Point and line feature-based observer design on SL(3) for Homography estimation and its application to image stabilization

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

▶ **To cite this version:**

Minh-Duc Hua, Jochen Trumpf, Tarek Hamel, Robert Mahony, Pascal Morin. Point and line feature-based observer design on SL(3) for Homography estimation and its application to image stabilization. [Research Report] I3S, Université Côte d'Azur; CNRS, University of Nice Sophia Antipolis, I3S, UMR 7271, COATI, Inria, 06900 Sophia Antipolis, France. 2017. hal-01628175v1

HAL Id: hal-01628175

https://hal.science/hal-01628175v1

Submitted on 2 Nov 2017 (v1), last revised 21 Sep 2018 (v2)

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.

Point and line feature-based observer design on $\mathrm{SL}(3)$ for Homography estimation and its application to image stabilization

Minh-Duc Hua, Jochen Trumpf, Tarek Hamel, Robert Mahony, 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 a monocular camera. The approach taken exploits the underlying Special Linear group $\mathrm{SL}(3)$ structure of the set of homographies along with gyrometer measurements and direct point- and line-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 even in the case of very fast camera motion (relative to frame rate), and in presence of severe occlusion, specular reflection, image blur, and light saturation.**

I. Introduction

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 [15], [16]. Navigation of robotic vehicles has been developed based on homography sequences [3], [5] and one of the most successful visual servo control paradigms uses homographies [11], [12]. Homography-based methods are particularly well suited for navigation of UAV [2] where the ground terrain is viewed from a distance from which the relief of surface features is negligible compared to the vehicles's distance to the scene.

Computing homographies from image point- and line-feature correspondences has been extensively studied in the last fifteen years [7]. The quality of the homography estimate obtained depends heavily on the number and quality of the data features 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 [14] and references therein). For a scene with poor texture, and consequently few reliable feature 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. In [17] image flow computed from a pair of images is used to compute the relative homography, although this method still only considers isolated pairs of images.

In recent work [9], [12] by some authors of this paper, a nonlinear observer for homography estimation was proposed based on the group structure of the set of homographies, the Special Linear group $\mathrm{SL}(3)$. This observer uses velocity information to interpolate across a sequence of images and improve the individual homography estimates. The observer proposed in [9], [12] still requires individual image homographies to be computed for each image, which are then smoothed using filtering techniques. Although this approach 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 our prior work [6], the question of deriving an observer for a sequence of image homographies that takes image point-feature correspondences directly as input has been considered. The approach in [6] 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 observer is posed on the Special Linear group $\mathrm{SL}(3)$, that is in one-to-one correspondence with the group of homographies, and uses velocity measurements to propagate the homography estimate and fuse this with new data as it becomes available.

In this paper by considering the same problem of [6] we extend the previous observer by also incorporating image line-feature correspondences (in addition to point-feature correspondences) directly as input in the design of the observer innovation. Instead of using a Lyapunov-based technique as in [6], the observer here proposed is derived based on a recent advanced theory for nonlinear observer design directly on the output space [10]. A key advance of both the observer of [6] and the observer of this paper with respect to the observer proposed in [9], [12] is the formulation of a point/line-feature innovation for the observer that incorporates point/line 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 feature tracking software. In addition, the algorithm is well posed even when there is insufficient data for full reconstruction of a homography using algebraic techniques. In such situations, the proposed observer will continue to operate, incorporating

M.-D. Hua and T. Hamel are with Université Côte d'Azur, CNRS, I3S, Sophia Antipolis, France. email: hua(thamel)@i3s.unice.fr.

J. Trumpf and R. Mahony are with the Research School of Engineering, Australian National University, Canberra, Australia. email: Robert.Mahony(Jochen.Trumpf)@anu.edu.au.

P. Morin is with ISIR UPMC–CNRS, Paris, France. email: morin@isir.upmc.fr.

what information is available and relying on propagation of prior estimates where necessary. Moreover, 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.

The paper is organized as follows. Section II provides technical background. In Section III a nonlinear observer on $\mathrm{SL}(3)$ is proposed using direct 2D point and line correspondences and the knowledge of the group velocity. 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. In Section V, as a complement contribution, an application of our approach to a real world problem in image stabilization is presented. A video, showing the experiment results, is provided as supplementary material.

