Visuo-inertial data fusion for pose estimation and self-calibration

Glauco Garcia Scandaroli

UNIVERSITÉ DE NICE–SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE STIC

Sciences Technologies de l'Information et de la Communication

T H È S E

pour l'obtention du grade de

Docteur en Sciences

de l'Université de Nice-Sophia Antipolis

Mention: Automatique, Traitement du Signal et des Images

Présentée par

Glauco Garcia SCANDAROLI

Fusion de données visuo-inertielles pour l'estimation de pose et l'autocalibrage

Thèse préparée à INRIA Sophia Antipolis-Méditéranée

Jury:

Gildas BESANÇON	Professeur, Grenoble-INP	*Rapporteur*
Robert MAHONY	Professeur, Australian National University	*Rapporteur*
Silvère BONNABEL	Maître-assistant, Mines ParisTech	*Examinateur*
Tarek HAMEL	Professeur, UNSA et I3S–CNRS	*Examinateur*
Eric MARCHAND	Professeur, Université de Rennes 1	*Examinateur*
Pascal MORIN	Professeur contractuel, ISIR–UPMC	*Directeur de thèse*

Visuo-inertial data fusion for pose estimation and self-calibration

Glauco Garcia SCANDAROLI

CONTENTS

NOTATION

- Matrices are represented by capital letters, while vectors and scalars are represented by the minuscules and lower greek letters. The difference between vector and scalars should be noticeable from the context.
- $\mathbb{M}(n,m)$ denotes the set of matrices with real elements and dimension $n \times m$, and for the sake of simplicity, $\mathbb{M}(n) = \mathbb{M}(n,n)$.
- The identity of $\mathbb{M}(n)$ is denoted by I_n.
- The matrix $0_{m\times n} \in \mathbb{M}(m,n)$ is the matrix with all elements equal to zero.
- The time derivative of a function $f(t)$, *i.e.* $\frac{df}{dt}$, is represented by the dot $\dot{f}$.
- The partial derivative of a function $f(x)$ with respect to a vector x, *i.e.* $\frac{\partial f}{\partial x}$, is represented by the shorter $\partial_x f$.

PROLOGUE

Combining information gathered from a multiple sources is ubiquitous. This process is known as data fusion, and it is performed very intuitively, for example, by Humans, who use information from balance, motion and joint position to walk. Furthermore, the sense of vision warns us about surrounding dangers, while hearing aids to identify threats outside the field of view. In some cases, we must resort to data fusion in order to extract information that cannot be obtained by a single sensor.[1] In other situations, even though information can be already obtained by a single sensor, complementary characteristics of multiple sensors improve the quality of system's description and robustness to other impairments. On that account, multi-sensory data fusion supports diverse advantages with respect to using information collected from a single sensor.

Data fusion has an important role in mobile, *i.e.* air, ground, and underwater, robotics. This "mobile" concept requires motion information, as mobile robots often employ linear and angular velocities to perform very low level control. Additionally, robots can execute large displacement in a short time, thus the employed data fusion technique should be robust to fast movements, i.e. high angular and linear velocities and their accelerations. Knowing the current pose, *i.e.* position and orientation of an object with respect to some reference, is a prerequisite for many applications. It is, indeed, a critical requirement in problem of aerial robot stabilization. Concerning most indoor and outdoor applications, pose measurement can usually be performed with either high frequency or high precision. Both of these characteristics are required, for instance, to achieve safe and high quality control of aerial robots in unfavorable environments.

A multi-sensory system to identify body pose must handle data obtained in various coordinates systems. The resulting dynamics is more complex than pure body motion, because a multi-coordinate system dynamics will take into account other parameters to relate each system. However, those parameters are but seldom known accurately in advance, and we often need to identify them prior to field applications. This data fusion problem requires persistent motion conditions to distinguish between the target pose and the additional parameters, which is an underlying difficulty to developing high quality techniques. Remarkably, the non-linear nature of the motion equations present an difficult task. Moreover, identification and use of the inherent observability conditions make that data fusion problem challenging.

