Rotation Free Active Vision

Omar Tahri, Paolo Robuffo Giordano and Youcef Mezouar

Abstract— **Incremental Structure from Motion (SfM) algorithms require, in general, precise knowledge of the camera linear and angular velocities in the camera frame for estimating the 3D structure of the scene. Since an accurate measurement of the camera own motion may be a non-trivial task in several robotics applications (for instance when the camera is onboard a UAV), we propose in this paper an *active* SfM scheme fully independent from the camera angular velocity. This is achieved by considering, as visual features, some rotational invariants obtained from the projection of the perceived 3D points onto a virtual unitary sphere (unified camera model). This feature set is then exploited for designing a *rotation-free* active SfM algorithm able to optimize *online* the direction of the camera linear velocity for improving the convergence of the structure estimation task. As case study, we apply our framework to the depth estimation of a set of 3D points and discuss several simulations and experimental results for illustrating the approach.**

I. INTRODUCTION

Structure from Motion (SfM) is a classical and well-studied problem in computer and robot vision. One possibility is to exploit some prior knowledge of the scene such as, for instance, the known size of the tracked objects. Otherwise, one can exploit presence of different points of view of the scene (with known displacement) for recovering the missing 3D information. When considering the case of a moving camera, one can often recast SfM as an iterative/filtering problem by processing the consecutive gathered images together with the (assumed known) camera displacement across frames. In this context, a number of filtering techniques have been proposed over the last years based on, e.g., an Extended Kalman Filter in [1]–[4], or an Unscented Kalman Filter in [5] for explicitly dealing, to some extent, with presence of measurement noise and other uncertainties. Other approaches have instead considered the use of deterministic nonlinear state estimation in, e.g., [6]–[13]. A recent experimental comparison of a EKF solution versus a deterministic nonlinear filter in the context of SfM for a quadrotor UAV can also be found in [14].

Whatever the adopted estimation strategy, it is well-known that the camera motion (in particular, its translation/linear velocity over the scene) plays a fundamental role for a successful 3D structure estimation. Roughly speaking, the camera motion must be 'exciting' enough for allowing recovering the scene structure: as a trivial example, the 3D structure of a single point feature (its depth Z) cannot be estimated by a camera traveling along the feature projection ray. The strong dependance of the 3D structure estimation performance on the chosen camera trajectory has motivated several works in the context of *active* vision/SFM whose goal is to optimze (either online or offline) the camera motion for improving the convergence rate/accuracy of the estimated quantities. In this respect, [15] has proposed a general framework for deterministic *active* SfM amenable to a large variety of scene geometries, and with strong and well-characterized convergence behavior for the estimation process. The framework has then been successfully applied to several case studies, involving SfM of some geometrical primitives [16] and planar patches from measured image moments [17], [18]. A possible coupling between active SfM and realization of a visual servoing task has also been considered in [19].

In all these works (and similar ones), an implicit requirement is the possibility to measure/estimate the camera motion (linear and angular velocity) in its own frame for then feeding this information to the estimation scheme. While this can be easily achieved for, e.g., eye-in-hand cameras carried by robot manipulators fixed to the ground, some practical difficulties can instead arise when dealing with, e.g., cameras mounted on mobile robots (e.g., quadrotor UAVs with onboard cameras) since obtaining a reliable measurement/estimation of the camera motion can not be a straightforward task. Furthermore, in all typical cases, the accuracy in measuring the camera self-motion depends on several calibration parameters (e.g., the eye-to-hand relative pose) which must be obtained independently form the SfM task. With respect to these issues, the goal of this paper is to propose a SfM framework *fully invariant* to a camera rotation in space, so that a measurement of the camera angular velocity is no longer needed for the estimation convergence. This is achieved by using features invariant to rotation computed from spherical projection using the unified model for central imaging systems [20]. The unified model consists in modeling the imaging system by two consecutive projections: spherical and then perspective. It is valid for a wide range of sensors (conventional and catadioptric cameras [20], some fisheye cameras [21]). Note that, once the camera is calibrated (using for instance [22]), the projection onto the sphere can be recovered for any point in the image plane. Besides, spherical projection has already been used to design visual servoing schemes and pose estimation strategies with nice decoupling properties between rotational and translational motions [23]–[27]. In this paper, we will see how these ideas can be used to design a fully rotation free active vision scheme.

The remaining of this paper is organized as follows:

O. Tahri and Y. Mezouar are with the Institut Pascal, BP 10448, F-63000 Clermont-Ferrand,France otahri@gmail.com, youcef.mezouar@ifma.fr.

P. Robuffo Giordano is with CNRS at Irisa and Inria Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France prg@irisa.fr.

the next section gives an overview of the active structure from motion scheme; in Section 3, the rotation free active vision scheme is presented; in Section 4, simulation and experimental results are discussed.

