Robust Structure from Motion observer: Input to State Stability approach

Hichem Arioui, Lamri Nehaoua, Hicham Hadj-Abdelkader

Robust Structure from Motion observer: Input to State Stability approach

Hichem Arioui * **Lamri Nehaoua** * **Hicham Hadj-Abdelkader** *

* *Laboratoire IBISC, Université Paris-Saclay, 40 Rue de Pelvoux, Essonne 91080 France (e-mail: hichem.arioui@univ-evry.fr)*

Abstract: The authors present a novel nonlinear Thau-Luenberger observer for estimating Structure from Motion using a calibrated camera. Accurate reconstruction of the 3D structure of a scene relies on precise estimation of the camera's translational and angular velocities, which can be challenging for cameras on mobile platforms. The proposed observer aims to estimate these velocities robustly in the presence of measurement noise, with stability characterized through Input to State Stability analysis and Lyapunov theory. The stability conditions are determined using optimization techniques based on Linear Matrix Inequalities. The performance of the proposed approach is validated through simulation and experimental data, demonstrating its effectiveness in recovering the depth of tracked features and its robustness against disturbances.

Keywords: Nonlinear Observer, Structure from Motion, Linear Matrix Inequalities.

1. INTRODUCTION

The 3D structure of a scene can be reconstructed from the depth information of features in images captured by a moving camera. This process, known as *Structure from Motion* (SfM), remains a significant area of research in computer vision. Currently, proposed solutions for SfM can be categorized as either *offline*, also referred to as *batch*, or *online* approaches (Oliensis, 2000). In batch schemes, the structure is estimated by processing all the information acquired from an image sequence using nonlinear optimization, commonly referred to as Bundle Adjustment (Agarwal et al., 2010), or matrix factorization (Wang and Jonathan Wu, 2011). However, these methods are computationally intensive and pose challenges for real-time applications. Additionally, they often require subsequent iterative or online methods to ensure convergence.

Online methods formulate SfM as a differential equation, where the image dynamics are derived from point sets extracted from a continuous image sequence. Estimates are obtained using recursive filtering methods or deterministic observers. Filtering techniques, such as the Kalman Filtering, Extended Kalman Filter (EKF) (Civera et al., 2012), Unscented Kalman Filter (Omari and Ducard, 2013), and Particle Filter (Pupilli and Calway, 2006), have been shown to be efficient methods for estimating 3D structure, and have been developed in the context of Visual-Monocular Simultaneous Localization And Mapping (VM-SLAM). As a result, the distinction between SfM and VM-SLAM is ambiguous, depending on the field where these terms are used: SfM is more popular in the computer vision community, whereas VM-SLAM is gaining increased interest in robotics research (Chhaya et al., 2016). However, most filtering techniques do not guarantee convergence, and their performance relies on an accurate a priori noise model. Furthermore, their linearized dynamics do not sufficiently approximate highly nonlinear systems. To overcome these limitations, different approaches have been proposed, relying on the design of nonlinear deterministic observers with theoretical stability guarantees. For example, in (Lowe, 2004), feature point depth estimation is addressed using an active strategy to optimize the convergence rate of the error by acting on the camera translational velocity. Similarly, geometric primitives such as lines (Oliensis, 2000), spherical and cylindrical targets (Zhang et al., 2006), and moments (Agarwal et al., 2010) have been introduced in the context of active SfM.

Active SfM typically require *a priori* knowledge of camera velocity for accurate reconstruction of the 3D structure of a scene. Camera velocity can be precisely measured and controlled using methods such as differential kinematics, joint encoders, and motors when the camera is mounted on the end effector of a fixed base robotic manipulator. However, this becomes considerably more challenging for mobile or flying robots, as they are subject to significant dynamics and non-holonomic constraints (Spica, 2015). To address this issue, several approaches have been proposed, including obtaining partial measurements/estimations of camera linear/angular velocity (Keshavan and Humbert, 2017), or using rotation-invariant features to mitigate the effects of disturbances on angular velocity.

In this work, a nonlinear deterministic observer is designed for feature depth estimation. The stability conditions derived are formulated in terms of LMI constraints. The robustness of the observer to the estimation error uses the Input-to-State (ISS) stability paradigm to minimize the measurement perturbations of the camera velocity input and unreliable feature tracking. The performance of the proposed scheme is validated through experimental results using a mobile platform (Spica, 2015). Additionally, the performance is compared with the recent work presented in (Tahri et al., 2017) which exploits rotationally invariant features to design a robust active vision scheme based on two transformations: the first transformation efficiently decouples the nonlinear model, whereas the second transforms the decoupled model into *Brunovsky's linear form.*

The paper is organized as follows: Section 2 summarizes the vision model expressed here in quasi-Linear Parameter Varying (qLPV) form and discusses the observability conditions. Section 3 addresses the design of the non-linear observer to estimate the depth information. The simulation and experimental results demonstrating the performance of the presented work are evaluated in section 4. Finally section 5 summarizes the conclusions of this work.

