Active Structure from Motion: Application to Point, Sphere and Cylinder

Riccardo Spica, Paolo Robuffo Giordano, and François Chaumette

***Abstract*—In this paper, we illustrate the application of a nonlinear *active* structure estimation from motion (SfM) strategy to three problems, namely, 3D structure estimation for (*i*) a point, (*ii*) a sphere and, (*iii*) a cylinder. In all three cases, an appropriate parametrization reduces the problem to the estimation of a single quantity. Knowledge of this estimated quantity and of the available measurements allows for then retrieving the full 3D structure of the observed objects. Furthermore, in the point feature case, two different parametrizations based on either a planar or a spherical projection model are critically compared: indeed, the two models yield, somehow unexpectedly, to different convergence properties for the SfM estimation task. The reported simulative and experimental results fully support the theoretical analysis and clearly show the benefits of the proposed active estimation strategy, which is in particular able to impose a *desired transient response* to the estimation error equivalent to that of a reference linear second-order system with assigned poles.**

I. INTRODUCTION

THE problem of Structure from Motion (SfM), i.e., how to recover the missing structure of the observed scene from images taken by a moving camera, is a very classical and well-studied topic in computer and robot vision. One solution is to rely on prior knowledge of the scene as, e.g., known size of a tracked object. Alternatively, one can exploit the possibility of observing the same scene from different points of view, and fuse together the acquired images with the known camera displacement among them. When processing consecutive images over time, a possibility is to treat SfM as a recursive/filtering task: 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* 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].

Within the first class of (recursive) methods, a vast literature exists for addressing SfM: for instance, as a non-exhaustive list, Extended Kalman Filter (EKF)-based solutions have been proposed in [2]–[5], and, along similar lines, an Unscented Kalman Filter was exploited in [6]. All these strategies have the considerable advantage of being 'aware', to some extent, of the measurement and process noise (when modeled as Gaussian distributions). On the other hand, they require a certain level of approximation of the system dynamics which may affect the estimation performance. Other approaches exploiting tools from (deterministic) nonlinear observation can instead be found in [7]–[14] and references therein, while [15] has tackled the more challenging problem of structure *and* motion estimation, i.e., how to simultaneously recover the missing structure along with the (partially) unknown camera velocity. This class of methods does not typically involve any linearization of the system dynamics and allows for some formal characterization of the estimation error convergence. However, presence of noise is not explicitly taken into account, with the filter design being developed in a fully deterministic setting. 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 [16].

While all these works study the general issue of structure estimation from motion in different contexts, much less attention has been devoted to the problem of *actively imposing* a desired (e.g., optimized) convergence behavior to a SfM estimation task by acting on the *motion imposed to the camera* and on the *employed estimation gains*. For instance, in [17] 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 [18], 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). In [19] an EKF-based SfM estimation scheme for a UAV is integrated with a path planning strategy aiming at minimizing the covariance matrix of the estimated states at the end of the motion. Nevertheless, one needs to assume full pre-knowledge of the surrounding environment (e.g., obstacles) so as to numerically propagate the EKF filter along all the edges of a randomly constructed roadmap (the method is, thus, only amenable for an off-line/planning use).

With respect to this state-of-the-art, this paper then tackles the problem of designing an *online* and *active* algorithm for structure from *known* and *controlled* motion, i.e., assuming that the camera velocity can be measured and actively modified by the robot actuators (as it is often the case in robotic applications). The active component of the scheme makes it possible to impose an estimation error transient response equivalent to that of a reference linear second-order system with *desired* poles. The developments build upon the theoretical framework presented in [20]: in a nonlinear context, the observability properties of the states under consideration are not (in general) time-invariant but may depend on the current state and on the

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

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

current inputs applied to the system. It is then possible to simultaneously act on the estimation gains and system inputs (i.e., the camera velocity for SfM) in order to optimize the observation process.

The methodology proposed in [20] can be applied to all those systems in which an *invertible function* of the unknown states can appear linearly in the system dynamics, as it is indeed the case for SfM problems. We then exploit this fact and propose three concrete active SfM applications: 3D structure estimation for (*i*) a point feature, (*ii*) a spherical target and, (*iii*) a cylindrical target. The estimation of the depth of a point feature has already been well studied in the past literature (see, e.g., [2], [7], [9], [11], [21], [22]) although by never considering the *active* perspective taken in this work. On the other hand, the machinery proposed for the spherical and cylindrical objects represents a novel contribution also in terms of the chosen parameterization. Indeed, we show that a suitable transformation of the observed features allows to express the 3D sphere/cylinder structures in terms of image measurements and of only one unknown *constant* parameter (the sphere/cylinder radius) rather than the classical (and time-varying) scaled orientation of the limb surface in the camera frame. This, of course, significantly simplifies the SfM task.

We conclude by highlighting that the ability of both characterizing and optimizing the transient response of the estimation error brings several added values compared to more classical inactive estimation strategies: for instance, it allows obtaining 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. 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 point features or specific 3D geometrical primitives). Finally, we note the many similarities between the SfM approach adopted in this work and the notion of "sensor-based" or "ego-centric" Visual SLAM, see, e.g., [23] for a recent overview. In both cases, a robot/camera builds a 3D model of the environment in its own body/sensor frame via a filtering technique: an EKF in [23] and similar works, and the deterministic filter (but with a fully *characterized* and *actively optimizable* transient response) derived from [20] in this paper.

The rest of the paper is organized as follows: Sect. II reviews the SfM problem in the context of nonlinear state observation and briefly summarizes the methodology developed in [20] for actively imposing a desired transient behavior to the estimation error. Section III then focuses on the three SfM problems considered in this work. Subsequently, Sect. IV reports the corresponding simulative and experimental results obtained with a manipulator equipped with an eye-in-hand camera. Finally, Sect. V concludes the paper and discusses some future directions.

II. PRELIMINARIES

In this section, we briefly summarize the active estimation framework originally proposed in [20]. This is then applied to SfM case studies discussed in the next Sect. III.

A. A nonlinear observation scheme

Let $(\boldsymbol{s}\,, \boldsymbol{\chi}) \in \mathbb{R}^{m+p}$ be the state of a dynamical system in the form

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

where $\boldsymbol{s} \in \mathbb{R}^m$ and $\boldsymbol{\chi} \in \mathbb{R}^p$ represent, respectively, a *measurable* and *unmeasurable* component of $\boldsymbol{x}$, and $\boldsymbol{u} \in \mathbb{R}^v$ is the system input vector. In formulation (1) vector $\boldsymbol{\chi}$ is required to appear *linearly* in the dynamics of $\boldsymbol{s}$ (first equation). Furthermore, matrix $\boldsymbol{\Omega}(t) \in \mathbb{R}^{p\times m}$ and vectors $\boldsymbol{f}_m(\cdot) \in \mathbb{R}^m$ and $\boldsymbol{f}_u(\cdot) \in \mathbb{R}^p$ are assumed to be generic but *known* and sufficiently smooth functions w.r.t. their arguments which are all available apart from the unknown value of $\boldsymbol{\chi}$ in $\boldsymbol{f}_u(\cdot)$.

SfM problems can be recast to formulation (1) by taking $\boldsymbol{s}$ as a set of visual features measured in the image, $\boldsymbol{u} = (\boldsymbol{v},\, \boldsymbol{\omega})$ as the camera linear/angular velocity in camera frame, and $\boldsymbol{\chi}$ as a suitable (and locally invertible) function of the unknown structure of the scene to be estimated. For instance, in the point feature case, $\boldsymbol{\chi}$ can be taken as the *inverse* of the feature depth [9], and, for image moments of planar scenes, $\boldsymbol{\chi}$ can be taken as the normal vector of the observed plane scaled by its distance from the camera optical center [10]. Furthermore, in SfM one has $\boldsymbol{\Omega}(t) = \boldsymbol{\Omega}(\boldsymbol{s}(t),\, \boldsymbol{v}(t))$ with, in particular, $\boldsymbol{\Omega}(\boldsymbol{s},\, \boldsymbol{0}) \equiv \boldsymbol{0}$: the camera linear velocity $\boldsymbol{v}(t)$ plays a key role for the resolution of SfM problems[1].

For a system in form (1), a possible estimation scheme can be devised as follows [9], [20]: let $(\hat{\boldsymbol{s}}\,, \hat{\boldsymbol{\chi}}) \in \mathbb{R}^{m+p}$ be the estimated state, $\boldsymbol{\xi} = \boldsymbol{s} - \hat{\boldsymbol{s}}$, $\boldsymbol{z} = \boldsymbol{\chi} - \hat{\boldsymbol{\chi}}$, $\boldsymbol{e} = (\boldsymbol{\xi},\, \boldsymbol{z})$ be the total error vector, and consider the following *observer*

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

where $\boldsymbol{H} > 0$, $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^T > 0$ and $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$ are positive definite gain matrices. Note that observer (2) is function of only measured/known quantities, with in particular a feedback action on the measurable error component $\boldsymbol{\xi}$. The corresponding estimation error dynamics is then given by

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

with $\boldsymbol{g}(\boldsymbol{e},\, t)$ being a 'perturbation term' vanishing w.r.t. the error vector $\boldsymbol{e}$, i.e., such that $\boldsymbol{g}(\boldsymbol{0},\, t) = \boldsymbol{0}$, $\forall t$. The origin of (3) can be proven to be locally exponentially stable if and only if (iff) the following *Persistency of Excitation* (PE) condition holds

$$\int_t^{t+T} \boldsymbol{\Omega}(\tau)\,\boldsymbol{\Omega}^T(\tau)\,d\tau \geq \gamma \boldsymbol{I}_p > 0, \quad \forall t \geq t_0, \tag{4}$$

[1]This is due to the well-known fact that, under perspective and spherical projection, the motion in the image induced by pure rotations of the camera (i.e., when $\boldsymbol{v} = \boldsymbol{0}$) does not depend on the structure of the scene.

for some $T > 0$ and $\gamma > 0$, with $\boldsymbol{I}_n$ representing the $n \times n$ identity matrix[2].

Remark II.1. *We note that the* local *stability properties of the error dynamics (3) are due to the perturbation term $\boldsymbol{g}(\boldsymbol{e}, t)$ which affects an otherwise globally exponentially stable error system. Indeed, in the special case $\dot{\boldsymbol{\chi}} = \boldsymbol{0}$ (unknown but* constant *parameters), one has $\boldsymbol{g}(\boldsymbol{e}, t) \equiv \boldsymbol{0}$ and* global *exponential convergence for the error system (3). This is, for instance, the case of the structure estimation problems for spherical and cylindrical objects considered in Sects. III-B and III-C. We stress, however, that the estimation scheme (2) is not restricted to this particular situation but can be applied (with, in this case, only local convergence guarantees) to the more general case of* state observation *problems in which the unknown $\boldsymbol{\chi}$ is subject to a non-negligible dynamics as in (1). The depth estimation for a point feature discussed in the following Sect. III-A falls in this class.*

The PE condition (4) plays the role of an *observability* criterium: convergence of the estimation error $\boldsymbol{e}(t) \to \boldsymbol{0}$ is possible iff the square matrix $\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t) \in \mathbb{R}^{p\times p}$ remains full rank in the integral sense of (4). We note that if $m \geq p$, that is, if the number of independent measurements $\boldsymbol{s}$ is larger or equal to the number of estimated quantities $\boldsymbol{\chi}$, then it is possible to *instantaneously* satisfy (4) by enforcing

$$\boldsymbol{\Omega}(t)\,\boldsymbol{\Omega}^T(t) \geq \frac{\gamma}{T}\boldsymbol{I}_p, \quad \forall t. \tag{5}$$

In the rest of this work we will only consider this (more restrictive) observability condition.

