Feature-based Recursive Observer Design for Homography Estimation and its Application to Image Stabilization

Minh-Duc Hua, Jochen Trumpf, Tarek Hamel, Robert Mahony, Pascal Morin

Feature-based Recursive Observer Design for Homography Estimation and its Application to Image Stabilization

Minh-Duc Hua, Jochen Trumpf, Tarek Hamel, Robert Mahony, and Pascal Morin

***Abstract*—This paper presents a new algorithm for online estimation of a sequence of homographies applicable to image sequences obtained from robotic vehicles equipped with vision sensors. The approach taken exploits the underlying Special Linear group structure of the set of homographies along with gyroscope measurements and direct point-feature correspondences between images to develop temporal filter for the homography estimate. Theoretical analysis and experimental results are provided to demonstrate the robustness of the proposed algorithm. The experimental results show excellent performance and robustness even in the case of very fast camera motions (relative to frame rate) and severe occlusions.**

I. Introduction

When a robotic vehicle equipped with a vision sensor is manoeuvering in a built environment, consisting primarily of planar or near planar surfaces, then the nature of the environment can be exploited in the vision processing algorithms. Different images of the same planar surface are related by homography mappings, and homographies have been used extensively in robotic applications as a vision primitive. Homography based algorithms have been used for estimation of the rigid-body pose of a vehicle equipped with a camera [23], [20], [22]. Navigation of robotic vehicles has been developed based on homography sequences [5], [8], [11] and one of the most successful visual servo control paradigms uses homographies [15], [16]. Homography based methods are particularly well suited to navigation of unmanned aerial vehicles [3], [19], [4] where the ground terrain is viewed from a distance and the relief of surface features is negligible compared to the vehicles distance from the scene.

Computing homographies from point correspondences has been extensively studied in the last fifteen years [10]. The quality of the homography estimate obtained depends heavily on the number and quality of the data points used in the estimation as well as the algorithm employed. For a well textured scene, the state-of-the-art methods can provide high quality homography estimates at the cost of significant computational effort (see [18] and references therein). For a scene with poor texture, and consequently few reliable feature point correspondences, existing homography estimation algorithms perform poorly. Robotic vehicle applications, however, provide temporal sequences of images and it seems natural to exploit the temporal correlation rather than try to compute individual raw homographies for each pair of frames. Zhang *et al* [25] used image flow computed from a pair of images to compute the relative homography, although this method still only considers isolated pairs of images. In recent work by the authors [16], [13] a nonlinear observer [2], [12] for homography estimation was proposed based on the group structure of the set of homographies, the Special Linear group $\mathrm{SL}(3)$ [1]. This observer uses velocity information to interpolate across a sequence of images and improve the individual homography estimates. The observer proposed in [16], [13] still requires individual image homographies to be algebraically computed for each image, which are then smoothed using filtering techniques. Although this approach [16], [13] provides improved homography estimates, it comes at the cost of running both a classical homography algorithm as well as the temporal filter algorithm, and only functions if each pair of images has sufficient data available to compute a raw homography.

In this paper, we consider the question of deriving an observer for a sequence of image homographies that takes image point-feature correspondences directly as input. The proposed approach takes a sequence of images associated with a continuous variation of the reference image, the most common case being a moving camera viewing a planar scene, a situation typical of robotic applications. The proposed nonlinear observer is posed on the Special Linear group $\mathrm{SL}(3)$, that is in one-to-one correspondence with the group of homographies [1], and uses velocity measurements to propagate the homography estimate and fuse this with new data as it becomes available [16], [13]. A key advance on prior work by the authors is the formulation of a point feature innovation for the observer that incorporates point correspondences directly in the observer without requiring reconstruction of individual image homographies. This saves considerable computational resources and makes the proposed algorithm suitable for embedded systems with simple point tracking software. In addition, the algorithm is well posed even when there is insufficient data for full reconstruction of a homography. For example, if the number of corresponding points between two images drops below four it is impossible to algebraically reconstruct an image homography and the existing algorithms fail [10]. In such

M.-D. Hua is with I3s, Université Côte d'Azur, CNRS, Sophia Antipolis, France, hua@i3s.unice.fr

J. Trumpf and R. Mahony are with the Australian Centre for Robotic Vision, Australian National University, Canberra, Australia, Robert.Mahony(Jochen.Trumpf)@anu.edu.au

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

P. Morin is with Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France pascal.morin@sorbonne-universite.fr

situations, the proposed observer will continue to operate, incorporating what information is available and relying on propagation of prior estimates where necessary. Finally, even if a homography can be reconstructed from a small set of feature correspondences, the estimate is often unreliable and the associated error is difficult to characterize. The proposed algorithm integrates information from a sequence of images, and noise in the individual feature correspondences is filtered through the natural low-pass response of the observer, resulting in a highly robust estimate. As a result, the authors believe that the proposed observer is ideally suited for estimation of homographies based on small windows of image data associated with specific planar objects in a scene, poorly textured scenes, and real-time implementation; all of which are characteristic of requirements for homography estimation in robotic vehicle applications.

The paper is organized into seven sections including the introduction and the concluding sections. Section II presents a brief recap of the Lie group structure of the set of homographies. In Section III, based on a recent advanced theory for nonlinear observer design [14], a nonlinear observer on $\mathrm{SL}(3)$ is proposed using direct measurements of known inertial directions and the knowledge of the group velocity. A rigourous stability analyses is provided in this section. In Section IV, the homography and associated homography velocity are related to rigid-body motion of the camera and two observers are derived for the case where only the angular velocity of the camera is known, a typical scenario in robotic applications. Simulation results are provided in Section V. Section VI provides an application of our approach to a real world problem in image stabilization. In this section, offline and real-time experimental validations are presented together with some useful practical implementation aspects. Four video links, showing the results of Sections VI, are also provided as supplementary material. The results show excellent performance and robustness even in the case of very fast camera motions (relative to frame rate) and severe occlusions.

A preliminary version of the theoretical results in this paper was presented at a conference [9]. The present version provides a more formal derivation of the observer based on a recent theory [14]. In particular, the M-estimator-like observer design proposed in Subsection IV-C is the new result that improves significantly the robustness of the proposed approach with respect to unavoidable feature correspondence outliers encountered in practice. Finally, extensive experimental validations for the image stabilization application as reported in Section VI have also been recently carried out.

II. Preliminary Material

A. Camera projections

Visual data is obtained via a projection of an observed scene onto the camera image surface. The projection is parameterised by two sets of parameters: intrinsic ("internal" parameters of the camera such as the focal length, the pixel aspect ratio, etc.) and extrinsic (the pose, i.e. the position and orientation of the camera). Let $\mathring{\mathcal{A}}$ (resp. $\mathcal{A}$) denote projective coordinates for the image plane of a camera $\mathring{A}$ (resp. A), and let $\{\mathring{A}\}$ (resp. $\{A\}$) denote its (right-hand) frame of reference. Let $\xi \in \mathbb{R}^3$ denote the position of the frame $\{A\}$ with respect to $\{\mathring{A}\}$ expressed in $\{\mathring{A}\}$. The orientation of the frame $\{A\}$ with respect to $\{\mathring{A}\}$, is given by a rotation matrix $R \in \mathrm{SO}(3)$, where $R\colon \{A\} \to \{\mathring{A}\}$ as a mapping. The pose of the camera determines a rigid body transformation from $\{A\}$ to $\{\mathring{A}\}$ (and visa versa). One has

$$\mathring{P} = RP + \xi \tag{1}$$

as a relation between the coordinates of the same point in the reference frame ($\mathring{P} \in \{\mathring{A}\}$) and in the current frame ($P \in \{A\}$). The camera internal parameters, in the commonly used approximation [10], define a 3×3 matrix K so that we can write[1]:

$$\mathring{p} \cong K\mathring{P}, \quad p \cong KP, \tag{2}$$

where $p \in \mathcal{A}$ is the image of a point when the camera is aligned with frame $\{A\}$, and can be written as $(x, y, w)^\top$ using the homogeneous coordinate representation for that 2D image point. Likewise, $\mathring{p} \in \mathring{\mathcal{A}}$ is the image of the same point viewed when the camera is aligned with frame $\{\mathring{A}\}$.

If the camera is calibrated (the intrinsic parameters are known), then all quantities can be appropriately scaled and the equation is written in a simple form:

$$\mathring{p} \cong \mathring{P}, \quad p \cong P. \tag{3}$$

B. Homographies

Since homographies describe image transformations of *planar* scenes, we begin by fixing a plane that contains the points of interest (target points).

Assumption 1: Assume a calibrated camera and that there is a planar surface Π containing a set of n target points ($n \geq 4$) so that

$$\Pi = \left\{ \mathring{P} \in R^3 : \mathring{\eta}^\top \mathring{P} - \mathring{d} = 0 \right\},$$

where $\mathring{\eta}$ is the unit normal to the plane expressed in $\{\mathring{A}\}$ and $\mathring{d}$ is the distance of the plane to the origin of $\{\mathring{A}\}$.
From the rigid-body relationships (1), one has $P = R^\top \mathring{P} - R^\top \xi$. Define $\zeta = -R^\top \xi$. Since all target points lie in a single planar surface one has

$$P_i = R^\top \mathring{P}_i + \frac{\zeta \mathring{\eta}^\top}{\mathring{d}} \mathring{P}_i, \quad i = \{1, \ldots, n\}, \tag{4}$$

and thus, using (3), the projected points obey

$$p_i \cong \left(R^\top + \frac{\zeta \mathring{\eta}^\top}{\mathring{d}} \right) \mathring{p}_i, \quad i = \{1, \ldots, n\}. \tag{5}$$

The projective mapping $H : \mathcal{A} \to \mathring{\mathcal{A}}$, $H :\cong \left(R^\top + \frac{\zeta \mathring{\eta}^\top}{\mathring{d}} \right)^{-1}$ is termed a homography and it relates the images of points on the plane Π when viewed from two poses defined by the coordinate systems $\mathcal{A}$ and $\mathring{\mathcal{A}}$, respectively. It is straightforward to verify that the homography H can be written as follows:

$$H \cong \left(R + \frac{\xi \eta^\top}{d} \right), \tag{6}$$

[1]Most statements in projective geometry involve equality up to a multiplicative constant denoted by $\cong$.

where η is the normal to the observed planar surface expressed in the frame $\{A\}$ and d is the orthogonal distance of the plane to the origin of $\{A\}$. One can verify that [1]:

$$\eta = R^\top \mathring{\eta} \tag{7}$$

$$d = \mathring{d} - \mathring{\eta}^\top \xi = \mathring{d} + \eta^\top \zeta. \tag{8}$$

The homography matrix contains the pose information (R, ξ) of the camera from the frame $\{A\}$ (termed current frame) to the frame $\{\mathring{A}\}$ (termed reference frame). However, since the relationship between the image points and the homography is a projective relationship, it is only possible to determine H up to a scale factor (using the image points relationships alone).