II. ACTIVE STRUCTURE FROM MOTION

Let $\boldsymbol{s} \in \mathbb{R}^m$ be the set of visual features *measured* on the image plane, $\boldsymbol{\chi} \in \mathbb{R}^p$ the *unmeasurable* 3D structure of the scene to be estimated by the SfM algorithm, and $\boldsymbol{u} = (\boldsymbol{v}, \boldsymbol{\omega}) \in \mathbb{R}^6$ the camera linear/angular velocity expressed in the camera frame. With these choices, one can show [16], [28] that the SfM dynamics takes the following general form linear w.r.t. the unknown vector $\boldsymbol{\chi}$

$$\begin{cases} \dot{\boldsymbol{s}} = \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{\omega}) + \boldsymbol{\Omega}^\top(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\chi} \\ \dot{\boldsymbol{\chi}} = \boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{u}) \end{cases}. \tag{1}$$

System (1) possesses the following relevant properties:

- matrix $\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v}) \in \mathbb{R}^{m\times p}$ is a function of known quantities (visual features $\boldsymbol{s}$ and camera linear velocity $\boldsymbol{v}$) and such that $\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{0}) = \boldsymbol{0}$. Therefore, the camera linear velocity plays a key role in estimating the 3D structure of the scene, since vector $\boldsymbol{\chi}$ has no effects on the feature dynamics whenever $\boldsymbol{v} = \boldsymbol{0}$;
- $\boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{\omega}) \in \mathbb{R}^m$ is a function of known quantities (visual features $\boldsymbol{s}$ and camera angular velocity $\boldsymbol{\omega}$);
- $\boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{u}) \in \mathbb{R}^p$ represents a generic smooth dependance of the time variation of the 3D structure on the system states and inputs (its expression depends on the particular geometry of the scene and chosen visual features, see Sect. III-A).

Remark 2.1: We note that model (1) is specific to central cameras since for the non central sensors the effect of angular speed can be a function of the depth information.

Let now $(\hat{\boldsymbol{s}}, \hat{\boldsymbol{\chi}})$ be the estimated state and $\boldsymbol{e} = (\boldsymbol{\xi}, \boldsymbol{z})$ be the error vector where $\boldsymbol{\xi} = \boldsymbol{s} - \hat{\boldsymbol{s}}$ and $\boldsymbol{z} = \boldsymbol{\chi} - \hat{\boldsymbol{\chi}}$. The following estimation scheme has been proposed in [15], [16] for recovering the unknown $\boldsymbol{\chi}(t)$ from the measured $\boldsymbol{s}(t)$ and (assumed known) $\boldsymbol{u}(t)$

$$\begin{cases} \dot{\hat{\boldsymbol{s}}} = \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{\omega}) + \boldsymbol{\Omega}^\top(\boldsymbol{s}, \boldsymbol{v})\hat{\boldsymbol{\chi}} + \boldsymbol{H}\boldsymbol{\xi} \\ \dot{\hat{\boldsymbol{\chi}}} = \boldsymbol{f}_u(\boldsymbol{s}, \hat{\boldsymbol{\chi}}, \boldsymbol{u}) + \alpha\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\xi} \end{cases} \tag{2}$$

where $\boldsymbol{H} > 0$ and $\alpha > 0$ are suitable gains. By coupling observer (2) with system (1), one obtains the following error dynamics

$$\begin{cases} \dot{\boldsymbol{\xi}} &= -\boldsymbol{H}\boldsymbol{\xi} + \boldsymbol{\Omega}^\top(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{z} \\ \dot{\boldsymbol{z}} &= -\alpha\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{u}) \end{cases} \tag{3}$$

with $\boldsymbol{g}(\boldsymbol{e}, t) = \boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{u}) - \boldsymbol{f}_u(\boldsymbol{s}, \hat{\boldsymbol{\chi}}, \boldsymbol{u})$ being a vanishing perturbation term ($\boldsymbol{g}(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{u}) \to \boldsymbol{0}$ as $\boldsymbol{e} \to \boldsymbol{0}$). As discussed in [15], the error system (3) can be proven to be locally exponentially stable provided the $p \times p$ square matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^\top$ remains full rank during motion[1]. In particular, the convergence rate of the estimation error $\boldsymbol{z}(t)$ is dictated by the norm of the square matrix $\alpha\boldsymbol{\Omega}\boldsymbol{\Omega}^\top$, in particular by its smallest eigenvalue $\alpha\sigma_1^2$. Since $\sigma_1^2 = \sigma_1^2(\boldsymbol{s}, \boldsymbol{v})$, one can affect the convergence rate of (3) by, e.g., optimizing *online* vector $\boldsymbol{v}$ in order to maximize the value of σ_1^2, see [15]. This possibility represents the *active* component of the SfM algorithm.

We note that the scheme (2) does not require knowledge of $\dot{\boldsymbol{s}}$ (i.e., measurement of the feature velocity on the image plane), but it includes a 'feedback' action on the error $\boldsymbol{\xi} = \boldsymbol{s} - \hat{\boldsymbol{s}}$ between measured and estimated visual features. We also note that the term $\boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{\omega})$ depends on the camera angular velocity, and the term $\boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{u})$ depends (in general) on both the linear and angular velocities. As explained before, the linear velocity $\boldsymbol{v}$ plays a key role on the estimation of the unmeasurable state $\boldsymbol{\chi}$ (see also the next Section). Conversely, the angular speed $\boldsymbol{\omega}$ essentially acts as a perturbation term which can be (partially) compensated for if a measurement of $\boldsymbol{\omega}$ is available (see, e.g., [16], [17]). The goal of the next Section is to propose a SfM scheme (2) completely *invariant* to the camera rotational motion by suitably designing the visual feature vector $\boldsymbol{s}$ based on a spherical-projection model.

