2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

Active Depth Estimation: Stability Analysis and its Applications

Rômulo T. Rodrigues[1], Pedro Miraldo[2], Dimos V. Dimarogonas[2], and A. Pedro Aguiar[1]

Abstract— **Recovering the 3D structure of the surrounding environment is an essential task in any vision-controlled Structure-from-Motion (SfM) scheme. This paper focuses on the theoretical properties of the SfM, known as the incremental active depth estimation. The term incremental stands for estimating the 3D structure of the scene over a chronological sequence of image frames. Active means that the camera actuation is such that it improves estimation performance. Starting from a known depth estimation filter, this paper presents the stability analysis of the filter in terms of the control inputs of the camera. By analyzing the convergence of the estimator using the Lyapunov theory, we relax the constraints on the projection of the 3D point in the image plane when compared to previous results. Nonetheless, our method is capable of dealing with the cameras' limited field-of-view constraints. The main results are validated through experiments with simulated data.**

I. INTRODUCTION

Structure-from-Motion (SfM) aims at recovering the 3D structure of the environment from a moving camera. It is used when the motion of the camera and its intrinsic parameters are known. This is one of the more important modules in applications such as: autonomous navigation [1], UAV flight control [2], robot hand-eye calibration [3], topographic surveying [4], and multi-robot relative pose estimation [5]. The SfM problem has been studied for the last three decades by the roboticists and computer vision researchers. Below, we categorize available solutions as geometric/filtering based methods and passive/active techniques.

Geometric-based techniques [6], [7], [8] often apply triangulation for estimating the depth of the points from two or more different viewpoints. The frames do not need to be consecutive, and this method is usually followed by an offline non-linear refinement such as bundle adjustment [9]. Geometric-based techniques provide accurate results but suffer from small baseline camera displacements. On the other hand, filter or incremental-based methods, such as [10], [11], [12], explicitly consider the dynamics of projected 3D points into a sequence of continuously acquired images. Incremental strategies focus on efficient computation and take advantage of the small continuous motions of the camera (small displacements). Besides, incremental-based techniques aim at getting a robust estimation of the model uncertainties.

The works mentioned in the previous paragraph are passive, i.e., the camera motion is not used to the goal of mapping the 3D environment. In the last decade, some authors have been studying the use of active vision techniques to assist the structure-from-motion modules. The authors in [13] propose the use of 3D reconstruction goals in the control loop. They use the proposed method in the reconstruction of 3D points, cylinders, straight lines, and spheres. In [14], the authors address an active strategy for tuning the transient response of a particular class of nonlinear observers that are well suited for active SfM problems. The technique is applied to a 3D point active SfM scheme. The framework was later used for the cases of cylinder, spheres (see [15]), 3D planes (in [16]), and 3D straight lines (see [17], [18]). There are also works on high level controllers based on SfM. For example, [19] presents a method to actively ensure the presence of good features in a structure-from-motion module, and [20] proposes an optimal path planning framework that maximizes the visual information during navigation.

In this paper, we study the stability analysis for an incremental active SfM using point features. The goal is to understand under what conditions it is possible to obtain an online estimation of the unknown depth of a point feature, from any initial condition. We resort to the knowledge of the motion of the camera and the 2D image plane coordinates of the projected 3D point. Our work builds on top of the incremental depth estimator addressed in [14], [15], where some guarantees for its stability and how to maximize its convergence speed were studied. However, for a point feature, the asymptotic stability result in [14], [15] only holds if 1) the camera motion drives the projection of the point to the origin of the image frame, and 2) the depth (unknown parameter being estimated) is constant after a transient. As a consequence, some issues arise in practical applications. For example, in [21], the results of [14] are applied to the coupled depth estimation and visual servo control problem. The strategy strives to increase the convergence speed, but the convergence properties are not met. This results from the requirement of translating the projection of a point to the origin of the image frame, which conflicts with the visual servoing goal.

In our work, we take a step back to first analyze the camera actuation policies that provide asymptotic stability guarantees on the depth estimation of a single feature. In

*This work was supported by PDMA-NORTE-08-5369-FSE-000061, UIDB/00147/2020 SYSTEC through the FCT/MCTES (PIDDAC); FCT projects POCI-01-0145-FEDER-031823; IMPROVE, POCI-01-0145-FEDER-031411-HARMONY; STRIDE NORTE-01-0145-FEDER-000033; LARSyS - FCT Plurianual funding 2020-2023; the Swedish research council (VR); Swedish Foundation for Strategic Research (SSF); and the Knut och Alice Wallenberg Foundation (KAW).

[1] R. T. Rodrigues and A. P. Aguiar are with the Research Center for Systems and Technologies (SYSTEC), Faculty of Engineering, University of Porto, Porto, Portugal.
E-Mail:rtr@fc.up.pt and pedro.aguiar@fe.up.pt.

[2] P. Miraldo and D. V. Dimarogonas are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden.
E-Mail:{miraldo,dimos}@kth.se.

978-1-7281-7395-5/20/$31.00 ©2020 IEEE

contrast to previous works with similar stability properties, we do not require the tracked feature to lie in (or visit) the origin of the image frame. Moreover, the unknown depth is not necessarily constant throughout the estimation process.

The next section presents the notations and background work. Section III presents the stability analysis of the active filter. Then, Sec. IV discusses its use in a single 3D point mapping application. Simulation results are shown in Sec. V, and Sec. VI concludes the paper.

II. PRELIMINARIES

This section presents notations and background work that support the remainder of this document.

A. Notation

Scalars are written in lower case letters and column vectors typed in bold symbol lower case letters. A vector can be split into smaller pieces using the notation $\mathbf{v}_{(i:j)} := [v_i, v_{i+1}, \dots, v_j]^T$. Matrices are printed in upper case letter, as well as the coordinates of a 3D Point.