2. MODEL DESCRIPTION

This work adopts the following notations: matrices are represented in bold uppercase $\mathbf{X}$ and vectors in bold lowercase $\mathbf{x}$. The symmetric matrix property is expressed as $\mathbf{X} = \mathbf{X}^\top$ and $(.)^{-1}$ is the inverse of a square matrix. $\mathbf{X} \succ 0$ (resp. $\mathbf{X} \prec 0$) means that $\mathbf{X}$ is a positive definite (resp. negative definite) matrix. $\mathbf{0}$ and $\mathbf{I}$ denote to null and identity matrices of appropriate dimensions. $[.]_\times$ is the skew-symmetric matrix operator. $\|(.)\|_2$ and $\|(.)\|_\infty$ are the Euclidean and infinity norms respectively. Lastly, $\bar{\mathbf{x}}$ is the representation of $\mathbf{x}$ in homogeneous coordinates.
This section introduces the *Takagi-Sugeno* transformation utilized for observer synthesis. Subsequently, an analysis of the uniform observability conditions is presented.

2.1 Camera-Object Relative Motion Model

Consider a 3D point $\mathbf{p}$ of coordinates defined in the camera frame $\mathcal{F}_c$ as ${}^c\mathbf{p} = (X^c\ Y^c\ Z^c)^\top \in \mathbb{R}^3$ and defined in the world frame $\mathcal{F}_w$ as ${}^w\bar{\mathbf{p}} = (X\ Y\ Z\ 1)^\top$. Its projection into a 2D point on the image plane with normalized homogeneous image coordinates, given by $\bar{\mathbf{m}} = (x\ y\ 1)^\top$. The projection is expressed by the pinhole equations:

$$x = \frac{X^c}{Z^c} \qquad y = \frac{Y^c}{Z^c} \tag{1}$$

The *pose* of $\mathcal{F}_c$ with respect $\mathcal{F}_w$ is defined in terms of the *camera extrinsic parameters*, encapsulated by the homogeneous transformation matrix $\mathbf{T}_{wc}$.

$$\mathbf{T}_{wc} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \tag{2}$$

where $\mathbf{R} \in \mathbf{SE}(3)$ is the rotation matrix $\mathcal{F}_c$ between and $\mathcal{F}_w$ and $\mathbf{t} \in \mathbb{R}^3$ represents transalational displacement between $\mathcal{F}_c$ and $\mathcal{F}_w$ The inverse of $\mathbf{T}_{wc}$, denoted by $\mathbf{T}_{cw} = \mathbf{T}_{wc}^{-1}$ takes the following form:

$$\mathbf{T}_{cw} = \begin{pmatrix} \mathbf{R}^\top & -\mathbf{R}^\top \mathbf{t} \\ 0 & 1 \end{pmatrix} \tag{3}$$

Then, the mapping coordinates from $\mathcal{F}_w$ to $\mathcal{F}_c$ is given by:

$${}^c\bar{\mathbf{p}} = \mathbf{T}_{cw}\,{}^w\bar{\mathbf{p}} \tag{4}$$

Time derivative of (4) to obtain the velocity of $\mathbf{p}$ in $\mathcal{F}_c$:

$$\begin{aligned} {}^c\dot{\bar{\mathbf{p}}} &= \dot{\mathbf{T}}_{cw}\,{}^w\bar{\mathbf{p}} + \mathbf{T}_{cw}\,{}^w\dot{\bar{\mathbf{p}}} \\ &= \dot{\mathbf{T}}_{cw}\,\mathbf{T}_{cw}\,{}^c\bar{\mathbf{p}} + \mathbf{T}_{cw}\,{}^w\dot{\bar{\mathbf{p}}} \end{aligned} \tag{5}$$

which can be written as:

$${}^c\dot{\bar{\mathbf{p}}} = \bar{\mathbf{V}}^c\ \bar{\mathbf{p}} + \mathbf{T}_{cw}\,{}^w\dot{\bar{\mathbf{p}}} \tag{6}$$

where ${}^w\dot{\bar{\mathbf{p}}}$ is the homogeneous representation of the absolute velocity of $\mathbf{p}$ in $\mathcal{F}_w$, and $\bar{\mathbf{V}}^c = \dot{\mathbf{T}}_{cw}\,\mathbf{T}_{cw}{}^c$, called *twist*, is an element of the Lie group $\mathbf{SE}(3)$ representing the velocity of $\mathcal{F}_w$ with respect to $\mathcal{F}_c$ in $\mathcal{F}_c$. Hence, the homogeneous matrix representation of $\bar{\mathbf{V}}^c$ is given by:

$$\bar{\mathbf{V}}^c = \begin{pmatrix} \dot{\mathbf{R}}^\top \mathbf{R} & -\mathbf{R}^\top \dot{\mathbf{t}} \\ 0 & 1 \end{pmatrix} \tag{7}$$

$$= \begin{pmatrix} [\omega]_\times & -\upsilon \\ 0 & 1 \end{pmatrix} \tag{8}$$

with $\upsilon = (\upsilon_x\ \upsilon_y\ \upsilon_z)^\top$ and $\omega = (\omega_x\ \omega_y\ \omega_z)^\top$ are the instantaneous translational and angular velocity of the camera frame origin. Thus, (6) is equivalent to:

$${}^c\dot{\bar{\mathbf{p}}} = \begin{pmatrix} \mathbf{T}_{cw}\,{}^w\dot{\mathbf{p}} - \upsilon + [\omega]_\times\,\mathbf{p} \\ 0 \end{pmatrix}. \tag{9}$$

with ${}^w\dot{\mathbf{p}}$ representing the translational velocity of $\mathbf{p}$ in $\mathcal{F}_w$ which is always assumed to be zero since the feature is considered fixed in the scene. Expanding (9), the following matrix is obtained where $\mathbf{u} = \begin{pmatrix}\upsilon^\top\ \omega^\top\end{pmatrix}^\top$ is the measured spatial velocity of the camera motion:

$${}^c\dot{\mathbf{p}} = \begin{pmatrix} -1 & 0 & 0 & 0 & -Z^c & Y^c \\ 0 & -1 & 0 & Z^c & 0 & -X^c \\ 0 & 0 & -1 & -Y^c & X^c & 0 \end{pmatrix} \mathbf{u} \tag{10}$$