III. ROTATION-FREE ACTIVE STRUCTURE FROM MOTION

A. Choice of Visual Features

Let $\boldsymbol{X}_i$ be the coordinates of a 3D point in the camera frame and $r_i = \|\boldsymbol{X}_i\|$ denote its distance to the camera center. The projection of a point onto the unit sphere is defined by

$$\boldsymbol{X}_{si} = \frac{\boldsymbol{X}_i}{r_i}. \tag{4}$$

Note that $\boldsymbol{X}_{si}$ is a *measurable* quantity on the image plane. The time variation of $\boldsymbol{X}_{si}$ is given by $\dot{\boldsymbol{X}}_{si} = \boldsymbol{L}_{X_{si}}\,[\boldsymbol{v},\ \boldsymbol{\omega}]^\top$, where $\boldsymbol{L}_{X_{si}}$ is the interaction matrix related to $\boldsymbol{X}_{si}$ [29]:

$$\boldsymbol{L}_{X_{si}} = \begin{bmatrix} \dfrac{-\boldsymbol{I}_3 + \boldsymbol{X}_{si}\boldsymbol{X}_{si}^\top}{r_i} & [\boldsymbol{X}_{si}]_\times \end{bmatrix} \tag{5}$$

with $\boldsymbol{I}_3$ being the identity matrix and $[\boldsymbol{X}_{si}]_\times \in \mathbb{R}^{3\times 3}$ the skew symmetric matrix associated to vector $\boldsymbol{X}_{si}$.

Let now c_{ij} be the dot product between two projected points on the sphere $\boldsymbol{X}_{si}$ and $\boldsymbol{X}_{sj}$, i.e.,

$$c_{ij} = \boldsymbol{X}_{si}^\top\ \boldsymbol{X}_{sj}. \tag{6}$$

Obviously, c_{ij} is a measurable quantity as well. The time derivative of c_{ij} can be obtained from $\dot{\boldsymbol{X}}_{si}$ and $\dot{\boldsymbol{X}}_{sj}$ as

$$\dot{c}_{ij} = \boldsymbol{X}_{si}^\top\ \dot{\boldsymbol{X}}_{sj} + \boldsymbol{X}_{sj}^\top\ \dot{\boldsymbol{X}}_{si}. \tag{7}$$

By combining (7) with (5), and exploiting $\boldsymbol{X}_{si}^\top\,[\boldsymbol{X}_{sj}]_\times + \boldsymbol{X}_{sj}^\top\,[\boldsymbol{X}_{sj}]_\times = \boldsymbol{0}$, one can easily see that c_{ij} is an *invariant* to rotation, i.e., not affected by the angular velocity $\boldsymbol{\omega}$. Indeed, by combining (5) and (7), we obtain:

$$\dot{c}_{ij} = \left(\frac{-\boldsymbol{X}_{sj}^\top + \boldsymbol{X}_{sj}^\top\boldsymbol{X}_{si}\boldsymbol{X}_{si}^\top}{r_i} + \frac{-\boldsymbol{X}_{si}^\top + \boldsymbol{X}_{si}^\top\boldsymbol{X}_{sj}\boldsymbol{X}_{sj}^\top}{r_j} \right) \boldsymbol{v}, \tag{8}$$

which shows a dependence on the sole camera linear velocity $\boldsymbol{v}$.

[1] This then requires presence of $m \geq p$ independent measurements for p quantities to be estimated.

Noting that $\boldsymbol{X}_{sj}^\top \boldsymbol{X}_{si} = \boldsymbol{X}_{si}^\top \boldsymbol{X}_{sj} = c_{ij}$, one can rewrite (8) as

$$\dot{c}_{ij} = \left(\frac{-\boldsymbol{X}_{sj}^\top + c_{ij}\boldsymbol{X}_{si}^\top}{r_i} + \frac{-\boldsymbol{X}_{si}^\top + c_{ij}\boldsymbol{X}_{sj}^\top}{r_j} \right) \boldsymbol{v} \quad (9)$$

which can be further expanded in

$$\dot{c}_{ij} = \left(-\boldsymbol{X}_{sj}^\top \boldsymbol{v} + c_{ij}\boldsymbol{X}_{si}^\top \boldsymbol{v}\right) \frac{1}{r_i} + \left(-\boldsymbol{X}_{si}^\top \boldsymbol{v} + c_{ij}\boldsymbol{X}_{sj}^\top \boldsymbol{v}\right) \frac{1}{r_j}. \quad (10)$$

Note that the only unknown quantities in (10) are the two distances from the camera center r_i and r_j.

Let now

$$\boldsymbol{\chi} = [\chi_1, ..., \chi_p] = \left[\frac{1}{r_1}, \ \frac{1}{r_2}, ..., \frac{1}{r_p} \right]$$

be the 3D structure to be estimated and define

$$\alpha_i = \boldsymbol{X}_{si}^\top \boldsymbol{v}, \qquad i = 1, ..., p$$

as an additional measurable quantity (assuming $\boldsymbol{v}$ is known). Equation (10) takes the form

$$\dot{c}_{ij} = (-\alpha_j + c_{ij}\alpha_i)\chi_i + (-\alpha_i + c_{ij}\alpha_j)\chi_j. \quad (11)$$

Since vector χ_i appears linearly in the dynamics, one can recognize that (11) has the same structural form of the first row of (1) but it lacks a dependence on $\boldsymbol{\omega}$.

As for vector $\boldsymbol{\chi}$, one has

$$\dot{\chi}_i = \frac{\boldsymbol{X}_i^\top}{r_i^3}\boldsymbol{v} = \frac{\boldsymbol{X}_{si}^\top}{r_i^2}\boldsymbol{v} = \boldsymbol{X}_{si}^\top \boldsymbol{v} \ \chi_i^2 = \alpha_i \ \chi_i^2, \quad (12)$$

thus showing that $\boldsymbol{\chi}$ is invariant to rotations too. The dynamics (12) plays then the role of the second row of (1), but, again, with the relevant difference of being *independent* from $\boldsymbol{\omega}$.