II. Technical background

A. Notation

- The Special Linear group $\mathrm{SL}(3)$ and its algebra $\mathfrak{sl}(3)$ are
$$\mathrm{SL}(3) := \{H \in \mathbb{R}^{3\times 3} \mid \det H = 1\},$$
$$\mathfrak{sl}(3) := \{U \in \mathbb{R}^{3\times 3} \mid \mathrm{tr}(U) = 0\}.$$
- The adjoint operator is a mapping $\mathrm{Ad} : \mathrm{SL}(3) \times \mathfrak{sl}(3) \to \mathfrak{sl}(3)$ defined by
$$\mathrm{Ad}_H U := HUH^{-1},\ H \in \mathrm{SL}(3), U \in \mathfrak{sl}(3).$$
- Let $\langle \cdot, \cdot \rangle : \mathfrak{sl}(3) \times \mathfrak{sl}(3) \longrightarrow \mathbb{R}$ be an 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
$$\langle U_1 H, U_2 H \rangle_H := \langle U_1, U_2 \rangle,\ H \in \mathrm{SL}(3), U_1, U_2 \in \mathfrak{sl}(3).$$

B. Homographies

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 $R \in \mathrm{SO}(3)$, with $R\colon \{A\} \to \{\mathring{A}\}$ as a mapping. In addition, we denote by $\mathring{d}$ (resp. d) and $\mathring{\eta}$ (resp. η) respectively the distance from the origin of $\{\mathring{A}\}$ (resp. $\{A\}$) to the planar scene and the normal to the scene expressed in $\{\mathring{A}\}$ (resp. $\{A\}$).

The image homography matrix that maps pixel coordinates from $\mathcal{A}$ to $\mathring{\mathcal{A}}$ is given by
$$G = \gamma K (R + \frac{\xi \eta^\top}{d}) K^{-1},$$
with $K \in \mathbb{R}^{3\times 3}$ the camera calibration matrix and γ a scale factor. Without loss of generality, γ can be chosen so that $G \in \mathrm{SL}(3)$. In this case, γ is equal $(d/\mathring{d})^{\frac{1}{3}}$ and corresponds to the second singular value of G [8]. The Euclidean homography $H \in \mathrm{SL}(3)$ is related to G by
$$H = K^{-1} G K = \gamma (R + \frac{\xi \eta^\top}{d}).$$
Note that the homography matrix H maps Euclidean coordinates of the scene's points from $\{A\}$ to $\{\mathring{A}\}$.

III. Nonlinear observer design on $\mathrm{SL}(3)$ based on 2D point and line correspondences

In this section the design of an observer for $H \in \mathrm{SL}(3)$ is based on a recent advanced theory for nonlinear observer design directly on the output space [10].

A. Kinematics and measurements

Consider the kinematics of $\mathrm{SL}(3)$ given by
$$\dot{H} = F(H, U) := HU, \tag{1}$$
with $U \in \mathfrak{sl}(3)$ the group velocity. Assume that U is measured. Moreover, we consider a set of $n\ (\geq 0)$ point measurements $p_i \in \mathbb{S}^2$, representing calibrated image points re-normalized onto the unit sphere, and/or a set of $m\ (\geq 0)$ line measurements $l_j \in \mathbb{S}^2$, corresponding to the normal directions to the planes containing a given line in the image and the camera focal point. All measurements are expressed in the camera current frame:
$$p_i = h_p(H, \mathring{p}_i) := \tfrac{H^{-1}\mathring{p}_i}{|H^{-1}\mathring{p}_i|},\quad i = \{1, \dots, n\}, \tag{2}$$
$$l_j = h_l(H, \mathring{l}_j) := \tfrac{H^\top \mathring{l}_j}{|H^\top \mathring{l}_j|},\quad j = \{1, \dots, m\}, \tag{3}$$
where $\mathring{p}_i \in \mathbb{S}^2$ and $\mathring{l}_j \in \mathbb{S}^2$ are constant and known. For later use, define
$$\mathring{p} := (\mathring{p}_1, \cdots, \mathring{p}_n),\quad p := (p_1, \cdots, p_n),$$
$$\mathring{l} := (\mathring{l}_1, \cdots, \mathring{l}_m),\quad l := (l_1, \cdots, l_m).$$
Let us introduce the following observability assumption.

Assumption 1 *Assume that the union set $\mathcal{S} := \mathcal{S}_p^n \bigcap \mathcal{S}_l^m$, with $\mathcal{S}_p^n$ the set of $n\ (\geq 0)$ observed constant points $\mathring{p}_i \in \mathbb{S}^2$ and $\mathcal{S}_l^m$ the set of $m\ (\geq 0)$ observed constant lines $\mathring{l}_j \in \mathbb{S}^2$, satisfies one of the four following cases*[1]*:*