The motion of the feature $\mathbf{m}$ is related to the velocity of the camera by an interaction matrix which itself depends on the current value of the features and other 3D information (depth). From the equations (1) and (10), one can deduce that the time derivative of the image point $\mathbf{m}$ is linked to the camera spatial velocity $\mathbf{u}$ by the following interaction matrix (Tahri et al., 2015):

$$\dot{\mathbf{m}} = \begin{pmatrix} -\frac{1}{Z^c} & 0 & \frac{x}{Z^c} & x\,y & -(1+x^2) & y \\ 0 & -\frac{1}{Z^c} & \frac{y}{Z^c} & (1+y^2) & -x\,y & -x \end{pmatrix} \mathbf{u} \tag{11}$$

The dynamics of the depth information $\frac{1}{Z}$ denoted by χ is deduced from the equation (10):

$$\dot{\chi} = \begin{pmatrix} 0 & 0 & \chi^2 & y\chi & -x\chi & 0 \end{pmatrix} \mathbf{u} \tag{12}$$

Let $\mathrm{x} = (\mathbf{m}^\top, \chi)$ denote the state vector where $\mathbf{m} = (x\ y)^\top$ is a *measurable* vector, and $\chi = \frac{1}{Z^c}$ is the *unmeasurable* 3D information to be estimated. A moving camera observing a scene induces a motion in the image plane and hence in the features. Point correspondences (the *measurable* variables) in an image sequence can be computed using feature-tracking techniques (Bay et al., 2006).

2.2 T-S Representation

The objective of this work is to design a nonlinear observer which robustly estimates the depth information χ during the camera motion in the presence of measurement noise. For this purpose, a new system representation of the state vector is adopted x expressed as a qLPV model:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{A}(\mathbf{x},\mathbf{u})\,\mathbf{x} + \mathbf{B}(\mathbf{y})\,\omega \\ \mathbf{y} = \mathbf{C}\mathbf{x} \end{cases} \tag{13}$$

where:

$$\mathbf{A}(\mathbf{x},\mathbf{u}) = \begin{pmatrix} 0 & 0 & -v_x + x v_z \\ 0 & 0 & -v_y + y v_z \\ 0 & 0 & \chi v_z + y w_x - x w_y \end{pmatrix}$$

$$\mathbf{B}(\mathbf{y}) = \begin{pmatrix} xy & -(1+x^2) & y \\ 1+y^2 & -xy & -x \\ 0 & 0 & 0 \end{pmatrix}$$

and $\mathbf{y}$ represents the output of the system:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Furthermore, the qLPV system is transformed into TS form with unmeasured premise variables. Following the sector nonlinearity approach (Kazuo and Wang, 2001), the TS model of the system is formulated as:

$$\begin{cases} \dot{\mathbf{x}} = \sum_{i=1}^{r} \mu_i(\mathbf{x})\big(\mathbf{A}_i\,\mathbf{x} + \mathbf{B}(\mathbf{y})\,\omega\big) \\ \mathbf{y} = \mathbf{C}\mathbf{x} \end{cases} \tag{14}$$

where $A_i \in \mathbb{R}^{3\times3}$, $B \in \mathbb{R}^{3\times3}$ and $\mathbf{x} \in \mathbb{R}^3$ is the state vector. The weighting functions μ_i, $i = 1,\ldots,r$ satisfy the following convex sum property:

$$\begin{cases} 0 \leqslant \mu_i \leqslant 1 \\ \sum_{i=1}^{r} \mu_i = 1 \end{cases} \tag{15}$$

with r is the number of the sub-models defined by the following three system nonlinearities:

$$\begin{cases} h_1 = -v_x + xv_z \\ h_2 = -v_y + yv_z \\ h_3 = \chi v_z + yw_x - xw_y \end{cases} \tag{16}$$

These terms are assumed to be bounded, so that the TS model obtained is a weighted sum of 8 sub-models $r = 2^3$ corresponding to the nonlinear model on the compact set.

3. OBSERVER DESIGN

This section develops the estimation of the unmeasured state by considering the given polytopic observer:

$$\begin{cases} \dot{\hat{\mathbf{x}}} = \sum_{i=1}^{8} \mu_i(\hat{\mathbf{x}})\big(\mathbf{A}_i\hat{\mathbf{x}} + \mathbf{L}_i(\mathbf{y} - \hat{\mathbf{y}})\big) + \mathbf{B}\,\omega \\ \hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} \end{cases} \tag{17}$$

The estimated state and output vector are respectively denoted by $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. A state estimator is developed for the system in equation (13) under the following assumptions.

Assumption 1. We consider tha a frozen values of parameters (Sename and Fergani, 2017):

(1) the state vector and the system inputs are bounded.
(2) the system is structurally uniformly observable.