C. Homographies as elements of the Special Linear Group SL(3)

Recall that the Special Linear group $\mathrm{SL}(3)$ is defined as the set of all real valued 3×3 matrices with unit determinant:

$$\mathrm{SL}(3) = \{H \in \mathbb{R}^{3\times 3} \mid \det H = 1\}.$$

The Lie-algebra $\mathfrak{sl}(3)$ for $\mathrm{SL}(3)$ is the set of matrices with trace equal to zero:

$$\mathfrak{sl}(3) = \{X \in \mathbb{R}^{3\times 3} \mid \mathrm{tr}(X) = 0\}.$$

The adjoint operator is a mapping $\mathrm{Ad} : \mathrm{SL}(3) \times \mathfrak{sl}(3) \to \mathfrak{sl}(3)$ defined by:

$$\mathrm{Ad}_H X = HXH^{-1}, \quad H \in \mathrm{SL}(3), X \in \mathfrak{sl}(3).$$

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

$$\bar{H} = \frac{1}{\det(H)^{\frac{1}{3}}} H \tag{9}$$

such that $\det(\bar{H}) = 1$. Every such matrix $\bar{H} \in \mathrm{SL}(3)$ occurs as a homography. Moreover, the map

$$w\colon \mathrm{SL}(3) \times \mathbb{P}^2 \longrightarrow \mathbb{P}^2,$$
$$(H, p) \mapsto w(H, p) \cong \frac{Hp}{|Hp|}$$

is a group action of $\mathrm{SL}(3)$ on the projective space $\mathbb{P}^2$ since

$$w(H_1, w(H_2, p)) = w(H_1 H_2, p),\ w(I, p) = p,$$

where H_1, H_2 and $H_1 H_2 \in \mathrm{SL}(3)$ and I is the identity matrix, the unit element of $\mathrm{SL}(3)$. The geometrical meaning of the above property is that the 3D motion of the camera between views $\mathcal{A}_0$ and $\mathcal{A}_1$, followed by the 3D motion between views $\mathcal{A}_1$ and $\mathcal{A}_2$ is the same as the 3D motion between views $\mathcal{A}_0$ and $\mathcal{A}_2$. As a consequence, we can think of homographies as described by elements of $\mathrm{SL}(3)$.

Since any homography is defined up to a scale factor, we assume from now on that $H \in \mathrm{SL}(3)$:

$$H = \gamma \left(R + \frac{\xi \eta^\top}{d} \right). \tag{10}$$

There are numerous approaches for determining H, up to this scale factor, cf. for example [17]. Note that direct computation of the determinant of H in combination with the expression of d (8) and using the fact that $\det(H) = 1$, shows that $\gamma = (\frac{d}{\mathring{d}})^{\frac{1}{3}}$.

Extracting R and $\frac{\xi}{d}$ from H is in general quite complex [1], [6], [17], [27], [24] and is beyond the scope of this paper.

III. Nonlinear observer design on SL(3) based on direct measurements

In this section, the design of an observer for $H \in \mathrm{SL}(3)$ is based on a recent theory for nonlinear observer design directly on the output space [14].

A. Kinematics and measurements

Consider the kinematics of $\mathrm{SL}(3)$ given by

$$\dot{H} = F(H, U) := HU, \tag{11}$$

with $U \in \mathfrak{sl}(3)$ the group velocity. Assume that U is measured. Furthermore, we dispose of a set of n measurements $p_i \in \mathbb{P}^2$ in the body-fixed frame:

$$p_i = h(H, \mathring{p}_i) := \frac{H^{-1}\mathring{p}_i}{|H^{-1}\mathring{p}_i|}, \quad i = \{1 \dots n\}, \tag{12}$$

where $\mathring{p}_i \in \mathbb{P}^2$ are constant and known. For later use, define

$$\mathring{p} := (\mathring{p}_1, \cdots, \mathring{p}_n), \quad p := (p_1, \cdots, p_n).$$

Definition 1: A set $\mathcal{M}_n$ of $n \geq 4$ vector directions $\mathring{p}_i \in \mathbb{P}^2$, with $i = \{1 \dots n\}$, is called *consistent*, if it contains a subset $\mathcal{M}_4 \subset \mathcal{M}_n$ of 4 constant vector directions such that all vector triplets in $\mathcal{M}_4$ are linearly independent.

This definition implies that if the set $\mathcal{M}_n$ is consistent then, for all $\mathring{p}_i \in \mathcal{M}_4$ there exists a unique set of three non vanishing scalars $b_j \neq 0$ $(j \neq i)$ such that

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

We verify that $\mathrm{SL}(3)$ is a symmetry group with group actions $\phi : \mathrm{SL}(3) \times \mathrm{SL}(3) \longrightarrow \mathrm{SL}(3)$, $\psi : \mathrm{SL}(3) \times \mathfrak{sl}(3) \longrightarrow \mathfrak{sl}(3)$ and $\rho : \mathrm{SL}(3) \times \mathbb{P}^2 \longrightarrow \mathbb{P}^2$ defined by

$$\begin{aligned} \phi(Q, H) &:= HQ, \\ \psi(Q, U) &:= Ad_{Q^{-1}} U = Q^{-1} U Q, \\ \rho(Q, p) &:= \tfrac{Q^{-1} p}{|Q^{-1} p|}. \end{aligned}$$

In effect, it is straightforward to verify that ϕ, ψ, and ρ are *right group actions* in the sense that $\phi(Q_2, \phi(Q_1, H)) = \phi(Q_1 Q_2, H)$, $\psi(Q_2, \psi(Q_1, U)) = \psi(Q_1 Q_2, U)$, and $\rho(Q_2, \rho(Q_1, p)) = \rho(Q_1 Q_2, p)$, for all $Q_1, Q_2, H \in \mathrm{SL}(3)$, $U \in \mathfrak{sl}(3)$, and $p \in \mathbb{P}^2$. Clearly,

$$\rho(Q, h(H, \mathring{p}_i)) = \frac{Q^{-1} \frac{H^{-1}\mathring{p}_i}{|H^{-1}\mathring{p}_i|}}{\left|Q^{-1} \frac{H^{-1} p_i}{|H^{-1}\mathring{p}_i|}\right|} = h(\phi(Q, H), \mathring{p}_i),$$

$$\begin{aligned} d\phi_Q(H)[F(H, U)] &= HUQ = (HQ)(Q^{-1}UQ) \\ &= F(\phi(Q, H), \psi(Q, U)). \end{aligned}$$

Thus, the kinematics are *right equivariant* (see [14]). This is an important property so that the observer design framework proposed in [14] can be used.

B. Nonlinear observer design

Let $\hat{H} \in \mathrm{SL}(3)$ denote the estimate of H. Define the right group error $E = \hat{H}H^{-1} \in \mathrm{SL}(3)$ and the output errors $e_i \in \mathbb{P}^2$:

$$e_i := \rho(\hat{H}^{-1}, p_i) = \frac{\hat{H}p_i}{|\hat{H}p_i|} = \frac{E\mathring{p}_i}{|E\mathring{p}_i|}. \tag{13}$$

The proposed observer takes the form

$$\dot{\hat{H}} = \hat{H}U - \Delta(\hat{H}, p)\hat{H}, \tag{14}$$

where $\Delta(\hat{H}, p) \in \mathfrak{sl}(3)$ is the innovation term to be designed and must be *right equivariant* in the sense that for all $Q \in \mathrm{SL}(3)$:

$$\Delta(\hat{H}Q, \rho(Q, p_1), \cdots, \rho(Q, p_n)) = \Delta(\hat{H}, p_1, \cdots, p_n).$$

Interestingly, if the innovation term Δ is right equivariant, the dynamics of the canonical error E is autonomous and given by [14, Th. 1]:

$$\dot{E} = -\Delta(E, \mathring{p})E. \tag{15}$$

In order to determine $\Delta(\hat{H}, p)$, we need a *non-degenerate right-invariant* cost function. To this purpose, let us first define individual *degenerate right-invariant costs* at $\mathring{p}_i$ on the output space $\mathbb{P}^2$ as follows:

$$\begin{aligned} &\mathcal{C}^i_{\mathring{p}_i} : \mathrm{SL}(3) \times \mathbb{P}^2 \longrightarrow \mathbb{R}^+, \\ &\qquad (\hat{H}, p_i) \;\mapsto\; \mathcal{C}^i_{\mathring{p}_i}(\hat{H}, p_i) := \frac{k_i}{2}\left|\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right|^2 \end{aligned}$$

with k_i positive gains, chosen depending on the relative confidence in the measurements. One verifies that $\mathcal{C}^i_{\mathring{p}_i}(\hat{H}, p_i)$ are right-invariant in the sense that $\mathcal{C}^i_{\mathring{p}_i}(\hat{H}Q, \rho(Q, p_i)) = \mathcal{C}^i_{\mathring{p}_i}(\hat{H}, p_i)$ for all $Q \in \mathrm{SL}(3)$. The costs $\mathcal{C}^i_{\mathring{p}_i}(\hat{H}, p_i)$ are degenerate since by taking $p_i = \mathring{p}_i$ there exists an infinity of $\hat{H}$ such that $\mathcal{C}^i_{\mathring{p}_i}(\hat{H}, \mathring{p}_i) = 0$.