B. Background

Consider a camera moving freely in space and let $\{C\}$ be the coordinate frame attached to the origin of the sensor. The camera observes a static 3D point described in $\{C\}$ as $\mathbf{p} := [X, Y, Z]^T \in \mathbb{R}^3$. Let $\mathbf{s} := [x, y]^T = [X/Z, Y/Z]^T \in \mathbb{R}^2$ be the projection of $\mathbf{p}$ into the camera's normalized image plane and consider the change of variable $\chi = 1/Z$. Applying the new variables in the well-known optical flow equation [22] gives

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\chi}\end{bmatrix} = \begin{bmatrix}-\chi & 0 & x\chi & xy & -(1+x^2) & y\\ 0 & -\chi & y\chi & 1+y^2 & -xy & -x\\ 0 & 0 & \chi^2 & y\chi & -x\chi & 0\end{bmatrix}\begin{bmatrix}\mathbf{v}\\ \mathbf{w}\end{bmatrix}, \tag{1}$$

where $\mathbf{v} := [v_x, v_y, v_z]^T \in \mathbb{R}^3$ and $\mathbf{w} := [w_x, w_y, w_z]^T \in \mathbb{R}^3$ are the camera linear and angular velocities described in $\{C\}$. The dynamics of the system can be stated in compact form

$$\begin{cases}\dot{\mathbf{s}} &= J_v\mathbf{v}\chi + J_w\mathbf{w}\\ \dot{\chi} &= J_q\mathbf{v}\chi^2 + J_l\mathbf{w}\chi\end{cases}, \tag{2}$$

where

$$\begin{cases} J_v = \begin{bmatrix}-1 & 0 & x\\ 0 & -1 & y\end{bmatrix}\\ J_w = \begin{bmatrix}xy & -(1+x^2) & y\\ 1+y^2 & -xy & -x\end{bmatrix}\\ J_q = \begin{bmatrix}0 & 0 & 1\end{bmatrix}, \qquad J_l = \begin{bmatrix}y & -x & 0\end{bmatrix}\end{cases}. \tag{3}$$

Given $\mathbf{s}, \mathbf{v}$, and $\mathbf{w}$, we want to estimate the unknown depth described by χ (also denoted as unmeasurable variable). For that, consider the following notations. The estimation variables are $\hat{\mathbf{s}}$ and $\hat{\chi}$. The respective estimation errors are $\tilde{\mathbf{s}} = \mathbf{s} - \hat{\mathbf{s}}$ and $\tilde{\chi} = \chi - \hat{\chi}$. The state estimation problem addressed here uses an observer similar to [14]:

$$\begin{cases}\dot{\hat{\mathbf{s}}} = J_v\mathbf{v}\hat{\chi} + J_w\mathbf{w} + k_s\tilde{\mathbf{s}}\\ \dot{\hat{\chi}} = J_q\mathbf{v}\hat{\chi}^2 + J_l\mathbf{w}\hat{\chi} + k_\chi(J_v\mathbf{v})^T\tilde{\mathbf{s}}\end{cases}, \tag{4}$$

where $k_s, k_\chi \in \mathbb{R}^+$ are the control gains. The corresponding estimation error dynamics is

$$\begin{cases}\dot{\tilde{\mathbf{s}}} = J_v\mathbf{v}\tilde{\chi} - k_s\tilde{\mathbf{s}}\\ \dot{\tilde{\chi}} = \tilde{\chi}(J_q\mathbf{v}(\chi + \hat{\chi}) + J_l\mathbf{w}) - k_\chi(J_v\mathbf{v})^T\tilde{\mathbf{s}}\end{cases}. \tag{5}$$

III. CONVERGENCE OF THE ESTIMATOR

In this section we provide the stability analysis of the depth estimation filter. The goal is to provide guarantees for the convergence of the unmeasurable depth for recovering the 3D structure of the world given by $\mathbf{p} = [\mathbf{s}, 1]/\chi$.

Assumption 1. *The observed 3D point cannot lie behind the camera. Consequently, we restrict our analysis to the domain where χ is positive, that is, we assume $\chi \geq 0, \forall t$.*

This assumption has an explicit physical meaning. In fact, cameras are not able to observe 3D points that are behind them. This would require a negative depth.

Theorem 1. *Consider the estimator* (4) *for the dynamic system* (2) *under Assumption 1. The equilibrium point $(\tilde{\mathbf{s}}, \tilde{\chi}) = \mathbf{0}$ is stable and the estimation error converges to zero as $t \to \infty$ provided that $\forall t \geq t_0$ the following constraints hold simultaneously:*

1) $J_l\mathbf{w} \leq 0$;

2) $\begin{cases} J_q\mathbf{v} \leq 0 \text{ , if } \hat{\chi} > 0\\ J_q\mathbf{v} = 0 \text{ , otherwise}\end{cases}$;

3) $\sigma^2 = (xv_z - v_x)^2 + (yv_z - v_y)^2 > 0$;

where $\mathbf{v}$, $\mathbf{w}$, and their time-derivatives are bounded signals.

Proof. Consider the Lyapunov function candidate

$$V(\tilde{\mathbf{s}}, \tilde{\chi}) = \frac{1}{2}\|\tilde{\mathbf{s}}\|^2 + \frac{1}{2k_\chi}\tilde{\chi}^2, \tag{6}$$

with $k_\chi > 0$, and its time-derivative

$$\dot{V} = \tilde{\mathbf{s}}^T\dot{\tilde{\mathbf{s}}} + \frac{1}{k_\chi}\tilde{\chi}\dot{\tilde{\chi}} \tag{7}$$

Substituting (5) in the previous equation:

$$\dot{V} = \tilde{\mathbf{s}}^T(J_v\mathbf{v}\tilde{\chi} - k_s\tilde{\mathbf{s}}) + \\ + \frac{1}{k_\chi}\tilde{\chi}(J_q\mathbf{v}(\chi + \hat{\chi})\tilde{\chi} + J_l\mathbf{w}\tilde{\chi} - k_\chi(J_v\mathbf{v})^T\tilde{s}) \tag{8}$$

$$= -\tilde{\mathbf{s}}^T k_s\tilde{\mathbf{s}} + \frac{1}{k_\chi}\tilde{\chi}J_q\mathbf{v}(\chi + \hat{\chi})\tilde{\chi} + \frac{1}{k_\chi}\tilde{\chi}J_l\mathbf{w}\tilde{\chi}. \tag{9}$$

By combining Assumption 1 and the input constraints stated in Theorem 1, we have that the three terms in the right-hand side of (9) are non-positive. Hence, $\dot{V} \leq 0$ and the equilibrium point $(\tilde{\mathbf{s}}, \tilde{\chi}) = \mathbf{0}$ is stable. We also conclude that $V(t) \leq V(t_0)$, and therefore, that the signals $\tilde{\mathbf{s}}$ and $\tilde{\chi}$ are bounded.

