Active Structure from Motion for Spherical and Cylindrical Targets

Riccardo Spica, Paolo Robuffo Giordano, and François Chaumette

***Abstract*— Structure estimation from motion (SfM) is a classical and well-studied problem in computer and robot vision, and many solutions have been proposed to treat it as a recursive filtering/estimation task. However, the issue of *actively* optimizing the transient response of the SfM estimation error has not received a comparable attention. In this paper, we provide an experimental validation of a recently proposed nonlinear active SfM strategy via two concrete applications: 3D structure estimation for a spherical and a cylindrical target. The experimental results fully support the theoretical analysis and clearly show the benefits of the proposed active strategy. Indeed, by suitably acting on the camera motion and estimation gains, it is possible to *assign* the error transient response and make it equivalent to that of a reference linear second-order system with *desired* poles.**

I. INTRODUCTION

The problem of Structure from Motion (SfM), i.e., how to reconstruct a 3D observed scene from images taken by a moving camera, is a very classical and well-studied topic in computer and robot vision. When acquiring consecutive images over time, a possibility is to treat SfM as a recursive/filtering process: images and camera motion can be elaborated *online* for obtaining an incremental estimation of the scene structure. Other approaches (e.g., bundle adjustment) rely instead on *global/offline* optimization methods meant to solve SfM problems by processing altogether information acquired over an extended time period. A recent discussion about the pros/cons of both approaches in the context of Visual SLAM can be found in [1]. Furthermore, within the first class of (online) methods, a vast literature exists for addressing SfM: for instance, as a non-exhaustive list, Extended Kalman Filter-based solutions have been proposed in [2]–[5], and other approaches exploiting techniques from (deterministic) nonlinear observation can be found in [6]–[13] and references therein. Finally, in [14] the authors nicely discuss the advantages of a sensor-centered recursive SLAM algorithm sharing the same theoretical setting of what presented in this work.

Much less attention has instead been devoted to the problem of *actively optimizing the convergence rate* of a SfM estimation task by acting on the *motion imposed to the camera* and on the *employed estimation gains*. For instance, in [15] an active strategy for minimizing the effects of image noise and discretization errors was proposed and experimentally tested, but without the aim of *also* imposing a desired estimation transient response. In [16], the problem of actively selecting which features to track for improving the indoor localization of a wheeled mobile robot is successfully addressed: however, no attempt is made to actively shape the robot motion so as to optimize the SfM convergence (the robot navigates in an 'uninformed' way w.r.t. the estimation task).

With respect to these and other prior works, in [17] we have recently proposed a general framework for designing a class of nonlinear observers (which includes SfM as particular case) with an estimation error transient response that can be (i) explicitly characterized and (ii) *actively shaped* so as to match that of a reference linear second-order system with assigned poles. This is achieved by simultaneously acting on the estimation gains and on the system inputs (i.e., the camera linear velocity in the SfM case).

The methodology proposed in [17] applies to all those problems in which an *invertible function* of the unknown states can appear linearly in the system dynamics, as it is indeed the case for SfM. In this paper, we exploit this fact and consider a concrete application involving SfM for two 3D objects, namely, a spherical and a cylindrical target. A first contribution of our work is the development of two *novel minimal parameterizations* of the sphere and cylinder 3D geometries which allow to reduce the SfM problem to the estimation of the sole sphere/cylinder radius, with all the remaining quantities obtained from image measurements and camera velocities. A second contribution is the reported successful experimental validation which (i) results in excellent agreement with the developed theoretical analysis and (ii) demonstrates the feasibility of our approach in real conditions.

We finally note that the ability to both characterize and optimize the transient behavior of SfM problems brings a significant added value compared to more classical 'inactive' estimation strategies: for instance, it allows to obtain the 'best' estimation error convergence when subject to real-world constraints such as limited camera velocity or upper bounds on the estimation gains due to noise, discretization, or other typical non-idealities. As shown in [18], one can also embed the active SfM strategy within the execution of typical image-based visual servoing tasks. Furthermore, from a more theoretical perspective, the proposed methodology can also be used to get insights into the 'optimal' camera trajectories needed to estimate the scene structure for particular classes of SfM problems (e.g., when dealing with points features, planar objects or specific 3D geometrical primitives). We

R. Spica is with the University of Rennes 1 at Irisa and Inria Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France riccardo.spica@irisa.fr

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

F. Chaumette is with Inria Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France francois.chaumette@irisa.fr

believe the reported results well support these considerations.

The rest of the paper is organized as follows: Sec. II reviews the SfM problem in the context of nonlinear state observation and briefly summarizes the methodology developed in [17] for actively imposing a desired transient response to the estimation error. Section III then focuses on the two SfM problems considered in this work, i.e., radius estimation for spherical and cylindrical targets. Subsequently, Sec. IV reports the corresponding experimental results obtained with a manipulator equipped with an eye-in-hand camera. Finally, Sec. V concludes the paper and discusses some future directions.

II. AN ACTIVE ESTIMATION STRATEGY

Let $\boldsymbol{s} \in \mathbb{R}^m$ be the set of visual features *measured* on the image plane of a (assumed calibrated) camera, $\boldsymbol{\chi} \in \mathbb{R}^p$ a suitable (and locally invertible) function of the unknown 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 that the SfM dynamics takes the general form

$$\left\{ \begin{array}{rcl} \dot{\boldsymbol{s}} & = & \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{u}) + \boldsymbol{\Omega}^T(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\chi} \\ \dot{\boldsymbol{\chi}} & = & \boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}, \boldsymbol{u}) \end{array} \right. \tag{1}$$

where matrix $\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v}) \in \mathbb{R}^{p \times m}$ is a *known* quantity such that $\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{0}) \equiv \boldsymbol{0}$. Let now $(\hat{\boldsymbol{s}}\,, \hat{\boldsymbol{\chi}}) \in \mathbb{R}^{m+p}$ be the estimated state, and define $\boldsymbol{\xi} = \boldsymbol{s} - \hat{\boldsymbol{s}}$ as the 'visual feedback' error (measured $\boldsymbol{s}$ vs. estimated $\hat{\boldsymbol{s}}$) and $\boldsymbol{z} = \boldsymbol{\chi} - \hat{\boldsymbol{\chi}}$ as the 3D structure estimation error. An estimation scheme for system (1) meant to recover the unmeasurable $\boldsymbol{\chi}(t)$ from the measured $\boldsymbol{s}(t)$ can be devised as