- ***Case 1 (at least 4 points):*** *There exists a subset $\mathcal{S}_p^4 \subset \mathcal{S}_p^n$ of 4 points such that all vector triplets in $\mathcal{S}_p^4$ are linearly independent.*
- ***Case 2 (at least 4 lines):*** *There exists a subset $\mathcal{S}_l^4 \subset \mathcal{S}_l^n$ of 4 lines such that all vector triplets in $\mathcal{S}_l^4$ are linearly independent.*
- ***Case 3 (at least 3 points and 1 line):*** *There exist 1 line $\mathring{l}_j$ and a subset $\mathcal{S}_p^3 \subset \mathcal{S}_p^n$ of 3 linearly independent points that do not lie on the line $\mathring{l}_j$, i.e. $\mathring{l}_j^\top \mathring{p}_i \neq 0,\ \forall \mathring{p}_i \in \mathcal{S}_p^3$.*
- ***Case 4 (at least 1 point and 3 lines):*** *There exist a subset $\mathcal{S}_l^3 \subset \mathcal{S}_l^m$ of 3 linearly independent lines and 1 point $\mathring{p}_i$ that does not lie on any line of $\mathcal{S}_l^3$, i.e. $\mathring{l}_j^\top \mathring{p}_i \neq 0,\ \forall \mathring{l}_j \in \mathcal{S}_l^3$.*

We verify that $\mathrm{SL}(3)$ is a symmetry group with group actions
$$\phi(Q, H) := HQ,\ \psi(Q, U) := Ad_{Q^{-1}} U = Q^{-1} U Q,$$
$$\rho(Q, p) := \tfrac{Q^{-1}p}{|Q^{-1}p|},\ \mu(Q, l) := \tfrac{Q^\top l}{|Q^\top l|},$$
which are *right group actions* in the sense of [10]. The kinematics are *right equivariant* in the sense of [10] since it is verified that
$$\rho(Q, h_p(H, \mathring{p}_i)) = h_p(\phi(Q, H), \mathring{p}_i),$$
$$\mu(Q, h_l(H, \mathring{l}_i)) = h_l(\phi(Q, H), \mathring{l}_i),$$
$$d\phi_Q(H)[F(H, U)] = F(\phi(Q, H), \psi(Q, U)).$$

[1]The homography for case of 2 points and 2 lines is not observable and cannot be algebraically reconstructed [7].

B. Observer design on $SL(3)$

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 $e_{pi} \in \mathbb{S}^2$, with $i \in \{1, \cdots, n\}$, and $e_{lj} \in \mathbb{S}^2$, with $j \in \{1, \cdots, m\}$, as:

$$e_{pi} := \rho(\hat{H}^{-1}, p_i) = \frac{\hat{H}p_i}{|\hat{H}p_i|} = \frac{E\mathring{p}_i}{|E\mathring{p}_i|}, \tag{4}$$

$$e_{lj} := \mu(\hat{H}^{-1}, l_j) = \frac{\hat{H}^{-\top}l_j}{|\hat{H}^{-\top}l_j|} = \frac{E^{-\top}\mathring{l}_j}{|E^{-\top}\mathring{l}_j|}. \tag{5}$$

Inspired by [10], the proposed observer takes the form

$$\dot{\hat{H}} = \hat{H}U - \Delta(\hat{H}, p, l)\hat{H} \tag{6}$$

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

$$\Delta(\hat{H}Q, \rho(Q, p_1), \cdots, \rho(Q, p_n), \mu(Q, l_1), \cdots, \mu(Q, l_m)) = \Delta(\hat{H}, p_1, \cdots, p_n, l_1, \cdots, l_m),\ \forall Q \in \mathrm{SL}(3).$$

Interestingly, if Δ is equivariant, the dynamics of E is autonomous and given by [10, Th. 1]:

$$\dot{E} = -\Delta(E, \mathring{p}, \mathring{l})E \tag{7}$$

To determine $\Delta(\hat{H}, p, l)$, a *non-degenerate right-invariant* cost function is needed. To this purpose, we first define individual *degenerate right-invariant costs* at $\mathring{p}_i$ or $\mathring{l}_j$ on the output space $\mathbb{S}^2$ as follows:

$$\mathcal{C}_{\mathring{p}_i}(\hat{H}, p_i) := \tfrac{k_i}{2}\left|\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right|^2,$$

$$\mathcal{C}_{\mathring{l}_j}(\hat{H}, l_j) := \tfrac{\kappa_j}{2}\left|\frac{\hat{H}^{-\top}l_j}{|\hat{H}^{-\top}l_j|} - \mathring{l}_j\right|^2,$$

with $k_i, \kappa_j > 0$. One verifies that $\mathcal{C}_{\mathring{p}_i}(\hat{H}, p_i)$ and $\mathcal{C}_{\mathring{l}_j}(\hat{H}, l_j)$ are right-invariant in the sense $\mathcal{C}_{\mathring{p}_i}(\hat{H}Q, \rho(Q, p_i)) = \mathcal{C}_{\mathring{p}_i}(\hat{H}, p_i)$ and $\mathcal{C}_{\mathring{l}_j}(\hat{H}Q, \mu(Q, p_i)) = \mathcal{C}_{\mathring{l}_j}(\hat{H}, l_j)$, $\forall Q \in \mathrm{SL}(3)$. Then, the aggregate cost defined as the sum of all the individual costs