Then, the aggregate cost is defined as the sum of all the individual costs as follows:

$$\begin{aligned} &\mathcal{C}_{\mathring{p}} : \mathrm{SL}(3) \times (\mathbb{P}^2 \times \cdots \times \mathbb{P}^2) \longrightarrow \mathbb{R}^+, \\ &\qquad (\hat{H}, p) \;\mapsto \mathcal{C}_{\mathring{p}}(\hat{H}, p) := \sum_{i=1}^{n} \frac{k_i}{2}\left|\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right|^2 \end{aligned} \tag{16}$$

It is straightforward that $\mathcal{C}_{\mathring{p}}(\hat{H}, p)$ is right-invariant. According to [14, Lem. 3], the aggregate cost is *non-degenerate* if

$$\bigcap_{i=1}^{n} \mathrm{stab}_\rho(\mathring{p}_i) = \{I\}$$

where the stabilizer $\mathrm{stab}_\rho(p)$ of an element $p \in \mathbb{P}^2$ is defined by $\mathrm{stab}_\rho(p) = \{Q \in \mathrm{SL}(3) \mid \rho(Q, p) = p\}$. In fact, $\bigcap_{i=1}^n \mathrm{stab}_\rho(\mathring{p}_i) = \{I\}$ is equivalent to $\bigcap_{i=1}^n \mathfrak{s}_i = \{0\}$, where $\mathfrak{s}_i = \ker(\mathrm{d}\rho_{\mathring{p}_i}(I))$ is the Lie-algebra associated with $\mathrm{stab}_\rho(\mathring{p}_i)$.

Lemma 1: Assume that the set $\mathcal{M}_n$ of the measured directions $\mathring{p}_i$ is consistent. Then, the aggregate cost $\mathcal{C}_{\mathring{p}}(\hat{H}, p)$ defined by (16) is non-degenerate. As a consequence, $(I, \mathring{p})$ is a global minimum of the aggregate cost $\mathcal{C}_{\mathring{p}}$.

Proof: One computes the derivative

$$\begin{aligned} \mathrm{d}\rho_{\mathring{p}_i}(H)[HU] &= \mathrm{d}\left(\tfrac{H^{-1}\mathring{p}_i}{|H^{-1}\mathring{p}_i|}\right)[HU] \\ &= \left(I - \tfrac{(H^{-1}\mathring{p}_i)(H^{-1}\mathring{p}_i)^\top}{|H^{-1}\mathring{p}_i|^2}\right)\tfrac{UH^{-1}\mathring{p}_i}{|H^{-1}\mathring{p}_i|} \end{aligned} \tag{17}$$

with some $U \in \mathfrak{sl}(3)$. From (17), one deduces that

$$\mathfrak{s}_i = \ker(\mathrm{d}\rho_{\mathring{p}_i}(I)) = \{U \in \mathfrak{sl}(3) \mid \pi_{\mathring{p}_i} U \mathring{p}_i = 0\}$$

with $\pi_x := (I - xx^\top)$ for all $x \in \mathbb{S}^2$. Thus,

$$\bigcap_{i=1}^{n} \mathfrak{s}_i = \{U \in \mathfrak{sl}(3) \mid \pi_{\mathring{p}_i} U \mathring{p}_i = 0, \forall i = 1, \cdots, n\}$$

Now, let us determine $U \in \mathfrak{sl}(3)$ such that $\pi_{\mathring{p}_i} U \mathring{p}_i = 0$, for all $i = 1, \cdots, n$. The relation $\pi_{\mathring{p}_i} U \mathring{p}_i = 0$ can be equivalently written as

$$U\mathring{p}_i = \lambda_i \mathring{p}_i,$$

with $\lambda_i := (\mathring{p}_i^\top U \mathring{p}_i)$. From here, one deduces that λ_i are eigenvalues of U and $\mathring{p}_i$ are the associated eigenvectors. Since the set $\mathcal{M}_n$ of the measured directions $\mathring{p}_i$ is consistent, without loss of generality we assume that the subset $\mathcal{M}_4 = \{\mathring{p}_1, \mathring{p}_2, \mathring{p}_3, \mathring{p}_4\}$ is consistent. Thus, there exist 3 non-null numbers b_1, b_2, and b_3 such that $\mathring{p}_4 = \sum_{i=1}^3 b_i \mathring{p}_i$. From here, one deduces that

$$\sum_{i=1}^{3} b_i \lambda_i \mathring{p}_i = \sum_{i=1}^{3} b_i \lambda_4 \mathring{p}_i, \tag{18}$$

using the fact that

$$U\mathring{p}_4 = U\sum_{i=1}^{3} b_1 \mathring{p}_1 = \sum_{i=1}^{3} b_1 U \mathring{p}_1 = \sum_{i=1}^{3} b_i \lambda_i \mathring{p}_i$$

and

$$U\mathring{p}_4 = \lambda_4 \mathring{p}_4 = \lambda_4 \sum_{i=1}^{3} b_i \mathring{p}_i = \sum_{i=1}^{3} b_i \lambda_4 \mathring{p}_i.$$

Since b_i (with $i = 1, 2, 3$) are non-null and the 3 unit vectors $\mathring{p}_i$ (with $i = 1, 2, 3$) are linearly independent, (18) directly yields $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$. Let λ denote the value of these four identical eigenvalues.

From here, one easily deduces that the *geometric multiplicity* of the eigenvalue λ (defined as the dimension of the eigenspace associated with λ) is equal to 3, since the 3 eigenvectors $\mathring{p}_i$ (with $i = 1, 2, 3$) associated with λ are linearly independent. Since the *algebraic multiplicity* of the eigenvalue λ is no less than the corresponding *geometric multiplicity*, one deduces that it is also equal to 3. This means that U has a triple eigenvalue λ. Since the number of linearly independent eigenvectors of U is equal to 3, the matrix U is diagonalizable. Then, the diagonalizability of U along with the fact that it has a triple eigenvalue implies that $U = \lambda I$. This in turn yields $\mathrm{tr}(U) = 3\lambda$, which is null since U is an element of $\mathrm{sl}(3)$. Consequently, $\lambda = 0$ and $U = 0$. One then deduces $\bigcap_{i=1}^n \mathfrak{s}_i = \{0\}$ which concludes the proof. ∎

Now that the non-degenerate right-invariant cost function $\mathcal{C}_{\mathring{p}}(\hat{H}, p)$ is defined, we compute the innovation term $\Delta(\hat{H}, p)$ as [14, Eq. (40)]

$$\Delta(\hat{H}, p) := (\mathrm{grad}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p))\hat{H}^{-1}, \tag{19}$$

where grad_1 is the *gradient* in the first variable, using a *right-invariant Riemannian metric* on $\mathrm{SL}(3)$. Let $\langle\cdot,\cdot\rangle : \mathfrak{sl}(3) \times \mathfrak{sl}(3) \longrightarrow \mathbb{R}$ be a positive definite inner product on $\mathfrak{sl}(3)$, chosen to be the Euclidean matrix inner product on $\mathbb{R}^{3\times 3}$. Then, a right-invariant Riemannian metric on $\mathrm{SL}(3)$ induced by the inner product $\langle\cdot,\cdot\rangle$ is defined by

$$\langle U_1 H, U_2 H\rangle_H := \langle U_1, U_2\rangle.$$

Lemma 2: The innovation term $\Delta(\hat{H}, p)$ defined by (19) is right equivariant and explicitly given by

$$\Delta(\hat{H}, p) = -\sum_{i=1}^{n} k_i \pi_{e_i} \mathring{p}_i e_i^\top, \quad \text{with } e_i = \frac{\hat{H} p_i}{|\hat{H} p_i|}. \tag{20}$$

Proof: The proof for $\Delta(\hat{H}, p)$ to be equivariant is a direct result of [14]. Now, using standard rules for transformations of Riemannian gradients and the fact that the Riemannian metric is right-invariant, one obtains

$$\begin{aligned}\mathcal{D}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p)[U\hat{H}] &= \langle \mathrm{grad}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p), U\hat{H}\rangle_H \\ &= \langle \mathrm{grad}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p)\hat{H}^{-1}\hat{H}, U\hat{H}\rangle_H \\ &= \langle \mathrm{grad}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p)\hat{H}^{-1}, U\rangle \\ &= \langle \Delta(\hat{H}, p), U\rangle,\end{aligned} \tag{21}$$

with some $U \in \mathfrak{sl}(3)$. Besides, in view of (16) one has

$$\begin{aligned}&\mathcal{D}_1 \mathcal{C}_{\mathring{p}}(\hat{H}, p)[U\hat{H}] = d_{\hat{H}} \mathcal{C}_{\mathring{p}}(\hat{H}, p)[U\hat{H}] \\ &= \textstyle\sum_{i=1}^{n} k_i \left(\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right)^\top \left(I - \frac{(\hat{H}p_i)(\hat{H}p_i)^\top}{|\hat{H}p_i|^2}\right) \frac{(U\hat{H})p_i}{|\hat{H}p_i|} \\ &= \textstyle\sum_{i=1}^{n} k_i (e_i - \mathring{p}_i)^\top (I - e_i e_i^\top) U e_i \\ &= \mathrm{tr}\left(\textstyle\sum_{i=1}^{n} k_i e_i (e_i - \mathring{p}_i)^\top (I - e_i e_i^\top) U\right) \\ &= -\mathrm{tr}\left(\textstyle\sum_{i=1}^{n} k_i e_i \mathring{p}_i^\top \pi_{e_i} U\right) \\ &= \left\langle -\textstyle\sum_{i=1}^{n} k_i \pi_{e_i} \mathring{p}_i e_i^\top, U\right\rangle.\end{aligned} \tag{22}$$

Finally, the explicit expression of $\Delta(\hat{H}, p)$ given by (20) is directly obtained from (21) and (22). ∎

One deduces from (20) that

$$\Delta(E, \mathring{p}) = -\sum_{i=1}^{n} k_i \pi_{e_i} \mathring{p}_i e_i^\top, \quad \text{with } e_i = \frac{E\mathring{p}_i}{|E\mathring{p}_i|},$$

and, consequently, from (15) that

$$\dot{E} = \left(\sum_{i=1}^{n} k_i \pi_{e_i} \mathring{p}_i e_i^\top\right) E = -\mathrm{grad}_1 \mathcal{C}_{\mathring{p}}(E, \mathring{p}). \tag{23}$$

Theorem 1: Consider the kinematics (11) and assume that the velocity group $U \in \mathfrak{sl}(3)$ is known. Consider the nonlinear observer defined by (14), with the innovation term $\Delta(\hat{H}, p) \in \mathfrak{sl}(3)$ defined by (20). Then, if the set $\mathcal{M}_n$ of the measured directions $\mathring{p}_i$ is consistent, then the equilibrium $E = I$ of the autonomous system (23) is locally asymptotically stable.