B. An active estimation strategy

As clear from (4) and (5) (some measure of) the norm of matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ determines the convergence properties of the error system (3). Furthermore, since in the SfM case it is $\boldsymbol{\Omega}(t) = \boldsymbol{\Omega}(\boldsymbol{s}(t), \boldsymbol{v}(t))$, it is meaningful to study how to optimize the camera linear velocity $\boldsymbol{v}$ in order to affect matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ and, as a consequence, to shape the transient response of the error vector $\boldsymbol{e}(t)$. The *active* strategy developed in [20] and summarized hereafter shows how to achieve this goal.

Let $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{\Omega}$ be the singular value decomposition of matrix $\boldsymbol{\Omega}$, where $\boldsymbol{\Sigma} = [\boldsymbol{S} \;\; \boldsymbol{0}]$, $\boldsymbol{S} = \mathrm{diag}(\sigma_i) \in \mathbb{R}^{p\times p}$, and $0 \leq \sigma_1 \leq \ldots \leq \sigma_p$ are the p singular values of $\boldsymbol{\Omega}$. Let also $\boldsymbol{Q} = \alpha\boldsymbol{I}_m$ and $\boldsymbol{\Lambda} = \beta\boldsymbol{I}_p$, with $\alpha > 0$, $\beta > 0$ (scalar gain matrices). By designing the gain matrix $\boldsymbol{H}$ in (2) as

$$\boldsymbol{H} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_2 \end{bmatrix}\boldsymbol{V}^T \tag{6}$$

with $\boldsymbol{D}_1 \in \mathbb{R}^{p\times p} > 0$, $\boldsymbol{D}_2 \in \mathbb{R}^{(m-p)\times(m-p)} > 0$, it is possible to show that, under the change of coordinates

$$\boldsymbol{\eta} = \frac{1}{\sqrt{\alpha\beta}}\boldsymbol{S}^{-1}\boldsymbol{U}^T\boldsymbol{z} \tag{7}$$

and in the approximation $\boldsymbol{S}^{-1}\boldsymbol{U}^T \approx const$, the behavior of vector $\boldsymbol{\eta}$ (and hence of the estimation error $\boldsymbol{z} = \boldsymbol{\chi} - \hat{\boldsymbol{\chi}}$) is governed by the following linear (and almost diagonal) dynamics

$$\ddot{\boldsymbol{\eta}} = (\boldsymbol{\Pi} - \boldsymbol{D}_1)\dot{\boldsymbol{\eta}} - \alpha\beta\boldsymbol{S}^2\boldsymbol{\eta}. \tag{8}$$

System (8) can be interpreted as a (unit-)mass-spring-damper system with diagonal stiffness matrix $\alpha\beta\boldsymbol{S}^2$ and damping matrix $\boldsymbol{D}_1$, together with an additional 'perturbing' term $\boldsymbol{\Pi}$ whose full expression can be found in [20].

The convergence rate of (8) is then related to its slowest mode dictated by the 'stiffness value' $\alpha\beta\sigma_1^2$, with σ_1^2 being the smallest eigenvalue of the square matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$. Therefore, for the sake of imposing a desired transient response to vector $\boldsymbol{\eta}(t)$ (i.e., to the estimation error $\boldsymbol{z}(t) = \boldsymbol{\chi}(t) - \hat{\boldsymbol{\chi}}(t)$), one can '*place the poles*' of (8) by (*i*) shaping the damping factor $\boldsymbol{D}_1$ in (6) (a free parameter), (*ii*) regulating the value of the smallest eigenvalue σ_1^2 by acting upon vector $\boldsymbol{v}$, and (*iii*) suitably choosing the gain $\alpha\beta$ (a free parameter).

For what concerns the design of matrix $\boldsymbol{D}_1$, we first note that, as explained in [20], matrix $\boldsymbol{\Pi}$ in (8) can be regarded as a second-order perturbation term affecting the dissipative action induced by $\boldsymbol{D}_1$. Therefore, neglecting the effects of matrix $\boldsymbol{\Pi}$ and choosing $\boldsymbol{D}_1 = \mathrm{diag}(c_i)$, $c_i > 0$, allows obtaining a completely decoupled transient behavior for (8)

$$\ddot{\eta}_i + c_i\dot{\eta}_i + \alpha\beta\sigma_i^2\eta_i = 0, \quad i = 1\ldots p. \tag{9}$$

One can then take, for instance, $c_i = c_i^* = 2\sqrt{\alpha\beta}\sigma_i$ in order to impose a critically damped evolution to the estimation error (coincident eigenvalues for (9)).

As for the regulation of $\sigma_1(t)$, being $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v})$, it is

$$\dot{(\sigma_i^2)} = \boldsymbol{J}_{v,i}\dot{\boldsymbol{v}} + \boldsymbol{J}_{s,i}\dot{\boldsymbol{s}} \tag{10}$$

where the Jacobian matrices $\boldsymbol{J}_{v,i} \in \mathbb{R}^{1\times v}$ and $\boldsymbol{J}_{s,i} \in \mathbb{R}^{1\times n}$ can be computed in *closed form*, see [20] for all the details. By inverting the differential mapping (10), vector $\dot{\boldsymbol{v}}$ can then be exploited so as to, e.g., asymptotically enforce $\sigma_1^2(t) \to \sigma_{1,des}^2$ for some desired value $\sigma_{1,des}^2 > 0$. We note that ensuring $\sigma_1^2(t) \to \sigma_{1,des}^2 > 0$ also automatically satisfies the observability condition (5).

Finally, the following considerations hold for the choice of gain $\alpha\beta$ in (8). 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 maximize the estimation convergence speed of (8) (dictated by $\alpha\beta\sigma_1^2$), one can equivalently (*i*) travel at a larger speed $\|\boldsymbol{v}\|$ for a given gain $\alpha\beta$, or (*ii*) increase the gain $\alpha\beta$ for a given $\|\boldsymbol{v}\|$. While increasing the gain $\alpha\beta$ 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 $\alpha\beta$, 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).

Remark II.2. *We note that the proposed strategy is an* active *one since, in the general case, inversion of (10) will result*

[2]The stability proof requires some additional technical assumptions on the regularity of the vanishing disturbance $\boldsymbol{g}$ (locally Lipschitz in a neighbourhood of the origin), on its growth bound w.r.t. $\|\boldsymbol{e}\|$ (which, since $\boldsymbol{g}(\cdot) \to \boldsymbol{0}$ if $(\boldsymbol{v}, \boldsymbol{\omega}) \to \boldsymbol{0}$, can always be made small enough by limiting $(\boldsymbol{v}, \boldsymbol{\omega})$), and on the norm of the initial error $\|\boldsymbol{e}(t_0)\|$. The interested reader can find in [9] a detailed derivation of the proof.

in a camera linear velocity $\boldsymbol{v}$ optimized as a function of the system measured state $\boldsymbol{s}$ in order to attain a desired $\sigma^2_{1,des}$ over time. We also highlight the role played by the gain matrix $\boldsymbol{H}$ weighting the feedback term in observer (2): the proposed machinery in fact relies on a suitable state-dependent design of the damping matrix $\boldsymbol{H}$ via the choice of $\boldsymbol{D}_1 = \text{diag}(c_i^*)$ with $c_i^* = 2\sqrt{\alpha\beta}\sigma_i$. For the interested reader, this state-dependent design is conceptually equivalent to the shaping of the damping factor *adopted in the context of impedance control for robot manipulator arms, see, e.g., [24].*

Remark II.3. *We also note that, in general, it is not possible to fully compensate for the term $\boldsymbol{J}_{s,i}\dot{\boldsymbol{s}}$ when inverting (10). Indeed, the expression in (1) implies a direct dependence of $\dot{\boldsymbol{s}}$ from the unmeasurable $\boldsymbol{\chi}$, so that an exact evaluation of $\dot{\boldsymbol{s}}$ is not obtainable in practice. A possible solution could be to use an approximation $\hat{\dot{\boldsymbol{s}}}$ of $\dot{\boldsymbol{s}}$ obtained by evaluating $\boldsymbol{f}_u(\cdot)$ on the current estimate $\hat{\boldsymbol{\chi}}$. Another simple workaround is, however, to just enforce $\dot{\boldsymbol{s}} \simeq 0$ by imposing a constraint on the camera motion. A combination of both strategies is, of course, also possible. The next sections will present some examples in this sense.*

Remark II.4. *It can be finally shown that, in the special situation $p = 1$ (only one quantity to be estimated), if $\sigma_1(t) \equiv const$ then $\boldsymbol{S}^{-1}\boldsymbol{U}^T \equiv const$ in (7) and matrix $\boldsymbol{\Pi}$ has no disturbing effects on (8). Therefore, in this case it is always possible to* exactly *enforce the ideal estimation error dynamics (9) by just keeping $\|\boldsymbol{\Omega}(t)\|^2 = \sigma_1^2(t) = const$ during the camera motion. This situation will apply to all the case studies discussed in the rest of the paper.*

III. APPLICATIONS TO STRUCTURE FROM MOTION

In this section we illustrate the application of the proposed active estimation framework to three concrete SfM problems: (*i*) estimation of the 3D coordinates of a point feature, (*ii*) estimation of the 3D position and radius of a spherical target, and (*iii*) estimation of the 3D position and radius of a cylindrical target.

In the point feature case, the effects of the adopted projection model on the estimation convergence are also explicitly considered by discussing the differences between the two popular choices of *planar* and *spherical* projection models. For the spherical and cylindrical targets, we instead propose two *novel* minimal parameterizations that allow to express the sphere/cylinder 3D structures in terms of measured visual features and of a single unknown parameter (the sphere/cylinder radius). This allows, in all three cases, to reduce the SfM task to the estimation of a single unknown quantity (point feature depth or sphere/cylinder radius), thus satisfying the requirements of Remark II.4 for *exactly* imposing the ideal dynamics (9) to the estimation error.

A. Depth estimation for a point feature

1) Planar projection model: Let $\boldsymbol{p} = (x, y, 1) = (X/Z, Y/Z, 1) \in \mathbb{R}^3$ be the perspective projection of a 3D point $\boldsymbol{P} = (X, Y, Z)$ onto the image plane of a calibrated pinhole camera. As it is well know [25], the differential relationship between the image motion of a point feature and the camera linear/angular velocity $\boldsymbol{u} = (\boldsymbol{v}, \boldsymbol{\omega}) \in \mathbb{R}^6$ expressed in camera frame is

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -\left(1+x^2\right) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix}\boldsymbol{u} \tag{11}$$

where Z is the *depth* of the feature point. The dynamics of Z is

$$\dot{Z} = \begin{bmatrix}0 & 0 & -1 & -yZ & xZ & 0\end{bmatrix}\boldsymbol{u}.$$

The expression in (11) is not linear in Z but it is linear in $1/Z$. Therefore, by defining $\boldsymbol{s} = (x, y) \in \mathbb{R}^2$ and $\chi = 1/Z$, with then $m = 2$ and $p = 1$, we obtain for (1)

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{u}, t) = \begin{bmatrix} xy & -\left(1+x^2\right) & y \\ 1+y^2 & -xy & -x \end{bmatrix}\boldsymbol{\omega} \\ \boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{v}) = \begin{bmatrix} xv_z - v_x & yv_z - v_y \end{bmatrix} \\ f_u(\boldsymbol{s}, \chi, \boldsymbol{u}, t) = v_z\chi^2 + (y\omega_x - x\omega_y)\chi \end{cases}, \tag{12}$$

with the perturbation term $g(\boldsymbol{e}, t)$ in (3) taking the expression

$$g(\boldsymbol{e}, t) = v_z\left(\chi^2 - \hat{\chi}^2\right) + (y\omega_x - x\omega_y)z, \tag{13}$$

so that $g(\boldsymbol{0}, t) = 0$ as expected. Note that, once χ has been estimated, one can obviously retrieve the 3D position of the point feature as $\boldsymbol{P} = \boldsymbol{p}/\chi$.