Let us now consider the minimal case of $m = 3$ observed points with $p = 3$ associated 3D quantities to be estimated (availability of $m \geq p$ independent measurements is indeed needed for allowing matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^\top$ to have full rank, see the previous Section): by defining

$$\mathbf{s} = [c_{12}, \ c_{13}, \ c_{23}]^\top, \quad \boldsymbol{\chi} = \left[\frac{1}{r_1}, \ \frac{1}{r_2}, \ \frac{1}{r_3} \right]^\top,$$

and using (11–12), a fully rotation-independent system can be obtained

$$\begin{cases} \dot{\boldsymbol{s}} = \begin{bmatrix} -\alpha_2 + c_{12}\alpha_1 & -\alpha_1 + c_{12}\alpha_2 & 0 \\ -\alpha_3 + c_{13}\alpha_1 & 0 & -\alpha_1 + c_{13}\alpha_3 \\ 0 & -\alpha_3 + c_{32}\alpha_2 & -\alpha_2 + c_{23}\alpha_3 \end{bmatrix} \boldsymbol{\chi} \\ \quad = \boldsymbol{\Omega}^\top(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\chi} \\ \dot{\boldsymbol{\chi}} = \begin{bmatrix} \alpha_1\chi_1^2 \\ \alpha_2\chi_2^2 \\ \alpha_3\chi_3^2 \end{bmatrix} = \boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{v}) \end{cases}. \quad (13)$$

The next Section details an active strategy similar to what proposed in [16] for optimizing the estimation of the 3D structure $\boldsymbol{\chi}$ but, contrarily to [16], *without* requiring knowledge of the camera rotational motion.

B. Optimization of the 3D Structure Estimation

Let $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{\Omega}$ be the singular value decomposition of matrix $\boldsymbol{\Omega}$, where $\boldsymbol{\Sigma} = [\boldsymbol{S} \ \ \mathbf{0}]$, $\boldsymbol{S} = diag(\sigma_i) \in \mathbb{R}^{p\times p}$, and $0 \leq \sigma_1 \leq ... \leq \sigma_p$ are the singular values of $\boldsymbol{\Omega}$. In [16], it has been proposed to choose gain $\boldsymbol{H}$ in (2) as

$$\boldsymbol{H} = \boldsymbol{V} \begin{bmatrix} \boldsymbol{D}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2 \end{bmatrix} \boldsymbol{V}^\top \quad (14)$$

with $\boldsymbol{D}_1 \in \mathbb{R}^{p\times p} > 0$, $\boldsymbol{D}_2 \in \mathbb{R}^{(m-p)\times(m-p)} > 0$, and to further design matrix $\boldsymbol{D}_1$ as a function of the singular values of $\boldsymbol{\Omega}$ by setting $\boldsymbol{D}_1 = diag(c_i), c_i > 0$ with $c_i = 2\sqrt{\alpha}\sigma_i$. Indeed, this choice allows to impose to the estimation error $\boldsymbol{z}(t)$ a transient behavior approximately equivalent to that of a linear critically-damped second-order system.

Since, in our case, $m = p = 3$, eq. (14) reduces to $\mathbf{H} = \boldsymbol{V}\boldsymbol{D}_1\boldsymbol{V}^\top$. Furthermore, as explained, the convergence speed of $\boldsymbol{z}(t)$ is dictated by the scalar quantity $\alpha\sigma_1^2$ (with $\sigma_1^2(\boldsymbol{s}, \boldsymbol{v})$ being the smallest eigenvalue of $\boldsymbol{\Omega}\boldsymbol{\Omega}^\top$). Therefore, the estimation convergence can be affected by either increasing gain α or by selecting a suitable linear velocity $\boldsymbol{v}$ in order to maximize σ_1^2. Increasing the gain α is a straightforward possibility but at the cost of practical issues such as an increased sensitivity to noise. On the other hand, since the norm of $\boldsymbol{\Omega}\boldsymbol{\Omega}^\top$ is highly dependent on the norm of the linear velocity $\|\boldsymbol{v}\|$, the value of σ_1^2 can also be increased by traveling at larger speed for a constant α. However, increasing $\|\boldsymbol{v}\|$ will also increase the traveled distance and the overall 'control effort'. Therefore, in this work we choose to keep $\|\boldsymbol{v}\|$ at a reasonable constant value, and optimize the direction of $\boldsymbol{v}$ in order to maximize σ_1^2. Sect. IV explains how to achieve this take in an effective way.

IV. RESULTS

In this section, simulation results are firstly presented. Then experimental results obtained using fisheye camera mounted on an Afma6 robot will be discussed.

A. Simulation results

In the following simulations, a fish-eye like camera obeying the unified projection model (refer to [20]) with focal scaling factors $F_x = F_y = 600$ pixels, coordinates of the principal point $u_x = 300$ and $u_y = 400$ pixels and distortion parameter $k = 1.6$ is used to generate the image point coordinates. Additionally, a Gaussian white noise with standard deviation equal to 0.5 pixel is added to the coordinates of each point in the image during the estimation process.

The coordinates of the three points are defined in the camera initial frame by:

$$\boldsymbol{X}_0 = \begin{bmatrix} -0.4 & 0.4 & 0.25 \\ 0.2 & -0.4 & 0.4 \\ 1 & 1 & 0.8 \end{bmatrix} \quad (15)$$

We now discuss how to implement the optimization of the SfM scheme in our particular case.