$$\left\{ \begin{array}{rcl} \dot{\hat{\boldsymbol{s}}} & = & \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{u}) + \boldsymbol{\Omega}^T(\boldsymbol{s}, \boldsymbol{v})\hat{\boldsymbol{\chi}} + \boldsymbol{H}\boldsymbol{\xi} \\ \dot{\hat{\boldsymbol{\chi}}} & = & \boldsymbol{f}_u(\boldsymbol{s}, \hat{\boldsymbol{\chi}}, \boldsymbol{u}) + \alpha\boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v})\boldsymbol{\xi} \end{array} \right. \tag{2}$$

where $\boldsymbol{H} > 0$ and $\alpha > 0$ are suitable gains. We note that the scheme (2) *does not* require knowledge of $\dot{\boldsymbol{s}}$ (i.e., measurement of velocities on the image plane), but it only needs measurement of $\boldsymbol{s}$ (the 'visual features') and of $(\boldsymbol{v}, \boldsymbol{\omega})$ (the camera linear/angular velocity in the camera frame).

From (1–2), the estimation error dynamics is given by

$$\left\{ \begin{array}{rcl} \dot{\boldsymbol{\xi}} & = & -\boldsymbol{H}\boldsymbol{\xi} + \boldsymbol{\Omega}^T(t)\,\boldsymbol{z} \\ \dot{\boldsymbol{z}} & = & -\alpha\boldsymbol{\Omega}(t)\,\boldsymbol{\xi} + (\boldsymbol{f}_u(\boldsymbol{s}, \boldsymbol{\chi}\,\boldsymbol{u}) - \boldsymbol{f}_u(\boldsymbol{s}, \hat{\boldsymbol{\chi}}\,\boldsymbol{u})) \\ & = & -\alpha\boldsymbol{\Omega}(t)\,\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{e}, t) \end{array} \right. \tag{3}$$

with $\boldsymbol{g}(\boldsymbol{e}, t)$ being a vanishing term w.r.t. the error vector $\boldsymbol{e}$, i.e., such that $\boldsymbol{g}(\boldsymbol{0}, t) = \boldsymbol{0}$, $\forall t$. Assuming $m \geq p$, that is, more measurements $\boldsymbol{s}$ are available than the number of estimated quantities $\boldsymbol{\chi}$, the origin of (3) can be proven to be locally exponentially stable if

$$\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t) \geq \frac{\gamma}{T}\boldsymbol{I}, \quad \forall t, \tag{4}$$

see [17].

Remark 2.1: In the special situation $\dot{\boldsymbol{\chi}} = \boldsymbol{0}$ (unknown but *constant* parameters), one has $\boldsymbol{g}(\boldsymbol{e}, t) \equiv \boldsymbol{0}$ and *global* exponential convergence for the error system (3). The sphere/cylinder cases considered in the next sections meet this condition. However, we stress that the estimation scheme (2) is not restricted to this particular situation but it can be applied to the general case of *state observation* problems in which the unknown $\boldsymbol{\chi}$ is subject to a non-negligible dynamics as in (1), see [17] for some examples.

It is further possible to show that, by degisning the estimation gain $\boldsymbol{H}$ as a function of the current $\boldsymbol{s}$ and $\boldsymbol{v}$, one can obtain a completely decoupled transient behavior for the estimation error $\boldsymbol{z}$ (expressed in some suitable coordinates $\boldsymbol{\eta} \in \mathbb{R}^p$)

$$\ddot{\eta}_i + c_i\dot{\eta}_i + \alpha\sigma_i^2\eta_i = 0, \quad i = 1 \ldots p, \tag{5}$$

where σ_i^2 are the eigenvalues of martix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ and c_i are the scalar damping factors embedded in $\boldsymbol{H}$. By then taking, for instance, $c_i = c_i^* = 2\sqrt{\alpha}\sigma_i$ a critically damped state evolution is imposed to the estimation error behavior.

From (5) it also follows that the convergence speed of the estimation error $\boldsymbol{z}(t)$ results dictated by the smallest eigenvalue σ_1^2 of $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$. Being $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v})$, one has

$$\dot{(\sigma_1^2)} = \boldsymbol{J}_{v,1}\dot{\boldsymbol{v}} + \boldsymbol{J}_{s,1}\dot{\boldsymbol{s}}, \tag{6}$$

with the Jacobian matrices $\boldsymbol{J}_{v,1}$ and $\boldsymbol{J}_{s,1}$ available in *closed form* and function of known quantities. This relationship can then be inverted w.r.t. vector $\dot{\boldsymbol{v}}$ for affecting *online* $\sigma_1^2(t)$ during motion, e.g., in order to maximize its value for increasing the convergence rate of $\boldsymbol{z}(t)$. We note that this step represents the *active* component of the estimation strategy since, in the general case, inversion of (6) will yield a camera velocity $\boldsymbol{v}(t)$ function of the system measured state $\boldsymbol{s}(t)$.

Remark 2.2: As discussed in [17], in the general case the 'ideal' estimation error dynamics (5) can result perturbed by some (unavoidable) disturbing terms. However, in the special case $p = 1$ (only one quantity to be estimated), if $\sigma_1(t) \equiv const$ then the disturbing effects can be fully compensated for and it is always possible to *exactly* enforce the "ideal" decoupled estimation error dynamics.

We then conclude with the following considerations on the choice of gain α in (2). In the SfM context, the norm of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ is strongly related to the norm of the camera linear velocity $\boldsymbol{v}$. Roughly speaking, the 'faster' the motion ($\sim$ larger $\|\boldsymbol{v}\|$), the 'larger' the value of σ_1^2 ($\sim$ larger $\|\boldsymbol{\Omega}\boldsymbol{\Omega}^T\|$). Therefore, in order to obtain a desired estimation convergence speed for (2) (dictated by $\alpha\sigma_1^2$) one can equivalently (i) travel at a larger speed $\|\boldsymbol{v}\|$ or (ii) increase the gain α. While increasing the gain α may always appear more convenient in terms of reduced control effort, practical issues such as noise, discretization or quantization errors, may impose an upper limit on the possible value of α, thus necessarily requiring a larger $\|\boldsymbol{v}\|$ for obtaining the desired convergence speed. Furthermore, as in all SfM problems, a $\|\boldsymbol{v}\| \neq 0$ is also mandatorily required for guaranteeing $\sigma_1^2 > 0$ (a non-translating camera cannot estimate the scene structure).