The aim is to compute the gain $\mathbf{L}_i$ to minimize the following estimation error.

$$\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t) \tag{18}$$

The time dependency is omitted for brevity. The dynamics model of the error is:

$$\begin{aligned} \dot{\mathbf{e}} &= \dot{\mathbf{x}} - \dot{\hat{\mathbf{x}}} \\ &= \sum_{i=1}^{8} \mu_i\big(\mathbf{x}\big)\big(\mathbf{A}_i\mathbf{x} + \mathbf{B}\,\omega\big) - \sum_{i=1}^{8} \mu_i(\hat{\mathbf{x}})\big(\mathbf{A}_i\hat{\mathbf{x}} + \mathbf{B}\,\omega + \mathbf{L}_i(\mathbf{y} - \hat{\mathbf{y}})\big) \\ &= \sum_{i=1}^{8} \mu_i(\mathbf{x})\mathbf{A}_i\mathbf{x} - \sum_{i=1}^{8} \mu_i(\hat{\mathbf{x}})\big(\mathbf{A}_i\hat{\mathbf{x}} + \mathbf{L}_i(\mathbf{y} - \hat{\mathbf{y}})\big) \end{aligned} \tag{19}$$

it follows:

$$\begin{aligned} \dot{\mathbf{e}} &= \sum_{i=1}^{8} \mu_i(\hat{\mathbf{x}})(\mathbf{A}_i - \mathbf{L}_i\mathbf{C})\mathbf{e} + \sum_{i=1}^{8}(\mu_i(\mathbf{x}) - \mu_i(\hat{\mathbf{x}}))\mathbf{A}_i\mathbf{x} \\ &= \mathscr{A}_{\mathbf{e}}\mathbf{e} + \Delta(\hat{\mathbf{x}}, \mathbf{x}) \end{aligned} \tag{20}$$

To satisfy the asymptotic stability of (19), let the following assumptions hold true (Bergsten and Palm, 2000):

(1) the system $\mathscr{A}_{\mathbf{e}}$ is Hurwitz.
(2) $\Delta(\hat{\mathbf{x}}, \mathbf{x})$ is a vanishing disturbance i.e:

$$\Delta(\hat{\mathbf{x}}, \mathbf{x}) \to 0 \text{ when } \hat{\mathbf{x}} \to \mathbf{x}$$

Theorem 1. Assuming that the system nonlinearities are bounded, the error dynamics in (19) are ISS with respect to $\Delta(\hat{\mathbf{x}}, \mathbf{x})$ with minimized ISS gain ϕ_2:

$$\big\|(\mathbf{e}(t))\big\|_2 < \phi_1 \big\|(\mathbf{e}(0))\big\|_2\, e^{-\frac{\sigma}{2}t} + \phi_2 \big\|(\Delta(\hat{\mathbf{x}}, \mathbf{x})\big\|_\infty \tag{21}$$

Thus, there exists a positive definite symmetric matrix $\mathbf{P} \in \mathbb{R}^{3\times3}$, matrices $\mathbf{F_i} \in \mathbb{R}^{3\times3}$, positive scalars $\bar{\eta}$ and ϵ such that the LMIs constraints (22c),(22d) and (22e) are satisfied. The resulting observer gains are given by $\mathbf{L_i} = \mathbf{P}^{-1}\mathbf{F_i}$. If it exists a symmetric positive definite matrix $\mathbf{P} \in \mathbb{R}^{3\times3}$ with the minimal and maximal eigenvalues denoted by $\mathcal{X}_1$ and $\mathcal{X}_2$, matrices $\mathbf{F_i} \in \mathbb{R}^{3\times3}$ such that the following LMI conditions for a given σ, α, β and $\mathcal{X}_1$ hold $\forall\ i = 1,\ldots,8$:

$$\min \gamma \text{ s.t.} \tag{22a}$$

$$\mathcal{X}_1\,\mathbf{I} \leq \mathbf{P} \tag{22b}$$

$$\begin{pmatrix} \mathbf{A}_i^\top\mathbf{P} + \mathbf{C}^\top\mathbf{F}_i^\top + \mathbf{P}\mathbf{A}_i + \mathbf{F}_i\mathbf{C} + \sigma\,\mathbf{P} & \mathbf{P} \\ \mathbf{P} & -\gamma\,\mathbf{I} \end{pmatrix} < 0 \tag{22c}$$

$$\begin{pmatrix} \beta\mathbf{P} & \mathbf{P}\mathbf{A}_i + \mathbf{F}_i\mathbf{C} \\ \mathbf{A}_i^T\mathbf{P} + \mathbf{C}^T\mathbf{F}_i^T & \beta\mathbf{P} \end{pmatrix} > 0 \tag{22d}$$

$$\mathbf{A}_i^T\mathbf{P} + \mathbf{C}^T\mathbf{F}_i^T + \mathbf{P}\mathbf{A}_i + \mathbf{F}_i\mathbf{C} + 2\alpha\mathbf{P} < 0 \quad i = 1,\ldots,8 \tag{22e}$$

with: $\mathbf{F_i} = \mathbf{P}\mathbf{L_i}$, $\phi_1 = \sqrt{\frac{\mathcal{X}_2}{\mathcal{X}_1}}$, $\phi_2 = \sqrt{\frac{\gamma}{\alpha\mathcal{X}_1}}$, and $\mathcal{X}_1\,\mathbf{I} \leqslant \mathbf{P} \leqslant \mathcal{X}_2\,\mathbf{I}$.