In the point feature case matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ reduces to its single eigenvalue which, for a planar projection model, takes the expression

$$\sigma_1^2 = \|\boldsymbol{\Omega}\|^2 = (xv_z - v_x)^2 + (yv_z - v_y)^2. \tag{14}$$

Furthermore, using (14), the Jacobian $\boldsymbol{J}_{v,1}$ in (10) is given by

$$\boldsymbol{J}_{v,1} = 2\begin{bmatrix} v_x - xv_z \\ v_y - yv_z \\ (xv_z - v_x)x + (yv_z - v_y)y \end{bmatrix}^T. \tag{15}$$

Since σ_1^2 does not depend on $\boldsymbol{\omega}$, it is then possible to freely exploit the camera angular velocity for fulfilling additional goals of interest without interfering with the regulation of $\sigma_1^2(t)$ (only affected by $\boldsymbol{v}$). For instance, as in [17], one can use $\boldsymbol{\omega}$ for keeping $\boldsymbol{s} \simeq const$ so as to make the effects of $\dot{\boldsymbol{s}}$ negligible when inverting (10) w.r.t. $\dot{\boldsymbol{v}}$, see Remark II.3.

We now note that σ_1^2 in (14) depends on both the camera linear velocity $\boldsymbol{v}$ and on the location $\boldsymbol{p}$ of the feature point on the image plane. Since the value of σ_1^2 directly affects the convergence speed of the estimation error, it is interesting to study what conditions on $\boldsymbol{p}$ and $\boldsymbol{v}$ result in the largest possible σ_1^2 (i.e., the fastest possible convergence for a given gain $\alpha\beta$). Letting $\boldsymbol{e}_3 = (0, 0, 1)$ being the camera optical axis, it is (by inspection)

$$\begin{bmatrix}\boldsymbol{\Omega}^T \\ 0\end{bmatrix} = [\boldsymbol{e}_3]_\times [\boldsymbol{p}]_\times \boldsymbol{v}$$

where $[\boldsymbol{v}_1]_\times$ is the skew-symmetric matrix representing the cross product operator for 3D vectors (i.e., $[\boldsymbol{v}_1]_\times \boldsymbol{v}_2 = \boldsymbol{v}_1 \times$

$\boldsymbol{v}_2$). Therefore,

$$\sigma_1^2 = \begin{bmatrix}\boldsymbol{\Omega} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\Omega}^T \\ 0\end{bmatrix} = \| [\boldsymbol{e}_3]_\times [\boldsymbol{p}]_\times \boldsymbol{v}\|^2$$
$$= \|\boldsymbol{p}\|^2\|\boldsymbol{v}\|^2 \sin^2(\theta_{\boldsymbol{p},\boldsymbol{v}}) \sin^2\left(\theta_{\boldsymbol{e}_3,[\boldsymbol{p}]_\times \boldsymbol{v}}\right)$$

where $\theta_{\boldsymbol{p},\boldsymbol{v}}$ and $\theta_{\boldsymbol{e}_3,[\boldsymbol{p}]_\times \boldsymbol{v}}$ represent the angles between vectors $(\boldsymbol{p},\, \boldsymbol{v})$ and vectors $(\boldsymbol{e}_3,\, [\boldsymbol{p}]_\times \boldsymbol{v})$, respectively. The maximum attainable value for σ_1^2 is then

$$\sigma_{max}^2 = \max_{\boldsymbol{p},\boldsymbol{v}} \sigma_1^2 = \|\boldsymbol{p}\|^2\|\boldsymbol{v}\|^2. \tag{16}$$

This maximum is obtained when the camera linear velocity $\boldsymbol{v}$ is such that $\boldsymbol{p} \perp \boldsymbol{v}$ and $\boldsymbol{e}_3 \perp [\boldsymbol{p}]_\times \boldsymbol{v}$, i.e., rearranging in matrix form

$$\begin{bmatrix}\boldsymbol{p}^T \\ \boldsymbol{e}_3^T [\boldsymbol{p}]_\times\end{bmatrix}\boldsymbol{v} = \begin{bmatrix}x & y & 1 \\ -y & x & 0\end{bmatrix}\boldsymbol{v} = 0. \tag{17}$$

If $\boldsymbol{p} \neq \boldsymbol{e}_3$ (point feature *not* at the center of the image plane), system (17) has (full) rank 2 and admits the unique solution (up to a scalar factor)

$$\boldsymbol{v} = \delta\, [\boldsymbol{p}]_\times^2\, \boldsymbol{e}_3, \quad \delta \in \mathbb{R}.$$

This requires the linear velocity $\boldsymbol{v}$ to be orthogonal to $\boldsymbol{p}$ and to lie on the plane defined by vectors $\boldsymbol{p}$ and $\boldsymbol{e}_3$ (i.e., $\boldsymbol{v}$ must belong to a straight line as shown in Fig. 1a).

If $\boldsymbol{p} = \boldsymbol{e}_3$ (point feature at the center of the image plane), system (17) loses rank and any $\boldsymbol{v} \perp \boldsymbol{e}_3$ is a valid solution, see Fig. 1b.

It is then possible to draw the following conclusions: for a given norm of the linear velocity $\|\boldsymbol{v}\|$ (i.e., the amount of 'control effort'), system (17) determines the direction of $\boldsymbol{v}$ resulting in $\sigma_1^2 = \sigma_{max}^2$ (maximization of σ_1^2). These conditions are summarized in Figs. 1a and 1b. The value of σ_{max}^2 is, however, also a function of the feature point location $\boldsymbol{p}$ which can be arbitrarily positioned on the image plane. In particular, $\sigma_{max}^2 = \|\boldsymbol{v}\|^2$ for $\boldsymbol{p} = \boldsymbol{e}_3$ and $\sigma_{max}^2 = \|\boldsymbol{p}\|^2\|\boldsymbol{v}\|^2 > \|\boldsymbol{v}\|^2$ $\forall\, \boldsymbol{p} \neq \boldsymbol{e}_3$, with $\lim_{\|\boldsymbol{p}\|\to\infty} \sigma_{max}^2(\boldsymbol{p}) = \infty$. The value of $\|\boldsymbol{p}\|$ (distance of the point feature from the image center) thus acts as an amplification factor for σ_{max}^2. Therefore,

1) the smallest σ_{max}^2 (i.e., the *slowest* 'optimal' convergence for the depth estimation error) is obtained for the smallest value of $\|\boldsymbol{p}\|$, i.e., when $\boldsymbol{p} = \boldsymbol{e}_3 \implies \|\boldsymbol{p}\| = 1$ (feature point at the center of the image plane). It is worth noting that in this case $v_z = 0$ (from the condition $\boldsymbol{v} \perp \boldsymbol{p}$) and $\sigma_{max}^2 = \|\boldsymbol{v}\|^2 = v_x^2 + v_y^2$: the camera moves on the surface of a sphere with a constant radius (depth) pointing at the feature point. Also, being in this case $\dot{\chi} = \dot{Z}/Z^2 = 0$, one has $g(\boldsymbol{e},\, t) \equiv 0$ and global convergence for the estimation error (see Remark II.1);
2) the largest σ_{max}^2 (i.e., the *fastest* 'optimal' convergence for the depth estimation error) is obtained for the largest possible value of $\|\boldsymbol{p}\|$. In the usual case of a rectangular image plane centered at the origin, this translates into keeping the feature point positioned at one of the four image corners. However, compared with the previous case, this results in a $\boldsymbol{g}(\boldsymbol{e},\, t) \neq \boldsymbol{0}$ and only local convergence for the estimation error.

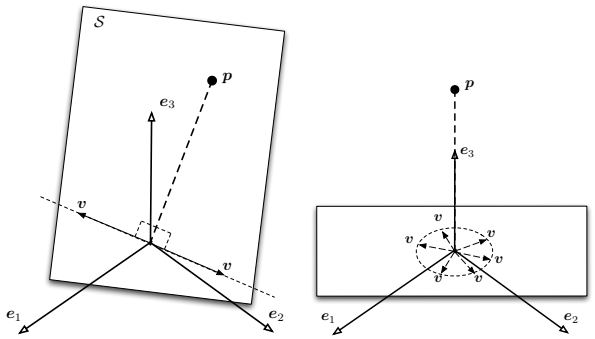


Fig. 1: Optimality conditions for the camera linear velocity $\boldsymbol{v}$ as dictated by system (17). (a): when $\boldsymbol{p} \neq \boldsymbol{e}_3$, vector $\boldsymbol{v}$ must be orthogonal to $\boldsymbol{p}$ and lie on the plane $\mathcal{S}$ spanned by $\boldsymbol{p}$ and $\boldsymbol{e}_3$ (that is, $\boldsymbol{v}$ must belong to a specific straight line). (b): when $\boldsymbol{p} = \boldsymbol{e}_3$, any $\boldsymbol{v} \perp \boldsymbol{e}_3$ is a valid solution to (17).

2) Spherical projection model: We now develop the depth estimation machinery for the spherical projection model. In this case, the following quantity is taken as visual feature measured on the image plane

$$\boldsymbol{s} = \frac{\boldsymbol{p}}{\|\boldsymbol{p}\|} = \frac{\boldsymbol{P}}{\|\boldsymbol{P}\|} \in \mathbb{S}^2,$$

where $\mathbb{S}^2$ represents the unit sphere and, as well-known [26],

$$\dot{\boldsymbol{s}} = \left[\frac{1}{\|\boldsymbol{P}\|}\left(\boldsymbol{s}\boldsymbol{s}^T - \boldsymbol{I}_3\right) \quad [\boldsymbol{s}]_\times\right]\boldsymbol{u},$$

and

$$\frac{d}{dt}\left(\frac{1}{\|\boldsymbol{P}\|}\right) = -\frac{1}{\|\boldsymbol{P}\|^2}\frac{d\|\boldsymbol{P}\|}{dt} = -\frac{\boldsymbol{s}^T\dot{\boldsymbol{P}}}{\|\boldsymbol{P}\|^2} = \frac{\boldsymbol{s}^T\boldsymbol{v}}{\|\boldsymbol{P}\|^2}.$$

Hence by taking $\chi = 1/\|\boldsymbol{P}\|$ one obtains for (1)

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{s},\, \boldsymbol{u}) = [\boldsymbol{s}]_\times \boldsymbol{\omega} \\ \boldsymbol{\Omega}(\boldsymbol{s},\, \boldsymbol{v}) = -\boldsymbol{v}^T\left(\boldsymbol{I}_3 - \boldsymbol{s}\boldsymbol{s}^T\right) \\ f_u(\boldsymbol{s},\, \chi, \boldsymbol{u}) = \chi^2 \boldsymbol{s}^T\boldsymbol{v} \end{cases} \tag{18}$$

with $m = 3$, $p = 1$, and $g(\boldsymbol{e},\, t) = (\chi^2 - \hat{\chi}^2)\boldsymbol{s}^T\boldsymbol{v}$ for the perturbation term in (3). We note that, although in this case $m = 3$, vector $\boldsymbol{s}$ is subject to the constraint $\|\boldsymbol{s}\| = 1$, thus resulting in only two independent measurements (as in the previous case of planar projection). Moreover, from the estimated χ one can easily retrieve $\boldsymbol{P} = \boldsymbol{s}/\chi$.

For the spherical projection model, the eigenvalue determining the convergence of the estimation error is

$$\sigma_1^2 = \boldsymbol{\Omega}\boldsymbol{\Omega}^T = \boldsymbol{v}^T\boldsymbol{v} - (\boldsymbol{s}^T\boldsymbol{v})^2,$$

with thus

$$\boldsymbol{J}_{v,1} = 2\boldsymbol{v}^T(\boldsymbol{I}_3 - \boldsymbol{s}\boldsymbol{s}^T). \tag{19}$$

As before, σ_1^2 does not depend on $\boldsymbol{\omega}$ which can then be exploited to fulfil any additional task of interest (e.g., keeping $\boldsymbol{s} \simeq const$ during motion).