$$\mathcal{C}(\hat{H}, p, l) := \sum_{i=1}^{n}\frac{k_i}{2}\left|\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right|^2 + \sum_{j=1}^{m}\frac{\kappa_j}{2}\left|\frac{\hat{H}^{-\top}l_j}{|\hat{H}^{-\top}l_j|} - \mathring{l}_j\right|^2 \tag{8}$$

is also right-invariant. According to [10, Lem. 3], the aggregate cost is *non-degenerate* if

$$\left(\bigcap_{i=1}^{n}\mathrm{stab}_\rho(\mathring{p}_i)\right) \bigcap \left(\bigcap_{j=1}^{m}\mathrm{stab}_\mu(\mathring{l}_j)\right) = \{I\}, \tag{9}$$

where the stabilizer $\mathrm{stab}_f(y)$ (with f standing for either ρ or μ) of an element $y \in \mathbb{S}^2$ is defined by $\mathrm{stab}_f(y) = \{Q \in \mathrm{SL}(3) \mid f(Q, y) = y\}$. In fact, (9) is equivalent to

$$\left(\bigcap_{i=1}^{n}\mathfrak{s}_{\rho i}\right) \bigcap \left(\bigcap_{j=1}^{m}\mathfrak{s}_{\mu j}\right) = \{0\}, \tag{10}$$

with $\mathfrak{s}_{\rho i} = \ker(\mathrm{d}\rho_{\mathring{p}_i}(I))$, $\mathfrak{s}_{\mu j} = \ker(\mathrm{d}\mu_{\mathring{l}_j}(I))$, respectively, the Lie-algebra associated with $\mathrm{stab}_\rho(\mathring{p}_i)$ and $\mathrm{stab}_\mu(\mathring{l}_j)$.

Lemma 1 *Under Assumption 1, the aggregate cost $\mathcal{C}(\hat{H}, p, l)$ defined by* (8) *is non-degenerate. As a consequence, $(I, \mathring{p}, \mathring{l})$ is a global minimum of the aggregate cost $\mathcal{C}(\hat{H}, p, l)$.*

The proof, which consists in proving condition (10), is omitted but can be requested to the authors. Then, the innovation term $\Delta(\hat{H}, p, l)$ is computed as [10, Eq. (40)]

$$\Delta(\hat{H}, p, l) = (\mathrm{grad}_1\mathcal{C}(\hat{H}, p, l))\hat{H}^{-1}, \tag{11}$$

with grad_1 the *gradient* in the first variable, using a *right-invariant Riemannian metric* on $\mathrm{SL}(3)$. As a direct result of [10], the innovation term $\Delta(\hat{H}, p)$ defined by (11) is right equivariant. Using standard rules for transformations of Riemannian gradients and the fact that the Riemannian metric is right-invariant, we obtain

$$\begin{aligned}\mathcal{D}_1\mathcal{C}(\hat{H}, p, l)[U\hat{H}] &= \langle \mathrm{grad}_1\mathcal{C}(\hat{H}, p, l), U\hat{H}\rangle_H \\ &= \langle \mathrm{grad}_1\mathcal{C}(\hat{H}, p, l)\hat{H}^{-1}, U\rangle \\ &= \langle \Delta(\hat{H}, p, l), U\rangle,\end{aligned} \tag{12}$$