This thesis addresses the problem of visuo-inertial data fusion for pose estimation. A visio-inertial system consists of a video camera together with an inertial measurement unit (IMU) attached to the same rigid body. The objective is to combine highly accurate pose measurements extracted by camera information with angular velocity and proper acceleration obtained by the IMU. On the one hand, the inertial sensors provide incremental displacement measurements that can initialize computer vision algorithms or compensate for momentary loss of sight. Measurements provided by the IMU, however are corrupted by an additive offset, also known as measurement bias, and noise caused by manufacturing characteristics. These characteristics can be significant for many low-cost sensors employed in robotics applications, and estimates based on purely on IMU measurements drift quickly. Therefore, pose information obtained using information provided by the camera limit the drift associated with direct

1. I do throughout this manuscript the common habit of referring to we as a generic third person. Sometimes, we can refer to myself only, or the reader and myself, or the research community, etc. The meaning should be clear from the context of each sentence.

integration of inertial data. It is convenient to estimate IMU bias, since these variables may vary due to several exogenous factors, *e.g.* temperature or battery level. Another source of difficulty concerns various parameters associated to the use of different coordinate frames, *e.g.* reference image, camera and IMU frames.

The results here discussed follow from two fundamental domains: computer vision and nonlinear control. In the prior discipline, we are mostly interested in direct visual tracking methods that can be applied for pose estimation, and these results are independent of the data fusion methods further presented. The latter domain is employed in the development of new methods in order to fuse pose and incremental information obtained by the cameras and IMU, respectively. Despite the fact that classical data fusion methods have been already employed to solve this specific problem, those techniques do not guarantee the convergence of the estimates under fast motion of the sensor. We analyze the data fusion problem using control theory, because, first, we can guarantee convergence of the estimates under fast motion, besides, we are also granted with simple expressions for movements that distinguish between pose and system parameters.

OBJECTIVES AND RESULTS

The objective of this thesis is to propose new techniques for pose estimation using visual and inertial information. Initially, inertial and visual data can be inspected separately, *i.e.* the computer vision methods for pose estimation do not depend on information given by inertial sensors, nor the data fusion techniques rely on any specific computer vision method. These two different domains, computer vision and data fusion, share the main contributions of this thesis:

1. Concerning computer vision, we propose a new direct visual tracking method based on the normalized cross-correlation, that implements region and pixel-wise weighting together with a Newton-like optimization. This method can accurately estimate pose even under severe illumination changes.
2. The main contributions of this thesis concern the data fusion process. We propose new nonlinear observers for pose estimation, IMU bias and sensor-to-sensor parameter calibration. The data fusion design includes a thorough observability analysis for the system. This analysis provides expression of the movements under which we can distinguish pose from the other system parameters. We obtain the stability of the observer under explicit conditions of body motion.

The prosed techniques are compared with the literature using synthetic (simulated) and real data. For instance, the visual tracking method is compared against the state-of-the-art using a benchmark dataset that evaluates the accuracy, and real world sequences with challenging illumination changes. Moreover, the proposed nonlinear observers are compared with classical methods using simulated data that exploit the obtained observability conditions. We conclude by presenting experiments with the integration of the proposed visual tracking method with the proposed nonlinear observers. Part of the work described in this thesis is already published in three major international conferences.

- (Scandaroli and Morin, 2011) presents the design of a nonlinear observer for pose estimation and calibration of IMU bias, together with a gain tuning procedure. We show that we can obtain globally stable nonlinear observers with a gain choice that is independent of the angular velocity.
- (Scandaroli et al., 2011) presents the design of a nonlinear observer for pose estimation, and calibration of IMU bias and sensor-to-sensor rotation. This paper also presents exper-

imental results using the fusion of the proposed nonlinear observer with a direct visual tracking method.
- (Scandaroli et al., 2012) proposes a novel direct visual tracking method that uses normalized cross-correlation as similarity measure. This method implements two weighting techniques: grid-wise and pixel-wise in order to avoid problems due to severe illumination changes. The method is complemented by a Newton-style optimization that improves the computation of the solution.

The results concerning the observability analysis and the observers for gravity and sensor-to-sensor translation are unpublished by the date of submission of this thesis.

THESIS ORGANIZATION

This thesis is organized as follows.
- Chapter 1 presents the necessary background information for the comprehension of this thesis. We first discuss the basic definition of pose, *i.e.* orientation and position, employed in this thesis, as well as technical details on camera modeling and inertial sensors concluding with four different visual-inertial systems.
- Chapter 2 discusses different state-of-the art solutions for the direct visual tracking problem. This chapter also presents a novel visual tracking method based on the normalized cross correlation to cope with complex illumination variations. The proposed technique is compared to the state of the art using a benchmark dataset and challenging sequences.
- Chapter 3 states the main results from this thesis: the analysis of the observability and design of new nonlinear observers for the simultaneous estimation of pose, numerous parameters, *e.g.* rate gyroscope and accelerometer bias, camera-to-IMU orientation, direction of the gravitational acceleration. We also compare the the proposed techniques to state-of-the-art methods using simulation data. In order to simplify the presentation of the results, the algebraic development of the observability analysis and the stability proofs for the nonlinear observers are presented separately in Appendix A.
- Chapter 4 presents experimental results using the techniques proposed in the thesis.