The critical case that precludes asserting asymptotically stability from (9) occurs when $J_q\mathbf{v} = 0$ and $J_l\mathbf{w} = 0$, and consequentially, $\dot{V} = -\tilde{\mathbf{s}}^T k_s\tilde{\mathbf{s}}$. Let $N(\cdot)$ denote the nullspace of a matrix, then $J_l\mathbf{w} = J_q\mathbf{v} = 0$ either because the feature lies in the origin of the image plane ($\mathbf{s} = [0, 0]^T$), or because $\mathbf{v} \in N(J_q)$ and $\mathbf{w} \in N(J_l)$ simultaneously. For $J_q\mathbf{v} = 0$ and

$J_l\mathbf{w} = 0$, the second derivative of the Lyapunov candidate function is

$$\ddot{V} = -2\tilde{\mathbf{s}}^T k_s \dot{\tilde{\mathbf{s}}} = -2\tilde{\mathbf{s}}^T k_s (J_v \mathbf{v}\tilde{\chi} - k_s\tilde{\mathbf{s}}). \tag{10}$$

As $\tilde{\mathbf{s}}$, $\tilde{\chi}$, and $\mathbf{v}$ (by definition) are bounded, the function $\ddot{V}$ is also bounded. Thus, $\dot{V}$ is uniformly continuous and from Barbalat's Lemma [23], we have that $\tilde{\mathbf{s}} \to 0$ as $t \to \infty$. Now, for the asymptotic behaviour of $\tilde{\chi}$ when $J_q\mathbf{v} = 0$ and $J_l\mathbf{w} = 0$, from (5) we have, as $t \to \infty$,

$$\begin{cases} \lim\limits_{t\to\infty} \dot{\tilde{\mathbf{s}}} = \lim\limits_{t\to\infty} J_v\mathbf{v}\tilde{\chi} \\ \lim\limits_{t\to\infty} \dot{\tilde{\chi}} = 0 \end{cases}. \tag{11}$$

The second equation states that the depth estimation error becomes a constant, but not necessarily zero. To show that indeed it will converge to zero, we first show that $\ddot{\tilde{\mathbf{s}}}$ is uniformly bounded because its time derivative given by

$$\begin{aligned} \ddot{\tilde{\mathbf{s}}} &= \dot{J}_v\mathbf{v}\tilde{\chi} + J_v\dot{\mathbf{v}}\tilde{\chi} + J_v\mathbf{v}\dot{\tilde{\chi}} - k_s\dot{\tilde{\mathbf{s}}} & (12)\\ &= \dot{J}_v\mathbf{v}\tilde{\chi} + J_v\dot{\mathbf{v}}\tilde{\chi} - k_\chi J_v\mathbf{v}(J_v\mathbf{v})^T\tilde{\mathbf{s}} - k_s J_v\mathbf{v}\tilde{\chi} + k_s^2\tilde{\mathbf{s}} & (13) \end{aligned}$$

is a function of bounded signals. Thus, since $\tilde{\mathbf{s}}$ converges to the origin and $\ddot{\tilde{\mathbf{s}}}$ is uniformly bounded, we conclude that $\dot{\tilde{\mathbf{s}}} \to 0$ as $t \to \infty$. Consequently, we have that

$$\lim_{t\to\infty} \dot{\tilde{\mathbf{s}}} = \lim_{t\to\infty} J_v\mathbf{v} \lim_{t\to\infty} \tilde{\chi} = 0. \tag{14}$$

It must be the case that either $\tilde{\chi} \to 0$ or $J_v\mathbf{v} \to 0$. If the function $J_v\mathbf{v}$ is persistently exciting through all time, then the depth estimation error converges to zero. The signal $J_v\mathbf{v}$ is persistently exciting if the integral

$$\int_{t_0}^{t} (J_v\mathbf{v})^T J_v\mathbf{v}\, d\tau \tag{15}$$

is positive definite $\forall t \geq t_0$. Hence, the persistency of excitation (PE) condition holds if

$$\sigma^2 = (J_v\mathbf{v})^T J_v\mathbf{v} > 0, \tag{16}$$

which is the case from condition (3) in Theorem 1.

Thus, one can now conclude that the equilibrium point $(\tilde{\mathbf{s}}^T, \tilde{\chi}) = \mathbf{0}$ is asymptotically stable. □

IV. CONSTRAINED ACTIVE DEPTH ESTIMATION

Any vision-based control scheme has to consider an important limitation of image sensors, its limited field of view. While tracking the projected 3D point (related to the unknown depth to be estimated), one needs to make sure the projection does not leave the image space. To achieve that, we have to include constraints on the motion of the camera. This section explores the theoretical stability guarantees derived in Sec. III for active depth estimation, while ensuring the tracked projected point does not leave the image space.

To address the constraints on the camera motion, we introduce the continuous and smooth desired signal $\mathbf{s}_{des}(t)$ and define the tracking error

$$\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{s}_{des}(t). \tag{17}$$

The signal $\mathbf{s}_{des}$ is chosen such that the feature remains within the field of view of the camera during the depth estimation process. Assume that the feedback control law $\boldsymbol{\pi}(t, \mathbf{s}, \mathbf{s}_{des})$ drives the tracking error to the origin[1], i.e., $\dot{\mathbf{s}} = \boldsymbol{\pi}, \forall t \geq t_0 \implies \mathbf{e} \to 0$ as $t \to \infty$. From inspection of (2), in addition to the camera's linear and angular velocities, $\dot{\mathbf{s}}$ depends on the unknown depth χ. Thus, it is only possible to shape the dynamics of $\dot{\mathbf{s}}$ up to an estimation error. That being said, the goal is to design a control law for $(\mathbf{v}, \mathbf{w})$ such that $\dot{\mathbf{s}}(\hat{\chi}, \mathbf{v}, \mathbf{w})$ tracks the signal $\boldsymbol{\pi}(t, \mathbf{s}, \mathbf{s}_{des})$, while:

(i) imposing the constraints stated in Theorem 1, to assure that the stability property holds;

(ii) improving the performance of the estimator, by maximizing σ^2 as defined in (16); and

(iii) accounting for the kinodynamics constraints of the camera described by $\|\mathbf{v}\| \leq v_{\max}$ and $\|\mathbf{w}\| \leq w_{\max}$, where $v_{\max}$ and $w_{\max}$ are the maximum linear and angular speed of the camera, respectively.