with some $U \in \mathfrak{sl}(3)$. In addition, in view of (8) we have

$$\begin{aligned}&\mathcal{D}_1\mathcal{C}(\hat{H}, p, l)[U\hat{H}] = d_{\hat{H}}\mathcal{C}(\hat{H}, p, l)[U\hat{H}] \\ &= \textstyle\sum_{i=1}^{n} k_i\left(\frac{\hat{H}p_i}{|\hat{H}p_i|} - \mathring{p}_i\right)^\top\left(I - \frac{(\hat{H}p_i)(\hat{H}p_i)^\top}{|\hat{H}p_i|^2}\right)\frac{(U\hat{H})p_i}{|\hat{H}p_i|} \\ &+ \textstyle\sum_{j=1}^{m}\kappa_j\left(\frac{\hat{H}^{-\top}l_j}{|\hat{H}^{-\top}l_j|} - \mathring{l}_j\right)^\top\left(I - \frac{(\hat{H}^{-\top}l_j)(\hat{H}^{-\top}l_j)^\top}{|\hat{H}^{-\top}l_j|^2}\right)\frac{(-U^\top\hat{H}^{-\top})l_j}{|\hat{H}^{-\top}l_j|} \\ &= \textstyle\sum_{i=1}^{n} k_i(e_{pi} - \mathring{p}_i)^\top(I - e_{pi}e_{pi}^\top)Ue_{pi} \\ &- \textstyle\sum_{j=1}^{m}\kappa_i(e_{lj} - \mathring{l}_j)^\top(I - e_{lj}e_{lj}^\top)U^\top e_{lj} \\ &= -\mathrm{tr}\left(\textstyle\sum_{i=1}^{n} k_i\pi_{e_{pi}}\mathring{p}_i e_{pi}^\top U^\top\right) + \mathrm{tr}\left(\textstyle\sum_{j=1}^{m}\kappa_j e_{lj}\mathring{l}_j^\top\pi_{e_{lj}}U^\top\right) \\ &= \left\langle -\textstyle\sum_{i=1}^{n} k_i\pi_{e_{pi}}\mathring{p}_i e_{pi}^\top + \sum_{j=1}^{m}\kappa_j e_{lj}\mathring{l}_j^\top\pi_{e_{lj}}, U\right\rangle,\end{aligned} \tag{13}$$

where we have used the projection $\pi_x := (I - xx^\top)$, $\forall x \in \mathbb{S}^2$.

Then, we directly deduce from (12) and (13) the expression of the innovation term $\Delta(\hat{H}, p, l)$ as

$$\Delta(\hat{H}, p, l) = -\sum_{i=1}^{n} k_i\pi_{e_{pi}}\mathring{p}_i e_{pi}^\top + \sum_{j=1}^{m}\kappa_j e_{lj}\mathring{l}_j^\top\pi_{e_{lj}} \tag{14}$$

We deduce from (14) that $\Delta(E, \mathring{p}, \mathring{l}) = \Delta(\hat{H}, p, l)$ and, consequently, from (7) that

$$\dot{E} = \Big(\sum_{i=1}^{n} k_i\pi_{e_{pi}}\mathring{p}_i e_{pi}^\top - \sum_{j=1}^{m}\kappa_j e_{lj}\mathring{l}_j^\top\pi_{e_{lj}}\Big)E. \tag{15}$$

From here, the convergence and stability of the proposed observer is stated next.

Theorem 1 *Consider the autonomous error system* (15)*. Assume that Assumption 1 is satisfied. Then, the equilibrium $E = I$ of System* (15) *is locally asymptotically stable.*

This theorem can be seen as a direct result of Theorem 2 in [10] and Lemma 1.

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 nonlinear 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.

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 kinematics of (R, ξ) are given by:

$$\dot{R} = R\Omega_\times \tag{16}$$

$$\dot{\xi} = RV \tag{17}$$

where $\Omega_\times$ is the skew symmetric matrix associated with the vector cross-product, i.e. $\Omega_\times y = \Omega \times y$, for all $y \in \mathbb{R}^3$. Consider a camera attached to the frame $\mathcal{A}$ moving with kinematics (16)–(17) 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 (1), is given by [9, Lem. 5.3]

$$U = \Omega_\times + \tfrac{V\eta^\top}{d} - \tfrac{\eta^\top V}{3d} I. \tag{18}$$

Note that the group velocity U induced by camera motion depends on the additional variables η and d, which are unmeasurable and cannot be extracted directly from the measurements. In the sequel, we rewrite

$$U = \Omega_\times + \Gamma = \Omega_\times + \Gamma_1 - \tfrac{1}{3}\mathrm{tr}(\Gamma_1) I, \tag{19}$$

with $\Gamma := \frac{V\eta^\top}{d} - \frac{\eta^\top V}{3d} I$ and $\Gamma_1 := \frac{V\eta^\top}{d}$.

Since $\{\mathring{A}\}$ is stationary by assumption, the vector Ω can be directly obtained from the set of embedded gyrometers. 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 $V = R^\top \dot{\xi}$, it is follows:

$$\dot{\Gamma} = [\Gamma, \Omega_\times], \tag{20}$$

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:

$$\dot{\Gamma}_1 = \Gamma_1 \Omega_\times. \tag{21}$$

B. Observer with partially known velocity of the rigid body

Assume that the part Γ (resp. Γ_1) of the group velocity U in (19) is not available to measurement. 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 (6)):

$$\dot{\hat{H}} = \hat{H}(\Omega_\times + \hat{\Gamma}) - \Delta(\hat{H}, p, l)\hat{H} \tag{22}$$

$$\dot{\hat{\Gamma}} = [\hat{\Gamma}, \Omega_\times] - k_I \mathrm{Ad}_{\hat{H}^\top} \Delta(\hat{H}, p, l) \tag{23}$$

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

$$\dot{\hat{H}} = \hat{H}(\Omega_\times + \hat{\Gamma}_1 - \frac{1}{3}\mathrm{tr}(\hat{\Gamma}_1) I) - \Delta(\hat{H}, p, l)\hat{H} \tag{24}$$

$$\dot{\hat{\Gamma}}_1 = \hat{\Gamma}_1 \Omega_\times - k_I \mathrm{Ad}_{\hat{H}^\top} \Delta(\hat{H}, p, l) \tag{25}$$

with some positive gain k_I and $\Delta(\hat{H}, p, l)$ given by (14).