The epilogue of the thesis reviews the achieved results and remarks some possible future work. Two appendixes present additional material:
- Appendix A presents the proofs of observability properties of systems and stability of the nonlinear observers proposed in Chapter 3. This appendix complements the main contributions of this thesis stated in Chapter 3.
- Appendix B discusses known results in parameter estimation robust to outliers. These results are employed in Chapter 2.

1

BACKGROUND AND PROBLEM STATEMENT

This chapter reviews basic notation and sensory structure necessary for the full comprehension of this work. The first section describes the notation and some properties of rigid-body motion. The following section discusses the perception of linear and angular motion using accelerometers and gyroscopes, also the pitfalls faced with these sensors. The third section discusses some preliminaries of pose estimation using computer vision, *i.e.* image formation and geometry of two views using perspective cameras. The fourth and fifth sections refer to the problem of state estimation. We first review general solutions to state estimation, then follow with a more specific discussion on attitude estimation. We conclude the chapter presenting the visuo-inertial sensor, and examine the operation and two different modes: pose estimation, and sensor-to-sensor self-calibration.

1.1 MOTION REPRESENTATION

It is quite common in robotics to use multiple frames in order to present vectors and points in space. In this thesis, we model the world using classic Euclidean geometry, *c.f.* , for example, (Ma et al., 2003). We define a point m in space, and a vector is given by two points m, n and a directive arrow that connects m to n, denoting $\vec{v} = \overrightarrow{mn}$.

A frame $\mathcal{R}$ is given by the quadruplet ${}^{\mathcal{R}}\{p, \vec{\imath}, \vec{\jmath}, \vec{k}\}$, also written without distinction as $\{{}^{\mathcal{R}}p$, ${}^{\mathcal{R}}\vec{\imath}$, ${}^{\mathcal{R}}\vec{\jmath}$, ${}^{\mathcal{R}}\vec{k}\}$. The point ${}^{\mathcal{R}}p$ defines the origin of $\mathcal{R}$, and the vector triplet ${}^{\mathcal{R}}\vec{\imath}$, ${}^{\mathcal{R}}\vec{\jmath}$, ${}^{\mathcal{R}}\vec{k}$ provides an orthonormal basis of the Euclidean space. Additionally, a right-handed frame also satisfies $\vec{\imath} \times \vec{\jmath} = \vec{k}$, $\vec{\jmath} \times \vec{k} = \vec{\imath}$, and $\vec{k} \times \vec{\imath} = \vec{\jmath}$, *c.f.* Figure 1.1. Three coordinates ${}^{\mathcal{R}}m_1$, ${}^{\mathcal{R}}m_2$, ${}^{\mathcal{R}}m_3$ are sufficient to describe a point m using the frame $\mathcal{R}$ via

$$ {}^{\mathcal{R}}m = {}^{\mathcal{R}}p + {}^{\mathcal{R}}m_1 {}^{\mathcal{R}}\vec{\imath} + {}^{\mathcal{R}}m_2 {}^{\mathcal{R}}\vec{\jmath} + {}^{\mathcal{R}}m_3 {}^{\mathcal{R}}\vec{k}, $$

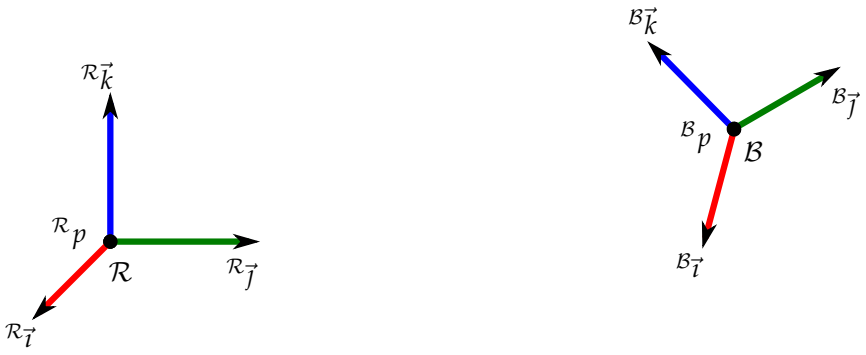


Figure 1.1: Examples of right-handed frames.

