and \tilde{p}_i are noncollinear $\forall i \neq j$, it follows directly that $\lambda_R = 0$. One concludes that the attitude R is uniformly observable with a minimum of two source points.

D. Depth estimation

In practice, the distance between the reference camera and the scene \tilde{d} is unknown. The proposed observer has an interesting property making it possible to estimate the latter along with the state when the linear velocity V is measured.

To this end, a new state variable $\rho = \frac{1}{d}$ and its estimate $\hat{\rho}$ are defined, such that

$$\begin{cases} \dot{\zeta} = -\Omega_{\times} \bar{\zeta} + \rho V \\ \dot{\rho} = 0 \end{cases} \quad (37)$$

Note that including the estimate of ρ in the Riccati observer framework would render the system naturally unobservable when V is equal to zero, it is therefore preferable to estimate it separately and cascade the two observers.

Proposition 1. *Consider the general solution to the Continuous Riccati Equation (6) as the time-varying symmetric positive definite matrix $P(t)$, if the measured direction vectors form a consistent set \mathcal{M}_N , then by setting*

$$\dot{\hat{\rho}} = k_{\rho} V^{\top} \bar{P} C^{\dagger} \mathbf{y} \quad (38)$$

with $\bar{P} = [0_{3 \times 5} \ I_3] P^{-1}$, $C^{\dagger} = [C^{\top} C]^{-1} C^{\top}$ and $k_{\rho} > 0$, the equilibrium $(\mathbf{x}, \tilde{\rho}) = (0, 0)$, with $\tilde{\rho} = \rho - \hat{\rho}$, is locally asymptotically stable if $S(t)$ and $D(t)$ are chosen larger than some positive matrix.

Proof. Define the following candidate Lyapunov function

$$L(\mathbf{x}, t) = \mathbf{x}^{\top} P^{-1} \mathbf{x} + \frac{1}{k_{\rho}} |\tilde{\rho}|^2 \quad (39)$$

Using the property $\frac{d}{dt} P^{-1}(t) = -P^{-1} \dot{P}(t) P^{-1}$, one finds by differentiating L

$$\begin{aligned} \dot{L} = & -\mathbf{x}^{\top} [C^{\top} S C + P^{-1} D P^{-1}] \mathbf{x} + \\ & 2\tilde{\rho} V^{\top} [0_{3 \times 5} \ I_3] P^{-1} \mathbf{x} - \frac{2}{k_{\rho}} \tilde{\rho} \dot{\tilde{\rho}} \end{aligned}$$

Since the direction vectors form a consistent set, the Riccati observability condition is satisfied and the matrix $C^{\top} C$ is therefore invertible, which yields

$$\mathbf{x} = [C^{\top} C]^{-1} C^{\top} \mathbf{y}$$

It follows that

$$\begin{aligned} \dot{L} = & -\mathbf{x}^{\top} [C^{\top} S C + P^{-1} D P^{-1}] \mathbf{x} \\ & - 2\tilde{\rho} \left[\frac{1}{k_{\rho}} \dot{\tilde{\rho}} - V^{\top} \bar{P} [C^{\top} C]^{-1} C^{\top} \mathbf{y} \right] \end{aligned}$$

Finally, introducing the expression (38) of $\dot{\hat{\rho}}$ yields

$$\dot{L} = -\mathbf{x}^{\top} [C^{\top} S C + P^{-1} D P^{-1}] \mathbf{x}$$

By choosing $S(t)$ and $D(t)$ positive definite, it follows that $(\mathbf{x}, \tilde{\rho}) = (0, 0)$ is locally asymptotically stable. \square

Remark 3. *If the linear velocity in the body-fixed frame is unmeasurable, and constant or slowly time-varying then, by defining $s = \frac{V}{d} = \rho V$, and proceeding as in proposition 1:*

$$\begin{cases} \dot{\zeta} = \Omega_{\times} \hat{\zeta} + \hat{s} - \sigma_{\zeta} \\ \dot{\hat{s}} = k_s \bar{P} C^{\dagger} \mathbf{y}, \quad k_s > 0 \end{cases} \quad (40)$$

one ensures that the equilibrium $(\mathbf{x}, \tilde{s}) = (0, 0)$ with $\tilde{s} = s - \hat{s}$ is locally asymptotically stable. Analogously, if the

linear velocity is rather constant or slowly time varying in the inertial frame, then by defining $s = \frac{v}{d}$ and

$$\begin{cases} \dot{\xi} = \hat{s} - \sigma_{\xi} \\ \dot{\hat{s}} = k_s \bar{P} C^{\dagger} \mathbf{y}, \quad k_s > 0 \end{cases} \quad (41)$$

one also ensures that the equilibrium $(\bar{\mathbf{x}}, \tilde{s}) = (0, 0)$ with $\bar{\mathbf{x}} = [\lambda_{\bar{Q},1}, \lambda_{\bar{Q},2}, \lambda_{\bar{R}}, \tilde{\xi}]^{\top}$ is locally asymptotically stable if conditions of proposition 1 are satisfied.