Since constraints are most commonly not addressed when designing a control law to track the reference signal $\mathbf{s}_{des}$, simultaneously tracking $\boldsymbol{\pi}$ and respecting all the forementioned constraints can lead to an infeasible problem. A workaround is proposed by introducing a scale factor $\lambda_\pi \in [0, 1]$ such that $\dot{\mathbf{s}}(\hat{\chi}, \mathbf{v}, \mathbf{w})$ is required to track the reference $\lambda_\pi\boldsymbol{\pi}$. As the depth converges, tracking the scaled vector $\lambda_\pi\boldsymbol{\pi}$ – rather than minimizing a norm error – ensures that the path of the feature in the image frame follows the assignment specified by $\boldsymbol{\pi}$. This allows us to design a path for the feature that does not visit the origin of the image frame. The problem is formulated next:

$$\begin{aligned} \underset{\mathbf{v},\mathbf{w},\lambda_\pi}{\text{maximize}} \quad & \lambda_\pi \\ \text{subject to} \quad & J_v\hat{\chi}\mathbf{v} + J_w\mathbf{w} = \lambda_\pi\boldsymbol{\pi} \\ & 0 \leq \lambda_\pi \leq 1 \\ & \text{constraints (i), (ii), and (iii)} \end{aligned} . \tag{18}$$

This problem is addressed in two configurations. The estimation strategy proposed in Section IV-A does not implicitly impose the unknown depth to be constant. In contrast, Sec. IV-B addresses the particular case that requires null depth rate. Both cases take advantage of the following Theorem:

Theorem 2. *Consider the non-convex problem:*

$$\begin{aligned} \underset{\lambda_1,\lambda_2,\mathbf{v}_r}{\textit{maximize}} \quad & \lambda_1 \\ \textit{subject to} \quad & \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 = r\mathbf{v}_r \\ & \|\mathbf{v}_r\| = 1 \\ & 0 \leq \lambda_1 \leq 1 \\ & -b \leq \lambda_2 \leq b \end{aligned} , \tag{19}$$

where $r, b \in \mathbb{R}^+, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, $\|\mathbf{v}_1\| > 0$, and $\|\mathbf{v}_2\| = 1$. The problem is always feasible if $r \leq b$.

[1] For instance, if $\mathbf{s}_{des}$ is constant, then the proportional controller $\boldsymbol{\pi} = -k_p(\mathbf{s} - \mathbf{s}_{des})$, where $k_p \in \mathbb{R}^+$ ensures the desired behaviour.

Due to the lack of space, the reader is referred to [24] for the proof of Theorem 2 and a closed form solution for the problem in (19), which is employed here. The solution does not impose restrictions on the feature coordinates, except the origin of the image frame, i.e., $\mathbf{s} = [0, 0]^T$, which is a singularity.

A. *Case:* $\mathbf{s} \neq \mathbf{0}, \forall t \geq t_0$

In this first scenario, $J_q\mathbf{v} = 0$ and $J_w\mathbf{w} \leq 0$. This allows us to to take advantage of Theorem 2, while still respecting the requirements for asymptotic convergence stated in Theorem 1. The PE condition of (16) simplifies to $\sigma^2 = v_x^2 + v_y^2 = \|\mathbf{v}_{(1:2)}\|^2$ and its maximum attainable value is limited by the kinodynamic constraint of the camera, $\sigma^2_{\max} = v_{\max}$. Under this scenario, the problem in (18) can be formulated as

$$\begin{aligned} \underset{\mathbf{v},\mathbf{w},\lambda_\pi}{\text{maximize}} \quad & \lambda_\pi \\ \text{subject to} \quad & \dot{\mathbf{s}}(\hat{\chi}, \mathbf{v}, \mathbf{w}) = \lambda_\pi \boldsymbol{\pi} \\ & 0 < \lambda_\pi \leq 1 \\ & J_q\mathbf{v} = 0,\ J_l\mathbf{w} \leq 0 \\ & \|\mathbf{v}\| = v_{\max},\ \|w\| \leq w_{\max} \end{aligned} \quad , \tag{20}$$

and solved with the following proposition:

Proposition 1. *Let the camera control input be*

$$\mathbf{v} = v_{\max} \begin{bmatrix} \mathbf{v}_r \\ 0 \end{bmatrix} \text{ and } \mathbf{w} = \begin{bmatrix} S\boldsymbol{\lambda}_s/\|\mathbf{s}\| \\ 0 \end{bmatrix}, \tag{21}$$

and S, $J_{\bar{w}}$, and $\boldsymbol{\lambda}_s$ be defined as follows:

$$\begin{cases} S = \begin{bmatrix} -\frac{\mathbf{s}_\perp}{\|\mathbf{s}_\perp\|} & \frac{\mathbf{s}}{\|\mathbf{s}\|} \end{bmatrix} \\ J_{\bar{w}} = \begin{bmatrix} xy & -(1+x^2) \\ 1+y^2 & -xy \end{bmatrix} \\ \boldsymbol{\lambda}_s = \begin{bmatrix} \lambda_{s_\perp} & \lambda_s \end{bmatrix}^T \end{cases} , \tag{22}$$

where $\lambda_{s_\perp} \in \mathbb{R}^+$, $\lambda_s \in \mathbb{R}$, and $\mathbf{s}_\perp = [-y, x]^T$ is a vector perpendicular to $\mathbf{s}$. *In particular, define $\boldsymbol{\lambda}_s$ as*

$$\boldsymbol{\lambda_s} = \begin{cases} \lambda_w \|\mathbf{s}\| (J_{\bar{w}}S)^{-1}\boldsymbol{\pi}/\|\boldsymbol{\pi}\|, \textit{ if } (\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} < 0 \\ \lambda_w \|\mathbf{s}\| [0, 1]^T, \textit{ otherwise} \end{cases} . \tag{23}$$