and we say equivalently that m has coordinates ${}^{\mathcal{R}}m = \begin{bmatrix} {}^{\mathcal{R}}m_1 & {}^{\mathcal{R}}m_2 & {}^{\mathcal{R}}m_3 \end{bmatrix}^T \in \mathbb{R}^3$ with respect to frame $\mathcal{R}$. Let us now consider two frames: $\mathcal{R}$ and $\mathcal{B}$. Frame $\mathcal{B}$ can be described with respect to $\mathcal{R}$ by the pair ${}^{\mathcal{R}}(p, R)_{\mathcal{B}}$, that is also written without distinction $({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}R_{\mathcal{B}})$. The element ${}^{\mathcal{R}}p_{\mathcal{B}} \in \mathbb{R}^3$ denotes the coordinates of the origin ${}^{\mathcal{B}}p$ with respect to $\mathcal{R}$, and ${}^{\mathcal{R}}R_{\mathcal{B}} \in \mathbb{M}(3)$ is a matrix whose columns are given by the coordinates of the triplet ${}^{\mathcal{B}}\{\vec{i}, \vec{j}, \vec{k}\}$ with respect to $\mathcal{R}$. Furthermore, since the columns of $R = {}^{\mathcal{R}}R_{\mathcal{B}}$ are orthonormal, we can directly verify that

$$R^T R = I_3,\ R^{-1} = R^T \text{ and } \det(R) = 1.$$

We can relate the coordinates of a point m in frames $\mathcal{R}$ and $\mathcal{B}$, via ${}^{\mathcal{R}}(p, R)_{\mathcal{B}}$, *i.e.* :

$${}^{\mathcal{R}}m = {}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}m + {}^{\mathcal{R}}p_{\mathcal{B}}\,. \tag{1.1}$$

and from the definition of vectors, we obtain the coordinate transformation of a vector as

$${}^{\mathcal{R}}v = {}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}v\,. \tag{1.2}$$

Remark that the pair ${}^{\mathcal{R}}(p, R)_{\mathcal{B}}$ completely defines the frame $\mathcal{B}$ with respect to $\mathcal{R}$. Now, let us consider a third frame $\mathcal{C}$ such that ${}^{\mathcal{B}}(p, R)_{\mathcal{C}}$ denotes $\mathcal{C}$ with respect to $\mathcal{B}$. Using, (1.1) and (1.2), we obtain that the frame $\mathcal{C}$ writes with respect to $\mathcal{R}$ writes

$${}^{\mathcal{R}}p_{\mathcal{C}} = {}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}p_{\mathcal{C}} + {}^{\mathcal{R}}p_{\mathcal{B}}, \quad {}^{\mathcal{R}}R_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{C}}\,. \tag{1.3}$$

The *pose* of a rigid-body is defined after two frames: $\mathcal{B}$ attached to some point of the rigid-body, and a reference frame $\mathcal{R}$ conveniently defined. The pose of a rigid-body is represented then by the frame $\mathcal{B}$ in $\mathcal{R}$ coordinates, *i.e.* ${}^{\mathcal{R}}(p, R)_{\mathcal{B}}$ depicted in Figure 1.2, where the elements of the pair ${}^{\mathcal{R}}(p, R)_{\mathcal{B}}$ denote the body position and orientation respectively. In this thesis, we consider an inertial reference frame in order to define the pose. The main characteristic of in inertial frame is that the frame moves with constant velocity, *i.e.* ${}^{\mathcal{R}}\ddot{p} = 0$, moreover ${}^{\mathcal{R}}\dot{p} = {}^{\mathcal{R}}v$ imply ${}^{\mathcal{R}}\dot{\vec{i}} = 0$, ${}^{\mathcal{R}}\dot{\vec{j}} = 0$, and ${}^{\mathcal{R}}\dot{\vec{k}} = 0$ from the definition of vectors.

The *position* of the body is defined by the origin of the associated frame, we can describe the dynamics of the position $p(t)$ by

$$\dot{p}(t) = v(t), \qquad \dot{v}(t) = a(t), \tag{1.4}$$

where $v(t) \in \mathbb{R}^3$ and $a(t) \in \mathbb{R}^3$ denote the linear velocity and acceleration.