IV. SIMULATION RESULTS AND EXPERIMENTAL EVALUATION

A. Simulation setup

In this section, the performance of the proposed observer is investigated through simulation. The camera is represented in 3D space, observing four points of interest belonging to the horizontal plane. The camera is initially at the center with $z = 5m$, representing the true value of d , and pointing downward, such that $\hat{\eta} = e_3$. For this scenario, the camera motion follows a circular path with velocity in the body-fixed frame $V(t) = [0.5 \sin(0.5t), -0.5 \cos(0.5t), 0]^{\top}$ and angular velocity $\Omega(t) = [0.1 \sin(0.5t), 0.1 \cos(0.5t), 0.1]^{\top}$. The attitude and normal vector estimates are expressed directly on the group of unit quaternions $(\hat{\mathbf{q}}_R, \hat{\mathbf{q}}_Q)$ associated with \hat{R} and \hat{Q} , respectively. Considering that V is known, the observer is implemented with depth estimation as described in Proposition 1.

The following initial state estimates are considered: $\hat{\zeta}(0) = [0.2, 0.2, 0.2]^{\top}$, $\hat{\mathbf{q}}_R(0) = [0.9509, 0.1503, 0.2250, 0.1503]^{\top}$, corresponding to errors in roll, pitch and yaw of $30(deg)$, $\hat{\mathbf{q}}_Q(0) = [0.924, 0.3827, 0, 0]$ corresponding to an angle of $30(deg)$ between $\hat{\eta}(0)$ and $\hat{\eta}(0)$, and $\hat{\rho}(0) = 0.3$, with $P(0) = \text{diag}(1I_5; 2I_3)$, $D = 100I_{3N}$, $S = 0.5I_8$. The initial estimates are randomly perturbed in each component by sampling from Gaussian distributions of mean 0 and standard deviation 0.2, $15(deg)$, $15(deg)$, and 0.2, respectively. The results of 100 Monte Carlo trials are shown in Figures 2 and 3; the shaded area shows 5th to 95th percentile.

The presented simulation results are consistent with the theoretical results and validate the performance of the Riccati observer proposed in this work.

B. Image stabilisation experiment

To support the simulation results and further evaluate the performance of the Riccati observer, an image stabilization experiment is conducted involving a sequence of images and angular velocity measurements collected from a camera-IMU system moving rapidly along a planar scene. The homography resulting from the implemented observer estimates is used to track a region of interest in the moving image by means of image warping with respect to the reference.

The results obtained (available at <https://shorturl.at/jGMX9>) demonstrate that the observer performs properly and manages to track the region of interest even in the presence of strong occlusion, severe

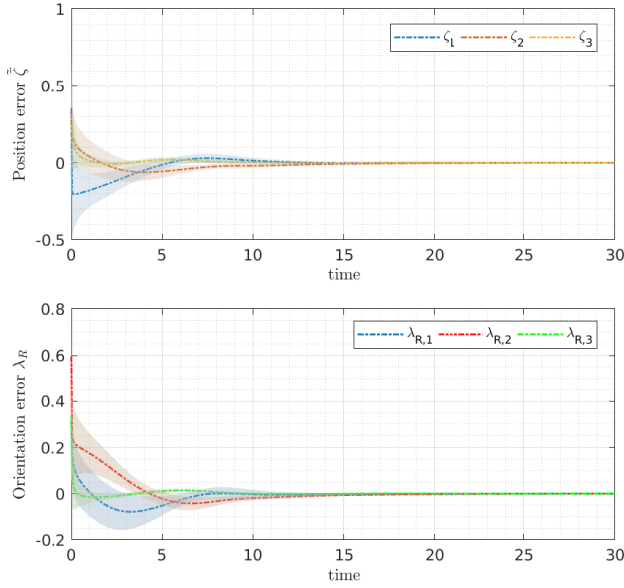


Fig. 2. Scaled position and attitude estimation errors, respectively.