Proof: This theorem can be seen as a direct result of Theorem 2 in [14], but it can also be proved using classical Lyapunov theory. The candidate Lyapunov function under consideration is $\mathcal{L}_0 := \mathcal{C}_{\mathring{p}}(E, \mathring{p})$. Using the consistency of the set $\mathcal{M}_n$, one can ensure that $\mathcal{L}_0$ is locally a positive definite function of E. The time-derivative of $\mathcal{L}_0$ along the error flow (23) verifies

$$\begin{aligned}\dot{\mathcal{L}}_0 &= \tfrac{\mathrm{d}}{\mathrm{d}t} \textstyle\sum_{i=1}^{n} \frac{k_i}{2} \left|\frac{E\mathring{p}_i}{|E\mathring{p}_i|} - \mathring{p}_i\right|^2 \\ &= \textstyle\sum_{i=1}^{n} k_i \left(\frac{E\mathring{p}_i}{|E\mathring{p}_i|} - \mathring{p}_i\right)^\top \left(I - \frac{(E\mathring{p}_i)(E\mathring{p}_i)^\top}{|E\mathring{p}_i|^2}\right) \frac{\dot{E}\mathring{p}_i}{|E\mathring{p}_i|} \\ &= -\textstyle\sum_{i=1}^{n} k_i (e_i - \mathring{p}_i)^\top \pi_{e_i} \Delta e_i \\ &= \mathrm{tr}\left(\textstyle\sum_{i=1}^{n} k_i e_i \mathring{p}_i^\top \pi_{e_i} \Delta\right) \\ &= -||\Delta(E, \mathring{p})||^2,\end{aligned}$$

with $||\cdot||$ denoting the Frobenius norm defined by $||A|| = \sqrt{\langle A, A\rangle}$ for any real valued square matrix A. From here, one ensures that E is locally bounded. Moreover, by application of LaSalle's theorem, one deduces that $\Delta(E, \mathring{p})$ converges to zero. From the definitions of Δ (20) and e_i (13), one deduces that

$$\Delta E^{-\top} = \sum_{i=1}^{n} \left(I - \frac{E\mathring{p}_i \mathring{p}_i^\top E^\top}{|E\mathring{p}_i|^2}\right) \frac{\mathring{p}_i \mathring{p}_i^\top}{|E\mathring{p}_i|}.$$

Computing the trace of $\Delta E^{-\top}$, it follows:

$$\mathrm{tr}(\Delta E^{-\top}) = \sum_{i=1}^{n} \frac{1}{|E\mathring{p}_i|^3} \left(|E\mathring{p}_i|^2 |\mathring{p}_i|^2 - ((E\mathring{p}_i)^\top \mathring{p}_i)^2\right).$$

Define $X_i = E\mathring{p}_i$ and $Y_i = \mathring{p}_i$, and it is straightforward to verify that

$$\mathrm{tr}(\Delta E^{-\top}) = \sum_{i=1}^{n} \frac{1}{|X_i|^3} \left(|X_i|^2 |Y_i|^2 - (X_i^\top Y_i)^2\right) \geq 0$$

Using the fact that $\Delta = 0$ at the equilibrium and therefore $\mathrm{tr}(\Delta E^{-\top}) = 0$, as well as the Cauchy-Schwarz inequality, it follows that $X_i^\top Y_i = \pm |X_i||Y_i|$ and consequently one has:

$$(E\mathring{p}_i)^\top \mathring{p}_i = \pm |E\mathring{p}_i||\mathring{p}_i|, \ \forall i = \{1, \cdots, n\},$$

which in turn implies the existence of some scalars $\lambda_i = \pm |E\mathring{p}_i|$ such that

$$E\mathring{p}_i = \lambda_i \mathring{p}_i. \tag{24}$$

Using (24) and exploiting the consistency of the set $\mathcal{M}_n$, one can proceed analogously to the proof of Lemma 1 to deduce that E has a triple eigenvalue λ and $E = \lambda I$. Then, evoking the fact that $\det(E) = 1$, one deduces that $\lambda = 1$ and $E = I$. Consequently, E converges asymptotically to the identity I. ∎

Remark 1: The boundaries of the stability domain associated with Theorem 1 are extremely difficult, and probably impossible, to analytically characterise. The nature of the Lyapunov function $\mathcal{L}_0$ is always well conditioned in the neighbourhood of the correct homography matrix, but the global geometry of $\mathrm{SL}(3)$ is complex and there will be critical points and non-convex cost structure when the rotation component of the homography matrix has errors approaching π rads from the true homography. The authors believe, based on extensive simulation and experimental studies (cf. Sections V and VI) and our intuition for such problems, that the stability domain is very large in practice, encompassing all realistic scenarios where the camera actually observes the desired scene (rotations of less than $\pi/2$ rads and moderate displacements).

IV. Application to Robotic Systems

A. Homography kinematics from a camera moving with rigid-body motion

In this section, we consider the case where a sequence of homographies is generated by a moving camera viewing a stationary planar surface. The goal is to develop a non-linear filter for the image homography sequence using the velocity associated with the rigid-body motion of the camera rather than the group velocity of the homography sequence, as was assumed in Section III. In fact, any group velocity (infinitesimal variation of the homography) must be associated with an instantaneous variation in measurement of the *current* image $\mathcal{A}$ and not with a variation in the *reference* image $\mathring{\mathcal{A}}$. This imposes constraints on two degrees of freedom in the homography velocity, namely those associated with variation of the normal to the reference image, and leaves the remaining six degrees of freedom in the homography group velocity depending on the rigid-body velocities of the camera.

Denote the rigid-body angular velocity and linear velocity of $\{A\}$ with respect to $\{\mathring{A}\}$ expressed in $\{A\}$ by Ω and V, respectively. The rigid body kinematics of (R, ξ) are given by:

$$\dot{R} = R\Omega_\times \tag{25}$$
$$\dot{\xi} = RV \tag{26}$$

where $\Omega_\times$ is the skew symmetric matrix associated with the vector cross-product, i.e. $\Omega_\times y = \Omega \times y$, for all $y \in \mathbb{R}^3$.

Recalling (8), it is easily verified that:

$$\dot{d} = -\eta^\top V, \qquad \frac{d}{dt}\mathring{d} = 0.$$

This constraint on the variation of η and $\mathring{d}$ is precisely the velocity constraint associated with the fact that the reference image is stationary.

Consider a camera attached to the moving frame $\mathcal{A}$ moving with kinematics (25) and (26) viewing a stationary planar scene. The group velocity $U \in \mathfrak{sl}(3)$ induced by the rigid-body motion, and such that the dynamics of H satisfies (11), is given by [13, Lem. 5.3]

$$U = \left(\Omega_\times + \frac{V\eta^\top}{d} - \frac{\eta^\top V}{3d} I\right).$$

Note that the group velocity U induced by camera motion depends on the additional variables η and d that define the scene geometry at time t as well as the scale factor γ. Since these variables are unmeasurable and cannot be extracted directly from the measurements, in the sequel, we rewrite:

$$U := (\Omega_\times + \Gamma), \text{ with } \Gamma = \frac{V\eta^\top}{d} - \frac{\eta^\top V}{3d} I. \tag{27}$$

Since $\{\mathring{A}\}$ is stationary by assumption, the vector Ω can be directly obtained from the set of embedded gyroscopes. The term Γ is related to the translational motion expressed in the current frame $\{A\}$. If we assume that $\frac{\dot{\xi}}{d}$ is constant (*e.g.* the situation in which the camera moves with a constant velocity parallel to the scene or converges exponentially toward it), and using the fact that $V = R^\top \dot{\xi}$, it is straightforward to verify that

$$\dot{\Gamma} = [\Gamma, \Omega_\times], \tag{28}$$

where $[\Gamma, \Omega_\times] = \Gamma\Omega_\times - \Omega_\times\Gamma$ is the Lie bracket.

However, if we assume that $\frac{V}{d}$ is constant (the situation in which the camera follows a circular trajectory over the scene or performs an exponential convergence towards it), it follows that

$$\dot{\Gamma}_1 = \Gamma_1\Omega_\times, \text{ with } \Gamma_1 = \frac{V}{d}\eta^\top. \tag{29}$$

B. Observer with partially known velocity of the rigid body

In this section we assume that the group velocity U in (27) is not available, but the angular velocity Ω is. The goal is to provide an estimate $\hat{H} \in \mathrm{SL}(3)$ for $H \in \mathrm{SL}(3)$ to drive the group error $E\,(= \hat{H}H^{-1})$ to the identity matrix I and the error term $\tilde{\Gamma} = \Gamma - \hat{\Gamma}$ (resp. $\tilde{\Gamma}_1 = \Gamma_1 - \hat{\Gamma}_1$) to 0 if Γ (resp. Γ_1) is constant or slowly time varying. The observer when $\frac{\dot{\xi}}{d}$ is constant is chosen as follows (compare to (14)):

$$\dot{\hat{H}} = \hat{H}(\Omega_\times + \hat{\Gamma}) - \Delta(\hat{H}, p)\hat{H}, \tag{30}$$
$$\dot{\hat{\Gamma}} = [\hat{\Gamma}, \Omega_\times] - k_I \mathrm{Ad}_{\hat{H}^\top}\Delta(\hat{H}, p). \tag{31}$$

and the observer when $\frac{V}{d}$ is constant is defined as follows:

$$\dot{\hat{H}} = \hat{H}(\Omega_\times + \hat{\Gamma}_1 - \frac{1}{3}\mathrm{tr}(\hat{\Gamma}_1)I) - \Delta(\hat{H}, p)\hat{H}, \tag{32}$$
$$\dot{\hat{\Gamma}}_1 = \hat{\Gamma}_1\Omega_\times - k_I \mathrm{Ad}_{\hat{H}^\top}\Delta(\hat{H}, p). \tag{33}$$

with some positive gain k_I and $\Delta(\hat{H}, p)$ given by (20).