The matrix R defines the *orientation* (or *attitude*) of the body. We can verify for the dynamics of the orientation $R(t)$ under the orthogonality constraint, *i.e.* $RR^T = I_3$, that

$$\dot{R}(t)R^T(t) + R(t)\dot{R}^T(t) = 0, \qquad \dot{R}(t)R^T(t) = -(\dot{R}(t)R^T(t))^T.$$

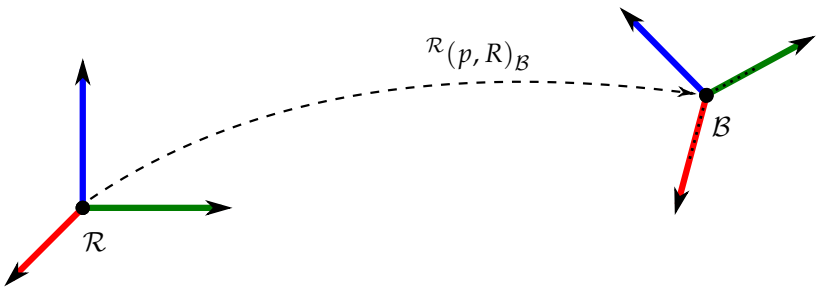


Figure 1.2: The pose of a rigid body

Notice that the term $\dot{R}(t)R^T(t)$ yields an anti-symmetric matrix Ω that stands for

$$\dot{R}(t)R^T(t) = \Omega. \tag{1.5}$$

More specifically, the orientation $R \in \mathbb{SO}(3)$, *i.e.* the special orthogonal group (Warner, 1987), and its associated Lie algebra $\mathfrak{so}(3)$ is the set of anti-symmetric matrices. The skew operator $\mathrm{S}(\cdot)\colon \mathbb{R}^3 \mapsto \mathfrak{so}(3)$ writes:

$$S(\omega) = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix},$$

where ω_i is the i-th element of any vector $\omega \in \mathbb{R}^3$. Letting $\omega \in \mathbb{R}^3$ be the rotational velocity in inertial coordinates and $\Omega = \mathrm{S}(\omega(t))$ in (1.5), we have that the orientation dynamics writes:

$$\dot{R}(t) = \mathrm{S}(\omega(t))R(t) \tag{1.6}$$

The operator $\mathrm{S}(\cdot)$ also represents the cross product $\mathrm{S}(u)v = u \times v$, $\forall\, u, v \in \mathbb{R}^3$, and denote its inverse map by $\mathrm{vex}(\cdot)$, *i.e.* $\mathrm{vex}(\mathrm{S}(u)) = u$, $\forall\, u \in \mathbb{R}^3$. Remark that we can write $R = \exp\{\Omega\}$ owing to the exponential map properties: $\exp\{\cdot\}\colon \mathfrak{so}(3) \mapsto \mathbb{SO}(3)$. We have that the following properties hold for any $u \in \mathbb{R}^3$, $A = A^T \in \mathbb{M}(3)$, and $R \in \mathbb{SO}(3)$:

$$\mathrm{S}(Ru) = R\mathrm{S}(u)R^T, \tag{1.7}$$
$$R\mathrm{vex}(\mathrm{P_a}(R)) = \mathrm{vex}(\mathrm{P_a}(R)), \tag{1.8}$$
$$\mathrm{S}(u)\mathrm{S}(v) = -u^T v I_3 + vu^T, \tag{1.9}$$
$$\mathrm{S}(u \times v) = \mathrm{P_a}(vu^T), \tag{1.10}$$
$$\mathrm{tr}(A\mathrm{S}(v)) = 0. \tag{1.11}$$

where $\mathrm{P_a}(M)\colon \mathbb{M}(3) \mapsto \mathfrak{so}(3)$, $\forall M \in \mathbb{M}(3)$ is the anti-symmetric matrix operator, *i.e.*

$$\mathrm{P_a}(R) = \frac{R - R^T}{2}. \tag{1.12}$$

There exist indeed other representations for rigid-body orientation, for instance, the unitary quaternions and Euler angles are often employed in the literature. We prefer the classic matrix representation because each element is unique and there are no singularities. The methods developed in this thesis can be extended to quaternion representations, we omit these discussions however. The reader interested in other attitude parametrizations can refer to, *e.g.* , (Shuster, 1993).