As for the conditions on $\boldsymbol{s}$ and $\boldsymbol{v}$ that yield maximization of σ_1^2, one clearly has

$$\sigma_1^2 = \sigma_{max}^2 = \max_{\boldsymbol{s},\,\boldsymbol{v}} \sigma_1^2 = \|\boldsymbol{v}\|^2 \tag{20}$$

iff $\boldsymbol{s}^T\boldsymbol{v} = 0$ (linear velocity orthogonal to the projection ray passing through $\boldsymbol{P}$). We also note that, in this case, one has $\dot{\chi} = 0$ and $g(\boldsymbol{e}, t) \equiv 0$ (*constant* unknown state and *global* convergence for the estimation error) *regardless* of the location of $\boldsymbol{s}$ on the image plane.

3) Comparison between planar and spherical projection models: for a spherical projection model, maximization of the eigenvalue σ_1^2 imposes only one condition for the linear velocity $\boldsymbol{v}$ ($\boldsymbol{s}^T\boldsymbol{v} = 0$). When this condition is met, one has $\sigma_1^2 = \sigma_{max}^2 = \|\boldsymbol{v}\|^2$ and global convergence for the estimation error *whatever* the location of the feature point $\boldsymbol{s}$. This is equivalent to what was obtained for the planar projection case in the special situation $\boldsymbol{p} = \boldsymbol{e}_3$ (indeed the two projection models coincide for $\boldsymbol{p} = \boldsymbol{s} = \boldsymbol{e}_3$). However, with a spherical projection model one also loses the possibility to increase the estimation convergence rate by suitably positioning the point feature $\boldsymbol{s}$ on the image plane (since in this case σ_{max}^2 does not depend on $\boldsymbol{s}$).

It is then worth noting the complementarity of the two cases: for a given $\|\boldsymbol{v}\|$, and provided the optimal condition $\boldsymbol{p}^T\boldsymbol{v} = 0$ is satisfied, the *planar* projection allows obtaining a faster error convergence at the price of local stability (increase of the perturbation g) by suitably positioning $\boldsymbol{s} = (x, y)$ (the larger $\|\boldsymbol{s}\|$ the faster the convergence). The *spherical* projection guarantees global error convergence for *any* location of the feature point but at the price of being always subject to the same convergence rate only function of the control effort $\|\boldsymbol{v}\|$.

B. Structure estimation for a spherical target

We now detail the application of the proposed estimation machinery to the case of 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

$$\mathcal{L}: \quad \boldsymbol{n}^T\boldsymbol{X} + d = 0$$

represent the *planar limb surface* associated to the sphere in the camera frame, where $\boldsymbol{X} \in \mathbb{R}^3$ is any 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 [27]. Figure 2 shows the quantities of interest.

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

$$\frac{1}{Z} = \frac{X_0}{K}x + \frac{Y_0}{K}y + \frac{Z_0}{K} = \boldsymbol{\chi}^T\boldsymbol{p}, \tag{21}$$

where $K = \boldsymbol{P}_0^T\boldsymbol{P}_0 - R^2$ and $\boldsymbol{\chi} = \boldsymbol{P}_0/K = -\boldsymbol{n}/d \in \mathbb{R}^3$ represent *unmeasurable* quantities (analogously to Z for the point feature case), see [28] for all the details. 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 [10], [27]. 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 (see Appendix A)

$$\dot{\boldsymbol{\chi}} = -\frac{\boldsymbol{v}}{K} - [\boldsymbol{\omega}]_\times\boldsymbol{\chi} + 2\boldsymbol{\chi}\boldsymbol{\chi}^T\boldsymbol{v}$$

and 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 $R = \sqrt{\boldsymbol{P}_0^T\boldsymbol{P}_0 - K}$.

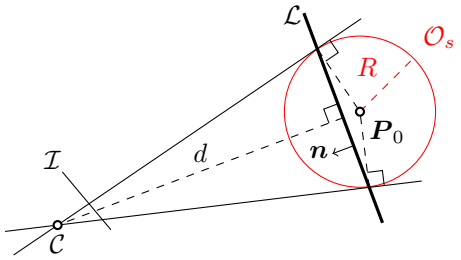


Fig. 2: Spherical target $\mathcal{O}_s$ and planar limb surface $\mathcal{L}$.

Although conceptually valid, this solution requires the concurrent estimation of *three time-varying quantities* (vector $\boldsymbol{\chi}(t)$). On the other hand, inspired by [29], we now describe a *novel representation* of the sphere projection on the image plane that allows to reformulate the structure estimation task in terms of a *single unknown constant parameter*, i.e., the sphere radius R.

To this end, define vector $\boldsymbol{s} = (s_x, s_y, s_z) \in \mathbb{R}^3$ as

$$\begin{cases} s_x &= \dfrac{x_g}{s_z a_1^2} \\ s_y &= \dfrac{y_g}{s_z a_1^2} \\ s_z &= \sqrt{\dfrac{1 + a_1^2}{a_1^2}} \end{cases}, \tag{22}$$

where $(x_g, y_g, n_{20}, n_{11}, n_{02})$ represent 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 is the minor axis of the observed ellipse with [27]

$$a_1^2 = 2\left(n_{20} + n_{02} - \sqrt{(n_{20} - n_{02})^2 + 4n_{11}}\right). \tag{23}$$

We thus note that vector $\boldsymbol{s}$ can be directly evaluated in terms of measured image quantities.

From [27], [29] one also has

$$x_g = \frac{X_0 Z_0}{Z_0^2 - R^2}, \quad y_g = \frac{Y_0 Z_0}{Z_0^2 - R^2}, \quad a_1^2 = \frac{R^2}{Z_0^2 - R^2} \tag{24}$$

which, when plugged in (22) and (23), result in the equivalent expression $\boldsymbol{s} = \boldsymbol{P}_0/R$. Since vector $\boldsymbol{s}$ can be computed from image measurements as in (22), estimation of the (unknown) sphere radius R allows to recover the 3D sphere center as $\boldsymbol{P}_0 = \boldsymbol{s}R$.

Exploiting now the results of [29], it is possible to show that

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\dfrac{1}{R}\boldsymbol{I}_3 & [\boldsymbol{s}]_\times \end{bmatrix}\boldsymbol{u}. \tag{25}$$

Since (25) is linear in $1/R$, we can define $\chi = 1/R$, with then $m = 3$ and $p = 1$, and obtain for (1) and (3)

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{u}) &= [\boldsymbol{s}]_\times \boldsymbol{\omega} \\ \boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{u}) &= -\boldsymbol{v}^T \\ f_u(\boldsymbol{s}, \chi, \boldsymbol{u}) &= 0 \\ g(\boldsymbol{e}, t) &= 0 \end{cases} . \tag{26}$$

We note that in this case it is always possible to obtain *global* convergence for the estimation error since $\dot{\chi} = 0$ and therefore $g(\boldsymbol{e}, t) = 0$ by construction (see Remark II.1). Furthermore, matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^T$ reduces again to its single eigenvalue $\sigma_1^2 = \|\boldsymbol{v}\|^2$ and, if $\sigma_1^2(t) \equiv const > 0$, the 'ideal' estimation error dynamics (9) can be exactly obtained. One also has $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{v})$ and $\boldsymbol{J}_{v,1} = 2\boldsymbol{v}^T$.

We finally note the following facts: first, contrarily to the previous cases, here $\dot{\boldsymbol{s}}$ has no effect on the regulation of σ_1^2 which is only function of the camera linear velocity $\boldsymbol{v}$. It is then of course still possible to 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, we note the strong similarities with the previous optimal results obtained for a point feature under a *spherical* projection model (σ_{max}^2 in (20)): in both cases the maximum estimation convergence rate for a given $\|\boldsymbol{v}\|$ does not depend on the position of the observed object on the image plane.

C. Structure estimation for a cylindrical target

We now finally consider the case of SfM for a 3D cylindrical object. 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 [17]). Moreover, analogously to the sphere, a cylinder is also associated with a planar limb surface $\mathcal{L}$ such that (21) holds for any point on $\mathcal{L}$ with projection $\boldsymbol{p} = (x, y, 1)$. Therefore, a possibility is to estimate the three unknown parameters of the limb plane $\mathcal{L}$ (vector $\boldsymbol{\chi}$) by exploiting (at least) three image measurements, see [17] and Appendix B for some details in this sense. However, following the previous developments, we now propose a *novel representation* of the cylinder projection on the image plane which, again, allows to obtain the cylinder parameters $(\boldsymbol{P}_0, \boldsymbol{a}, R)$ in terms of image measurements and of the *unknown but constant* cylinder radius R which, therefore, represents the only quantity to be estimated.

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, and

$$\boldsymbol{n}_1 = (\cos\theta_1, \sin\theta_1, -\rho_1),\ \boldsymbol{n}_2 = (\cos\theta_2, \sin\theta_2, -\rho_2) \tag{27}$$

be the normal vectors to the two planes passing through the origin of the camera frame and the two above-mentioned projected lines[3]. Figure 3 gives a graphical representation of the quantities of interest. Note that vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$

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

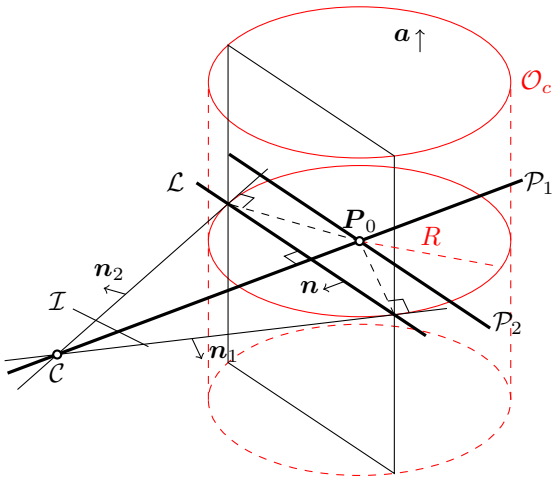


Fig. 3: Camera $\mathcal{C}$ and cylindrical target $\mathcal{O}_c$ with the planar limb surface $\mathcal{L}$ and the other planes of interest $\mathcal{P}_1$ and $\mathcal{P}_2$

can be directly evaluated from image measurements (the line parameters). We then define vector $\boldsymbol{s} \in \mathbb{R}^3$ as

$$\boldsymbol{s} = \frac{\boldsymbol{\Delta}}{\|\boldsymbol{\Delta}\|^2} \tag{28}$$

with

$$\boldsymbol{\Delta} = \frac{1}{2}\left(\frac{\boldsymbol{n}_1}{\|\boldsymbol{n}_1\|} + \frac{\boldsymbol{n}_2}{\|\boldsymbol{n}_2\|}\right). \tag{29}$$

Vector $\boldsymbol{s}$ is, thus, also directly obtainable in terms of image quantities.