Proposition 1: Consider a camera moving with kinematics (25) and (26) viewing a planar scene. Assume that $\mathring{\mathcal{A}}$ is stationary and that the orientation velocity $\Omega \in \{A\}$ is measured and bounded. Let $H : \mathcal{A} \to \mathring{\mathcal{A}}$ denote the calibrated homography (10) and consider the kinematics (11) along with (27). Assume that H is bounded and that Γ (resp. Γ_1) is such that it obeys (28) (resp. (29)).
Consider the nonlinear observer defined by (30–31), (resp. (32–33)) along with the innovation $\Delta(\hat{H}, p) \in \mathfrak{sl}(3)$ given by (20). Then, if the set $\mathcal{M}_n$ of the measured directions $\mathring{p}_i$ is consistent, the equilibrium $(E, \tilde{\Gamma}) = (I, 0)$ (resp. $(E, \tilde{\Gamma}_1) = (I, 0)$) is locally asymptotically stable.

Proof: We will consider only the situation where the estimate of Γ is used. The same arguments can also be used for the case where the estimate of Γ_1 is considered. Differentiating e_i (13) and using (30) yields

$$\dot{e}_i = -\pi_{e_i}(\Delta + \mathrm{Ad}_{\hat{H}}\tilde{\Gamma})e_i.$$

Define the following candidate Lyapunov function:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_0 + \tfrac{1}{2k_I}||\tilde{\Gamma}||^2 \\ &= \textstyle\sum_{i=1}^n \frac{k_i}{2}|e_i - \mathring{p}_i|^2 + \frac{1}{2k_I}||\tilde{\Gamma}||^2. \end{aligned} \tag{34}$$

Differentiating $\mathcal{L}$ and using the fact that $\mathrm{tr}\left(\tilde{\Gamma}^\top\left(\left[\tilde{\Gamma}, \Omega\right]\right)\right) = 0$, it follows:

$$\dot{\mathcal{L}} = \sum_{i=1}^n k_i(e_i - \mathring{p}_i)^\top \dot{e}_i + \mathrm{tr}\left(\tilde{\Gamma}^\top \mathrm{Ad}_{\hat{H}^T}\Delta\right).$$

Introducing the above expression of $\dot{e}_i$ and using the fact that $\text{tr}(AB) = \text{tr}(B^\top A^\top)$, it follows:

$$\begin{aligned}\dot{\mathcal{L}} &= -\sum_{i=1}^{n} k_i(e_i - \mathring{p}_i)^\top \pi_{e_i}(\Delta + \text{Ad}_{\hat{H}}\tilde{\Gamma})e_i + \text{tr}\left(\text{Ad}_{\hat{H}^{-1}}\Delta^\top\tilde{\Gamma}\right)\\ &= \sum_{i=1}^{n} k_i \mathring{p}_i^\top \pi_{e_i}(\Delta + \text{Ad}_{\hat{H}}\tilde{\Gamma})e_i + \text{tr}\left(\text{Ad}_{\hat{H}^{-1}}\Delta^\top\tilde{\Gamma}\right)\\ &= \text{tr}\left(\sum_{i=1}^{n} k_i e_i \mathring{p}_i^\top \pi_{e_i}(\Delta + \text{Ad}_{\hat{H}}\tilde{\Gamma}) + \text{Ad}_{\hat{H}^{-1}}\Delta^\top\tilde{\Gamma}\right)\\ &= \text{tr}\left(\sum_{i=1}^{n} k_i e_i \mathring{p}_i^\top \pi_{e_i}\Delta + \text{Ad}_{\hat{H}^{-1}}(\Delta^\top + \sum_{i=1}^{n} k_i e_i \mathring{p}_i^\top \pi_{e_i})\tilde{\Gamma}\right)\end{aligned}$$

Finally, introducing the expression of Δ (20), one gets

$$\dot{\mathcal{L}} = -\|\Delta\|^2.$$

The derivative of the Lyapunov function is negative semi-definite, and equal to zero when $\Delta = 0$. Given that Ω is bounded, it is easily verified that $\dot{\mathcal{L}}$ is uniformly continuous and Barbalat's Lemma can be used to prove asymptotic convergence of $\Delta \to 0$. Using the same arguments as in the proof of Theorem 1, it is straightforward to verify that $E \to I$. Consequently the left-hand side of the Lyapunov expression (34) converges to zero and $\|\tilde{\Gamma}\|^2$ converges to a constant.

Computing the time derivative of E and using the fact that Δ converges to zero and E converges to I, it is straightforward to show that

$$\lim_{t\to\infty} \dot{E} = -\text{Ad}_{\hat{H}}\tilde{\Gamma} = 0.$$

Using boundedness of H, one ensures $\lim_{t\to\infty} \tilde{\Gamma} = 0$. ∎

C. Outlier rejection with M-estimator-like nonlinear observer

The cost function (16) previously used for observer design is the sum of individual squared residuals r_i^2, with $r_i := \left|\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right|$. This function is, however, extremely unstable when outliers are present in the measurements. In our case, outliers are a result of wrong feature matching, which is almost unavoidable in practice. Inspired by the robust M-estimator techniques (see, e.g., [26] and the references therein), we consider the following cost function [instead of (16)]:

$$\mathcal{C}_{\mathring{p}}(\hat{H}, p) := \sum_{i=1}^{n} k_i \rho(r_i), \tag{35}$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen less increasing than square. From there, the proposed observer is still given by the general form (14) and the innovation term $\Delta(\hat{H}, p)$ is still derived from (19). From (35) and analogously to (22), one deduces

$$\begin{aligned}&\mathcal{D}_1\mathcal{C}_{\mathring{p}}(\hat{H}, p)[U\hat{H}] = d_{\hat{H}}\mathcal{C}_{\mathring{p}}(\hat{H}, p)[U\hat{H}]\\ &= \sum_{i=1}^{n} k_i w(r_i)\left(\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right)^\top \left(I - \frac{(\hat{H}p_i)(\hat{H}p_i)^\top}{|\hat{H}p_i|^2}\right)\frac{(U\hat{H})p_i}{|\hat{H}p_i|}\\ &= \left\langle -\sum_{i=1}^{n} k_i w(r_i)\pi_{e_i}\mathring{p}_i e_i^\top, U\right\rangle,\end{aligned} \tag{36}$$

where $w(\cdot)$ is the weight function defined by

$$w(x) := \frac{1}{x}\frac{d\rho(x)}{dx}.$$

Several possible choices of ρ and its associated w function are discussed in [26]. For instance, the Tukey-ρ function, which is chosen for the experimental Section VI, is given by

$$\rho(x) = \begin{cases} (c^2/6)(1 - [1 - (x/c)^2]^3) & \text{if} \quad |x| \le c \\ c^2/6 & \text{if} \quad |x| > c \end{cases}$$

and, thus,

$$w(x) = \begin{cases} [1 - (x/c)^2]^2 & \text{if} \quad |x| \le c \\ 0 & \text{if} \quad |x| > c \end{cases}$$

From (21) and (36) one obtains [instead of (20)]

$$\Delta(\hat{H}, p) = -\sum_{i=1}^{n} k_i w(r_i)\pi_{e_i}\mathring{p}_i e_i^\top. \tag{37}$$

From there, the same results as in Theorem 1 and Proposition 1 can be stated but with the innovation term $\Delta(\hat{H}, p)$ given by (37) instead of (20). The proofs proceed identically to the proofs of Theorem 1 and Proposition 1 where the function $\mathcal{L}_0$ is chosen equal to $\mathcal{C}_{\mathring{p}}(\hat{H}, p)$ defined in (35) instead of (16).

V. Simulation Results

In this section, we illustrate the performance and robustness of the proposed observers through simulation results. The camera is assumed to be attached to an aerial vehicle moving in a circular trajectory which stays in a plane parallel to the ground. The reference camera frame $\{\mathring{A}\}$ is chosen as the NED (North-East-Down) frame situated above four observed points on the ground. The four observed points form a square whose center lies on the Z-axis of the NED frame $\{\mathring{A}\}$. The vehicle's trajectory is chosen such that the term Γ_1 defined by (29) remains constant, and the observer (32-33) is applied with the following gains: $k_P = 4$, $k_I = 1$. Distributed noise of variance 0.01 is added on the measurement of the angular velocity Ω. The chosen initial estimated homography $\hat{H}(0)$ corresponds to *i)* an error of $\pi/2$ in both pitch and yaw angles of the attitude, and *ii)* an estimated translation equal to zero. The initial value of $\hat{\Gamma}_1$ is set to zero. From $40s$ to $45s$, we assumed that the measurements of two observed points are lost. Then, from $45s$ we regain the measurements of all four points as previously.

The results reported in Fig. 1 show a good convergence rate of the estimated homography to the real homography (see from 0 to $40s$ and from $45s$). The loss of point measurements marginally affects the global performance of the proposed observed. Note that in this situation, no existing method for extracting the homography from measurements of only two points is available.