Proof 1. Lyapunov theory is used to investigate the stability of the error dynamics in (19) by considering the following quadratic function:

$$V = \mathbf{e}^T\mathbf{P}\mathbf{e} \qquad \mathbf{P} = \mathbf{P}^T > 0 \tag{23}$$

Given that the Lyapunov function is positive definite, it is bounded in the following manner:

$$\mathcal{X}_1 \big\|(\mathbf{e}(t))\big\|^2 \leqslant V(t) \leqslant \mathcal{X}_2 \big\|(\mathbf{e}(t))\big\|^2 \tag{24}$$

where $\mathcal{X}_1$ and $\mathcal{X}_2$ are the minimum and maximum eigenvalues of the matrix $\mathbf{P}$. By taking the time derivative of the Lyapunov function and replacing the estimation error (19) derived previously, one obtains:

$$\begin{aligned} \dot{V} &= \dot{\mathbf{e}}^\top\mathbf{P}\mathbf{e} + \mathbf{e}^\top\mathbf{P}\dot{\mathbf{e}} \\ &= (\mathbf{e}^\top\mathscr{A}_{\mathbf{e}} + \Delta^\top(\hat{\mathbf{x}}, \mathbf{x}))\mathbf{P}\mathbf{e} + \mathbf{e}^\top\mathbf{P}(\mathscr{A}_{\mathbf{e}}\mathbf{e} + \Delta(\hat{\mathbf{x}}, \mathbf{x})) \\ &= \mathbf{e}^\top(\mathscr{A}_{\mathbf{e}}{}^\top\mathbf{P} + \mathbf{P}\mathscr{A}_{\mathbf{e}})\mathbf{e} + \Delta^T(\hat{\mathbf{x}}, \mathbf{x})\mathbf{P}\mathbf{e} + \mathbf{e}^T\mathbf{P}\Delta(\hat{\mathbf{x}}, \mathbf{x}) \end{aligned} \tag{25}$$

which is equivalent to:

$$\begin{aligned} \dot{V}(t) &= \mathbf{e}^\top\,\mathscr{A}_{\mathbf{e}}{}^\top\mathbf{P} + \mathbf{P}\mathscr{A}_{\mathbf{e}}\,\mathbf{e} + 2\,\mathbf{e}\,\mathbf{P}\,\Delta \\ &= \begin{pmatrix} \mathbf{e} \\ \Delta \end{pmatrix}^\top \begin{pmatrix} \mathscr{A}_{\mathbf{e}}{}^\top\mathbf{P} + \mathbf{P}\mathscr{A}_{\mathbf{e}} + \sigma\mathbf{P} & \mathbf{P} \\ \mathbf{P} & -\gamma\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{e} \\ \Delta \end{pmatrix} \\ &\quad -\sigma\,\mathbf{e}^\top\mathbf{P}\,\mathbf{e} + \gamma\,\Delta^\top\Delta \end{aligned} \tag{26}$$

where σ is a real positive constant.

Recalling the relations $\sum_{i=1}^{8} \mu_i(\hat{\mathbf{x}})(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}) = \mathscr{A}_{\mathbf{e}}$ and $\mathbf{F_i} = \mathbf{P}\mathbf{L_i}$, then if the feasibility of the LMIs expressed in (22c) $\forall i = 1,\ldots,8$ holds, this infers the following bounds of the Lyapunov function derivative:

$$\dot{V}(t) < -\sigma\, \mathbf{e}^{\top}\mathbf{P}\,\mathbf{e} + \gamma\, \Delta^{\top}\Delta \tag{27}$$

By integrating the previous inequality (27), one obtains:

$$\begin{aligned}\dot{V}(t) &< V(0)\, e^{-\sigma t} + \gamma \int_0^t e^{-\sigma(t-s)\, t} \left\|\Delta(s)\right\|^2 ds \\ &\leqslant V(0)\, e^{-\sigma t} + \frac{\gamma}{\sigma}\left\|\Delta(s)\right\|_{\infty}^2\end{aligned} \tag{28}$$

including the bounds of the Lyapunov function in (24) yields:

$$\left\|\mathbf{e}(t)\right\|^2 < \frac{\mathcal{X}_2}{\mathcal{X}_1}\left\|\mathbf{e}(0)\right\|^2 e^{-\sigma t} + \frac{\gamma}{\alpha\mathcal{X}_1}\left\|\Delta(t)\right\|_{\infty}^2 \tag{29}$$

which is equivalent to:

$$\left\|(\mathbf{e}(t))\right\| < \sqrt{\frac{\mathcal{X}_2}{\mathcal{X}_1}}\left\|\mathbf{e}(0)\right\| e^{-\frac{\sigma}{2} t} + \sqrt{\frac{\gamma}{\sigma\mathcal{X}_1}}\left\|(\Delta(t)\right\|_{\infty} \tag{30}$$

This inequality proves the ISS stability of the system (19) with respect to the disturbance Δ and with minimal ISS gain ensured by the condition (22a) defined by $\phi_2 = \sqrt{\gamma/\left(\sigma\mathcal{X}_1\right)}$. The pole placement of $(\mathbf{A}_i - \mathbf{L}_i\mathbf{C})$, $i = 1, \ldots, 8$ is selected to in order to maximize the convergence rate of the estimation. The eigenvalues are assigned in particular regions defined by the combination of a disk of radius β centered at (0,0) and the half-plane delimited by α:

$$S(\alpha, \beta) = \{z \in \mathbb{C},\ \Re(z) < -\alpha,\ |z| < \beta\} \tag{31}$$