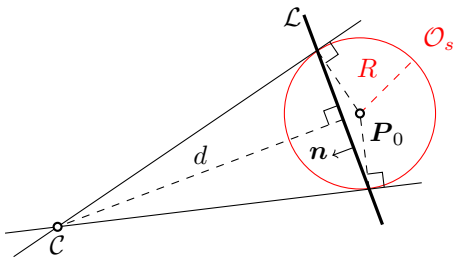


Fig. 1: Camera $\mathcal{C}$ and spherical target $\mathcal{O}_s$ with planar limb surface $\mathcal{L}$

III. STRUCTURE FROM MOTION FOR 3D PRIMITIVES

We now discuss the application of the proposed active estimation framework to two SfM problems involving a sphere and a cylinder. A main contribution of this section is the development of two *novel minimal parameterizations* of the sphere and cylinder 3D geometry that allow to considerably simplify the estimation task to the sole unknown sphere/cylinder radius: it is then possible to implement an estimator meeting the requirements of both Remarks 2.1 and 2.2 (estimation of a *single* and *constant* unknown state).

A. Spherical target

Consider a sphere $\mathcal{O}_s$ of radius R and let $\boldsymbol{P}_0 = (X_0,\, Y_0,\, Z_0)$ be the coordinates of its center in the camera frame. Let also $\mathcal{L} : \boldsymbol{n}^T\boldsymbol{E} + d = 0$ represent the *planar limb surface* associated to the sphere in the camera frame, where $\boldsymbol{E} \in \mathbb{R}^3$ is a 3D point on the plane, $\boldsymbol{n} \in \mathbb{S}^2$ is the plane unit normal vector and $d \in \mathbb{R}$ the plane distance to the camera center [19]. Figure 1 shows the quantities of interest.

From [19], the depth Z of any point $\boldsymbol{E}$ lying on $\mathcal{L}$ can be expressed in terms of its normalized image coordinates $\boldsymbol{p} = (x,\, y,\, 1)$ as

$$\frac{1}{Z} = \frac{X_0}{K}x + \frac{Y_0}{K}y + \frac{Z_0}{K} = \boldsymbol{\chi}^T\boldsymbol{p}, \tag{7}$$

where $K = \boldsymbol{P}_0^T\boldsymbol{P}_0 - R^2$ and $\boldsymbol{\chi} = \boldsymbol{P}_0/K = -\boldsymbol{n}/d \in \mathbb{R}^3$ represents an *unmeasurable* quantity. The interaction matrix of a generic $(i,\, j)$-th order moment m_{ij} evaluated on the image of $\mathcal{O}_s$ depends linearly on $\boldsymbol{\chi}$, see [9], [20]. Therefore, a first possibility to retrieve the sphere 3D parameters $(\boldsymbol{P}_0,\, R)$ would be to implement the estimation scheme (2) with $\boldsymbol{s}$ being a suitable collection of image moments (e.g., area and barycenter). It is in fact possible to show that K can be expressed in terms of image moments and of vector $\boldsymbol{\chi}$ itself, so that, having estimated $\boldsymbol{\chi}$, one can consequently retrieve $\boldsymbol{P}_0 = \boldsymbol{\chi}K$ and $R = \sqrt{\boldsymbol{P}_0^T\boldsymbol{P}_0 - K}$.

Although conceptually valid, this solution requires the concurrent estimation of *three time-varying quantities* (vector $\boldsymbol{\chi}(t)$). On the other hand, we now propose a *novel representation* that allows to reduce the estimation task to a *single unknown constant parameter*, i.e., the sphere radius R.

To this end, let $\boldsymbol{s} = (s_x,\, s_y,\, s_z) = \boldsymbol{P}_0/R \in \mathbb{R}^3$, with

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\frac{1}{R}\boldsymbol{I} & [\boldsymbol{s}]_\times \end{bmatrix}\boldsymbol{u} \tag{8}$$

and $[\boldsymbol{v}_1]_\times\boldsymbol{v}_2 = \boldsymbol{v}_1 \times \boldsymbol{v}_2$, see [21]. Let also $(x_g,\, y_g,\, n_{20},\, n_{11},\, n_{02})$ be the barycenter and normalized centered moments of order 2 measured from the elliptical projection of the sphere $\mathcal{O}_s$ on the image plane, and a_1 be length of the minor axis of the observed ellipse. From [20], [21], one has

$$x_g = \frac{X_0Z_0}{Z_0^2 - R^2}, \quad y_g = \frac{Y_0Z_0}{Z_0^2 - R^2}, \quad a_1^2 = \frac{R^2}{Z_0^2 - R^2} \tag{9}$$

and, additionally,

$$a_1^2 = 2\left(n_{20} + n_{02} - \sqrt{(n_{20} - n_{02})^2 + 4n_{11}}\right). \tag{10}$$

Using (9) and the definition of vector $\boldsymbol{s}$, it then follows

$$s_x = \frac{x_g}{s_z a_1^2}, \quad s_y = \frac{y_g}{s_z a_1^2}, \quad s_z = \sqrt{\frac{1 + a_1^2}{a_1^2}}, \tag{11}$$

which, exploiting (10), shows that vector $\boldsymbol{s}$ can be equivalently expressed in terms of sole *image quantities* (i.e., measured image moments). Therefore, having obtained $\boldsymbol{s}$ from image moments, the only unknown quantity left is the sphere radius R which, if available, allows recover the 3D sphere parameters as $\boldsymbol{P}_0 = \boldsymbol{s}R$.

Since (8) is linear in $1/R$, we can define $\chi = 1/R$, with then $m = 3$ and $p = 1$, and obtain for (1)

$$\begin{cases} \dot{\boldsymbol{s}} &= [\boldsymbol{s}]_\times\boldsymbol{\omega} - \boldsymbol{v}\chi \\ \dot{\chi} &= 0 \end{cases}. \tag{12}$$

We note that, being in this case $\dot{\chi} = 0$, it is $g(\boldsymbol{e},\, t) = 0$ thus resulting in a global convergence for the error system (3). Furthermore, matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ reduces to its single eigenvalue $\sigma_1^2 = \|\boldsymbol{v}\|^2$: therefore by keeping a $\|\boldsymbol{v}\|^2 \equiv const > 0$, one can (i) satisfy condition (4) and (ii) *exactly* enforce the 'ideal' estimation error dynamics (5), see Remark 2.2. Moreover, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{v}) = \boldsymbol{v}^T$ with the Jacobians in (6) taking the expressions $\boldsymbol{J}_{v,1} = 2\boldsymbol{v}^T, \boldsymbol{J}_{s,1} = \boldsymbol{0}$.