VI. Application to Image Stabilization with Real-Time Implementation

A. Experimental implementation aspects

Feature detection and matching: Our code has been implemented in C++ with OpenCV library. Due to real-time constraints, feature detection and descriptor extraction in

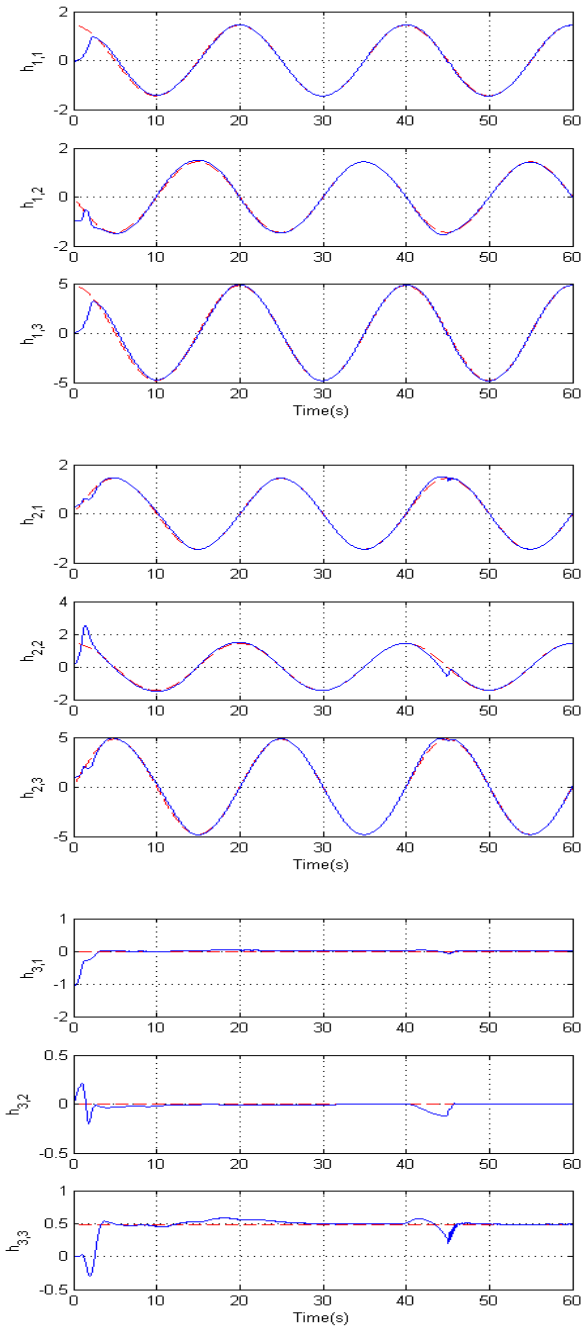

Fig. 1. Estimated homography components (solid line) and true homography ones (dashed line) vs. time.

images are carried out using the FAST Feature Detector and ORB Descriptor Extractor algorithms already implemented in the OpenCV library. Then, feature matching is performed using OpenCV's brute-force matcher routine with L_2-norm. We have purposefully avoided using more sophisticated image processing routines in order to demonstrate the raw performance of our observer.

It is quite unrealistic to track one and the same set of point-features through a long image sequence. We have hence opted to match point-features between the reference image and each subsequent image frame separately. To do this, we first compute a predicted homography estimate $\hat{H}_k^+$ that will be used to transform the current image I_k (i.e. warp the current image using $\hat{H}_k^+$ to obtain a prediction of the reference image) using the OpenCV's `warpPerspective` function before applying feature extraction and matching. More precisely, in case where the angular velocity measurements are available, $\hat{H}_k^+$ is obtained by forward integrating the observer equations (30)–(31) during the time period $[t_{k-1}, t_k]$ using only the gyrometer measurements (i.e. setting the observer gains k_i $(i = 1, \cdots, n)$ and k_I to zero) and using $(\hat{H}_{k-1}, \hat{\Gamma}_{k-1})$ as initial conditions. Note that the brute-force matching algorithm is well suited to this approach since it favors translational motion over rotational motion, and most of the rotational motion has already been compensated for by forward integrating the angular velocity. On the other hand, in case of absence of angular velocity measurements, $\hat{H}_k^+$ is simply set equal to $\hat{H}_{k-1}$.

Note that the procedure described above does not satisfy the assumptions of our theoretical convergence results, not least because different point-features are selected for each video frame. However, it is intuitively clear that such a procedure will still work as long as the selected features provides a consistent set. Indeed, our experimental results strongly support this claim.

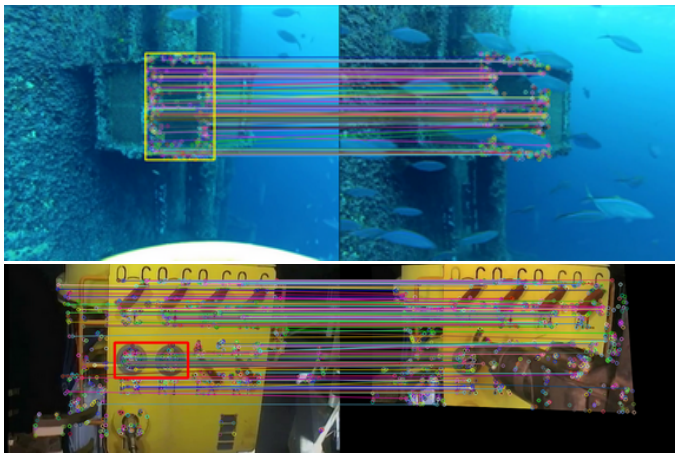

Fig. 2. Good quality of feature matching between the warped image frame ♯420 **(Top-Right)** and the reference frame **(Top-Left)** of the video sequence ♯1 and between the warped image frame ♯2200 **(Bottom-Right)** and the reference frame **(Bottom-Left)** of the video sequence ♯2 when applying our outlier removal procedure despite severe occlusions by fishes or a manipulator arm.

Outlier removal: To remove matched point-feature outliers, we first compute the standard deviation (sd_u, sd_v) and mean values (m_u, m_v) of the differences of coordinates in pixel (du_k, dv_k) of the point correspondences and then keep only those satisfying

$$\begin{cases} m_u - \max(sd_u, S) \leq du_k \leq m_u + \max(sd_u, S) \\ m_v - \max(sd_v, S) \leq dv_k \leq m_v + \max(sd_v, S) \\ |du_k| \leq D, \ |dv_k| \leq D \end{cases}$$

with S, D pre-defined positive thresholds ($S = 30$, $D = 80$ in our experiments). Again, we have ***purposefully avoided*** the use of more sophisticated (and much more computationally expensive) alternative algorithms for outlier removal, such as RANSAC [7]. Our simple and fast outlier removal method has yielded quite good matching results (see, e.g., Fig. 2).

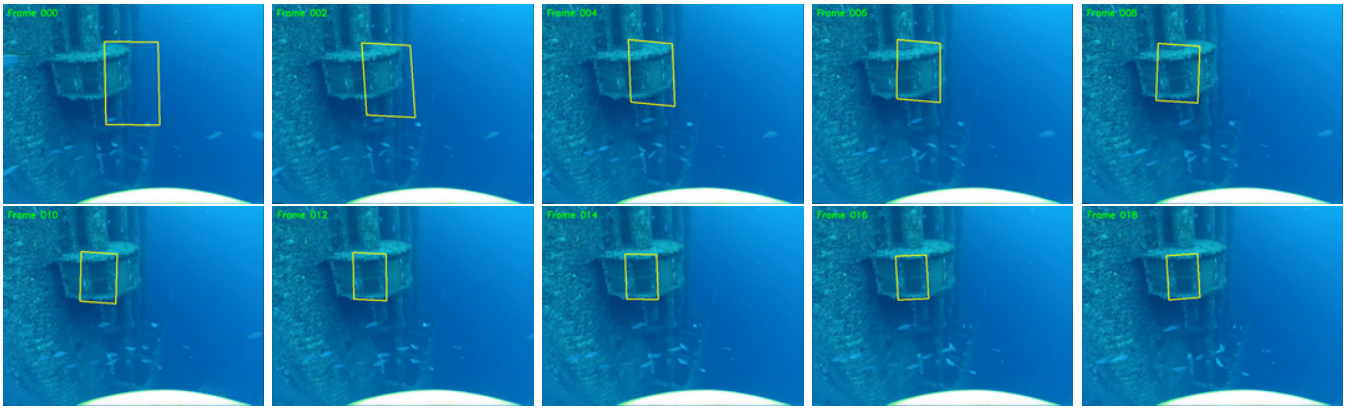

Fig. 3. *Video sequence* $\sharp 1$ *(offline validation).* Fast convergence of the image homography estimates to the real ones (less than $0.3s$), which can be appreciated by the positions of the yellow quadrangle (i.e. warped reference rectangle using the image homography estimate) in the reported image frames $\sharp 0, 2, 4, 6, 8, 10$.