1) Maximizing the smallest singular value of $\mathbf{\Omega\Omega}^\top$: As previously discussed, the smallest singular value of $\mathbf{\Omega\Omega}^\top$ could be increased by traveling faster in order to increase the convergence rate of the estimation error $\boldsymbol{z}(t)$. In the following we instead keep a constant $\|\boldsymbol{v}\|$ and just optimize the direction of $\boldsymbol{v}$. For this purpose, let us define the linear velocity using spherical angles by $\boldsymbol{v} = \|\boldsymbol{v}\|[cos(\phi)cos(\theta) \quad cos(\phi)sin(\theta) \quad sin(\phi)]$. Therefore, since $\sigma_1^2 = \sigma_1^2(\boldsymbol{s}, \boldsymbol{v})$ and $\|\boldsymbol{v}\| = const$, maximization of σ_1^2 reduces to a two dimensional optimization problem of the angles ϕ and θ for a given (measured) vector $\boldsymbol{s}$.

Since matrix $\mathbf{\Omega\Omega}^\top$ changes w.r.t. the camera poses, the multiplicity of σ_1^2 can vary during motion making a gradient-based optimization ill-conditioned in practice (because of the difficulties in evaluating derivatives of an eigenvalue when close to be repeated, see [30] for a discussion).

Let us consider the case of the three points defined in the camera frame by (15). Figure 2.a shows in color map the values of the smallest singular value σ_1^2 as a function of the direction of the linear speed. The points in red are the location on the sphere of the three tracked points. These results are obtained for the linear direction over a hemisphere, the results for the second part are perfectly symmetrical (since matrix $\mathbf{\Omega\Omega}^\top$ is unaffected by replacing $\|\boldsymbol{v}\|$ with $-\|\boldsymbol{v}\|$).

From Fig. 1 it can be seen that the optimization problem has several maxima. It can be further possible to prove that the optimal directions of the linear speed are invariant to rotations. Indeed, let us consider a set of points in 3D $\boldsymbol{X}'$ and $\boldsymbol{X}$ expressed in two camera poses (denoted by pose 1 and 2) and linked by a rotational motion $\boldsymbol{R}$. In this case, we have $\boldsymbol{X}' = \boldsymbol{R}\,\boldsymbol{X}$ and using (4) it can be shown that for the corresponding projection points on the sphere, we also have $\boldsymbol{X}'_s = \boldsymbol{R}\,\boldsymbol{X}_s$. From this, firstly, invariance to rotations of the inner products between projected points on the sphere is confirmed again since $c'_{ij} = \boldsymbol{X}'^\top_{si}\boldsymbol{X}'_{sj} = \boldsymbol{X}^\top_{si}\boldsymbol{R}^\top\boldsymbol{R}\boldsymbol{X}_{sj} = c_{ij}$. Secondly, for any value of $\boldsymbol{v}$ in the camera pose 1, if the linear velocity for the second pose $\boldsymbol{v}'$ is defined by $\boldsymbol{v}' = \boldsymbol{R}\,\boldsymbol{v}$, then we obtain $\alpha'_i = \boldsymbol{X}'^\top_{si}\boldsymbol{v}' = \boldsymbol{X}^\top_{si}\boldsymbol{R}^\top\boldsymbol{R}\boldsymbol{v} = \alpha_i$. Since $\mathbf{\Omega}$ is nothing but a function of c_{ij} and α_i, one can conclude that rotational motions do not change the optimal directions with respect to the 3D points. This will also be confirmed by the next results.

In the following, the direction of the linear speed maximizing σ_1^2 is obtained as follows: the optimal direction is initialized by evaluating σ_1^2 over a regular grid on ϕ and θ defining a hemisphere as the one shown on Fig. 1. The values of ϕ and θ corresponding to a maximum for σ_1^2 are chosen as an initialization for a stepest-descent like method that keeps track of the optimal direction during the camera motion.

2) Estimation results obtained by maximizing the smallest singular value of $\mathbf{\Omega\Omega}^\top$: in the first simulations, the depth information is initialized as follows $\hat{\chi_1} = \chi_1 + 0.5, \quad \hat{\chi_2} = \chi_2 - 0.5, \quad \hat{\chi_3} = \chi_3 - 0.5$. Two cases has been considered: in the first case a pure translational motion is realized, and in the second case both translations and rotations are involved. In both cases, the linear velocity follows the direction maximizing σ_1 with a constant norm $\|\boldsymbol{v}\| = 0.05$ m/s. In the second case, a uniform angular velocity with norm $\omega = 2.5$ deg/s around a randomly generated axis is also added. The behaviors of the estimated χ is depicted in Figs. 2.a and 2.b for the two considered cases. From the plots, it can be noticed how the estimated $\hat{\boldsymbol{\chi}}(t)$ correctly converges towards $\boldsymbol{\chi}(t)$, and also how the estimation transient is almost identical in both cases despite the presence of a non-zero rotational motion in the second case. This can also be again verified by looking at the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$ in Figs. 2.c and 2.d. Finally, the values of $\sigma_1(t)$ during the camera motion are depicted on Figs. 2.c and 2.d. From these plots, one can again verify the approximately identical behavior of the smallest singular values $\sigma^2(t)$ in the two cases (pure translation, or concurrent translation and rotation).

In the second simulation, the behavior of the estimation scheme (2) is verified using a different initialization $\hat{\chi_1} = \chi_1 - 0.6, \quad \hat{\chi_2} = \chi_2 + 0.5, \quad \hat{\chi_3} = \chi_3 - 0.4$. The obtained results are reported on Fig. 3: Figs. 3.a and 3.b show the behavior of $\hat{\boldsymbol{\chi}}(\boldsymbol{t})$ with respect to $\hat{\boldsymbol{\chi}}(t)$, while Figs. 3.c and 3.d show the estimation errors $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}$(t). From the obtained plots, one can conclude the soundness of the reported theoretical analysis as the proposed framework clearly results invariant to rotations.