The following constraints define the chosen pole clustering:

$$\begin{pmatrix} \beta\mathbf{P} & (\mathbf{A}_i - \mathbf{L}_i\mathbf{C}) \\ (\mathbf{A}_i - \mathbf{L}_i\mathbf{C})^T\mathbf{P} & \beta\mathbf{P} \end{pmatrix} > 0 \tag{32a}$$

$$(\mathbf{A}_i - \mathbf{L}_i\mathbf{C})^T\mathbf{P} + \mathbf{P}(\mathbf{A}_i - \mathbf{L}_i\mathbf{C}) + 2\alpha\mathbf{P} < 0 \quad i = 1, \ldots, 8 \tag{32b}$$

To obtain feasible LMIs, the variables must appear linearly in the constraints. Hence, the following change of variables is performed: $\mathbf{F_i} = \mathbf{PL_i}$ The pole clustering in inequality (32) is equivalent to the following and this ends the proof.

$$\begin{pmatrix} \beta\mathbf{P} & \mathbf{P}\mathbf{A}_i + \mathbf{F}_i\mathbf{C} \\ \mathbf{A}_i^T\mathbf{P} + \mathbf{C}^T\mathbf{F}_i^T & \beta\mathbf{P} \end{pmatrix} > 0 \tag{33a}$$

$$\mathbf{A}_i^T\mathbf{P} + \mathbf{C}^T\mathbf{F}_i^T + \mathbf{P}\mathbf{A}_i + \mathbf{F}_i\mathbf{C} + 2\alpha\mathbf{P} < 0 \quad i = 1, \ldots, 8 \tag{33b}$$

The feasibility of the LMIs in (22) after imposing the parameters α, β, σ and $\mathcal{X}_1$ ensures the stability of the error dynamics in (19) with improved convergence rate.

4. VALIDATION RESULTS

The performance of the proposed observer is validated against simulation and experimental results. The proposed approach is compared to the active *Incremental Structure from Motion* (ISfM) scheme proposed in (Tahri et al., 2017) in the presence of strong measurement noise on the camera velocity. The latter observer was chosen for comparison due to its robust performance.

4.1 Simulation Results

Concerning the simulation results, three points defined with respect to the initial camera frame by the following coordinates were considered:

$$\mathbf{P_0} = \begin{pmatrix} -1 & .9 & 0.7 \\ 0.2 & 0.9 & 0.8 \\ 1.6 & 1.1 & 1.2 \end{pmatrix} \tag{34}$$

The *Root Mean Squared Error* (RMSE) is used to compare the performance of the proposed work with the results obtained in (Tahri et al., 2017) to quantify the performance of our estimator in the presence of measurement noise and modeling error where κ is the parameter to be estimated and n is the number of error samples in a given interval.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\kappa - \hat{\kappa}\right)_i^2} \tag{35}$$

The RMSE of the depth estimation error in the steady-state performance for both approaches is compared in order to conclude which observer delivers more robust performance with respect to camera velocity uncertainties. Hence, both the gains of both observers were tuned to have approximately the same rate convergence and the steady state was chosen to start at 0.5 seconds. The initial values of all estimated parameters were set to zero for the comparison test. The following four cases were considered:

Case 1 : Noise-free measurements.
Case 2 : The controlled angular velocity is affected by a measurement noise of zero mean and variance of 0.001 using the random number block function in Simulink.
Case 3 : Both translational and angular velocities are affected by measurement noise of zero mean and variance of 0.001.
Case 4 : Both translational and angular velocities are affected by measurement noise of zero mean and variance of 0.003.

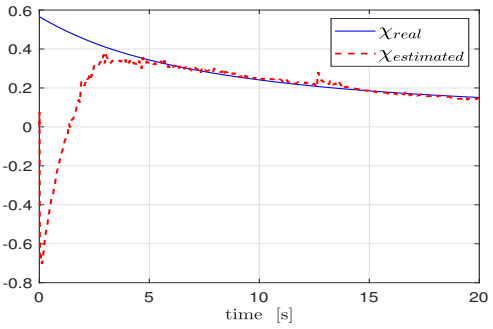
Observer	Observer in (Tahri et al., 2017)	Proposed Observer
Case 1	1.4781×10^{-5}	6.5898×10^{-4}
Case 2	5.0537×10^{-3}	0.0654
Case 3	0.2529	0.0854
Case 4	0.4247	0.1549

Table 1.

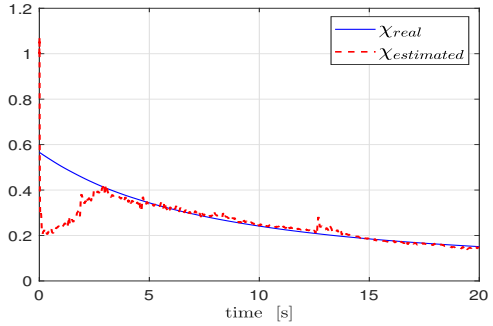
From the RMSE table, it can be seen that both observers perform almost identically for noise-free system. RMSE results obtained with the observer in (Tahri et al., 2017) in the second case show a lower value implying that it performs better in the presence of measurement noise on the angular velocity due to extracted features being rotation invariant. However, when the observer performance is evaluated for strong measurement noise in cases 3 and 4, one notices that the proposed observer has favourable noise rejection properties. Thus, our observer delivers better performance in the presence of measurement noise.