We now note that, from [30], an equivalent expression for vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ in terms of the cylinder 3D geometry can be obtained as

$$\boldsymbol{n}_1 = \frac{1}{N_1}\begin{bmatrix} R\frac{X_0}{\sqrt{K}} - \alpha \\ R\frac{Y_0}{\sqrt{K}} - \beta \\ R\frac{Z_0}{\sqrt{K}} - \gamma \end{bmatrix}, \quad \boldsymbol{n}_2 = \frac{1}{N_2}\begin{bmatrix} R\frac{X_0}{\sqrt{K}} + \alpha \\ R\frac{Y_0}{\sqrt{K}} + \beta \\ R\frac{Z_0}{\sqrt{K}} + \gamma \end{bmatrix} \tag{30}$$

with

$$\begin{cases} K &= \sqrt{\boldsymbol{P}_0^T\boldsymbol{P}_0 - R^2} \\ (\alpha, \beta, \gamma) &= [\boldsymbol{P}_0]_\times \boldsymbol{a} \\ N_1 &= \sqrt{\left(R\frac{X_0}{\sqrt{K}} - \alpha\right)^2 + \left(R\frac{Y_0}{\sqrt{K}} - \beta\right)^2} \\ N_2 &= \sqrt{\left(R\frac{X_0}{\sqrt{K}} + \alpha\right)^2 + \left(R\frac{Y_0}{\sqrt{K}} - \beta\right)^2} \end{cases}, \tag{31}$$

thus yielding

$$\begin{cases} \dfrac{\boldsymbol{n}_1}{\|\boldsymbol{n}_1\|} &= \dfrac{1}{\boldsymbol{P}_0^T\boldsymbol{P}_0}\begin{bmatrix} RX_0 - \alpha\sqrt{K} \\ RY_0 - \beta\sqrt{K} \\ RZ_0 - \gamma\sqrt{K} \end{bmatrix} \\ \dfrac{\boldsymbol{n}_2}{\|\boldsymbol{n}_2\|} &= \dfrac{1}{\boldsymbol{P}_0^T\boldsymbol{P}_0}\begin{bmatrix} RX_0 + \alpha\sqrt{K} \\ RY_0 + \beta\sqrt{K} \\ RZ_0 + \gamma\sqrt{K} \end{bmatrix} \end{cases} . \tag{32}$$

Plugging (32) in (29) results in the equivalent expression

$$\boldsymbol{\Delta} = \frac{R^2}{\boldsymbol{P}_0^T \boldsymbol{P}_0} \boldsymbol{s}$$

which, using (28), finally yields the following relationship between image quantities and cylinder 3D structure

$$\boldsymbol{s} = \frac{\boldsymbol{\Delta}}{\|\boldsymbol{\Delta}\|^2} = \frac{\boldsymbol{P}_0}{R}. \tag{33}$$

As for the cylinder axis $\boldsymbol{a}$, exploiting (30) one has

$$\begin{aligned} [\boldsymbol{n}_2]_\times \boldsymbol{n}_1 &= \frac{2R}{N_1 N_2 \sqrt{K}} \begin{bmatrix} Z_0\beta - Y_0\gamma \\ X_0\gamma - Z_0\alpha \\ Y_0\alpha - X_0\beta \end{bmatrix} = \frac{2R}{N_1 N_2 \sqrt{K}} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}_\times \boldsymbol{P}_0 \\ &= \frac{2R}{N_1 N_2 \sqrt{K}} \left[[\boldsymbol{P}_0]_\times \boldsymbol{a} \right]_\times \boldsymbol{P}_0 = \frac{2R \boldsymbol{P}_0^T \boldsymbol{P}_0}{N_1 N_2 \sqrt{K}} \boldsymbol{a} \end{aligned} \tag{34}$$

where in the last step the property $\boldsymbol{a}^T \boldsymbol{P}_0 = 0$ was used. Since $\|\boldsymbol{a}\| = 1$, from (34) it is

$$\boldsymbol{a} = \frac{[\boldsymbol{n}_2]_\times \boldsymbol{n}_1}{\left\| [\boldsymbol{n}_2]_\times \boldsymbol{n}_1 \right\|}. \tag{35}$$

The cylinder axis $\boldsymbol{a}$ can then be directly obtained in terms of only measured quantities.

We now note that, as in the sphere case, the only unknown left is the cylinder radius R: once known, the cylinder 3D structure can be fully recovered from image measurements as $\boldsymbol{P}_0 = R\boldsymbol{s}$ from (33) and $\boldsymbol{a}$ from (35).

An estimation scheme for R can be obtained exploiting the following differential relationship whose derivation is given in Appendix C

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\frac{1}{R} \left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T \right) & [\boldsymbol{s}]_\times \end{bmatrix} \boldsymbol{u}. \tag{36}$$

Note the similarity of (36) with (25) for the sphere case.

Being (36) linear in $1/R$, one can then apply observer (2) by choosing $\chi = 1/R$ with $m = 3$ and $p = 1$, and obtaining

$$\begin{cases} \boldsymbol{f}_m(\boldsymbol{s}, \boldsymbol{u}) &= [\boldsymbol{s}]_\times \boldsymbol{\omega} \\ \boldsymbol{\Omega}(\boldsymbol{s}, \boldsymbol{u}) &= -\boldsymbol{v}^T \left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T \right) \\ f_u(\boldsymbol{s}, \chi, \boldsymbol{u}) &= 0 \\ g(\boldsymbol{e}, t) &= 0 \end{cases}. \tag{37}$$

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

$$\sigma_1^2 = \boldsymbol{\Omega}\boldsymbol{\Omega}^T = \|\boldsymbol{v}\|^2 - (\boldsymbol{a}^T \boldsymbol{v})^2. \tag{38}$$

It is worth comparing (38) with the result obtained for the sphere ($\sigma_1^2 = \|\boldsymbol{v}\|^2$). In the cylinder case, the convergence rate of the estimation error 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 $\boldsymbol{a}$ ($\boldsymbol{a}^T \boldsymbol{v} = 0$) with, in this case, $\sigma_1^2 = \sigma_{max}^2 = \|\boldsymbol{v}\|^2$. Intuitively, any camera motion along the cylinder axis does not provide any useful information to the estimation task. Furthermore, as in all previous cases, keeping a $\sigma_1^2(t) = const$ allows to exactly enforce the ideal estimation error dynamics (9), see Remark II.4.

Finally, from (38) one has

$$\dot{(\sigma_1^2)} = \boldsymbol{J}_{v,1} \dot{\boldsymbol{v}} + \boldsymbol{J}_{a,1} \dot{\boldsymbol{a}} = \boldsymbol{J}_{v,1} \dot{\boldsymbol{v}} + \boldsymbol{J}_{a,1} [\boldsymbol{a}]_\times \boldsymbol{\omega} \tag{39}$$

with $\boldsymbol{J}_{v,1} = 2\boldsymbol{v}^T \left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T \right)$ and $\boldsymbol{J}_{a,1} = 2\boldsymbol{v}^T \boldsymbol{a} \boldsymbol{v}^T$. Although (39) 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 (39) w.r.t. $\dot{\boldsymbol{v}}$ as discussed in Sect. IV-D. 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. SIMULATION AND EXPERIMENTAL RESULTS

In this section we show some experimental and simulation 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 [31]. Some snapshots of the three experiments are shown in Fig. 4 where the result of the image processing is highlighted in red.

A video of the reported experiments (including the three cases of point, sphere and cylinder) is also attached to the paper for the reader's convenience.

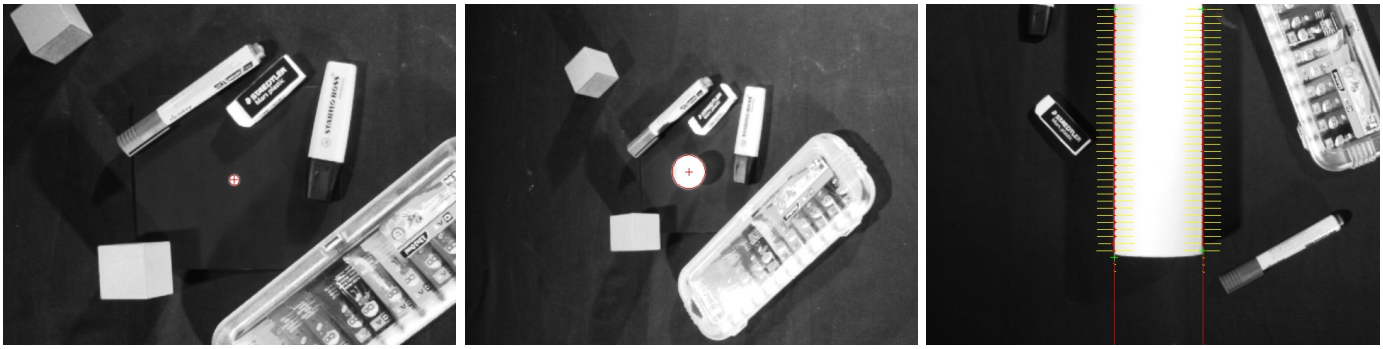
A. Comparison of planar and spherical projection models

We start by comparing via simulation results the effects of adopting a planar and spherical projection model for the depth estimation of a point feature as extensively discussed in Sect. III-A1 and Sect. III-A2. We considered three cases differing for the location on the image plane at which the point feature was (purposely) kept exploiting the camera angular velocity $\boldsymbol{\omega}$:

1) in case I the point feature was kept at the center of the image plane (red line in the following plots);
2) in case II the point feature was kept at one of the corners of an image plane with the same size of the camera used in the experiments (green line in the following plots);
3) in case III the point feature was kept at one of the corners of an image plane with a size five times larger than case II (blue line in the following plots).

In all cases, a constant camera velocity $\boldsymbol{v}(t) \equiv \boldsymbol{v}(t_0) = const$ was kept during motion, with the initial condition $\boldsymbol{v}(t_0)$ chosen so as to comply with the optimality conditions discussed in Sects. III-A1 and III-A2 for letting $\sigma_1^2 = \sigma_{max}^2$ (e.g., with $\boldsymbol{v}(t_0)$ being a solution of (17) in the planar projection case).

Figure 5a shows the behavior of $z(t)$ for the three cases when using a *planar* projection model. We can then note how the convergence rate of the estimation error increases from case I (slowest convergence) to case III (fastest convergence) as predicted by the theory (for the same $\|\boldsymbol{v}\|$ a larger $\|\boldsymbol{p}\|$ results in a larger σ_{max}^2). Similarly, Fig. 5b reports the behavior of $\sigma_1^2(t)$ for the three cases: as expected, $\sigma_1^2(t)$ results largest



(a) (b) (c)

Fig. 4: Camera snapshots for the point feature (a), the sphere (b) and the cylinder (c) experiment.

for case III. Note also how $\sigma_1^2(t)$ for case II (green line) is only *slightly* larger than case I (red line). This is due to relatively small size of the image plane of case II whose dimensions were set as those of the real camera used for the experiments. Finally, Fig. 5c shows the behavior of the perturbation term $g(\boldsymbol{e}, t)$ in the three cases: here, one can verify how $g = 0$ for case I, with then an increasing $|g|$ for cases II and III. Indeed, as discussed in Sect. III-A1, the 'amplification' effect on σ_{max}^2 obtained by increasing $\|\boldsymbol{p}\|$ comes at the price of an increased magnitude of the perturbation g. This is also evident in Fig. 5a where the ideal response of (9) is plotted with dashed lines for the three considered cases. We can thus note how $z(t)$ in case I presents a perfect match with its corresponding ideal response, with then an increasing (albeit very limited) mismatch in the other two cases due to the increased effect of the perturbation g.

As for the spherical projection model, Fig. 5d reports the behavior of the estimation error $z(t)$ for the three cases under consideration, together with the ideal response (9). Here, the symbol $z_s(t)$ is used to denote the estimation error in the spherical projection case in order to distinguish it from the error obtained with the planar projection model. All the plots result perfectly superimposed as expected from the analysis of Sect. III-A2. Indeed, in the spherical projection case, $\sigma_{max}^2 = \|\boldsymbol{v}\|^2$ regardless of the location of $\boldsymbol{p}$ and $g(t) \equiv 0$. However, absence of perturbation terms is obtained at the expense of the convergence rate of $z_s(t)$, which indeed results slower or equal to that of $z(t)$ in the planar projection case. This is shown in Fig. 5e where the behavior of $z(t) - z_s(t)$ is reported for the three cases. We can then note how $z(t) - z_s(t) = 0$ only in case I, as the planar and spherical models coincide when the feature point is at the center of the image plane.