A sub-optimal solution for the problem in (20) *can be obtained by casting it in the shape of the problem in* (19)*, where the input variables are written as*

$$\begin{cases} \mathbf{v}_1 = -\boldsymbol{\pi} \\ \mathbf{v}_2 = \begin{cases} \boldsymbol{\pi}/\|\boldsymbol{\pi}\|, \textit{ if } (\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} < 0 \\ \mathbf{s}_\perp/\|\mathbf{s}_\perp\|, \textit{ otherwise} \end{cases} \\ r = \hat{\chi}v_{\max},\ b = w_{\max} \end{cases} , \tag{24}$$

and the outputs mapped into

$$\begin{cases} \lambda_\pi = \lambda_1^*,\ \lambda_w = \lambda_2^*; \\ \mathbf{v}_r = \mathbf{v}_r^* \end{cases} . \tag{25}$$

Proof. First, we show that the control inputs are described as in (21). The constraint $J_q\mathbf{v} = 0$ implies that $v_z = 0$. Combining with $\|\mathbf{v}\| = v_{\max}$, the linear velocity vector can be written as $\mathbf{v} = v_{\max}[\mathbf{v}_r^T, 0]^T$, where $\mathbf{v}_r \in \mathbb{R}^2$ is a unit vector. For the angular velocity, re-write the constraint $J_l\mathbf{w} \leq 0$ using the slack variable $\lambda_{s_\perp}$, such that

$$J_l\mathbf{w} \leq 0 \implies \begin{cases} J_l\mathbf{w} & = \lambda_{s_\perp} \\ \lambda_{s_\perp} & \leq 0 \end{cases} . \tag{26}$$

From $J_l\mathbf{w} = \lambda_{s_\perp}$ one concludes that $w_y = (y/x)w_x - (1/x)\lambda_{s_\perp}$. Applying this result into $J_w\mathbf{w}$:

$$J_w\mathbf{w} = J_w \begin{bmatrix} w_x \\ (y/x)w_x - (1/x)\lambda_{s_\perp} \\ w_z \end{bmatrix} \tag{27}$$

$$= \begin{bmatrix} -y/x & y \\ 1 & -x \end{bmatrix} \begin{bmatrix} w_x \\ w_z \end{bmatrix} + \begin{bmatrix} (1/x + x) \\ y \end{bmatrix} \lambda_{s_\perp}. \tag{28}$$

The column space of the first matrix on the right hand side of the previous equation has dimension 1 and, consequentially, it can be generated assuming $w_z = 0$. Thus, the following equivalence holds:

$$J_l\mathbf{w} = \lambda_{s_\perp} \implies -\mathbf{s}_\perp^T\mathbf{w}_{(1:2)} = \lambda_{s_\perp}, \tag{29}$$

where $\mathbf{s}_\perp = [-y, x]^T$. For $w_z = 0$, we conclude that any feasible angular velocity can be described as

$$\mathbf{w}_{(1:2)} = -\frac{\mathbf{s}_\perp}{\|\mathbf{s}_\perp\|^2}\lambda_{s_\perp} + \frac{\mathbf{s}}{\|\mathbf{s}\|^2}\lambda_s \tag{30}$$

$$= \frac{1}{\|\mathbf{s}\|} \begin{bmatrix} -\frac{\mathbf{s}_\perp}{\|\mathbf{s}_\perp\|} & \frac{\mathbf{s}}{\|\mathbf{s}\|} \end{bmatrix} \begin{bmatrix} \lambda_{s_\perp} \\ \lambda_s \end{bmatrix} \tag{31}$$

$$= \frac{1}{\|\mathbf{s}\|} S\boldsymbol{\lambda}_s, \tag{32}$$

where S, $\boldsymbol{\lambda}_s$, and λ_s are as defined in (22). Within this setup the kinodynamics constraint $\|\mathbf{w}\| \leq w_{\max}$ is equivalent to $\|\boldsymbol{\lambda}_s\| \leq \|\mathbf{s}\|w_{\max}$:

$$\|\mathbf{w}\| = \|\mathbf{w}_{(1:2)}\| = \frac{1}{\|\mathbf{s}\|}\sqrt{\boldsymbol{\lambda}_s^T S^T S \boldsymbol{\lambda}_s} \tag{33}$$

$$= \frac{\|\boldsymbol{\lambda}_s\|}{\|\mathbf{s}\|} \leq w_{\max}. \tag{34}$$

This concludes the proof of (21) and (22).

Applying the control inputs into the first constraint of (20) and re-organizing the terms yields:

$$\lambda_\pi(-\boldsymbol{\pi}) + J_w \begin{bmatrix} S\boldsymbol{\lambda}_s/\|\mathbf{s}\| \\ 0 \end{bmatrix} = -\hat{\chi}v_{\max}J_v \begin{bmatrix} \mathbf{v}_r \\ 0 \end{bmatrix} \tag{35}$$

$$\lambda_\pi(-\boldsymbol{\pi}) + \frac{1}{\|\mathbf{s}\|}J_{\bar{w}}S\boldsymbol{\lambda}_s = \hat{\chi}v_{\max}\mathbf{v}_r. \tag{36}$$

Let $\boldsymbol{\nu} = (1/\|\mathbf{s}\|)J_{\bar{w}}S\boldsymbol{\lambda}_s$ and notice that if $\|\boldsymbol{\pi}\| > \hat{\chi}v_{\max}$, $\boldsymbol{\lambda}_s$ must be such that $\boldsymbol{\pi}^T\boldsymbol{\nu} > 0$. On the contrary, if $\|\boldsymbol{\pi}\| < \hat{\chi}v_{\max}$, then one has to ensure $(-\boldsymbol{\pi})^T\boldsymbol{\nu} > 0$. Maximizing the dot product in both cases requires that $\boldsymbol{\nu}$ and $\boldsymbol{\pi}$ to be parallel. Both vectors are aligned if

$$\boldsymbol{\lambda}_s \propto (J_{\bar{w}}S)^{-1}\boldsymbol{\pi}, \tag{37}$$

where the symbol $\propto$ denotes the relationship holds up to a scale factor. The matrix S is orthogonal and, therefore, full rank. The matrix $J_{\bar{w}}$ is also full rank since $\det(J_{\bar{w}}) = 1+x^2+y^2 \neq 0$. From the Sylvester rank inequality, we have

$$\operatorname{rank}(S) + \operatorname{rank}(J_{\bar{w}}) - 2 \leq \operatorname{rank}(J_{\bar{w}}S). \tag{38}$$

Since both S and $J_{\bar{w}}$ are 2×2 full rank matrices, one concludes that their product is also full rank (and invertible).