B. Experimental results

In this section, we detail experiments obtained with a fish-eye camera mounted on the end-effector of an AFMA 6 robot and the Visp software library [31] for image processing. The camera parameters are: focal scaling factors $F_x = 716.09, F_y = 715.31$ pixels, principal point $u_x = 304.45$ and $u_y = 396.83$ pixels and distortion parameter $k = 1.71$. Similarly to simulations, we will compare the behavior of the observer (2) in case of a pure translation to the case of a generic camera motion involving both translations and rotations.

Figure 4 shows the value of σ_1 in color map as a function of the direction of the linear velocity over a hemisphere. The points in red on the same figure represent the projected points on the unitary sphere for the camera initial pose. As in the simulation results, it can be seen that several directions of the linear velocity $\boldsymbol{v}$ allow to maximize the smallest singular value σ_1. In our experiment, from the values of σ_1 over the grid on ϕ and θ, we chose the direction corresponding to the global maximum as initialization for $\boldsymbol{v}$. Then, during the camera motion, this optimal direction is tracked iteratively starting from the previous direction to initialize the optimization procedure. The norm of the linear velocity is chosen as $\|\boldsymbol{v}\| = 0.02$ m/s during the experiments and when a generic motion is considered the angular velocity is setup to $\boldsymbol{\omega} = [0 \ -1 \ 1.5]$ degrees/s.

The obtained behaviors of the estimated depth informations are depicted on Figure 5. The results shown on Fig. 5.a are obtained using (16) to initialize the depth information. From the plots, it can be seen the almost identical behaviors in the case of pure translation and in the case of a general motion. Figure 5.b depicts the results obtained using (17) to initialize the depth information. Once again we note

that the behavior of the observer is rotation invariant. In addition, the plots on Figure 5.c show that the observer has converged to the same values for different initializations. Finally, the invariance to rotations is also confirmed by the plots depicting σ_1 during the estimation (refer to Fig. 5.d).

$$\hat{\chi}_1 = 1,\ \hat{\chi}_2 = 1.5,\ \hat{\chi}_3 = 0.8 \tag{16}$$

$$\hat{\chi}_1 = 0.3,\ \hat{\chi}_2 = .5,\ \hat{\chi}_3 = 1.5 \tag{17}$$

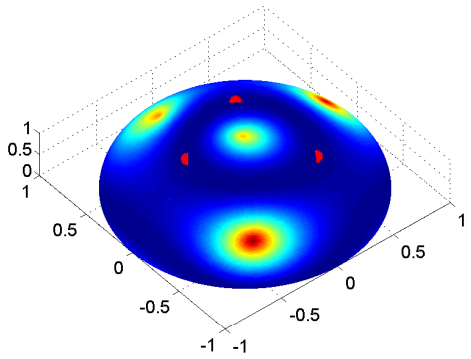

Fig. 1. Simulation results 1: σ_1 as function of the direction of $\boldsymbol{v}$

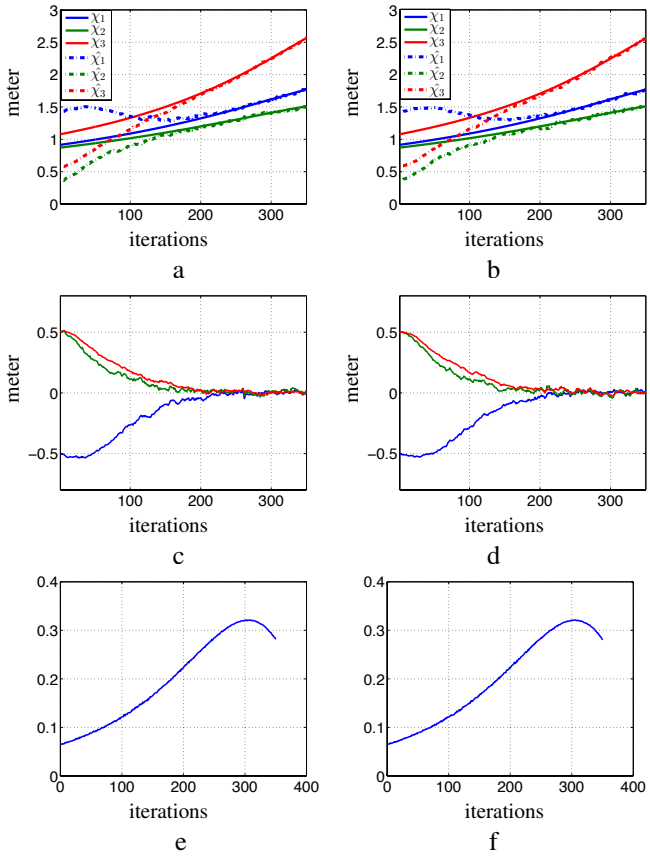

Fig. 2. Simulation results 1: a) behavior of the estimated and real $\boldsymbol{\chi}$ in case of a pure translation motion, b) behavior of the estimated and real $\boldsymbol{\chi}$ in case of a generic motion, c) behavior of the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$ in case of a pure translation motion, d) behavior of the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$ in case of a generic motion, e) behavior of σ_1 in case of pure translation motion, f) behavior of σ_1 in case of a generic motion

V. Conclusion and Perspectives

This paper has proposed a rotation-free active SfM strategy. More precisely, we have discussed a new estimation scheme fully invariant to the camera angular velocity for recovering the 3D structure of the scene. Thanks to this possibility, one is freed from the need of obtaining precise measurements/estimations of the camera angular which can be corrupted by miscalibration errors, noise or biases (e.g., cameras mounted on a UAV). Simulation and experimental results have shown the effectiveness of the proposed approach. The proposed observer exhibits similar performances whatever the angular velocities of the camera as expected. Future works will be devoted to extend our active rotation-free SfM scheme to features computed directly from photometric data.