These results then fully confirm the validity of the theoretical analysis reported in Sects. III-A1 and III-A2. However, we also note the marginal effects of the two projection models on the estimation performance when applied to an image plane of size comparable to that of the real camera used in our experimental setup. Therefore, in the following experimental results we will only consider the case of planar projection model.

B. Depth estimation for a point feature

We here report some experimental results for the depth estimation of a point feature under a planar projection model (Sect. III-A1). The following experiments are meant to demonstrate how the proposed active estimation framework can be exploited to select online the 'best' camera motion. As visual target, we made use of a circular white dot of 5 mm radius painted on a planar black surface and sufficiently far from the camera in order to safely consider it as a 'point feature'.

Figure 6a shows the evolution of the estimation error $z(t) = 1/Z(t) - 1/\hat{Z}(t)$ for two experiments[4] in which $\|\boldsymbol{v}(t)\| = \|\boldsymbol{v}_0\|$ but with its direction being either optimized in order to maximize the estimation convergence rate (case I, red line) or kept constant so that $\boldsymbol{v}(t) = \boldsymbol{v}_0 = const$ (case II, blue line). This effect was obtained by using the following control law[5]

$$\dot{\boldsymbol{v}} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|^2} k_1 \left(\kappa_{des} - \kappa\right) + k_2 \left(\boldsymbol{I}_3 - \frac{\boldsymbol{v}\boldsymbol{v}^T}{\|\boldsymbol{v}\|^2}\right) \boldsymbol{J}_{v,1}^T \qquad (40)$$

with $k_1 > 0$, $k_2 \geq 0$, $\kappa = \frac{1}{2}\boldsymbol{v}^T\boldsymbol{v}$, $\kappa_{des} = \frac{1}{2}\boldsymbol{v}_0^T\boldsymbol{v}_0$, and $\boldsymbol{J}_{v,1}$ given by (15). In fact, the first term in (40) enforces the constraint $\|\boldsymbol{v}(t)\| = \|\boldsymbol{v}_0\|$ (same control effort in both cases), while the second term allows to implement either case I ($k_2 > 0$) or case II ($k_2 = 0$) (maximization of σ_1^2) within the null-space of the first constraint. In both cases, the angular velocity $\boldsymbol{\omega}$ was exploited for keeping the point feature at the center of the image plane $(x, y) \rightarrow (0, 0)$. We note that, as discussed in Sect. III-A1, when $(x, y) = (0, 0)$ one has $\sigma_{max}^2 = v_x^2 + v_y^2$ from (16) and $\sigma_1^2 = \sigma_{max}^2$ iff $v_z = 0$ (circular motion around the point feature). The experiments were run with the following parameters: $\alpha\beta = 10^3$ for gains $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$, $c_1 = c_1^*$ for $\boldsymbol{D}_1$ in (6), $\boldsymbol{v}(t_0) = \boldsymbol{v}_0 = (0.03, 0, -0.04)$ m/s, $k_1 = 5$ and $k_2 = 10^4$, thus resulting in the maximum value $\sigma_{max}^2 = 0.0025$ for the eigenvalue σ_1^2.

As clear from Fig. 6b, while in case II the camera gets closer to the point feature, the use of the active strategy of case I results in a null component of $\boldsymbol{v}$ along the projection

[4]The ground truth $Z_0(t)$ was obtained from a previous offline estimation of the 3D position $\boldsymbol{P}_0$ in the world frame, and by then using the information on the camera position provided by the robot forward kinematics.

[5]The value of $\dot{\boldsymbol{v}}$ resulting from this (and following) optimizations was numerically integrated so as to obtain the commanded $\boldsymbol{v}(t)$ sent to the robot low-level controller.

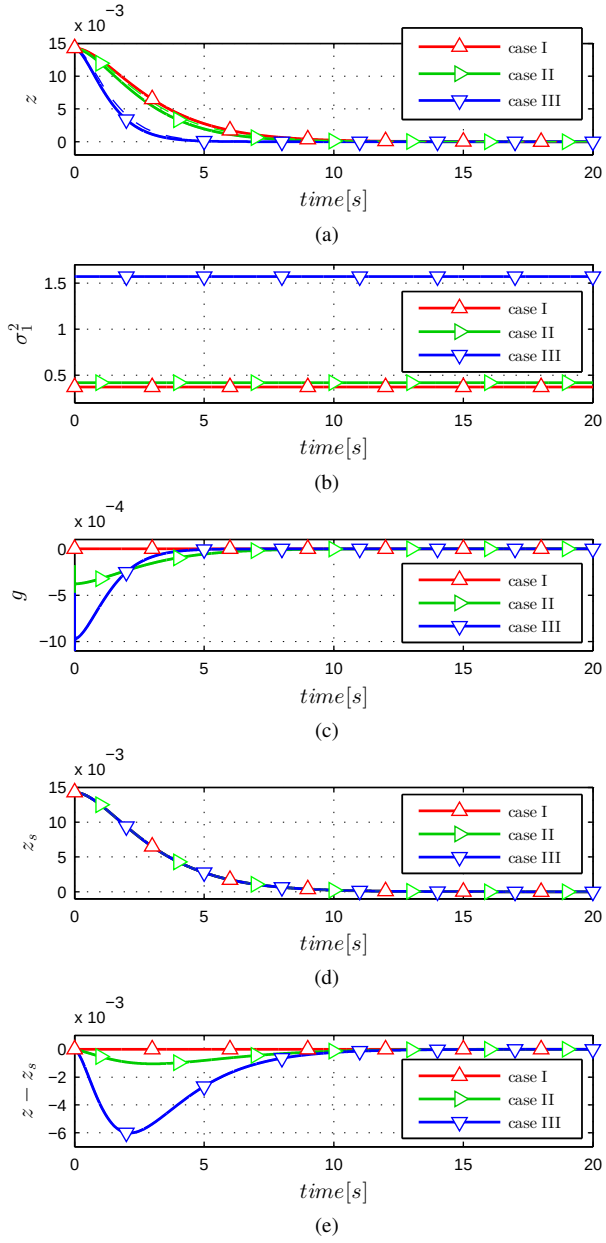


Fig. 5: Simulation results comparing the planar and spherical projection models for the depth estimation of a point feature. The following color coding is adopted for the three considered cases: red–case I, green–case II, blue–case III. (a) behavior of the estimation error $z(t)$ in the planar projection case (solid lines) with superimposed the corresponding ideal response (9) (dashed lines). (b) behavior of $\sigma_1^2(t)$ for the three cases with, again, the largest $\sigma_1^2(t)$ in case III. (c) behavior of the perturbation term $g(\boldsymbol{e},\, t)$ for the three cases. (d) behavior of the estimation error $z_s(t)$ for the spherical projection model in the three cases. (e) behavior of $z(t) - z_s(t)$.

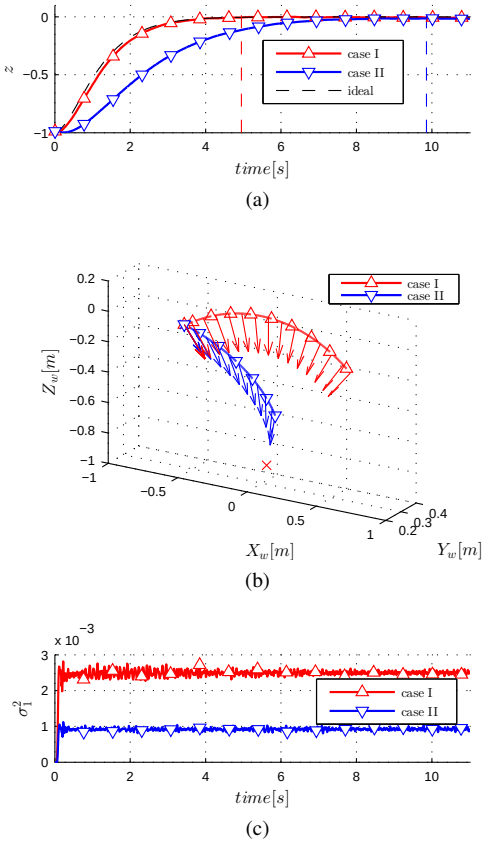


Fig. 6: Experimental results for the point feature case. (a): behavior of the estimation error for case I (solid red line) and case II (solid blue line), and for an 'ideal' second order system (9) with desired poles at σ_{max}^2 (dashed black line). The two vertical dashed lines indicate the times $T_1 = 4.95\,\mathrm{s}$ and $T_2 = 9.85\,\mathrm{s}$ at which the estimation error drops below the threshold of $5\,\mathrm{mm}$. (b): Camera trajectories for case I (red line) and case II (blue line) with arrows indicating the direction of the camera optical axis. (c): behavior of $\sigma_1^2(t)$ for case I (red line) and case II (blue line).

ray of the point feature (i.e., $v_z = 0$) and in an associated circular trajectory centered on the tracked point as predicted by the theoretical analysis of Sect. IV-A. This then allows to move faster in the 'useful' directions (while keeping the same constant $\|\boldsymbol{v}\|$), and, thus, to increase the value of σ_1^2 towards its theoretical maximum $\sigma_{max}^2 = 0.0025$ (Fig. 6c), resulting in an overall faster convergence for the estimation error (Fig. 6a). Furthermore, Fig. 6a also reports the ideal response of (9) with desired poles at σ_{max}^2 (dashed black line). We can then note the almost perfect match with case I (solid red line): indeed, as explained in Remark II.4, imposing a $\sigma_1^2(t) = const$ allows to exactly obtain the ideal behavior governed by (9). It is finally worth noting the accuracy of the reconstructed depth: Fig. 6a reports two vertical dashed lines indicating, for the two cases under consideration, the times $T_1 = 4.95\,\mathrm{s}$ and $T_2 = 9.85\,\mathrm{s}$ at which the estimation error $z(t)$ becomes smaller than $5\,\mathrm{mm}$. We then obtained a standard deviation of approx. 0.8 and $0.3\,\mathrm{mm}$ evaluated on a time window of $1\,\mathrm{s}$ after the

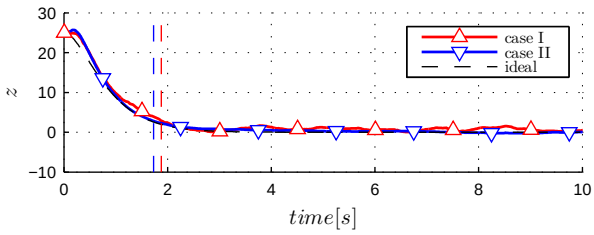
times T_1 and T_2, respectively. These results then also confirm the robustness of the proposed estimation approach despite the unavoidable presence of noise and discretization in the image acquisition. Note also that, as expected, the estimation error in the (active) case I reaches 'convergence' (i.e., drops below the threshold of 5 mm) significantly faster than case II ($T_1 < T_2$).

C. Structure Estimation for a Spherical Target

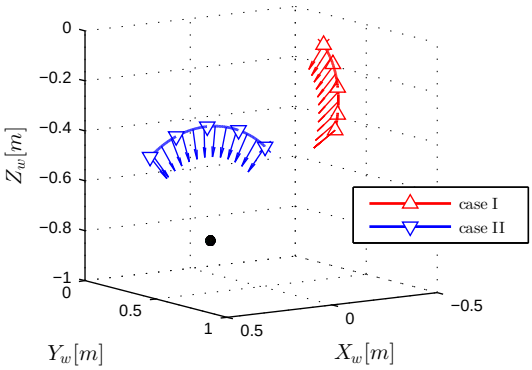
We now discuss some experimental results concerning the estimation of the radius of a spherical target: indeed, as explained in Sect. III-B, estimation of R allows to fully recover the sphere 3D position $\boldsymbol{P}_0 = \boldsymbol{s}R$ where vector $\boldsymbol{s}$ is directly obtainable from image measurements, see (22). As object to be tracked, we made use of a white table tennis ball placed on a black table and with a radius of 1.9 cm. As explained in Sect. III-B, 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: $\alpha\beta = 2 \cdot 10^3$ for gains $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$, $c_1 = c_1^* = 2\sqrt{\alpha\beta}\sigma_1$ for $\boldsymbol{D}_1$ in (6), and $\boldsymbol{v} = (-0.05,\, 0,\, 0)$ m/s for case I and $\boldsymbol{v} = (0,\, 0.045,\, 0.02)$ m/s for case II, with $\|\boldsymbol{v}\| = 0.05$ m/s in both cases. The camera angular velocity $\boldsymbol{\omega}$ was exploited to keep $(s_x,\, s_y) \simeq (0,\, 0)$ (centered sphere).