For feasibility, the first component of $\boldsymbol{\lambda}_s$ – corresponding to $\lambda_{s_\perp}$ – must be non-positive. Solving the right hand side of (37), $\lambda_{s_\perp}$ can be described as

$$\lambda_{s,\perp} \propto \begin{cases} \mathbf{s}^T\boldsymbol{\pi}, \text{ if } \|\boldsymbol{\pi}\| > \hat{\chi}v_{\max} \\ -\mathbf{s}^T\boldsymbol{\pi}, \text{ if } \|\boldsymbol{\pi}\| \leq \hat{\chi}v_{\max} \end{cases}. \tag{39}$$

If $\lambda_{s,\perp}$ is positive in either cases, it means that $\lambda_{s_\perp} = 0$ is the largest feasible value that maximizes the projection of $\boldsymbol{\nu}$ into $\boldsymbol{\pi}$ or $(-\boldsymbol{\pi})$. Using a compact notation:

$$\boldsymbol{\lambda}_s = \begin{cases} \lambda_w\|\mathbf{s}\|(J_{\bar{w}}S)^{-1}\frac{\boldsymbol{\pi}}{\|\boldsymbol{\pi}\|}, \text{ if } (\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} < 0 \\ \lambda_w\|\mathbf{s}\|[0,1]^T, \text{ otherwise} \end{cases}, \tag{40}$$

where $\lambda_w \in \mathbb{R}$. For the maximum feasible value of λ_w, compute the norm of the previous equation and compare with (34). When $(\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} > 0$, we have

$$\|\boldsymbol{\lambda}_s\| = \frac{\|\lambda_w\|\|\mathbf{s}\|}{\|\boldsymbol{\pi}\|}\|(J_{\bar{w}}S)^{-1}\boldsymbol{\pi}\| \leq \|\mathbf{s}\|w_{\max}. \tag{41}$$

The singular values of $(J_{\bar{w}}S)^{-1}$ are 1 and $1/(1+x^2+y^2)$. Since the maximum singular value is 1, the upper bound $\|(J_{\bar{w}}S)^{-1}\boldsymbol{\pi}\| \leq \|\boldsymbol{\pi}\|$ holds and

$$\|\boldsymbol{\lambda}_s\| \leq \|\lambda_w\|\|\mathbf{s}\| \leq \|\mathbf{s}\|w_{\max}, \tag{42}$$

$$\|\lambda_w\| \leq w_{\max}. \tag{43}$$

The same bound is obtained when $\boldsymbol{\lambda}_s = \lambda_w\|\mathbf{s}\|\begin{bmatrix}0 & 1\end{bmatrix}^T$ in (40):

$$\|\lambda_w\|\mathbf{s}\|\begin{bmatrix}0 & 1\end{bmatrix}^T\| \leq \|\mathbf{s}\|w_{\max} \Rightarrow \|\lambda_w\| \leq w_{\max}. \tag{44}$$

Finally, substituting (40) in (36):

$$\hat{\chi}v_{\max}\mathbf{v}_r = \begin{cases} \lambda_\pi(-\boldsymbol{\pi}) + \lambda_w\frac{\boldsymbol{\pi}}{\|\boldsymbol{\pi}\|}, \text{ if } (\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} > 0 \\ \lambda_\pi(-\boldsymbol{\pi}) + \lambda_w\frac{\mathbf{s}_\perp}{\|\mathbf{s}_\perp\|}, \text{ otherwise} \end{cases}, \tag{45}$$

which allows us to obtain a sub-optimal solution for the problem in (46) in the shape of the problem in (19) using the substitutions described by (48) and (49). □

The sub-optimality comes from the fact that the solution consists in projecting $\boldsymbol{\lambda}_s$ into $\boldsymbol{\pi}$ when $(\|\boldsymbol{\pi}\| - \hat{\chi}v_{\max})\mathbf{s}^T\boldsymbol{\pi} < 0$. The projection is done via the mapping $J_{\bar{w}}S$. The singular values of $J_{\bar{w}}S$ are 1 and $1+x^2+y^2$. Therefore, if $\mathbf{s} \neq [0,0]^T$, there can exist a $\boldsymbol{\lambda}_s$ that is not projected into $\boldsymbol{\pi}$, but the shear transformation performed by $J_{\bar{w}}S$ allows for a higher value of λ_π. Since in practical applications $1+x^2+y^2 \approx 1$, the solution obtained is not far from the optimal solution. The main advantage in our approach is that it is possible to compute a direction for $\boldsymbol{\lambda}_s$ in a closed-form.

B. *Case:* $\mathbf{s} \neq \mathbf{0}$ *and* $\dot{\chi} = 0, \forall t \geq t_0$

Now, consider the specific scenario where the depth must be kept constant throughout the entire estimation process. For an unknown χ in (2), setting $J_q\mathbf{v} = 0$ and $J_l\mathbf{w} = 0$ guarantees that $\dot{\chi} = 0$. Both aforementioned constraints are in accordance with Theorem 1. The problem, which is stated next:

$$\begin{aligned} \underset{\mathbf{v},\mathbf{w},\lambda_\pi}{\text{maximize}} \quad & \lambda_\pi \\ \text{subject to} \quad & \dot{\mathbf{s}}(\hat{\chi},\mathbf{v},\mathbf{w}) = \lambda_\pi\boldsymbol{\pi} \\ & 0 \leq \lambda_\pi \leq 1 \\ & J_q\mathbf{v} = 0,\ J_l\mathbf{w} = 0 \\ & \|\mathbf{v}\| = v_{\max},\ \|w\| \leq w_{\max} \end{aligned}, \tag{46}$$

is a particular case of problem (20). According to the following corollary, an optimal solution can be obtained using Theorem 2.