Acknowledgment

This work has been sponsored by the French Government research program Investissements d'avenir through the RobotEx Equipment of Excellence (ANR-10-EQPX-44) and the LabEx IMobS3 (ANR7107LABX716701), the European Union and the Région Auvergne.

References

[1] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. Journal of Computer Vision*, vol. 3, no. 3, pp. 209–238, 1989.

[2] S. Soatto, R. Frezza, and P. Perona, "Motion estimation via dynamic vision," *IEEE Trans. on Automatic Control*, vol. 41, no. 3, pp. 393–413, 1996.

[3] J. Civera, A. J. Davison, and J. Montiel, "Inverse Depth Parametrization for Monocular SLAM," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[4] J. Civera, Ó. Grasa, A. J. Davison, and J. M. M. Montiel, "1-Point RANSAC for EKF Filtering: Application to Real-Time Structure from Motion and Visual Odometry," *Journal of Field Robotics*, vol. 27, no. 5, pp. 609–631, 2010.

[5] S. Omari and G. Ducard, "Metric Visual-Inertial Navigation System Using Single Optical Flow Feature," in *Proc. of the 2013 European Control Conference (ECC'13)*, 2013, pp. 1310–1316.

[6] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range Identification for Perspective Vision Systems," *IEEE Trans. on Automatic Control*, vol. 48, no. 12, pp. 2232–2238, 2003.

[7] N. Metni and T. Hamel, "Visual Tracking Control of Aerial Robotic Systems with Adaptive Depth Estimation," *Int. Journal of Control, Automation, and Systems*, vol. 1, no. 5, pp. 51–60, 2007.

[8] A. De Luca, G. Oriolo, and P. Robuffo Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *Int. Journal of Robotics Research*, vol. 27, no. 10, pp. 1093–1116, 2008.

[9] F. Morbidi, G. L. Mariottini, and D. Prattichizzo, "Observer design via immersion and invariance for vision-based leader-follower formation control," *Automatica*, vol. 46, no. 1, pp. 148–154, 2010.

[10] P. Corke, "Spherical Image-Based Visual Servo and Structure Estimation," in *IEEE Int. Conf. on Robotics and Automation, ICRA'10*, 2010, pp. 5550–5555.

[11] M. Sassano, D. Carnevale, and A. Astolfi, "Observer design for range and orientation identification," *Automatica*, vol. 46, no. 8, pp. 1369–1375, 2010.

[12] A. Martinelli, "Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Trans. on Robotics*, vol. 1, no. 28, pp. 44–60, 2012.

[13] A. Eudes, P. Morin, R. Mahony, and T. Hamel, "Visuo-inertial fusion

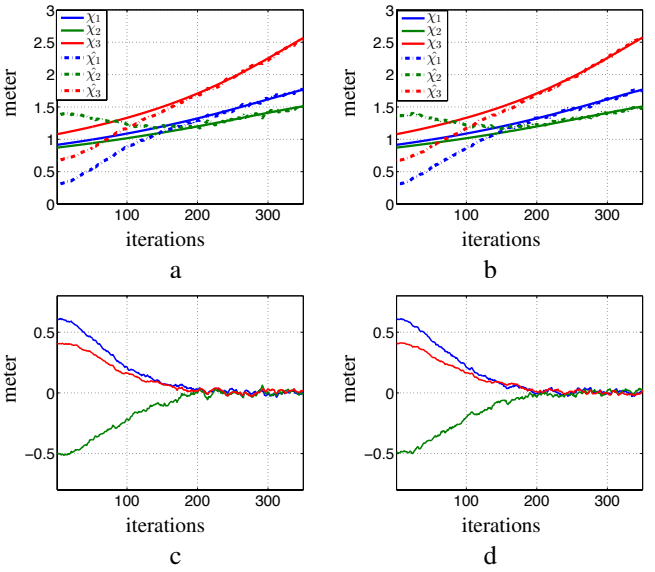

Fig. 3. Simulation results 2: a) behavior of the estimated and real $\boldsymbol{\chi}$ in case of a pure translation motion, b) behavior of the estimated and real $\boldsymbol{\chi}$ in case of a generic motion, c) behavior of the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$ in case of a pure translation motion, d) behavior of the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$ in case of a generic motion