4.2 An Experimental Case Study

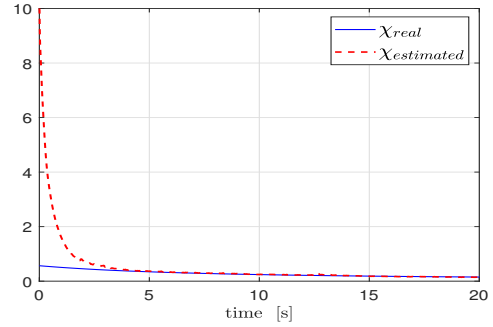
To validate the proposed approach experimentally, point depth estimation was performed on images captured from a Kinect camera mounted on a TurtleBot mobile robot platform. The 2D feature considered in this experiment is the center of a detected and tracked circle on the image. The estimate is compared to the ground truth collected using the indoor positioning tool VICON. After calibrating the VICON Motion Capture system, the Kinect fixed on



(a)



(b)



(c)

Fig. 1. Simulation results of the estimated $\hat{\chi}$ and real χ depth information for the following initial values: (a) $\hat{\chi} = 0$, (a) $\hat{\chi} = 1$ and (c) $\hat{\chi} = 10$.

the turtleBot and the center of the image circle were recognized by the VICON Tracker system (version 3.7.) using mounted markers. The ground truth (VICON) date was published to the ROS network and saved as `.bag` files created with the *rosbag* utility. Images were published at 30Hz while VICON readings were updated at 100 Hz. Let ${}^{v}\mathbf{p_c}$ be the position of the marker attached to the circle and ${}^{v}\mathbf{p_k}$ the position of the Kinect camera with respect to the VICON-fixed frame. The aim is to build a ground truth consisting of successive circle center positions with respect to the camera frame during camera motion. Hence, the position of the marker $\mathbf{p_c}^{v}$ can be expressed in the camera frame with coordinates $\mathbf{p_c}^{k}$ as follows:

$$\mathbf{p_c}^{k} = {}^{c}\mathbf{R_v}^{-1}\,\mathbf{p_c}^{v\top} - {}^{c}\mathbf{R_v}^{-1}\,\mathbf{p_k}^{v\top} \tag{36}$$

where: ${}^{c}\mathbf{R_v}$ is the rotation matrix from the VICON-fixed frame to the camera frame.

The feature considered here is the center of the circle on the image. The imposed parameters used in the optimization problem were chosen in order to minimize ISS gain between disturbances $\Delta(\hat{\mathbf{x}}, \mathbf{x})$ and the estimation error $\mathbf{e}$ while satisfying all LMI constraints. If the optimization problem did not permit a numerical solution, the imposed parameters σ and $\mathcal{X}_1$ were relaxed further and the optimization problem was re-run under these new LMI constraints. The choice of predefined parameters of the pole clustering is crucial: High values will result in a rapid convergence of the estimation error. However, the observer becomes more sensitive to measurement noise. Ultimately, there was a compromise in the selection of the imposed parameters between the accuracy of the estimate and the feasibility of the LMIs. After extensive testing, the following parameter set was used for feasibility reasons: $\sigma = 10$, $\mathcal{X}_1 = 10^{-5}$. After exploring different regions, the pole clustering was taken in the region defined by $\beta = 30$ and $\alpha = 5$ which resulted in a faster convergence rate and reduced sensitivity to measurement noise. The LMIs conditions in (22) are solved using YALMIP giving:

$$\gamma = 1.3171\,10^{-6}$$

$$\mathbf{P} = 10^{-9}\begin{pmatrix} 0.2193 & -0.0002 & -0 \\ -0.0002 & 0.2193 & -0 \\ -0 & -0 & 0.0004 \end{pmatrix}$$

The observability of $(\mathbf{A}_i, \mathbf{C})$ can be calculated from (14) and it is shown that the eight sub-models of the TS system are observable. Note that this does not necessarily guarantee the true observability of the nonlinear system, only that it satisfies the observability criterion used here. The experimental results show the satisfactory performance of the observer. In Fig. 2, the true and estimated depth Z_c^k converge after 4 seconds. Fig. 1 illustrates the behavior of the real and estimated depth information χ for different initial values of $\hat{\chi}$. In summary, the theoretical claims have been validated against extensive simulations and experimental results. Non-holonomic constraints and non-negligible dynamics strongly perturb estimation of the camera velocity, particularly the translational component. Thanks to the ISS property of the proposed observer, the impact of these disturbances is confined to a spherical region characterized by ϕ_2. Finally, the experimental results are highlighted in the video link https://t.ly/ub5K.