Proposition 1 *Consider a camera moving with kinematics* (16)–(17) *viewing a planar scene. Assume that $\mathring{\mathcal{A}}$ is stationary and that Ω is measured and bounded. Consider the kinematics* (1) *along with* (19)*. Assume that H is bounded and that Γ (resp. Γ_1) is such that it obeys* (20) *(resp.* (21)*). Consider the nonlinear observer defined by (22–23), (resp. (24–25)) along with the innovation $\Delta(\hat{H}, p, l) \in \mathfrak{sl}(3)$ given by* (14)*. Assume that Assumption 1 holds. Then, 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 (4) and using (22) yields

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

Define the following candidate Lyapunov function:

$$\mathcal{L} = \mathcal{L}_0 + \tfrac{1}{2k_I}||\tilde{\Gamma}||^2 = \textstyle\sum_{i=1}^n \frac{k_i}{2}\,|e_i - \mathring{p}_i|^2 + \frac{1}{2k_I}||\tilde{\Gamma}||^2. \tag{26}$$

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

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

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

$$\begin{aligned}
\dot{\mathcal{L}} &= -\textstyle\sum_{i=1}^n k_i (e_i - \mathring{p}_i)^\top \pi_{e_i}(\Delta + \mathrm{Ad}_{\hat{H}}\tilde{\Gamma}) e_i + \mathrm{tr}(\mathrm{Ad}_{\hat{H}^{-1}} \Delta^\top \tilde{\Gamma}) \\
&= \mathrm{tr}(\textstyle\sum_{i=1}^n k_i e_i \mathring{p}_i^\top \pi_{e_i}(\Delta + \mathrm{Ad}_{\hat{H}}\tilde{\Gamma}) + \mathrm{Ad}_{\hat{H}^{-1}} \Delta^\top \tilde{\Gamma}) \\
&= \mathrm{tr}(\textstyle\sum_{i=1}^n k_i e_i \mathring{p}_i^\top \pi_{e_i} \Delta + \mathrm{Ad}_{\hat{H}^{-1}}(\Delta^\top + \sum_{i=1}^n k_i e_i \mathring{p}_i^\top \pi_{e_i}) \tilde{\Gamma}) \\
&= -\|\Delta\|^2.
\end{aligned}$$

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 $\Delta \to 0$. Since the cost $\mathcal{C}(E, \mathring{p}, \mathring{l})$ is non-degenerate (Lemma 1), we evoke the same arguments as in the proof of Theorem 2 in [10] to ensure the existence of a set $\mathcal{B} \subseteq SL(3)$ such that for all $E(0) \in \mathcal{B}$, $E(t)$ converges to I. Consequently, the left-hand side of (26) 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 $\lim_{t\to\infty} \dot{E} = -\mathrm{Ad}_{\hat{H}}\tilde{\Gamma} = 0$. Using boundedness of H, one ensures the boundedness of $\hat{H}$ and $\hat{H}^{-1}$ and consequently $\lim_{t\to\infty} \tilde{\Gamma} = 0$. ■

V. Experimental results – Image stabilization

We present an application of our approach to image stabilization in the presence of very fast camera motion, severe occlusion, specular reflections, image blurring, and light saturation. The reported test has been conducted on a data set recorded by a prototype synchronized camera-IMU combination with an Aptina MT9V034 CMOS sensor and an Analog Devices ADIS16375 MEMS IMU. The IMU runs at 100 Hz, providing angular velocity measurements to the observer. The camera provides 20 frames per second at a resolution of 752×480 pixels. The estimated camera parameters are $(464.66, 462.75)$ pixels for the focal length and $(385.63, 227.53)$ pixels for the principle point.

Point-feature detection and matching: Code has been implemented in C++ with OpenCV library. Point-features are extracted using the `FastFeatureDetector` routine, with descriptors extracted using `OrbDescriptorExtractor`, and then matched using OpenCV's brute-force matcher

`BFMatcher` 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 the long video sequence, in particular given the low frame rate and comparatively rapid motion in our test sequence as well as the presence of severe occlusion, specular reflections, poor image quality due to blur or light saturation. We have hence opted to match point-features between the reference image and each subsequent image frame separately. To do this, we first forward integrate the observer equations (22)–(23) using only the gyrometer measurements, i.e. setting the observer gains k_i $(i = 1, \cdots, n)$, κ_j $(j = 1, \cdots, m)$, and k_I to zero (i.e. Prediction step). We then use the resulting predicted homography estimate $\hat{H}^+$ to transform the current image (i.e. warp the current image using the predicted homography $\hat{H}^+$ to obtain a prediction of the reference image) using the OpenCV's `warpPerspective` function before applying feature extraction and matching. 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.