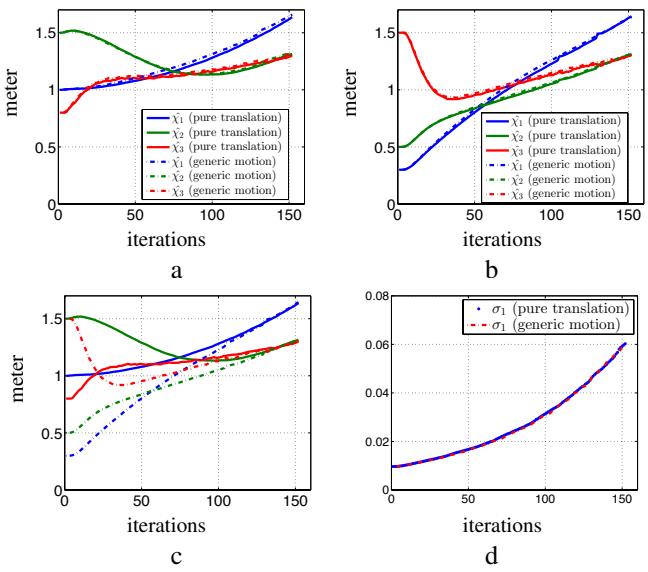

Fig. 5. Behavior of the estimated $\boldsymbol{\chi}$ in case of pure translation and generic motions: a) result obtained using (16) as initialization, b) result obtained using (17) as initialization c) comparison of the behavior using (16) (continuous plot) and using (17) (dashed plot) when generic motion is considered, d) behavior of σ_1 in the cases of a pure translational and a generic motions

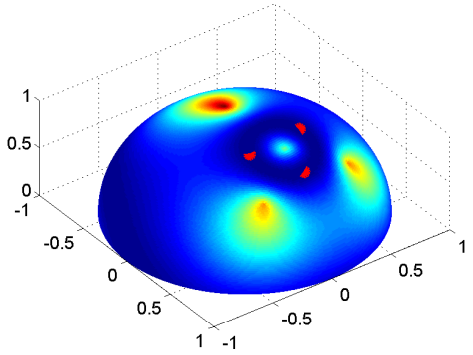

Fig. 4. Experimental results: σ_1 as function of $\boldsymbol{v}$ directions

for homography-based filtering and estimation," in *IEEE Int. Conf. on Intelligent Robots and Systems, IROS'13*, 2013, pp. 5186–5192.

[14] V. Grabe, H. H. Bülthoff, D. Scaramuzza, and P. Robuffo Giordano, "Nonlinear Ego-Motion Estimation from Optical Flow for Online Control of a Quadrotor UAV," *The International Journal of Robotics Research*, vol. 8, no. 34, pp. 1114–1135, 2015.

[15] R. Spica and P. Robuffo Giordano, "A framework for active estimation: Application to structure from motion," in *IEEE Conference on Decision and Control, CDC'13*, Firenze, Italy, December 2013.

[16] R. Spica, P. Robuffo Giordano, and F. Chaumette, "Active structure from motion: Application to point, sphere and cylinder," *IEEE Trans. on Robotics*, vol. 30, no. 6, pp. 1499–1513, December 2014.

[17] R. Spica, P. Robuffo Giordano, and F. Chaumette, "Plane estimation by active vision from point features and image moments," in *IEEE Int. Conf. on Robotics and Automation, ICRA'15*, 2015.

[18] P. Robuffo Giordano, R. Spica, and F. Chaumette, "Learning the shape of image moments for optimal 3d structure estimation," in *IEEE Int. Conf. on Robotics and Automation, ICRA'15*, 2015.

[19] R. Spica, P. Robuffo Giordano, and F. Chaumette, "Coupling Visual Servoing with Active Structure from Motion," in *IEEE Int. Conf. on Robotics and Automation, ICRA'14*, Hong Kong, China, May 2014, pp. 3090–3095.

[20] C. Geyer and K. Daniilidis, "Mirrors in Motion: Epipolar Geometry and Motion Estimation," *Int. Journal on Computer Vision*, vol. 45, no. 3, pp. 766–773, 2003.

[21] J. Courbon, Y. Mezouar, and P. Martinet, "Evaluation of the unified model on the sphere for fisheye cameras in robotic applications," *Advanced Robotics*, vol. 6, no. 8, pp. 947–967, 2012.

[22] C. Mei and P. Rives, "Single View Point Omnidirectional Camera Calibration from Planar Grids," in *IEEE Int. Conf. on Robotics and Automation*, April 2007, pp. 3945–3950.

[23] O. Tahri, H. Araujo, F. Chaumette, and Y. Mezouar, "Robust image-based visual servoing using invariant visual information," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1588–1600, December 2013.

[24] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke, "Decoupled image-based visual servoing for cameras obeying the unified projection model," *IEEE Trans. on Robotics*, vol. 26, no. 4, pp. 684–697, August 2010.

[25] R. T. Fomena, O. Tahri, and F. Chaumette, "Distance-based and orientation-based visual servoing from three points," *Robotics, IEEE Transactions on*, vol. 27, no. 2, pp. 256–267, 2011.

[26] O. Tahri, H. Araujo, Y. Mezouar, and F. Chaumette, "Efficient iterative pose estimation using an invariant to rotations," *IEEE Trans. on Cybernetics*, vol. 22, no. 4, pp. 199–207, February 2014.

[27] H. Hadj-Abdelkader, Y. Mezouar, P. Martinet, and F. Chaumette, "Catadioptric visual servoing from 3d straight lines," *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 652–665, June 2008.

[28] F. Chaumette and S. Hutchinson, "Visual servo control, Part I: Basic approaches," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[29] T. Hamel and R. Mahony, "Visual servoing of an under-actuated dynamic rigid-body system: an image-based approach," *Robotics and Automation, IEEE Transactions on*, vol. 18, no. 2, pp. 187–198, 2002.

[30] M. I. Friswell, "The derivatives of repeated eigenvalues and their associated eigenvectors," *Journal of Vibration and Acoustics*, vol. 118, pp. 390–397, 1996.

[31] E. Marchand, F. Spindler, and F. Chaumette, "Visp for visual servoing: a generic software platform with a wide class of robot control skills," *IEEE Robotics and Automation Magazine*, vol. 12, no. 4, pp. 40–52, December 2005.