We can finally note the following facts: first, since σ_1^2 is only function of the camera linear velocity $\boldsymbol{v}$, one can freely exploit the camera angular velocity $\boldsymbol{\omega}$ for, e.g., keeping the sphere at the center of the image by regulating $(s_x,\, s_y)$ to zero. Second, the direction of the camera motion has no influence on the estimation convergence: the convergence rate is only dictated by the norm of $\boldsymbol{v}$. However, for the sake of imposing a desired *transient response* to the estimation error, one still needs to properly shape the damping gain $\boldsymbol{H}$ as a function of σ_1^2 as explained in Sec. II.

B. Cylindrical target

A cylinder $\mathcal{O}_c$ can be described by its radius $R > 0$ and by its main axis $\boldsymbol{a} \in \mathbb{S}^2$ passing through a 3D point $\boldsymbol{P}_0 = (X_0,\, Y_0,\, Z_0)$, with $\|\boldsymbol{a}\| = 1$ and, w.l.o.g., $\boldsymbol{a}^T\boldsymbol{P}_0 = 0$ ($\boldsymbol{P}_0$ can be chosen as the closest point on $\boldsymbol{a}$ to the origin of the camera frame). Moreover, analogously to the sphere, a cylinder is also associated with a planar limb surface $\mathcal{L}$ such that (7) holds for any point on $\mathcal{L}$ with projection $\boldsymbol{p} = (x,\, y,\, 1)$. Therefore, as in the sphere case, a possibility is to estimate the three unknown parameters of the limb plane $\mathcal{L}$

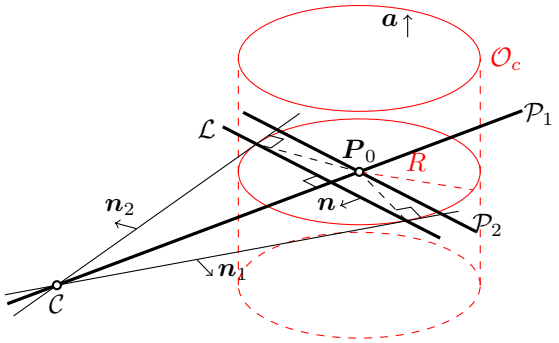


Fig. 2: Camera $\mathcal{C}$ and cylindrical target $\mathcal{O}_c$ with the planar limb surface $\mathcal{L}$ and the other planes of interest $\mathcal{P}_1$ and $\mathcal{P}_2$

(vector $\boldsymbol{\chi}$) by exploiting (at least) three image measurements, see [15] for a solution in this sense. However, we now propose a *novel representation* of the cylinder 3D geometry which, again, allows to reduce the estimation task of its parameters $(\boldsymbol{P}_0, \boldsymbol{a}, R)$ to the sole *unknown but constant* cylinder radius R.

Let (ρ_1, θ_1) and (ρ_2, θ_2) be the (*measured*) distance/angle parameters of the two straight lines resulting from the projection of the cylinder on the image plane. We define $\boldsymbol{n}_1 = (\cos\theta_1, \sin\theta_1, -\rho_1)$, $\boldsymbol{n}_2 = (\cos\theta_2, \sin\theta_2, -\rho_2)$ as the normal vectors to the two planes passing through the origin of the camera frame and the two above-mentioned projected lines[1]. Figure 2 gives a graphical representation of the quantities of interest. Note that vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ can be directly evaluated from image measurements (the line parameters). From [22] an equivalent expression for vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ can be obtained as

$$\boldsymbol{n}_1 = \frac{1}{N_1}\left(\frac{R\boldsymbol{P}_0}{\sqrt{K}} - \boldsymbol{b}\right), \quad \boldsymbol{n}_2 = \frac{1}{N_2}\left(\frac{R\boldsymbol{P}_0}{\sqrt{K}} + \boldsymbol{b}\right) \tag{13}$$

with

$$\begin{cases} K &= \boldsymbol{P}_0^T\boldsymbol{P}_0 - R^2 \\ \boldsymbol{b} &= (\alpha, \beta, \gamma) = [\boldsymbol{P}_0]_\times \boldsymbol{a} \\ N_1 &= \sqrt{\left(R\frac{X_0}{\sqrt{K}} - \alpha\right)^2 + \left(R\frac{Y_0}{\sqrt{K}} - \beta\right)^2} \\ N_2 &= \sqrt{\left(R\frac{X_0}{\sqrt{K}} + \alpha\right)^2 + \left(R\frac{Y_0}{\sqrt{K}} + \beta\right)^2} \end{cases}, \tag{14}$$

thus yielding

$$\frac{\boldsymbol{n}_1}{\|\boldsymbol{n}_1\|} = \frac{R\boldsymbol{P}_0 - \sqrt{K}\boldsymbol{b}}{\boldsymbol{P}_0^T\boldsymbol{P}_0}, \quad \frac{\boldsymbol{n}_2}{\|\boldsymbol{n}_2\|} = \frac{R\boldsymbol{P}_0 + \sqrt{K}\boldsymbol{b}}{\boldsymbol{P}_0^T\boldsymbol{P}_0} \tag{15}$$

By now defining

$$\boldsymbol{s} = \frac{\boldsymbol{P}_0}{R} \in \mathbb{R}^3, \tag{16}$$

from (15) one has

$$\boldsymbol{\Delta} = \frac{1}{2}\left(\frac{\boldsymbol{n}_1}{\|\boldsymbol{n}_1\|} + \frac{\boldsymbol{n}_2}{\|\boldsymbol{n}_2\|}\right) = \frac{R^2}{\boldsymbol{P}_0^T\boldsymbol{P}_0}\boldsymbol{s}$$

[1]The two planes are therefore tangent to the surface of the cylinder.

which can be exploited to obtain $\boldsymbol{s} = \boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|^2$. This then shows how $\boldsymbol{s}$, defined as in (16), can be directly evaluated in terms of only image measurements, being $\boldsymbol{\Delta}$ a function of only measurable quantities (vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$).