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 [4]. Our simple and fast outlier removal method has yielded quite remarkable matching results as for the test sequence there are either none or very few outliers (see Fig. 1 and the supplemental video).

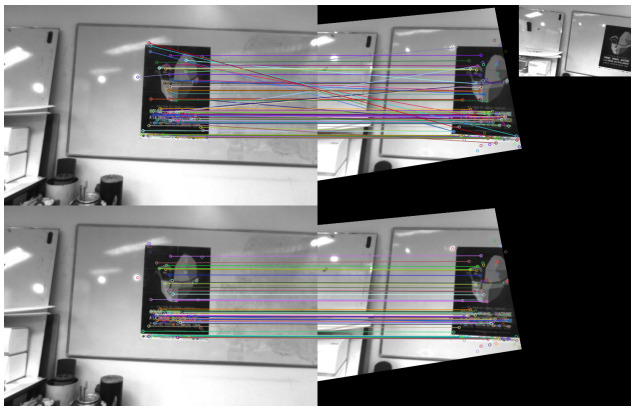

Fig. 1. Matching point correspondences between the warped image frame 264 (warped by the predicted homography) and the reference frame. **Poor matching (top) and excellent matching (bottom) before and after applying our outlier removal procedure.** Reference frame (left), warped current frame (right), current frame (small image on top right corner).

Line-feature detection and matching: Line-features are extracted using the probabilistic Hough transform [13] with the OpenCV's `HoughLinesP` routine. Each extracted line in image coordinates (given by two points P_1 and P_2 in homogeneous coordinates) is then transformed into the line representation used in this paper (i.e., the normal to the plane containing the scene's line and the camera focal point) as

$$l = \frac{(K^{-1}P_1) \times (K^{-1}P_2)}{|(K^{-1}P_1) \times (K^{-1}P_2)|} \in \mathbb{S}^2.$$

Fig. 2. Successful line matching between the image frame 169 (right) and the reference frame (left). Colorful points in both images are point correspondences used for our line matching algorithm.

Matching two sets of lines of the reference image and the current image is more involved and has been scarcely developed in literature (and in OpenCV) compared to the point matching problem. However, keeping in mind that for most of the time we have "good" matched point-features, the following simple line matching algorithm has been developed. Assume that two sets of good matched points (in homogeneous coordinates) $\mathcal{S}_p := \{P_i, \cdots, P_n\}$ from the current image and $\mathring{\mathcal{S}}_p := \{\mathring{P}_1, \cdots, \mathring{P}_n\}$ from the reference image are available for matching two sets of lines $\mathcal{S}_l := \{l_1, \cdots, l_{m_1}\}$ and $\mathring{\mathcal{S}}_l := \{\mathring{l}_1, \cdots, \mathring{l}_{m_2}\}$ from the same respective images. For any point and any line on the planar target, it is verified that $P_i = G^{-1}\mathring{P}_i$ and $l_j = \lambda_j G^\top \mathring{l}_j$, with $\lambda_j := \frac{1}{|G^\top \mathring{l}_j|}$. Therefore, for any two points (of index i_1 and i_2) and a line (of index j), one has

$$P_{i_1}^\top l_j = \lambda_j \mathring{P}_{i_1}^\top \mathring{l}_j, \quad P_{i_2}^\top l_j = \lambda_j \mathring{P}_{i_2}^\top \mathring{l}_j,$$

yielding the following equality

$$(P_{i_1}^\top l_j)(\mathring{P}_{i_2}^\top \mathring{l}_j) - (\mathring{P}_{i_1}^\top \mathring{l}_j)(P_{i_2}^\top l_j) = 0. \tag{27}$$