Figure 7a 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* (9) with the desired poles, i.e., by setting $\sigma_1^2 = \|\boldsymbol{v}\|^2 = const$ and $c_1 = c_1^*$ in (9) (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. 7b) which negatively affects the estimation task. The standard deviation of the radius estimation error, computed on a time window of 1 s after $z(t)$ has become smaller than 1 mm (vertical dashed lines in the plot), is 0.3 mm for case I and 0.2 mm for case II: we can note, again, the very satisfactory results obtained with the proposed estimation scheme in terms of accuracy of the reconstructed sphere radius. Note also how, in the two cases, the estimation error $z(t)$ drops below the threshold of 1 mm at essentially the same time, as expected (same error transient respose).

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 Sect. II-B clearly indicates the importance of choosing a proper value of c_1 for the damping matrix $\boldsymbol{D}_1$ in (6). 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 (9), respectively. The experimental results reported in Fig. 8 show that the behavior of the estimation error z (solid lines) has an excellent match with that of (9) (represented by dashed lines), thus fully confirming (*i*) the validity of the proposed theoretical analysis,



(a)



(b)

Fig. 7: 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). The vertical dashed lines indicate the times at which the estimation error $z(t)$ drops below the threshold of 1 mm. (b): camera trajectories for case I (blue line) and case II (red line) with arrows indicating the direction of the camera optical axis.

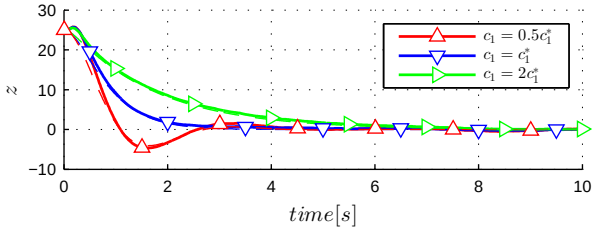


Fig. 8: Experimental results for the estimation of the radius of a sphere with $c_1 = c_1^*$ (blue line), $c_1 = 2c_1^*$ (green line) and $c_1 = 0.5c_1^*$ (red line). The dashed lines represent the response of an 'ideal' second order system with the corresponding poles. Note again the almost perfect match between the plots.

and (*ii*) the importance of choosing the 'right' damping matrix $\boldsymbol{D}_1$ for optimizing the convergence speed in addition to a proper regulation of σ_1^2.

D. Structure Estimation for a Cylindrical Target

In this final section we report some experimental results concerning the active estimation of the radius of a cylindrical object. Indeed, as in the sphere case, knowledge of R allows to fully recover the 3D point $\boldsymbol{P}_0 = R\boldsymbol{s}$, with vector $\boldsymbol{s}$ from (33)

and the cylinder axis $\boldsymbol{a}$ in (35) being directly obtainable from image measurements. For these experiments we used a white cardboard cylinder placed on a black table. The radius of the cylinder was approximately $4.2\,\text{cm}$.

In the cylinder case, the convergence rate of the estimation error depends both on the norm of the camera linear velocity $\boldsymbol{v}$ and on its direction w.r.t. the cylinder axis $\boldsymbol{a}$, see (38). 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 (39), maximization of $\sigma_1^2(t)$ w.r.t. vector $\boldsymbol{v}$ can be obtained by choosing

$$\dot{\boldsymbol{v}} = \boldsymbol{J}_{v,1}^T - \boldsymbol{J}_{v,1}^\dagger \boldsymbol{J}_{a,1} [\boldsymbol{a}]_\times \boldsymbol{\omega}, \tag{41}$$

with $\boldsymbol{A}^\dagger$ being the pseudoinverse of a 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, eq. (41) can be modified as

$$\dot{\boldsymbol{v}} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|^2} k_1 (\kappa_{des} - \kappa) + k_2 \left(\boldsymbol{I}_3 - \frac{\boldsymbol{v}\boldsymbol{v}^T}{\|\boldsymbol{v}\|^2} \right) (\boldsymbol{J}_{v,1}^T - \boldsymbol{J}_{v,1}^\dagger \boldsymbol{J}_{a,1} [\boldsymbol{a}]_\times \boldsymbol{\omega}), \tag{42}$$

with $k_1 > 0$ and $k_2 > 0$. Analogously to the point feature case, the first term in (42) asymptotically guarantees $\|\boldsymbol{v}(t)\| = \|\boldsymbol{v}_0\|$ while the second term projects (41) onto the null-space of the constraint $\|\boldsymbol{v}(t)\| = const$. 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 structured as follows: in the first experiment (case I), the update rule (42) 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 (42) 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 (42) is again fully implemented. This last case is meant to show how the convergence properties of the estimator are not affected by the direction of the camera linear velocity as long as it stays orthogonal to the cylinder axis $\boldsymbol{a}$.

The experiments were run with the following conditions: $\alpha\beta = 500$ for gains $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$, $c_1 = c_1^*$ for $\boldsymbol{D}_1$ in (6), $k_1 = 10$, $k_2 = 1$ for cases I and III, and $k_2 = 0$ for case II. As for the linear velocity, we set $\boldsymbol{v}(t_0) = \boldsymbol{v}_0 = (-0.01, 0.05, 0.05)$ m/s for cases I and II, and $\boldsymbol{v}(t_0) = \boldsymbol{v}_0 = (-0.05, 0.05, 0.01)$ m/s for case III (note how $\|\boldsymbol{v}_0\|^2 = 5.1 \times 10^{-3}$ m^2/s^2 in all three cases).

The behavior of $\sigma_1^2(t)$ is shown in Fig. 9a: as explained at the end of Sect. III-C, 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 $\boldsymbol{v}^T \boldsymbol{a} = 0$). 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 (42) results in a

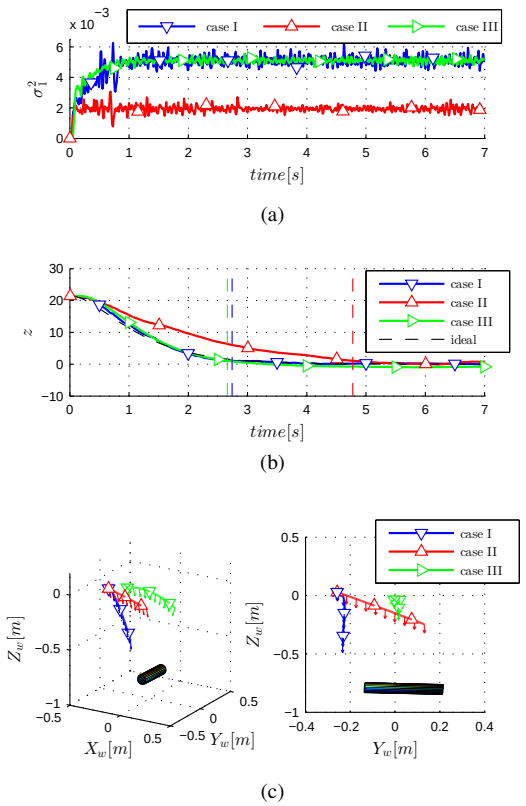


Fig. 9: Experimental results for the estimation of the radius of a cylinder with the following color coding: blue – case I, red – case II, green – case III. (a): behavior of $\sigma_1^2(t)$ for the three cases (coincident for cases I and III and larger than in case II). (b): behavior of $z(t)$. The three vertical dashed lines indicate the times $T_1 = 2.74\,\text{s}$, $T_2 = 4.78\,\text{s}$ and $T_3 = 2.66\,\text{s}$ at which the estimation error drops below the threshold of $2\,\text{mm}$. Note how $T_1 \approx T_3$ and $T_1 < T_2$ as expected. (c): two views of the camera trajectories for the three cases with arrows indicating the direction of the camera optical axis.

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. 9b: 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 (9) with desired poles (dashed black line). As expected, the response for case II (red line) is slower than in cases I and III. As in the point feature case, Fig. 9b reports, for the three cases under consideration, the times $T_1 = 2.74\,\text{s}$, $T_2 = 4.78\,\text{s}$ and $T_3 = 2.66\,\text{s}$ at which the estimation error drops below the threshold $2\,\text{mm}$ (vertical dashed lines). The standard deviation of the error evaluated on a time window of $1\,\text{s}$ after convergence has been 'reached' resulted in the values of 0.4, 0.6 and $0.7\,\text{mm}$, respectively. We can then appreciate, again, the high accuracy of the proposed approach in estimating the cylinder radius

R while also optimizing online for the camera motion. The higher estimation error in case III can be ascribed to the larger distance between the camera and the observed target, which increases the effect of discretization errors. Note also how $T_1 \approx T_3 < T_2$ thanks to the active optimization of the error convergence rate. Finally, Fig. 9c 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 $\boldsymbol{a}$.

V. CONCLUSIONS

In this paper we have addressed the problem of active SfM for recovering the 3D structure of a point feature and of spherical and cylindrical objects by exploiting a novel active estimation strategy tailored to the three cases under consideration. For the depth estimation of a point feature, two possibilities differing in the adopted projection model (planar or spherical) were proposed and critically compared. The results showed the complementarity of the two models in terms of attainable convergence rates and basin of attraction for the estimation error. In the spherical and cylindrical cases, we instead showed how an adequate choice of the measured visual features allows to reduce the SfM task to the estimation of a single unknown constant quantity (the sphere/cylinder radius R) in place of the classical (and time-varying) three parameters (scaled normal vector of the planar limb surface). Availability of this quantity allows to then retrieve the full 3D structure of the observed targets. The reported experimental results fully confirmed the validity of the theoretical analysis and, in particular, the ability of the proposed *active* estimation strategy to impose, in all three cases, a *desired transient response* to the estimation error equivalent to that of a reference linear second-order system with desired poles.

We are currently investigating the use of similar active strategies for dealing with more complex 3D scenes. A possibility in this sense could be to decompose the SfM problem in two phases by (*i*) extracting and classifying, possibly from an initial measurement in the form of a point cloud, a set of primitive shapes belonging to the classes described in our work (points, spheres, cylinders or other 3D geometries), or also to other classes such as 2D planar patches made of discrete/dense sets of points; (*ii*) performing an (active) estimation of the whole scene structure by applying the same strategy presented here and by either sequencing the estimation of single scene components, or considering an "extended" system obtained by concatenating the observable and unobservable components corresponding to each of the basic shapes/classes.

We are also investigating how to extend our solution to problems involving the estimation of more parameters than the number of available measurements (i.e., with $m < p$), thus requiring to fulfil the more general observability condition (4). Finally, we are also interested in the use of the proposed active strategy in the context of vision-based manipulation tasks. Some preliminary results in this context are reported in [32] where it is shown, and experimentally proven, that an online optimization of the estimation convergergence rate can improve the performance in executing visual servoing tasks.