A similar result can also be obtained for the cylinder axis $\boldsymbol{a}$: indeed, exploiting (13) it is

$$[\boldsymbol{n}_2]_\times \boldsymbol{n}_1 = \frac{2R}{N_1N_2\sqrt{K}}[\boldsymbol{b}]_\times \boldsymbol{P}_0 = \frac{2R\boldsymbol{P}_0^T\boldsymbol{P}_0}{N_1N_2\sqrt{K}}\boldsymbol{a} \tag{17}$$

where in the last step the property $\boldsymbol{a}^T\boldsymbol{P}_0 = 0$ was used. Since $\|\boldsymbol{a}\| = 1$, from (17) one then has

$$\boldsymbol{a} = \frac{[\boldsymbol{n}_2]_\times \boldsymbol{n}_1}{\left\|[\boldsymbol{n}_2]_\times \boldsymbol{n}_1\right\|}, \tag{18}$$

that is, again, an expression in terms of only measured quantities. Therefore, the only unknown left is the cylinder radius R: once known, the 3D cylinder parameters can be fully recovered as $\boldsymbol{P}_0 = R\boldsymbol{s}$ and $\boldsymbol{a}$ from (18).

In order to estimate R from image measurements, we seek an expression for $\dot{\boldsymbol{s}}$ in terms of the measured $(\boldsymbol{s}, \boldsymbol{a})$, of the camera linear/angular velocity $(\boldsymbol{v}, \boldsymbol{\omega})$, and of the unknown R. As shown in the Appendix, it is possible to show that

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\frac{1}{R}\left(\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^T\right) & [\boldsymbol{s}]_\times \end{bmatrix} \boldsymbol{u}. \tag{19}$$

Note the similarity of (19) with (8) for the sphere case.

Being (19) linear in $1/R$, we can then apply observer (2) by choosing $\chi = 1/R$ with $m = 3$ and $p = 1$, and having

$$\begin{cases} \dot{\boldsymbol{s}} = [\boldsymbol{s}]_\times \boldsymbol{\omega} + \left(\boldsymbol{a}\boldsymbol{a}^T - \boldsymbol{I}\right)\boldsymbol{v}\chi \\ \dot{\chi} = 0 \end{cases}. \tag{20}$$

Note how, again, being $\dot{\chi} = 0$ it is $g(\boldsymbol{e}, t) = 0$ (global convergence for the error system (3) as in the sphere case).

Matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ reduces to its single eigenvalue

$$\sigma_1^2 = \boldsymbol{\Omega}\boldsymbol{\Omega}^T = \|\boldsymbol{v}\|^2 - (\boldsymbol{a}^T\boldsymbol{v})^2. \tag{21}$$

It is worth comparing (21) with the result obtained for the sphere ($\sigma_1^2 = \|\boldsymbol{v}\|^2$). In the cylinder case, the convergence rate of the estimation scheme is affected by both the *norm* and the *direction* of the linear velocity $\boldsymbol{v}$. In particular, for a given $\|\boldsymbol{v}\| = const$, the maximum value for σ_1^2 is obtained when $\boldsymbol{v}$ has a null component along the cylinder axis $\boldsymbol{a}$, i.e., when $\boldsymbol{a}^T\boldsymbol{v} = 0$ with, in this case, $\sigma_1^2 = \sigma_{1,max}^2 = \|\boldsymbol{v}\|^2$. Furthermore, as with the sphere, keeping a $\sigma_1^2(t) = const$ allows to exactly enforce the ideal estimation error dynamics (5), see Remark 2.2.

Finally, from (21) one has

$$\dot{(\sigma_1^2)} = \boldsymbol{J}_{v,1}\dot{\boldsymbol{v}} + \boldsymbol{J}_{a,1}\dot{\boldsymbol{a}} = \boldsymbol{J}_{v,1}\dot{\boldsymbol{v}} + \boldsymbol{J}_{a,1}[\boldsymbol{a}]_\times\boldsymbol{\omega} \tag{22}$$

with $\boldsymbol{J}_{v,1} = 2\boldsymbol{v}^T\left(\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^T\right)$ and $\boldsymbol{J}_{a,1} = 2\boldsymbol{v}^T\boldsymbol{a}\boldsymbol{v}^T$. Although (22) also depends on the angular velocity $\boldsymbol{\omega}$, it is possible to fully compensate for the effects of $\boldsymbol{J}_{a,1}[\boldsymbol{a}]_\times\boldsymbol{\omega}$ (a known quantity) when inverting (22) w.r.t. $\dot{\boldsymbol{v}}$ as discussed in Sect. IV-B. Therefore, one can act on $\dot{\boldsymbol{v}}$ to regulate the value of $\sigma_1^2(t)$ and, at the same time and in a decoupled way, exploit the camera angular velocity $\boldsymbol{\omega}$ for implementing additional tasks of interest such as keeping the cylinder axis $\boldsymbol{a}$ at the center of the image plane by enforcing $(s_x, s_y) = \boldsymbol{0}$.

IV. EXPERIMENTAL RESULTS

In this section we show some experimental results meant to validate the theoretical developments of the previous sections. The experiments were run by employing a greyscale camera with a resolution of 640×480 px and a framerate of 30 fps. The camera was mounted on the end-effector of a 6-dofs Gantry robot commanded in velocity at a frequency of 100 Hz. All the image processing and feature tracking were implemented via the open-source ViSP library [23]. As objects to be tracked, for the spherical and cylindrical targets we made use of a white table-tennis ball and of a white cardboard cylinder with radius of 1.9 cm and 4.2 cm, respectively. A video of the experiments is attached to the paper.

A. Experiments with a Spherical Target

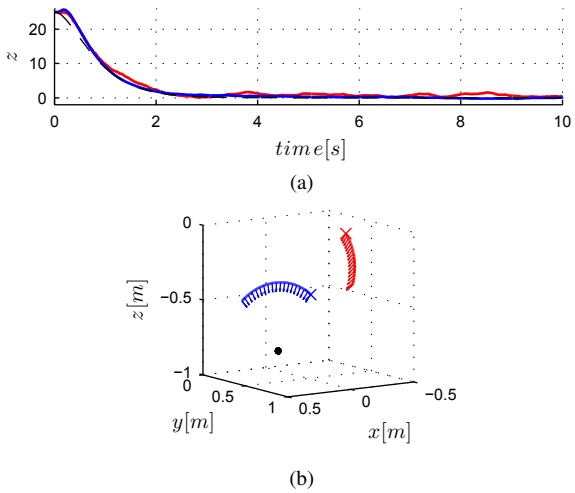
As explained in Sec. III-A, the convergence rate of the estimation error for the sphere case only depends on the norm of the linear velocity $\|\boldsymbol{v}\|$ and not on its direction. This fact is proven by the first experiment where the estimation task is run twice starting from two different positions and imposing two different camera velocities but with same norm. These values were used during the experiments: $\alpha = 2 \cdot 10^3$, $c_1 = c_1^* = 2\sqrt{\alpha}\sigma_1$ in (5) and $\boldsymbol{v} = (-0.05,\, 0,\, 0)$ for case I and $\boldsymbol{v} = (0,\, 0.045,\, 0.02)$ for case II, with $\|\boldsymbol{v}\| = 0.05$ in both cases. The camera angular velocity $\boldsymbol{\omega}$ was exploited to keep $(s_x,\, s_y) \simeq (0,\, 0)$ (centered sphere).