The pair position–orientation $(p(t), R(t))$ defines the *pose* $P(t)$ of a rigid body at a given instant with respect to an inertial reference coordinate system. Notice that we can either represent the pose by the pair itself $P = (p, R)$ or via the homogeneous matrix

$$P = \begin{bmatrix} R & p \\ 0_{1\times 3} & 1 \end{bmatrix} \in \mathbb{M}(4), \quad \text{with} \quad P^{-1} = \begin{bmatrix} R^T & -R^T p \\ 0_{1\times 3} & 1 \end{bmatrix},$$

and, more specifically, $P \in \mathbb{SE}(3)$, *i.e.* the special Euclidean group (Warner, 1987), with its associated Lie algebra $\mathfrak{se}(3)$ given by the set of twist matrices. Remark that we can write $P = \exp\{T\}$, with $T \in \mathfrak{se}(3)$, owing to the exponential map properties: $\exp\{\cdot\}\colon \mathfrak{se}(3) \mapsto \mathbb{SE}(3)$. Likewise the skew-matrix for the special orthogonal group, we can define the twist operator $T(\cdot,\cdot)\colon \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathfrak{se}(3)$:

$$T(v, \omega) = \begin{bmatrix} S(\omega) & v \\ 0_{1\times 3} & 0 \end{bmatrix}.$$

where $v, \omega \in \mathbb{R}^3$.

1.2 INERTIAL NAVIGATION SYSTEMS

Inertial navigation systems (INS) are examples where data fusion is a necessary procedure. A concise and comprehensive overview about INS is given in (Kuritsky and Goldstein, 1983). In order to understand inertial navigation, let us first wonder about navigation. The word navigation, itself, derives from the Latin words *navis* "ship" and *agere* "to drive", and the problem of navigation can be defined as directing the necessary movements from one point to another. This problem is somehow instinctive and it can be delineated by the following two questions:

– Where am I leaving from?
– Where do I want to go?

Answers to the these questions define a target origin and destination, which show different scales of navigation. One can see the problem as cell trajectories in the order of micrometers, up to space travels of hundreds of thousands kilometers. Those answers, however, embedded a third question:

– How did we define the target origin and destination?

The answer to this third question will define an appropriate reference coordinate system in which it makes sense to assign positions, velocities, and trajectories. Notice that the navigation problem will become more or less complex depending on the choice of the reference coordinates.

Inertial coordinate systems, *i.e.* coordinate systems with constant velocity, are the natural choice for navigation. As an illustration, consider the Earth-Centered, Earth-Fixed (ECEF) coordinate system. This coordinate system has the origin defined by the Earth's center of mass, its first axis points towards the international reference median and the third axis points towards the north pole, furthermore, the second axis is defined to make the system right handed; the position in these coordinates is thus provided in terms of latitude, longitude and altitude. This is considered an inertial coordinate system for many terrestrial applications. However, those coordinates are not strictly inertial, since the Earth rotates around the Sun, and, under "military" precision, ECEF inertial coordinates requires to compensate Earth's rotation. For most civilian research and commercial systems, however, that is a good approximation for an inertial coordinate system, as many sensors available to the the end consumer present larger measurement errors than the ones provided by the inertial Earth assumption. Depending on the target application, of course, stellar observations can be employed to define an astronomical reference. In such a case, all the involved frames must be integrated to obtain the current coordinates.

The inertial navigation problem could be solved by integrating the angular rate to establish the current angular position, with respect to the predefined reference coordinates, while integrating twice the linear acceleration. Notice that the current angular position is a requirement to integrate accelerations consistently, and obtain body position and velocity in reference coordinates. Gyroscopes and accelerometers are the sensors capable of measuring angular rate and linear acceleration. These instruments are known as inertial sensors, because they exploit the physical property of inertia, *i.e.* resistance to changes in angular momentum and linear motion. In general, we refer to the gyroscopes and accelerometers by the set composed by three components of each sensor, measuring angular rate and linear acceleration in each of the three axis from a coordinate system. A cluster of triaxial gyroscopes and accelerometers is called inertial measurement unit (IMU). Gyroscopes and accelerometers present an interesting similarity with organs from the vestibular system (Viéville and Faugueras, 1989), which is constituted by the otoliths and the semi-circular canals. The vestibular system provides movement information as well as sense of balance in most mammals. The otoliths are responsible for providing information about the tilt and linear motion of the head, and the semi-circular

canals are able to detect angular velocity of the head. (Viéville and Faugueras, 1989) and (Corke et al., 2007) present an intriguing comparison between the Human sensory system and the sensors from an IMU.