APPENDIX A
TIME-DERIVATIVE OF THE LIMB SURFACE PARAMETERS FOR A SPHERICAL TARGET

Differentiation of $\boldsymbol{\chi}$ from (21) w.r.t. time yields

$$\dot{\boldsymbol{\chi}} = \frac{\dot{\boldsymbol{P}}_0 K - \boldsymbol{P}_0 \dot{K}}{K^2} = \frac{\dot{\boldsymbol{P}}_0 K - 2\boldsymbol{P}_0 \boldsymbol{P}_0^T \dot{\boldsymbol{P}}_0}{K^2} \tag{43}$$

which, being $\dot{\boldsymbol{P}}_0 = -\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{P}_0$ and exploiting the property $\boldsymbol{P}_0^T [\boldsymbol{\omega}]_\times \boldsymbol{P}_0 = 0$, can be rewritten as

$$\dot{\boldsymbol{\chi}} = -\frac{\boldsymbol{v}}{K} - \frac{[\boldsymbol{\omega}]_\times \boldsymbol{P}_0}{K} + 2\frac{\boldsymbol{P}_0 \boldsymbol{P}_0^T \boldsymbol{v}}{K^2} = -\frac{\boldsymbol{v}}{K} - [\boldsymbol{\omega}]_\times \boldsymbol{\chi} + 2\boldsymbol{\chi}\boldsymbol{\chi}^T \boldsymbol{v}. \tag{44}$$

Letting $s_z = Z_0/R$ ($s_z > 1$), one then has

$$\boldsymbol{\chi}^T \boldsymbol{\chi} - \frac{1}{s_z^2}\chi_z^2 = \frac{X_0^2 + Y_0^2 + Z_0^2}{K^2} - \frac{R^2}{Z_0^2}\frac{Z_0^2}{K^2} = \frac{1}{K}. \tag{45}$$

This then shows how $1/K$ can be expressed in terms of $\boldsymbol{\chi}$ and of s_z^2, with s_z being directly obtainable from image measurements, see (22).

APPENDIX B
ESTIMATION OF THE LIMB SURFACE PARAMETER FOR A CYLINDRICAL TARGET

In order to estimate the parameters of the limb surface associated to a cylindrical object, one could consider as measurement the $2+2$ angle-distance parameters $(\theta_i,\, \rho_i)$ of the straight lines resulting from the projection of the cylinder on the image plane. From [17], [27], the interaction matrix in this case is given by:

$$\boldsymbol{L} = \begin{bmatrix} \lambda_{\rho_1} c_1 & \lambda_{\rho_1} s_1 & -\lambda_{\rho_1}\rho_1 & (1+\rho_1^2)s_1 & -(1+\rho_1^2)c_1 & 0 \\ \lambda_{\theta_1} c_1 & \lambda_{\theta_1} s_1 & -\lambda_{\theta_1}\rho_1 & -\rho_1 c_1 & -\rho_1 s_1 & -1 \\ \lambda_{\rho_2} c_2 & \lambda_{\rho_2} s_2 & -\lambda_{\rho_2}\rho_2 & (1+\rho_2^2)s_2 & -(1+\rho_2^2)c_2 & 0 \\ \lambda_{\theta_2} c_2 & \lambda_{\theta_2} s_2 & -\lambda_{\theta_2}\rho_2 & -\rho_2 c_2 & -\rho_2 s_2 & -1 \end{bmatrix} \tag{46}$$

with $s_i = \sin\theta_i$, $c_i = \cos\theta_i$, and

$$\begin{cases} \lambda_{\rho_i} &= -\left(\chi_x \rho_i c_i + \chi_y \rho_i s_i + \chi_z\right) \\ \lambda_{\theta_i} &= \chi_y c_i - \chi_x s_i \end{cases}. \tag{47}$$

Therefore, being (46) and (47) linear in the unknown $\boldsymbol{\chi}$, one can again apply the estimation scheme (2) with $\boldsymbol{s}$ taken as the vector of measured quantities on the image plane, i.e., $\boldsymbol{s} = (\rho_1, \theta_1, \rho_2, \theta_2)$.

As for the dynamics of $\boldsymbol{\chi}$, since (21) still holds for a cylindrical object (see [27]), one can again exploit (43) with, however, in this case

$$\dot{\boldsymbol{P}}_0 = -\left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T\right)\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{P}_0$$

and thus

$$\dot{\boldsymbol{\chi}} = -\left(\frac{1}{K}\boldsymbol{I}_3 - 2\boldsymbol{\chi}\boldsymbol{\chi}^T\right)\left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T\right)\boldsymbol{v} - [\boldsymbol{\omega}]_\times \boldsymbol{\chi}.$$

Finally, one can invoke (45) in order to express $1/K$ as a function of $\boldsymbol{\chi}$ and s_z^2, with s_z being the third element of vector $\boldsymbol{s}$ in (33).

APPENDIX C
DERIVATION OF EQUATION (36)

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

$$\boldsymbol{r}_1 = \frac{[\boldsymbol{a}]_\times \boldsymbol{P}_0}{\|\boldsymbol{P}_0\|},\ d_1 = 0, \quad \boldsymbol{r}_2 = -\frac{\boldsymbol{P}_0}{\|\boldsymbol{P}_0\|},\ d_2 = \|\boldsymbol{P}_0\|, \quad (48)$$

see Fig. 3. 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 $\boldsymbol{a}$, we have $R\boldsymbol{r}_i^T \boldsymbol{s} - d_i = 0$, $i = 1, 2$ (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

$$\boldsymbol{r}_i^T \dot{\boldsymbol{s}} = \frac{1}{R}\dot{d}_i - \boldsymbol{s}^T \dot{\boldsymbol{r}}_i, \quad i = 1, 2. \quad (49)$$

Since $\dot{\boldsymbol{r}}_i = [\boldsymbol{r}_i]_\times \boldsymbol{\omega}$ and $\dot{d}_i = \boldsymbol{r}_i^T \boldsymbol{v}$ (see [10]), eq. (49) can be rewritten as

$$\boldsymbol{r}_i^T \dot{\boldsymbol{s}} = \frac{1}{R}\boldsymbol{r}_i^T \boldsymbol{v} - \boldsymbol{s}^T [\boldsymbol{r}_i]_\times \boldsymbol{\omega}, \quad i = 1, 2. \quad (50)$$

Finally, from $\boldsymbol{a}^T \boldsymbol{P}_0 = 0$ and $\boldsymbol{P}_0 = R\boldsymbol{s}$ we have $\boldsymbol{a}^T \boldsymbol{s} = 0$ implying that

$$\boldsymbol{a}^T \dot{\boldsymbol{s}} = -\boldsymbol{s}^T \dot{\boldsymbol{a}} = -\boldsymbol{s}^T [\boldsymbol{a}]_\times \boldsymbol{\omega}. \quad (51)$$

We now note that equations (50) and (51) provide three linear constraints for $\dot{\boldsymbol{s}}$ which, by using (48), can be rearranged in matrix form as the following linear system

$$\begin{bmatrix} \dfrac{\boldsymbol{P}_0^T}{\|\boldsymbol{P}_0\|} \\ \boldsymbol{a}^T \\ \dfrac{\left([\boldsymbol{a}]_\times \boldsymbol{P}_0\right)^T}{\|\boldsymbol{P}_0\|} \end{bmatrix} \dot{\boldsymbol{s}} = \frac{1}{R} \begin{bmatrix} \dfrac{\boldsymbol{P}_0^T}{\|\boldsymbol{P}_0\|}\boldsymbol{v} \\ -\boldsymbol{P}_0^T [\boldsymbol{a}]_\times \boldsymbol{\omega} \\ \|\boldsymbol{P}_0\| \boldsymbol{a}^T \boldsymbol{\omega} + \dfrac{\left([\boldsymbol{a}]_\times \boldsymbol{P}_0\right)^T}{\|\boldsymbol{P}_0\|}\boldsymbol{v} \end{bmatrix}. \quad (52)$$

It is easy to verify that the 3×3 matrix on the left hand side of (52) is orthonormal: by then solving (52) for $\dot{\boldsymbol{s}}$ and performing some simplifications we finally obtain the sought result

$$\dot{\boldsymbol{s}} = \begin{bmatrix} -\dfrac{1}{R}\left(\boldsymbol{I}_3 - \boldsymbol{a}\boldsymbol{a}^T\right) & [\boldsymbol{s}]_\times \end{bmatrix} \boldsymbol{u}. \quad (53)$$

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] S. Omari and G. Ducard, "Metric Visual-Inertial Navigation System Using Single Optical Flow Feature," in *2013 European Control Conference*, 2013, pp. 1310–1316.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] A. P. Dani, N. R. Fischer, and W. E. Dixon, "Single Camera Structure and Motion," *IEEE Trans. on Automatic Control*, vol. 57, no. 1, pp. 238–243, 2012.

[16] V. Grabe, H. H. Bülthoff, and P. Robuffo Giordano, "A Comparison of Scale Estimation Schemes for a Quadrotor UAV based on Optical Flow and IMU Measurements," in *2013 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013.

[17] 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.

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

[19] M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "Path Planning for Motion Dependent State Estimation on Micro Aerial Vehicles," in *2013 IEEE Int. Conf. on Robotics and Automation*, 2013, pp. 3926–3932.

[20] 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, pp. 7647–7653.

[21] R. Mahony, A. von Brasch, P. Corke, and T. Hamel, "Adaptive depth estimation in image based visual servo control of dynamic systems," in *2005 IEEE Conf. on Decision and Control*, 2005, pp. 5372–5378.

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

[23] 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.

[24] A. Albu-Schaffer, C. Ott, U. Frese, and G. Hirzinger, "Cartesian impedance control of redundant robots: recent results with the DLR-light-weight-arms," in *2003 IEEE Int. Conf. on Robotics and Automation*, 2003, pp. 3704–3709.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

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

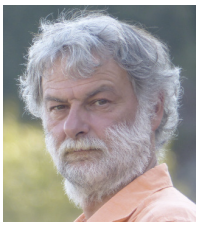
[31] E. Marchand, F. Spindler, and F. Chaumette, "ViSP for visual servoing: a generic software platform with a wide class of robot control skills," *IEEE Robotics and Automation Magazine*, vol. 12, no. 4, pp. 40–52, 2005.

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

Riccardo Spica (M'12) received his M.Sc. degree in Electronic Engineering from the University of Rome "La Sapienza" in 2012. He worked first as a Master's Student and later as a Graduate Research Assistant at the Max Planck Institute for Biological Cybernetics in Tuebingen, Germany for one year between 2012 an 2013. In December 2013 he started a Ph.D. in Signal Processing at the University of Rennes, France within the Lagadic group of Irisa and Inria. His research interests are in planning and control for robotics applications. In particular his current research deals with visual servoing and active structure from motion.

Paolo Robuffo Giordano (M'08) received his M.Sc. degree in Computer Science Engineering in 2001, and his Ph.D. degree in Systems Engineering in 2008, both from the University of Rome "La Sapienza". In 2007 and 2008 he spent one year as a PostDoc at the Institute of Robotics and Mechatronics of the German Aerospace Center (DLR), and from 2008 to 2012 he was Senior Research Scientist at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. Since 2012 he is a CNRS researcher in the Lagadic group at Irisa and Inria in Rennes, France.

Dr. Robuffo Giordano is currently Associate Editor of the IEEE Transactions on Robotics. His research interests span nonlinear control, robotics, planning, haptics and VR applications.



François Chaumette (M'02, SM'09, F'13) was graduated from École Nationale Supérieure de Mécanique, Nantes, France, in 1987. He received the Ph.D. degree in computer science from the University of Rennes, France, in 1990. Since 1990, he has been with Inria in Rennes where he is now senior research scientist and head of the Lagadic group (`http://www.irisa.fr/lagadic`). His research interests include robotics and computer vision, especially visual servoing and active perception.

Dr. Chaumette received the AFCET/CNRS Prize for the best French thesis in automatic control in 1991. He also received the 2002 King-Sun Fu Memorial Best IEEE Transactions on Robotics and Automation Paper Award. He has been Associate Editor of the IEEE Transactions on Robotics from 2001 to 2005 and is now in the Editorial Board of the Int. Journal of Robotics Research.