5. CONCLUSIONS

This paper presents a nonlinear observer design for structure and motion identification reconstruction using a calibrated monocular camera. Sufficient conditions were provided to ensure the uniform observability of the vision system modeled as a qLPV system. The convergence of the estimation error was investigated based on the ISS property and Lyapunov analysis and the observer gains were computed by solving the derived LMI constraints. The observer was implemented in MatLab and compared with a recently proposed rotation-invariant feature-based active Incremental Structure from Motion scheme (Tahri et al., 2017) against simulation data and experimental results. An interesting use case of the proposed approach would be navigation tasks for mobile robots: a robot could apply this approach to scene features while exploring an

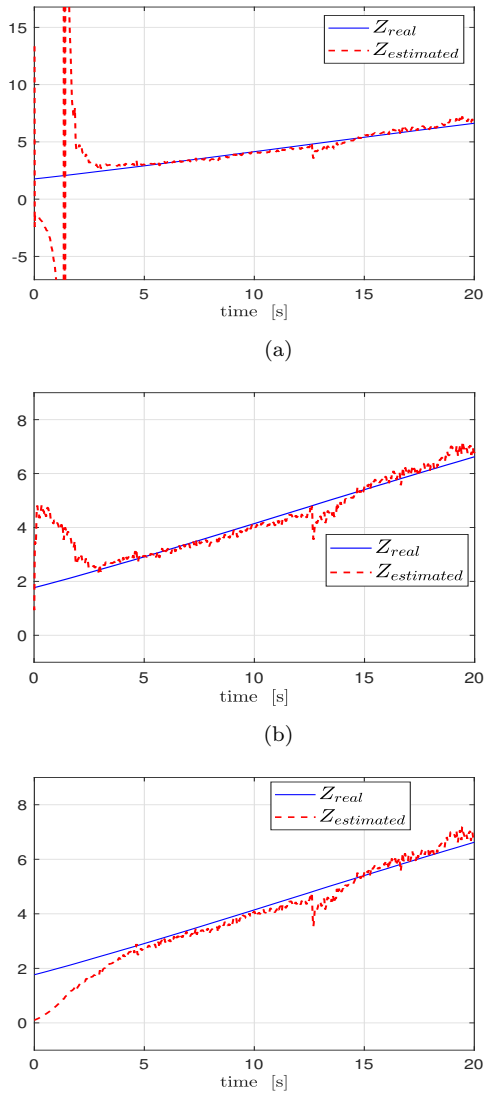


Fig. 2. Simulation results of the estimated $\hat{Z}$ and real Z_c^k depth for the following initial values of $\hat{\chi}$: (a) $\hat{\chi} = 0$, (a) $\hat{\chi} = 1$ and (c) $\hat{\chi} = 10$.

environment to create a depth map.

REFERENCES

Agarwal, S., Snavely, N., Seitz, S.M., and Szeliski, R. (2010). Bundle Adjustment in the Large. In K. Daniilidis, P. Maragos, and N. Paragios (eds.), *Computer Vision – ECCV 2010*, 29–42. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded Up Robust Features. In A. Leonardis, H. Bischof, and A. Pinz (eds.), *Computer Vision – ECCV 2006*, 404–417. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bergsten, P. and Palm, R. (2000). Thau-Luenberger observers for TS fuzzy systems. In *Ninth IEEE International Conference on Fuzzy Systems. FUZZ- IEEE 2000 (Cat. No.00CH37063)*, volume 2, 671–676.

Chhaya, F., Reddy, D., Upadhyay, S., Chari, V., Zia, M.Z., and Krishna, K.M. (2016). Monocular reconstruction of vehicles: Combining SLAM with shape priors. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 5758–5765.

Civera, J., Davison, A.J., and Montiel, J.M.M. (2012). *Structure from Motion using the Extended Kalman Filter*. Springer Berlin, Heidelberg.

Dani, A.P., Fischer, N.R., and Dixon, W.E. (2012). Single Camera Structure and Motion. *IEEE Transactions on Automatic Control*, 57(1), 238–243.

Jiang, Z.P., Teel, A.R., and Praly, L. (1994). Small-gain theorem for ISS systems and applications. *Mathematics of Control, Signals and Systems*, 7, 95–120.

Kazuo, T. and Wang, H.O. (2001). *Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach*. John Wiley & Sons, Inc.

Keshavan, J. and Humbert, J.S. (2017). An Analytically Stable Structure and Motion Observer Based on Monocular Vision. *Journal of Intelligent and Robotic Systems*.

Lasenby, J., Zisserman, A., Cipolla, R., Longuet-Higgins, H.C., Kanade, T., and Morris, D.D. (1998). Factorization methods for structure from motion. In *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, volume 356, 1153–1173.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*.

Oliensis, J. (2000). A Critique of Structure-from-Motion Algorithms. *Computer Vision and Image Understanding*, 80(2), 172–214.

Omari, S. and Ducard, G. (2013). Metric visual-inertial navigation system using single optical flow feature. In *2013 European Control Conference (ECC)*, 1310–1316.

Pupilli, M. and Calway, A. (2006). Real-Time Visual SLAM with Resilience to Erratic Motion. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, 1244–1249.

Sename, O. and Fergani, S. (2017). Linear Parameter Varying systems: from modelling to control. In *IVSS 2017 - Intelligent Vehicles International Summer School*. Compiègne, France.

Spica, R. (2015). *Contributions to active visual estimation and control of robotic systems*. Theses, Université Rennes 1.

Tahri, O., Boutat, D., and Mezouar, Y. (2017). Brunovsky's Linear Form of Incremental Structure From Motion. *IEEE Transactions on Robotics*, 33(6), 1491–1499.

Tahri, O., Giordano, P.R., and Mezouar, Y. (2015). Rotation free active vision. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3086–3091.

Wang, G. and Jonathan Wu, Q.M. (2011). *Guide to Three Dimensional Structure and Motion Factorization*. Springer London.

Zhang, J., Boutin, M., and Aliaga, D.G. (2006). Robust Bundle Adjustment for Structure from Motion. In *2006 International Conference on Image Processing*, 2185–2188.