Figure 3a shows the behavior of the estimation errors (solid blue and red lines): note how the error transient response for the two cases is essentially coincident, and also equivalent to that of the *reference second order system* (5) with the desired poles, i.e., by setting $\sigma_1^2 = \|\boldsymbol{v}\|^2 = const$ and $c_1 = c_1^*$ in (5) (dashed black line). The higher noise level in case II (red line) is due to the larger distance between the camera and the spherical target (see Fig. 3b) which negatively affects the image processing.

Since the direction of the velocity does not play any role in this case, no optimization of σ_1^2 can be performed under the constraint $\|\boldsymbol{v}\| = const$. On the other hand, the analysis of Sec. II clearly indicates the importance of choosing a proper value of c_1 in (5). To show this fact, we report here three experiments characterized by the same camera trajectory of the previous case I, but by employing three different values for c_1, that is, c_1^*, $2c_1^*$ and $0.5c_1^*$. These correspond to a critically damped, overdamped and underdamped response for the ideal system (5), respectively. The experimental results reported in Fig. 4 show that the behavior of the estimation error z (solid lines) has an excellent match with that of (5) (represented by dashed lines), thus fully confirming (i) the validity of the proposed theoretical analysis, and (ii) the importance of choosing the 'right' damping matrix $\boldsymbol{H}$ for optimizing the convergence speed in addition to a proper regulation of σ_1^2.

B. Experiments with a Cylindrical Target

In the cylinder case, the convergence rate of the estimation error depends both on the norm of the camera linear velocity



(a)

(b)

Fig. 3: Experimental results for the estimation of the radius of a sphere using different constant camera velocities with the same norm. (a): behavior of the estimation error $z(t)$ for the two cases (solid blue and red lines), and for an 'ideal' second order system with poles at the desired locations (dashed black line). Note the almost perfect match between the three plots. (b): camera trajectories for case I (blue line) and case II (red line) with arrows indicating the direction of the camera optical axis. Note that both trajectories are circular as expected, but (i) with a different motion direction and (ii) starting from a different initial position (larger distance in case II).

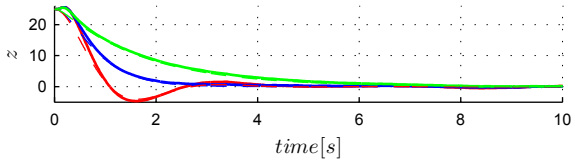


Fig. 4: Experimental results for the estimation of the radius of a sphere with $c_1 = c_1^*$ (blue line), $c_1 = 2c_1^*$ (green line) and $c_1 = 0.5c_1^*$ (red line). The dashed lines represent the response of an 'ideal' second order system with the corresponding poles. Note again the almost perfect match between the plots.

$\boldsymbol{v}$ and on its direction w.r.t. the cylinder axis $\boldsymbol{a}$, see (21). It is then interesting to optimize the direction of $\boldsymbol{v}$ under the constraint $\|\boldsymbol{v}\| = const$ for maximizing the eigenvalue σ_1^2 (i.e., so as to obtain the fastest convergence rate for a given 'control effort' $\|\boldsymbol{v}\|$).

From (22), maximization of $\sigma_1^2(t)$ w.r.t. vector $\boldsymbol{v}$ can be obtained by choosing

$$\dot{\boldsymbol{v}} = \boldsymbol{J}_{v,1}^T - \boldsymbol{J}_{v,1}^\dagger \boldsymbol{J}_{a,1} [\boldsymbol{a}]_\times \boldsymbol{\omega}, \tag{23}$$

with $\boldsymbol{A}^\dagger$ being the pseudoinverse of matrix $\boldsymbol{A}$, i.e., by following the gradient of σ_1^2 w.r.t. $\boldsymbol{v}$ and by compensating for the (known) effects of input $\boldsymbol{\omega}$. In order to additionally enforce the constraint $\|\boldsymbol{v}\| = const$ during the eigenvalue maximization, we let $\kappa = \frac{1}{2}\|\boldsymbol{v}\|^2$, $\kappa_{des} = \frac{1}{2}\|\boldsymbol{v}_0\|^2$ and modify (23) as

$$\dot{\boldsymbol{v}} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|^2} k_1 (\kappa_{des} - \kappa) + k_2 \left(\boldsymbol{I} - \frac{\boldsymbol{v}\boldsymbol{v}^T}{\|\boldsymbol{v}\|^2} \right) (\boldsymbol{J}_{v,1}^T - \boldsymbol{J}_{v,1}^\dagger \boldsymbol{J}_{a,1} [\boldsymbol{a}]_\times \boldsymbol{\omega}), \tag{24}$$

with $k_1 > 0$ and $k_2 > 0$. The first term in (24) asymptotically guarantees $\|\boldsymbol{v}(t)\|^2 = \|\boldsymbol{v}_0\|^2$ while the second term projects (23) onto the null-space of the constraint $\|\boldsymbol{v}\| = const$. As for the angular velocity $\boldsymbol{\omega}$, we exploited it for keeping the axis of the cylinder at the center of the image plane by regulating $(s_x,\, s_y)$ to $(0,\, 0)$.

We now present three experimental results for the cylinder case structured as follows: in the first experiment (case I), the update rule (24) is fully implemented ($k_1 > 0$, $k_2 > 0$) for actively optimizing the direction of $\boldsymbol{v}$. In the second experiment (case II), the camera starts from the same initial pose and velocity as in case I, but (24) is implemented with $k_1 > 0$ and $k_2 = 0$, i.e., without performing any optimization of σ_1^2. Finally, in the third experiment (case III), the camera starts from a different initial pose and with a different velocity direction (but same norm) w.r.t. the previous two cases, and (24) is again fully implemented. This last case is meant to show how the convergence properties of the estimator are not affected by the camera starting position but only by the norm and direction of $\boldsymbol{v}$ w.r.t. $\boldsymbol{a}$.