Corollary 1. *Let the camera control input be described as*

$$\mathbf{v} = v_{\max}\begin{bmatrix}\mathbf{v}_r \\ 0\end{bmatrix} \text{ and } \mathbf{w} = \lambda_w\begin{bmatrix}\mathbf{s}/\|\mathbf{s}\| \\ 0\end{bmatrix}. \tag{47}$$

Then, the problem in (46) *is equivalent to the problem in* (19)*, where*

$$\begin{cases} \mathbf{v}_1 = -\boldsymbol{\pi},\ \mathbf{v}_2 = \mathbf{s}_\perp/\|\mathbf{s}_\perp\| \\ r = \hat{\chi}v_{\max},\ b = w_{\max} \end{cases}; \tag{48}$$

and the outputs are mapped as:

$$\begin{cases} \lambda_\pi = \lambda_1^*,\ \lambda_w = \lambda_2^* \\ \mathbf{v}_r = \mathbf{v}_r^* \end{cases}. \tag{49}$$

The proof is similar to the one presented in Sec. IV-A by imposing $\lambda_{s_\perp} = 0$, that is, no slackness. In this case, the solution is optimal because the shear mapping is not involved.

V. EXPERIMENTS

The theoretical results derived in this work are validated using a numerical simulator. The following fixed parameters were employed: $v_{\max} = 0.1$ m/s, $w_{\max} = 0.15$ rad/s, $k_s = 10$, and $k_\chi = 2500$. The sampling time of the simulations is 0.05 ms. In Fig. 1, we compare the methods proposed in Sec. IV-A and Sec. IV-B with the one presented in [14], [15]. For asymptotic stability, the strategy described in [14], [15] (continuous red line) and denoted here as *Spica et al. (2014)*, requires the projection of the tracked 3D point to lie in the origin of the image plane and its corresponding depth to be constant, i.e., $\mathbf{s}_{des} = [0,0]^T$ and $\dot{\chi} = 0$. The method presented in Sec. IV-A (dashed green line) relaxes both requirements. The strategy described in Sec. IV-B (continuous blue line) is a particular case of the previous method which keeps the unknown depth constant throughout the trajectory of the camera. Aiming at a fair comparison, the initial visual servoing error and the inverse depth estimation error are the same in the three cases. The initial configurations are: $\|\mathbf{e}(t_0)\| = 0.2$ m and $\tilde{\chi}(t_0) = 0.9$ m^{-1} (with $\chi(t_0) = 1$ m^{-1} and $\hat{\chi}(t_0) = 0.1$ m^{-1}).

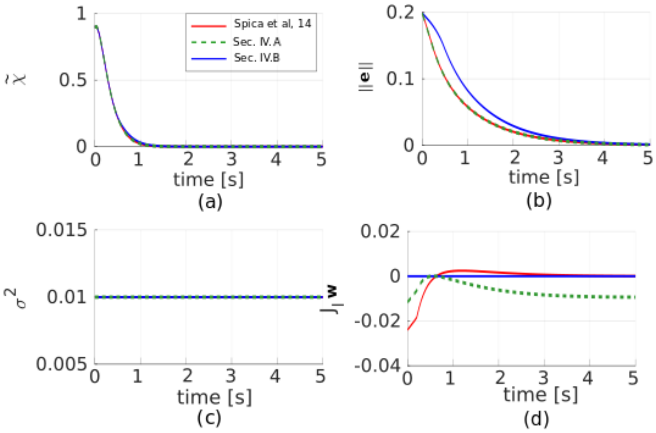

Fig. 1. Comparison of the estimation strategies described in [15] (*Spica et al. 14*), Sec. IV-A ($\dot{\chi} = 0$ relaxed), and Sec. IV-B ($\dot{\chi} = 0$) . The initial inverse depth estimation error is $\tilde{\chi} = 0.9$ m^{-1} and the initial tracking error is $\|\mathbf{e}\| = 0.2$ m. From top to bottom, it is shown the results of (a) the inverse depth estimation error, (b) the tracking error, (c) the persistence of excitation measurement σ^2, and (d) the constraint $J_l\mathbf{w}$ described in Theorem 1.

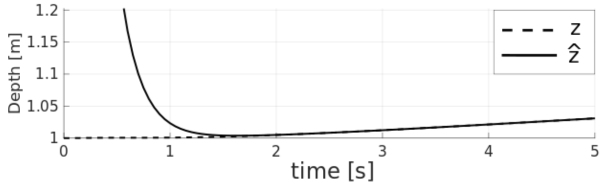

Fig. 2. True depth ($z = 1/\chi$) and its estimation ($\hat{z} = 1/\hat{\chi}$) using the strategy described in Sec. IV-A and the same setup as in Fig. 1

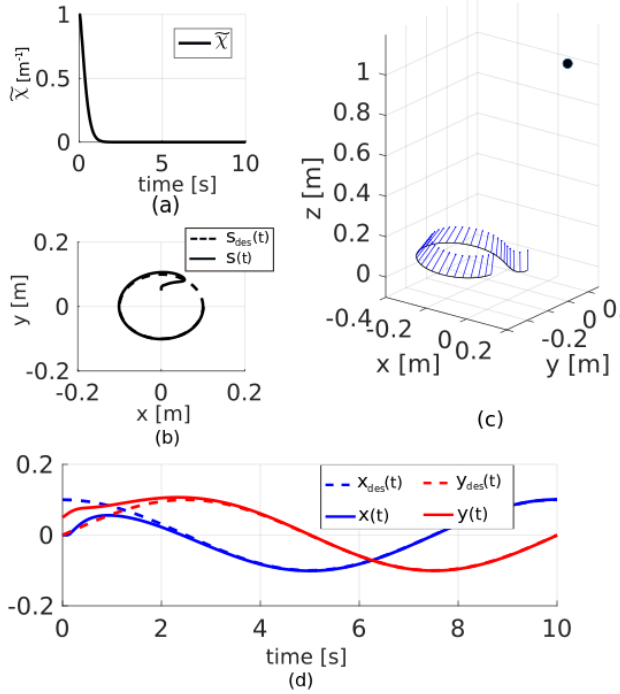

Fig. 3. Assessing the performance of the proposed depth estimation framework when the desired feature coordinates ($\mathbf{s}_{des}(t)$) is time-varying. (a) shows the depth estimation error, (b) shows the desired and the current projection of the 3D point in the image plane, (c) illustrates the trajectory of the camera in a black line, the z–axis in a blue arrow, and the 3D point in black, and (d) the two previous signals over time per axis.