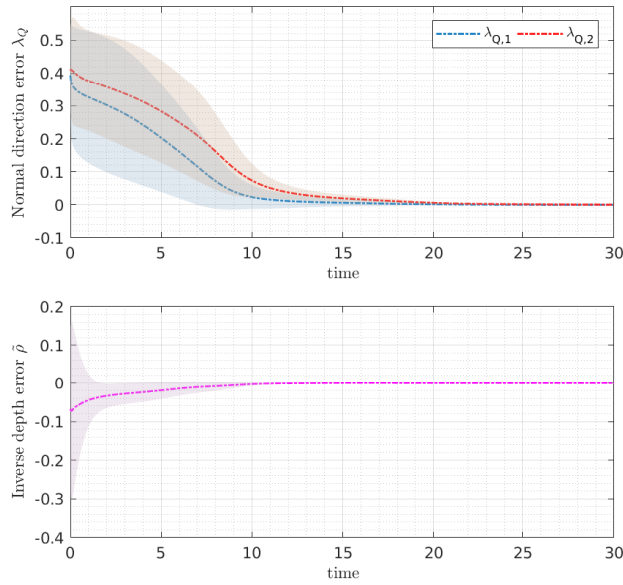


Fig. 3. Normal direction and inverse depth estimation errors, respectively.

image blur and light saturation, demonstrating its robustness and the good convergence of the estimates to the real values.

V. CONCLUSIONS

This work presents a novel approach for camera pose estimation when viewing a planar scene. The proposed solution uses the homography constraint to design a nonlinear observer based on the Riccati observer design framework. A sufficient condition to guarantee uniform observability is provided, along with an extension of the observer for depth estimation. The simulation and experimental results illustrate that the proposed approach appropriately and accurately estimate the homography components.

As part of future work, an online version of the Riccati observer will be implemented in C++ along with OpenCV for feature extraction and matching. An extension of the observer to bias estimation will also be investigated.

ACKNOWLEDGMENT

This work has been supported by the French government, through the EUR DS4H Investments in the Future project managed by the National French Agency (ANR) with the reference number ANR-17-EURE-0004.

REFERENCES

- [1] Alex M Andrew. Multiple view geometry in computer vision. *Kybernetes*, 2001.
- [2] Olivier D Faugeras and Francis Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(03):485–508, 1988.
- [3] Tarek Hamel, Minh-Duc Hua, and Claude Samson. Deterministic observer design for vision-aided inertial navigation. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1306–1313. IEEE, 2020.
- [4] Tarek Hamel, Robert Mahony, Jochen Trumpf, Pascal Morin, and Minh-Duc Hua. Homography estimation on the special linear group based on direct point correspondence. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 7902–7908. IEEE, 2011.
- [5] Tarek Hamel and Claude Samson. Riccati observers for the non-stationary pnp problem. *IEEE Transactions on Automatic Control*, 63(3):726–741, 2017.
- [6] Minh-Duc Hua, Simone De Marco, Tarek Hamel, and Randal W Beard. Relative pose estimation from bearing measurements of three unknown source points. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 4176–4181. IEEE, 2020.
- [7] Minh-Duc Hua, Jochen Trumpf, Tarek Hamel, Robert Mahony, and Pascal Morin. Feature-based recursive observer design for homography estimation and its application to image stabilization. *Asian Journal of Control*, 21(4):1443–1458, 2019.
- [8] Minh-Duc Hua and Guillaume Allibert. Riccati observer design for pose, linear velocity and gravity direction estimation using landmark position and imu measurements. In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1313–1318. IEEE, 2018.
- [9] CV Jawahar and PK Jain. Homography estimation from planar contours. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 877–884. IEEE, 2006.
- [10] Jeremy Yermiyahou Kaminski and Amnon Shashua. Multiple view geometry of general algebraic curves. *International Journal of Computer Vision*, 56(3):195–219, 2004.
- [11] Robert Mahony, Tarek Hamel, Pascal Morin, and Ezio Malis. Non-linear complementary filters on the special linear group. *International Journal of Control*, 85(10):1557–1573, 2012.
- [12] Ezio Malis, Francois Chaumette, and Sylvie Boudet. 2 1/2 d visual servoing. *IEEE Transactions on Robotics and Automation*, 15(2):238–250, 1999.
- [13] Ezio Malis and Manuel Vargas. *Deeper understanding of the homography decomposition for vision-based control*. PhD thesis, INRIA, 2007.
- [14] Ninad Manerikar, Minh-Duc Hua, Simone De Marco, and Tarek Hamel. Riccati observer design for homography decomposition. In *2020 European Control Conference (ECC)*, pages 1306–1311. IEEE, 2020.
- [15] Davide Scaramuzza and Roland Siegwart. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE transactions on robotics*, 24(5):1015–1026, 2008.
- [16] Zhongfei Zhang and Allen R Hanson. 3d reconstruction based on homography mapping. *Proc. ARPA96*, pages 1007–1012, 1996.