The experiments were run with the following conditions: $\alpha = 500$, $c_1 = c_1^*$, $k_1 = 10$, $k_2 = 1$ for cases I and III, and $k_2 = 0$ for case II. As for the linear velocity, we set $\boldsymbol{v}(t_0) = \boldsymbol{v}_0 = (-0.01, 0.05, 0.05)$ for cases I and II, and $\boldsymbol{v}(t_0) = \boldsymbol{v}_0 = (-0.05, 0.05, 0.01)$ for case III (note how $\|\boldsymbol{v}_0\| = 0.0714$ in all three cases).

We start showing in Fig. 5a the behavior of $\|\boldsymbol{v}(t)\|^2$ for the three experiments with the following color coding shared by all the next plots: blue – case I, red – case II, green – case III. One can note how, in all three cases, the same $\|\boldsymbol{v}(t)\|^2 = \|\boldsymbol{v}_0\|^2 = 5.1 \times 10^{-3}$ is attained thanks to the action of the first term in (24). The behavior of $\sigma_1^2(t)$ is shown in Fig. 5b: as explained at the end of Sect. III-B, under the constraint $\|\boldsymbol{v}\| = const$ one has $\max_{\boldsymbol{v}} \sigma_1^2 = \|\boldsymbol{v}\|^2$ as the largest possible value for σ_1^2 (obtained when $\boldsymbol{v}^T \boldsymbol{a} = 0$). By comparing Fig. 5b with Fig. 5a, it is then possible to verify that, indeed, $\sigma_1^2(t) \to \|\boldsymbol{v}_0\|^2$ in cases I and III despite the different initial conditions of the experiments (different camera pose and direction of $\boldsymbol{v}$). The optimization in (24) results in a null component of $\boldsymbol{v}$ along $\boldsymbol{a}$, thus allowing to move faster in the 'useful' directions (while keeping a constant $\|\boldsymbol{v}\|$), and to increase the value of σ_1^2 to its maximum possible value. Also, note how the value of $\sigma_1^2(t)$ for case II results smaller than in the other two cases (as expected) since no optimization is present in this case.

The behavior of the estimation error $z(t)$ is shown in Fig. 5c: again, we can note that the transient response for cases I and III results essentially coincident and in almost perfect agreement with that of the reference system (5) with desired poles (dashed black line). As expected, the response for case II (red line) is slower than in cases I and III, but still in agreement with the corresponding response of system (5) (dashed black line). Finally, Fig. 5d depicts the camera trajectories for the three experiments with arrows indicating the direction of the optical axis. In case II the camera simply travels along a straight line ($\boldsymbol{v}(t) \equiv \boldsymbol{v}_0$), while in cases I and III the direction of $\boldsymbol{v}$ is suitably modified resulting in a trajectory lying on a plane orthogonal to $\boldsymbol{a}$.

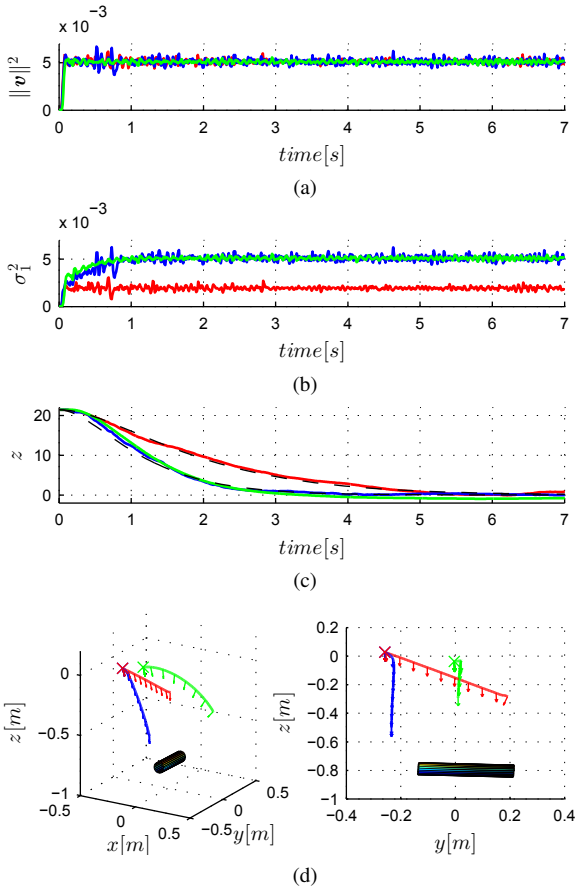


Fig. 5: Experimental results for the estimation of the radius of a cylinder with the following color coding: blue – case I, red – case II, green – case III. (a): behavior of $\|\boldsymbol{v}(t)\|^2$ over time. Note how the same *constant* velocity norm is reached and then kept in all cases. (b): behavior of $\sigma_1^2(t)$ for the three cases (coincident for cases I and III and larger than in case II). (c): behavior of $z(t)$. As expected, the convergence rate of the estimation error of cases I and III is faster than in case II. Note also how all the transient responses are in almost perfect agreement with the corresponding response of the reference system (5) with desired poles (dashed black lines). (d): two views of the camera trajectories for the three cases with arrows indicating the direction of the camera optical axis.

V. CONCLUSIONS

In this paper we have addressed the problem of active SfM for recovering the 3D structure of a spherical and a cylindrical object by exploiting a recently proposed active estimation strategy tailored to the two cases under consideration. As an additional contribution, we have also proposed a novel *minimal parameterization* of the sphere/cylinder 3D geometry able to reduce the estimation task to a single unknown constant quantity (the radius R) in place of the classical (and time-varying) three parameters (scaled normal vector of the sphere/cylinder planar limb surface). The reported experimental results fully confirmed the validity of the analysis and the ability to *actively* impose a desired transient behavior to the (nonlinear) estimation schemes, in particular

matching that of a reference linear second-order system with desired poles.

We are currently investigating the use of similar active strategies when dealing with more complex 3D scenes (e.g., planar images or other 3D primitives), and are also aiming at exploiting these techniques in the context of visual-based control of robot manipulators (see [18] for a first attempt in this direction).