Early steps towards inertial instruments date back to the early 1900s, but the course of the second world war yielded the dawn of inertial sensing and navigation. The works (Pitman, 1962, pp. 8-12), and (Draper, 1981) present an interesting historical view of early inertial sensors and navigation systems. The first successful IMUs were presented from 1948 and, since the beginning of IMU development, these units have been built under two different specifications: gimbaled and strap-down. These configurations operate similarly, although the latter computes orientation and pose numerically, while the prior performs the computation electromechanically. More specifically, the gimbaled version consists in a system where gyroscopes and accelerometers are attached to a servo controlled gimbal. Every angular motion about the axes is sensed by the gyroscopes, that provide the information to the servo controllers. The servo controllers, in turn, maintain the gimbals stabilized to coincide with the reference coordinates. The position can be directly obtained by integrating twice the accelerometers, since the gimbaled system is designed to maintain the reference coordinate system. The physical mount in the strap-down form is much simpler, because the two sensors are literally strapped down to the body. The reference orientation matrix, however, must be computed mathematically, instead of the mechanical compensation in gimbaled systems. Furthermore, this computed orientation matrix is used online to set the measured linear acceleration in reference coordinates. Attitude computation was a heavy burden for early digital computers, and that is, in part, the reason why mechanically complex gimbaled systems were utilized instead of strap-down mounts in initial INS.

Electro-mechanical transducers were employed initially in gimballed IMUs. These sensors emphasized on accuracy over cost, dimension and weight. Long military missions often employed even more accurate IMUs, that down-graded the integration errors via a closed loop system with orientation and position information obtained by star-tracking devices (Lerman, 1983). The fade of gimballed electro-mechanical IMUs owes mostly to the advent of ring-laser gyroscopes (RLG) (Silver, 1983), micro-machined electro-mechanical sensors (MEMS) (Yazdi et al., 1998) and the evolution of digital computers. The development of RLG in the 1970s provided a cheaper, lighter and accurate option to electro-mechanical gyroscopes. Furthermore, the development of computers that could easily process high rate attitude and position information made gimbaled systems give way to strap-down digital IMUs. The global positioning system (GPS) replaced star-tracking devices (Moustafa, 2001) in inertial navigation. GPS complemented the IMUs with a three-dimensional position measurement given in an earth based coordinated system, as well as the current time of the sensor. Position and time information are computed from radio signals transmitted from satellites that orbit the Earth (Dudek and Jenkin, 2008), where civilian devices often obtain 20 [m] accuracy. Still, the use of inertial sensors remained restricted to military and civilian aeronautics until the development of MEMS packages. This former class reduced not only size and weight of sensors, but also their cost which allowed their widespread presence nowadays.

Commercial low-cost MEMS provide significantly less accurate measurements compared to pioneer and military-graded systems. These sensors provide raw measurements impaired by different factors, *e.g.* misalignment, temperature or battery level, that can be compensated via a pre-treatment of the data. However, even after a pre-treatment of the data we do not obtain "perfect" measurements. In the literature, a commonly employed model consists of the target physical characteristic corrupted by additive noise and constant offset, also known as measurement bias (Lefferts et al., 1982). Thus, the pose obtained by merely integrating MEMS data drifts after a few seconds. The reduction of this drift can be achieved after a good calibration of sensor's bias. Even though this offset can be also obtained in a previous

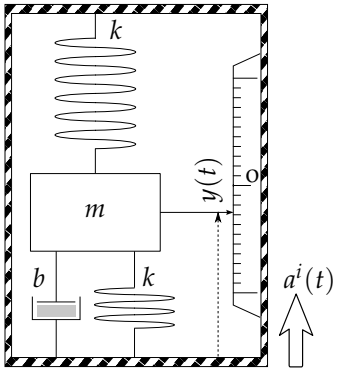


Figure 1.3: Description of a simplified mechanical accelerometer, adapted (Corke et al., 2007)

calibration procedure, the constant model is only valid for some periods of time, and an online method for estimating the measurement bias is indeed positive. Remark, however, that other sensors, that provide explicit or implicit attitude and position measurements, must be employed in order to cope with the bias online calibration and bound the resulting drift from IMU data integration.