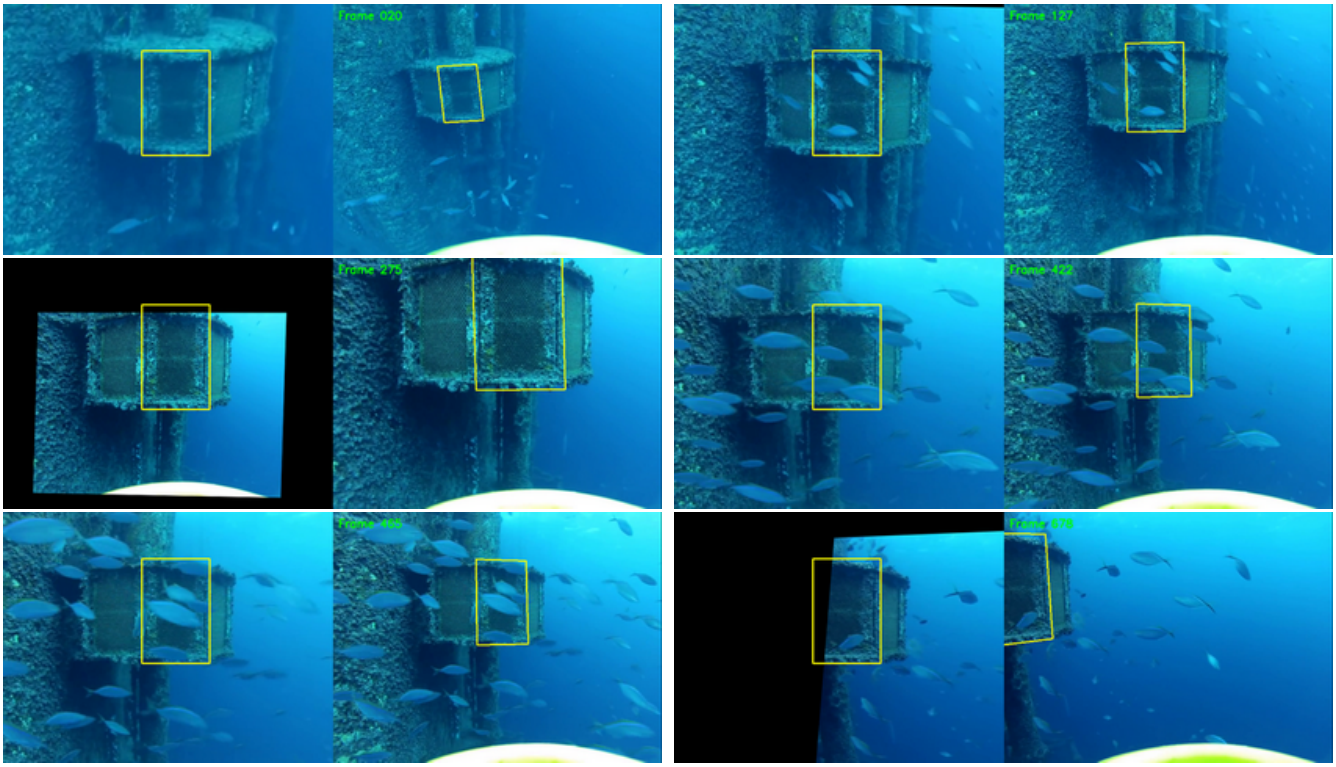

Fig. 4. *Warped and current image frames* $\sharp 20, 127, 275, 422, 465, 678$ *of the video sequence* $\sharp 1$ *(offline validation).* Excellent and robust performance of the proposed approach can be appreciated from either the warped images **(Right)** or the yellow quadrangle in the current images **(Left)** despite image low resolution, large camera motions (e.g. frames $\sharp 20, 275, 678$) and important occlusions due to fishes (e.g. frames $\sharp 127, 422, 465$). A complete video showing this experiment is available at https://youtu.be/-eMjWSWVm2A.

However, outliers cannot be completely removed especially in case of occlusions. This in turn justifies the usefulness of our M-estimator-like observer for outlier rejection.

Correction step of the observer: After the steps of feature detection and matching, we use the observer gains of $k_i = 80$ $(i = 1, \cdots, n)$, $k_I = k_i/10$ and the Tukey ρ function (see [26]) with parameter $c = 0.0025$ to rapidly iterate the observer equations 200 times per video frame[2]. The computational effort for this last step is negligible compared to the previous image processing steps.

[2]Note that the observer gains used for each iteration are divided by the number of iteration, i.e. $\bar{k}_i = k_i/N_{iter}$ and $\bar{k}_I = k_I/N_{iter}$ with $N_{iter} = 200$.

B. Offline validations using Youtube videos

Two video sequences (downloaded from Youtube) filmed from a camera mounted on an underwater robotic vehicle are used to test our algorithm for the classical image stabilization problem. Since the camera parameters are not known, we just make use of very rough (and certainly very erroneous) estimates of these parameters. The reported results presented hereafter indicate that, as long as the image homography estimate (used to warp the current image) is concerned, the stabilization performance of the proposed observer is very robust with respect to these parameter uncertainties.

These two video sequences show some realistic scenarios of inspection of underwater infrastructures using underwater

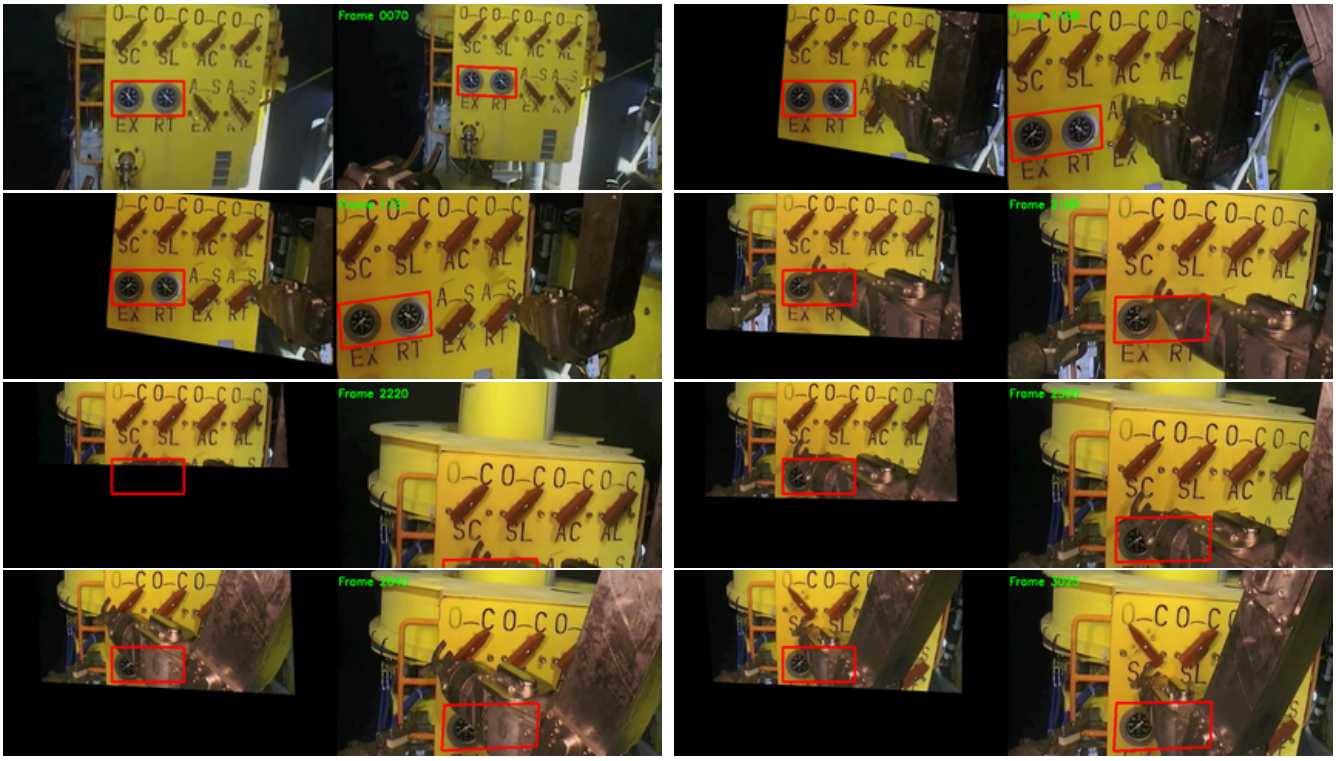

Fig. 5. *Warped and current image frames* ♯70, 1158, 1732, 2195, 2220, 2300, 2640, 3025 *of the video sequence* ♯2 *(offline validation).* Excellent and robust performance of the proposed approach is illustrated by stable warped images **(Right)** and the correct location of the red quadrangle (that warps the interested red rectangle in the reference image using the estimated homography) in the current images **(Left)** despite extreme occlusions due to the manipulator arm (e.g. frames ♯1158, 1732, 2195, 2300, 2640, 3025) and changes in the target due to turned valves (e.g. frames ♯1158, 1731, 3025). A complete video showing this experiment is available at https://youtu.be/XqZkOC01eys

vehicles. Some challenging issues such as occlusions (due to moving fishes or a robotic manipulator), low-light and low-textured scene, large translation motions, poor image quality, etc. allow us to test the robustness and performance of the proposed approach. Moreover, the absence of the angular velocity measurements in association with each image frame also renders the feature matching procedure more challenging.

The full validation videos of these two video sequences are available at:

- Video sequence ♯1: https://youtu.be/-eMjWSWVm2A
- Video sequence ♯2: https://youtu.be/XqZkOC01eys

1) Video sequence ♯1*:* In this video sequence ($30fps$ and $632 \times 480px$ of resolution), the underwater vehicle that carries the camera approaches a 3D complex structure composed of several roughly planar surfaces. The proposed homography observer is tested using the image frame ♯180 as the reference image. Feature detection is carried out within a small region of interest (yellow rectangle depicted in Fig. 2) that corresponds to one of the planar surfaces of the "box" structure. The initial homography estimate is set equal to the identity matrix, but the initial homography error is quite large since the camera's initial position is relatively far from the reference one (corresponding to the reference image). However, as shown in Fig. 3 and in view of the location of the yellow quadrangle (i.e. warped reference rectangle using the image homography estimate) in the reported images (i.e. frames ♯0, 2, 4, 6, 8, 10), one observes a very fast convergence of the image homography estimate to the real one. From the warped images in Fig. 4, very good and robust quality of image stabilization can be appreciated despite poor image quality, large motions of the camera (frames ♯20, 275, 678) and large occlusions due to fishes (frames ♯127, 422, 465) which in turn highlights the excellent performance of our approach. A complete video showing this offline experiment can be viewed at https://youtu.be/-eMjWSWVm2A where the second part of the video (starting from second 23) shows a comparison between our approach with the classical algebraic least-square homography algorithm using the OpenCV's `findHomography` function with RANSAC option activated for outlier removal. One easily observes that, by contrast with our algorithm, the performance and robustness of the classical least-square algorithm are very poor even when combined with the sophisticated RANSAC outlier removal procedure.

2) Video sequence ♯2*:* This video sequence ($25fps$ and $640 \times 360px$ of resolution) shows a scenario where an underwater vehicle is stabilized in front of an artificial panel and a mounted manipulator arm is teleoperated to turn valves on this panel. We apply our homography observer to this sequence using the image frame ♯530 as the reference image. The region of interest for feature detection is the whole image. Similarly to the first video sequence, we initially set $(\hat{H}_0, \hat{\Gamma}_0) = (I, 0)$. As the initial homography error is not very large like in the first test and the feature quality is good, the challenge is here essentially related to: 1) extreme and permanent occlusions due to the manipulator arm and/or camera rotation, and 2) the fact that some parts of the observed scene are modified

throughout the video sequence (i.e. valves are turned by the manipulator arm). Analogously to the first test, Fig. 5 shows an excellent and extremely robust performance of our approach despite the above-mentioned challenges related to occlusions and changes inside the target. A complete video showing this experiment is available at https://youtu.be/XqZkOC01eys

Fig. 6. Real-time experimental setup and validation.