APPENDIX

We note that the cylinder axis $\boldsymbol{a}$ can be determined by the intersection of two planes $\mathcal{P}_i : \boldsymbol{r}_i^T \boldsymbol{E} - d_i = 0$, $i = 1, 2$, with

$$\boldsymbol{r}_1 = \frac{[\boldsymbol{a}]_\times \boldsymbol{P}_0}{\|\boldsymbol{P}_0\|},\ d_1 = 0, \quad \boldsymbol{r}_2 = -\frac{\boldsymbol{P}_0}{\|\boldsymbol{P}_0\|},\ d_2 = \|\boldsymbol{P}_0\|, \tag{25}$$

see Fig. 2. In particular, plane $\mathcal{P}_1$ passes through the camera optical center, it is orthogonal to plane $\mathcal{P}_2$, and both planes contain the axis $\boldsymbol{a}$ passing through $\boldsymbol{P}_0$ (by construction).

Since $R\boldsymbol{s} = \boldsymbol{P}_0$ and $\boldsymbol{P}_0$ belongs to the cylinder axis $\boldsymbol{a}$, it is $R\boldsymbol{r}_i^T \boldsymbol{s} - d_i = 0$, $i = 1, 2$ (the point $R\boldsymbol{s}$ belongs to both planes $\mathcal{P}_i$). Taking the time derivative of these latter constraints (with $R = const$), one has

$$\boldsymbol{r}_i^T \dot{\boldsymbol{s}} = \frac{1}{R}\dot{d}_i - \boldsymbol{s}^T \dot{\boldsymbol{r}}_i, \quad i = 1, 2. \tag{26}$$

Since $\dot{\boldsymbol{r}}_i = [\boldsymbol{r}_i]_\times \boldsymbol{\omega}$ and $\dot{d}_i = \boldsymbol{r}_i^T \boldsymbol{v}$ (see [9]), eq. (26) can be rewritten as

$$\boldsymbol{r}_i^T \dot{\boldsymbol{s}} = \frac{1}{R}\boldsymbol{r}_i^T \boldsymbol{v} - \boldsymbol{s}^T [\boldsymbol{r}_i]_\times \boldsymbol{\omega}, \quad i = 1, 2. \tag{27}$$

Finally, from $\boldsymbol{a}^T \boldsymbol{P}_0 = 0$ and $\boldsymbol{P}_0 = R\boldsymbol{s}$ it is $\boldsymbol{a}^T \boldsymbol{s} = 0$ implying that (taking a time derivative)

$$\boldsymbol{a}^T \dot{\boldsymbol{s}} = -\boldsymbol{s}^T \dot{\boldsymbol{a}} = -\boldsymbol{s}^T [\boldsymbol{a}]_\times \boldsymbol{\omega}. \tag{28}$$

We now note that equations (27–28) provide three linear constraints for $\dot{\boldsymbol{s}}$ which, by using (25), can be rearranged in matrix form as the following linear system

$$\begin{bmatrix} \dfrac{\boldsymbol{P}_0^T}{\|\boldsymbol{P}_0\|} \\ \boldsymbol{a}^T \\ \dfrac{\left([\boldsymbol{a}]_\times \boldsymbol{P}_0\right)^T}{\|\boldsymbol{P}_0\|} \end{bmatrix} \dot{\boldsymbol{s}} = \frac{1}{R} \begin{bmatrix} \dfrac{\boldsymbol{P}_0^T}{\|\boldsymbol{P}_0\|}\boldsymbol{v} \\ -\boldsymbol{P}_0^T [\boldsymbol{a}]_\times \boldsymbol{\omega} \\ \|\boldsymbol{P}_0\| \boldsymbol{a}^T \boldsymbol{\omega} + \dfrac{\left([\boldsymbol{a}]_\times \boldsymbol{P}_0\right)^T}{\|\boldsymbol{P}_0\|}\boldsymbol{v} \end{bmatrix}. \tag{29}$$

It is easy to verify that the 3×3 matrix on the lhs of (29) is orthonormal: by then solving (29) for $\dot{\boldsymbol{s}}$ and performing some simplifications we finally obtain the sought result

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\dfrac{1}{R}\left(\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^T\right) & [\boldsymbol{s}]_\times \end{bmatrix} \boldsymbol{u}. \tag{30}$$

REFERENCES

[1] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Journal Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[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] P. Robuffo Giordano, A. De Luca, and G. Oriolo, "3D structure identification from image moments," in *2008 IEEE Int. Conf. on Robotics and Automation*, Pasadena, CA, may 2008, pp. 93–100.

[10] 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.

[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. Durand Petiteville, M. Courdesses, and V. Cadenat, "A New Predictor/Corrector Pair to Estimate the Visual Features Depth during a Vision-based Navigation Task in an Unknown Environment," in *Inte. Conf. on Informatics in Control, Automation and Robotics*, 2010, pp. 268–274.

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

[14] B. Guerreiro, P. Batista, C. Silvestre, and P. Oliveira, "Globally Asymptotically Stable Sensor-Based Simultaneous Localization and Mapping," *IEEE Trans. on Robotics*, vol. 29, no. 6, pp. 1380–1395, Dec 2013.

[15] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin, "Structure from controlled motion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 492–504, 1996.

[16] A. J. Davison and D. W. Murray, "Mobile Robot Localisation Using Active Vision," in *1998 European Conf. on Computer Vision*, 1998, pp. 809–825.

[17] R. Spica and P. Robuffo Giordano, "A Framework for Active Estimation: Application to Structure from Motion," in *52nd IEEE Conf. on Decision and Control*, 2013.

[18] R. Spica, P. Robuffo Giordano, and F. Chaumette, "Coupling Image-Based Visual Servoing with Active Structure from Motion," in *2014 IEEE Int. Conf. on Robotics and Automation*, Hong Kong, China, May 2014.

[19] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.

[20] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *IEEE Trans. on Robotics*, vol. 20, no. 4, pp. 713–723, 2004.

[21] R. T. Fomena and F. Chaumette, "Improvements on visual servoing from spherical targets using a spherical projection model," *IEEE Trans. on Robotics*, vol. 25, no. 4, pp. 874–886, 2009.

[22] F. Chaumette, "Visual servoing using image features defined upon geometrical primitives," in *1994 IEEE Conf. on Decision and Control*, 1994, pp. 3782–3787.

[23] 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, 2005.