The behaviour of the depth estimation error is almost the same for the three methods - see Fig. 1(a). In fact, as shown in Fig. 1(c), the three strategies continuously fulfill the PE condition, given by σ^2, at its maximum value. Figure 1(b) shows that the feature tracking error converges slower for the method described in Sec. IV-B. This is because the constraint $J_l\mathbf{w} = 0$ imposes severe limitations on the the angular velocity vector. *Spica et al. (2014)* guarantees asymptotic stability by driving the feature to the origin of the image frame, while the strategies proposed in this paper ensure that the constraints described in Theorem 1 hold throughout the entire estimation process regardless of the feature coordinate. In particular, the constraint associated to $J_l\mathbf{w}$ can be seen in Fig. 1(d). For the method in Sec. IV-A, $J_l\mathbf{w}$ is smaller or equal to zero. For the method in Sec. IV-B, the constraint is always zero.

For the same scenario, Fig. 2 shows the ground truth and the depth estimation using the method in Sec. IV-A. In contrast to other continuous estimation strategies presented in the literature (namely [15]), the method proposed in Sec. IV-A ensures the depth estimation error converges to zero even thought the depth of the point with respect to the camera is not constant throughout the entire estimation process.

In our formulation, the desired feature coordinate $\mathbf{s}_{des}$ can be time-varying. Figure 3 shows a scenario where the goal is to have the projection of the feature moving in a circular pattern. More specifically, we define $\mathbf{s}_{des} = 0.1[\cos(2\pi/10t), \sin(2\pi/10t)]^T$. As shown in Fig. 3(a), the speed of convergence of the depth estimation error does not change when compared to the previous case (constant $\mathbf{s}_{des}$). Finally, Fig. 3(b) and (c) show that while the depth estimation converges, the proposed control law is able to follow the time-varying signal $\mathbf{s}_{des}$.

VI. CONCLUSIONS

In this paper we analyze the required conditions for asymptotic stability of a class of depth estimation observers when the control inputs of the camera can be computed in an active manner. We applied the results for the depth estimation of a single 3D point. In contrast to previous works, our framework guarantees asymptotic stability when the feature coordinate does not converge to the origin of the image frame, nor its depth with respect to the camera is necessarily constant. We believe that relaxing the feature coordinates within the image frame while still providing asymptotic stability guarantees is paramount to apply incremental depth estimation in multiple point scenarios. Despite the relaxed constraints that allow a larger set of motions with theoretical guarantees for depth estimation, the numerical simulations shows that the proposed strategy performs similarly to related literature methods. In future work, we will extend our framework to multiple point and performs tests with a real robot/camera setup.

References

[1] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone, "Rover navigation using stereo ego-motion," *Robotics and Autonomous Systems (RAS)*, vol. 43, no. 4, pp. 215–229, 2003.

[2] N. H. M. Li and H. H. T. Liu, "Formation uav flight control using virtual structure and motion synchronization," in *American Control Conf. (ACC)*, 2008, pp. 1782–1787.

[3] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *The International Journal of Robotics Research (IJRR)*, vol. 20, no. 3, pp. 228–248, 2001.

[4] F. Clapuyt, V. Vanacker, and K. V. Oost, "Reproducibility of uav-based earth topography reconstructions based on structure-from-motion algorithms," *Geomorphology*, vol. 260, pp. 4–15, 2016.

[5] R. T. Rodrigues, P. Miraldo, D. V. Dimarogonas, and A. P. Aguiar, "A framework for depth estimation and relative localization of ground robots using computer vision," in *IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS)*, 2019.

[6] J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," *J. Opt. Soc. Am. A*, vol. 8, no. 2, pp. 377–385, 1991.

[7] A. Bartoli and P. Sturm, "Structure-from-motion using lines: Representation, triangulation, and bundle adjustment," *Computer Vision and Image Understanding (CVIU)*, vol. 100, no. 3, pp. 416–441, 2005.

[8] J. L. Schnberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104–4113.

[9] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment — a modern synthesis," in *Vision Algorithms: Theory and Practice*, 2000, pp. 298–372.

[10] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Trans. Robotics (T-RO)*, vol. 24, no. 5, pp. 932–945, 2008.

[11] A. D. Luca, G. Oriolo, and P. R. Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *The International Journal of Robotics Research (IJRR)*, vol. 27, no. 10, pp. 1093–1116, 2008.

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

[13] F. Chaumette, S. Boukir, P. Bouthemy, , and D. Juvin, "Structure from controlled motion," *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 18, no. 5, pp. 492–504, 1996.

[14] R. Spica and P. Robuffo Giordano, "A framework for active estimation: Application to structure from motion," in *IEEE Conf. Decision and Control (CDC)*, 2013, pp. 7647–7653.

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

[16] ——, "Plane estimation by active vision from point features and image moments," in *IEEE Int'l Conf. Robotics and Automation (ICRA)*, 2015, pp. 6003–6010.

[17] A. Mateus, O. Tahri, and P. Miraldo, "Active structure-from-motion for 3d straight lines," in *IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS)*, 2018, pp. 5819–5825.

[18] ——, "Active estimation of 3d lines in spherical coordinates," in *American Control Conf. (ACC)*, 2019, to appear.

[19] R. T. Rodrigues, M. Basiri, A. P. Aguiar, and P. Miraldo, "Low-level active visual navigation: Increasing robustness of vision-based localization using potential fields," *IEEE Robotis and Automation Letters (RA-L)*, vol. 3, no. 3, pp. 2079–2086, 2018.

[20] G. Costante, J. Delmerico, M. Werlberger, P. Valigi, and D. Scaramuzza, *Exploiting Photometric Information for Planning Under Uncertainty*. Springer, 2018, vol. 1, pp. 107–124.

[21] R. Spica, P. R. Giordano, and F. Chaumette, "Coupling active depth estimation and visual servoing via a large projection operator," *The International Journal of Robotics Research*, vol. 36, no. 11, pp. 1177–1194, 2017.

[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[23] J.-J. E. Slotine and W. Li, *Applied nonlinear control*. Prentice-Hall, 1991.

[24] R. T. Rodrigues, P. Miraldo, D. V. Dimarogonas, and A. P. Aguiar, "On the Guarantees of Incremental Depth Estimation and its Applications in Visual Servoing (Proof of Theorem 2) – available here: https://c2sr.fe.up.pt/TechReports/ReportICRA20.pdf," Universidade do Porto, Tech. Rep., 09 2019.