C. Real-time experimental validations using an IMU-Camera system

In this section, we show that real-time implementation of our algorithm is possible. We make use of a Visual-Inertial (VI) sensor developed by the Autonomous Systems Lab (ETH Zurich) and the company Skybotix [21]. This VI-sensor is composed of two cameras (Aptina MT9V034 CMOS) and an IMU (Analog Devices ADIS16375 MEMS). However, only the left camera is used to validate the proposed algorithm. The main reason for using the VI-sensor in this experimental setup is the possibility of obtaining perfectly time-synchronized images and IMU readings ($20Hz$ for the camera and $200Hz$ for the IMU). The implementation has been carried out on an Intel Core i7-6400 CPU running at 3.40Ghz. The transmission of data from the camera to the PC is carried out through a high speed ethernet cable. The PC has a Linux based operating system and is responsible for two major tasks: 1) interfacing with the camera hardware and acquisition of images and IMU data from the VI-sensor; and 2) real-time estimation of the homography at $20Hz$ using the proposed algorithm.

Real-time implementation of the proposed observer has been performed successfully, as illustrated by the two following videos (as depicted in Fig. 6) that show the excellent performance and robustness of our algorithm:

- Fast motion: https://youtu.be/PeoaUzDkyUo
- Very fast motion: https://youtu.be/cctG_jKelXo

As shown in Fig. 7, our algorithm is robust with respect to very fast translational and rotational motions, poor image quality, severe image blurs due to fast motions and extreme occlusions. Even when our algorithm selects wrong feature matches (e.g. from frame ♯1140 to frame ♯1190 of the first video), the observer continues to track the region of interest well and quickly recovers from any tracking errors.

VII. Conclusions

In this paper we developed a nonlinear observer for a sequence of homographies represented as elements of the Special Linear group $\mathrm{SL}(3)$. The observer directly uses point-feature correspondences from an image sequence without requiring explicit computation of the individual homographies between any two given images and fuses these measurements with measurements of angular velocity from onboard gyroscopes using the correct Lie group geometry. The stability of the observer has been proved for both cases of known full group velocity and known rigid-body velocities only. Even if the characterization of the stability domain still remains an open issue, simulation and experimental results have been provided as a complement to the theoretical approach to demonstrate a large domain of stability. A potential application to image stabilization in the presence of very fast camera motions and severe occlusions has been demonstrated with very encouraging results even for a relatively low video frame rate.

Acknowledgments

This work was supported by the FUI GreenExplorer project, the EQUIPEX Robotex project, and the *Chaire d'excellence en robotique RTE-UPMC*, and the *Australian Research Council* through the "Australian Centre of Excellence for Robotic Vision" CE140100016.

References

[1] S. Benhimane and E. Malis. Homography-based 2D visual tracking and servoing. *International Journal of Robotics Research*, 26(7): 661–676, 2007.

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

[3] F. Caballero, L. Merino, J. Ferruz, and A. Ollero. Homography based Kalman filter for mosaic building. Applications to UAV position estimation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2004–2009, 2007.

[4] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel. Visual servoing for underactuated VTOL UAVs: a linear homography-based approach. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3004-3010, 2011.

[5] Y. Fang, D. M. Dawson, W. E. Dixon, and M. S. de Queiroz. Homography-based visual servoing of wheeled mobile robots. In *IEEE Conference on Decision and Control (CDC)*, 2002.

[6] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recongnition and Artificial Intelligence*, 2(3): 485–508, 1988.

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6): 381-395, 1981.

[8] F. Fraundorfer, C. Engels, and D. Nister. Topological mapping, localization and navigation using image collections. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3872–3877, 2007.

[9] T. Hamel, R. Mahony, J. Trumpf, P. Morin, and M.-D. Hua. Homography estimation on the Special Linear group based on direct point correspondence. In *Proceedings of the 50th IEEE Conference on Decision and Control (CDC)*, pages 7902–7908, 2011.

[10] R. Hartley and A. Zisserman. *Multiple View Geomerty in Computer Vision*, Cambridge University Press, second edition, 2003.

[11] M.-D. Hua, G. Allibert, S. Krupoinski, and T. Hamel. Homography-based Visual Servoing for Autonomous Underwater Vehicles. In *Proceedings of the 19th IFAC World Congress*, pages 5726–5733, 2014.

[12] C. Lageman, J. Trumpf, and R. Mahony. Gradient-like observers for invariant dynamics on a Lie group. *IEEE Transactions on Automatic Control*, 55(2): 367–377, 2010.

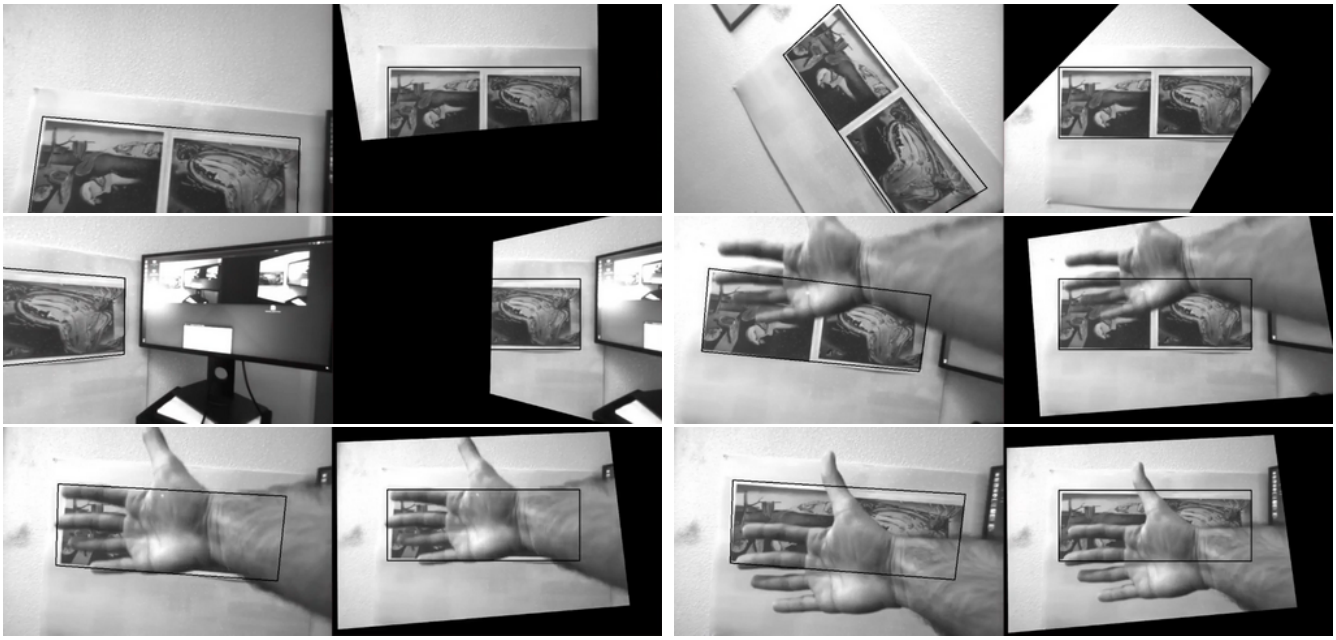

Fig. 7. Current **(Left)** and warped **(Right)** image frames $\sharp 114, 200, 482, 844, 964, 975$ of the real-time experiment video reported in https://youtu.be/PeoaUzDkyUo that show the excellent and robust performance of the proposed observer despite severe occlusions (e.g. frames $\sharp 482, 844, 964, 975$) and fast rotational motions (e.g. frames $\sharp 114, 200, 482$).

[13] R. Mahony, T. Hamel, P. Morin, and E. Malis. Nonlinear complementary filters on the special linear group. *International Journal of Control*, 85(10): 1557–1573, 2012.

[14] R. Mahony, J. Trumpf and T. Hamel. Observers for Kinematic Systems with Symmetry. In *9th IFAC Symposium on Nonlinear Control Systems*, pages 617–633, 2013.

[15] E. Malis and F. Chaumette and S. Boudet 2-1/2-D visual servoing. *IEEE Transactions on Robotics and Automation*, 15(2): 238–250, 1999.

[16] E. Malis, T. Hamel, R. Mahony, and P. Morin. Dynamic estimation of homography transformations on the special linear group for visual servo control. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1498–1503, 2009.

[17] E. Malis and M. Vargas, Deeper understanding of the homography decomposition for vision-based control. INRIA, Tech. Rep. 6303, 2007, available at http://hal.inria.fr/inria-00174036/fr/.

[18] C. Mei, S. Benhimane, E. Malis, and P. Rives. Efficient homography-based tracking and 3-D reconstruction for single-viewpoint sensors. *IEEE Transactions on Robotics*, 24(6): 1352–1364, 2008.

[19] I. F. Mondragón, P. Campoy, C. Martínez, and M. A. Olivares-Méndez. 3D pose estimation based on planar object tracking for UAVs control. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 35–41, 2010.

[20] F. Mufti, R. Mahony, and J. Kim. Super resolution of speed signs in video sequences. In *Proceedings of Digital Image Computing: Techniques and Applications (DICTA)*, pages 278–285, 2007.

[21] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P.T. Furgale, and R. Siegwart. A synchronized visual-inertial sensor system with FPGA pre-processing for accurate real-time SLAM. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 431437, 2014.

[22] D. Scaramuzza and R. Siegwart. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Transaction on Robotics*, 24: 1015–1026, 2008.

[23] H. Wang, K. Yuan, W. Zou, and Q. Zhou. Visual odometry based on locally planar ground assumption. In *Proceedings of the IEEE International Conference on Information Acquisition*, pages 59–64, 2005.

[24] J. Weng, N. Ahuja, and T. S. Huang. Motion and structure from point correspondences with error estimation: planar surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2691–2717, 1991.

[25] B. Zhang, Y. Li, and Y. Wu. Self-recalibration of a structured light system via plane-based homography. *Pattern Recognition*, 40: 1368–1377, 2007.

[26] Z. Zhang. Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):5976, 1997.

[27] Z. Zhang and A. R. Hanson. 3D reconstruction based on homography mapping. In *Proceedings of the ARPA Image Understanding Workshop*, pages 1007–1012, 1996.