In view of (27), two lines $l_{j_1} \in \mathcal{S}_l$ and $\mathring{l}_{j_2} \in \mathring{\mathcal{S}}_l$ can be matched if they satisfy

$$\begin{cases} C(l_{j_1}, \mathring{l}_{j_2}) \leq C(l_{j_1}, \mathring{l}_j), & \forall \mathring{l}_j \in \mathring{\mathcal{S}}_l \\ C(l_{j_1}, \mathring{l}_{j_2}) \leq C(l_j, \mathring{l}_{j_2}), & \forall l_j \in \mathcal{S}_l \\ C(l_{j_1}, \mathring{l}_{j_2}) \leq \varepsilon_c \end{cases}$$

where $\varepsilon_c > 0$ is a small threshold associated with the non-negative cost function $C(\cdot, \cdot)$ defined by

$$C(l_a, \mathring{l}_b) := \textstyle\sum_{i=1}^n f\big((P_1^\top l_a)(\mathring{P}_i^\top \mathring{l}_b) - (\mathring{P}_1^\top \mathring{l}_b)(P_i^\top l_a)\big),$$

with $l_a \in \mathcal{S}_l$, $\mathring{l}_b \in \mathring{\mathcal{S}}_l$, and $f(\cdot)$ an even convex non-negative function. In our experiments, $f(x) = \frac{1}{n}\sqrt{|x|}, \forall x \in \mathbb{R}$, and $\varepsilon_c = 0.025$. Excellent line matching results have been obtained for the reported test sequence (see Fig. 2 and the supplemental video).

Correction step of the observer: After the steps of feature detection and matching, we use the observer gains of $k_i = 200$ $(i = 1, \cdots, n)$, $\kappa_j = 100$ $(j = 1, \cdots, m)$, $k_I = 0.05$ to rapidly iterate the observer equations 200 times per video frame. The computational effort for this last step is negligible compared to the previous image processing steps.

Fig. 3. **Good and robust performance** of our algorithm for a very fast camera motion (relative to frame rate), and in presence of strong occlusion (e.g., frames 200, 2044), severe image blur (e.g., frames 1394, 1593) and light saturation (e.g., frame 363, 2044). The observer continues to operate even when temporarily no usable feature match is available (e.g., frames 1148, 1593). In each subplot of current frame, colorful points and lines are those successfully matched with the corresponding features in the reference image; and the green polygon represents a tracked region of interest (i.e., the poster) using the homography estimate. In the bottom right corner of each subplot, a crop of the warped current image is shown, telling us if the image is well stabilized or not. The full video of this experiment is available at *https://youtu.be/hlTkzjyENhg*.

Performance evaluation: The experimental results (cf. the video in the supplementary material and available at [1]) show good and robust performance throughout the entire video sequence, including the previously mentioned passages with severe occlusion, specular reflection, poor image quality due to blur and/or light saturation (see Fig. 3). Even when temporarily no usable feature correspondence is available (e.g., frames 1148, 1593), or when our algorithm selects a wrong feature matching set (e.g., frames $1474 - 1488$, $1571 - 1581$) the observer continues to track the region of interest well and quickly recovers from any tracking errors.

Experiment conducted on the video test sequence was run on a laptop equipped with a 2.9 GHz Intel(R) Core(TM) i7-4910MQ and a 32Go RAM. We are currently able to perform between 9 and 20 fps (and 14 fps in average) using standard OpenCV functions. Replacing these basic OpenCV functions with OpenCV GPU functions (based on CUDA library) for a more real-time implementation is now under development.

VI. Conclusions

A nonlinear observer for a sequence of homographies represented as elements of $\mathrm{SL}(3)$ has been proposed. It directly makes use of point- and line-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 gyrometers 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, 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 motion, severe occlusion, specular reflection, image blur, and light saturation has been demonstrated with very encouraging results even for a relatively low video frame rate.

Acknowledgement: This research was supported by the ANR ASTRID SCAR project (ANR-12-ASTR-0033), the ROBOTEX project (ANR-10-EQPX-44), the Australian Research Council through the "Australian Centre of Excellence for Robotic Vision" CE140100016, and the "Chaire d'excellence en Robotique RTE UPMC".

References

[1] Experiment video. URL: `https://youtu.be/hlTkzjyENhg`.

[2] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel. Visual servoing for underactuated VTOL UAVs: a linear homography-based approach. In *IEEE Int. Conf. on Robotics and Automation*, pages 3004-3010, 2011.

[3] Y. Fang, W. Dixon, D. Dawson, and P. Chawda. Homography-based visual servoing of wheeled mobile robots. *IEEE Trans. on Systems, Man, and Cybernetics - Part B*, 35: 1041–1050, 2005.

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

[5] F. Fraundorfer, C. Engels, and D. Nister. Topological mapping, localization and navigation using image collections. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 3872–3877, 2007.

[6] T. Hamel, R. Mahony, J. Trumpf, P. Morin, M.-D. Hua. Homography estimation on the Special Linear group based on direct point correspondence. In *IEEE Conf. on Dec. and Contr.*, pp. 7902–7908, 2011.

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

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

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

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

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

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

[13] J. Matas, C. Galambos, and J. V. Kittler. Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. *Computer Vision and Image Understanding*, 78(1): 119-137, 2000.

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

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

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

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