1.2.1 *Inertial sensors*

Inertial sensors are instruments that exploit the principle of inertia, *i.e.* resistance to a change in momentum, in order to measure motion changes with respect to an inertial reference frame. Angular rate gyroscopes and accelerometers are the inertial sensors responsible for measuring respectively angular velocity and (specific) linear acceleration. In this section we describe the basic structure of those sensors, and discuss physical characteristics that can impair pose estimation. The reader can find more information on the structure of the sensors, historical view and comparison to biological systems in (Corke et al., 2007; Dudek and Jenkin, 2008, and references therein).

1.2.2 *Accelerometers*

Accelerometers are the sensors responsible for measuring the linear acceleration of a body with respect to an inertial frame. The damped spring-mass system displayed in Figure 1.3 represents the basic structure for a simplified single-axis accelerometer. A proof mass m is supported by a sprint with elastic constant k and a dumper with stiffness b, and the following second order dynamics

$$ma^i = m\ddot{y} + b\dot{y} + ky \tag{1.13}$$

converts the acceleration $a^i(t)$ of the system into a displacement $y(t)$. Normally, a sensor like the one depicted in Figure 1.3 is attached to moving rigid body with an associated frame $\mathcal{B}$. Thus, a simplified sensor can measure the rigid-body acceleration ${}^{\mathcal{B}}a^i(t)$ along the axis i of $\mathcal{B}$ via the output position $y(t)$, and the response of the system is classified as overdamped, underdamped or critically damped depending on the choice of m, k and b. These characteristics are usually provided by the manufacturer, and although practical accelerometers can vary in design and technology, that previous structure is always present in essence.

In practice, we must employ a triplet of accelerometers in order to completely measure the acceleration ${}^{\mathcal{B}}a = [{}^{\mathcal{B}}a_1, {}^{\mathcal{B}}a_2, {}^{\mathcal{B}}a_3]^T$. We should, however, be aware that accelerometers measure the *specific* linear acceleration via $y(t)$, *i.e.* the expression of the *all* physical forces

applied to the body. Therefore, the effects arisen from the real body acceleration $a_{\mathcal{B}}$ and earth's gravitational field are measured as if they were the same force, *i.e.*

$$^{\mathcal{B}}a_{\mathrm{s}} = {}^{\mathcal{B}}a + {}^{\mathcal{B}}g$$

with $^{\mathcal{B}}a_{\mathrm{s}}$ denoting the specific body acceleration, $^{\mathcal{B}}a$ is the component due to linear acceleration and $^{\mathcal{B}}g$ the gravitational acceleration field in expressed in body coordinates.

1.2.3 *Gyroscopes*

Angular rate gyroscopes, or simply gyros, are the sensors responsible for measuring the angular velocity of a body with respect to an inertial frame. These instruments rely on the conservation of the angular momentum, *i.e.* a body tends to keep rotating at the same angular velocity about the same axis with the absence of external torque, *e.g.* we have the angular momentum of a spinning disk

$$L = I\omega \times v,$$

where I denotes the disk's moment of inertia, v the disk velocity and ω the angular velocity of the body. As the system is rotated, it applies an opposing moment that is measured by the gyros.

Vibrating structure gyroscopes are widely applied in practice. These sensors do not employ spinning disks, but a probe mass that moves in a straight line with linear velocity v, and experiences a Coriolis acceleration

$$a_{\mathrm{C}} = 2\omega \times v,$$

caused by an angular velocity ω. Accelerometer-like sensors then measure the Coriolis acceleration.

1.2.4 *Compensating for sensory characteristics*

Several applications these days employ micro-electro-mechanical sensors (MEMS). In this thesis, we consider that class of sensors since they have a good dynamical response, are lightweight, low-power consuming and most specially: they are mass produced and accessible to the end user. [1]

Several physical characteristics in MEMS differ from the aforesaid models. For instance, we must employ triads of gyros and another of accelerometers to measure the angular velocity and specific linear acceleration in 3-axis, thus, we often verify cross-coupling residual measurements due to misalignment of the sensors. MEMS are also subjected to nonlinear variations caused by temperature. Moreover, the analog data must be digitalized for numerical treatment, and, to avoid aliasing, the analog signal is filtered cutting frequencies, at least, higher than half of the sampling one. These are a few examples of deterministic effects. Furthermore, stochastic models on gyro and accelerometer measurements include a Gaussian additive measurement noise. However, we follow a deterministic approach in this thesis and neglect this former characteristic.

The calibration of accelerometers, *e.g.* (Batista et al., 2011, and references therein), (Beravs et al., 2012), and gyroscopes, *e.g.* , (Ojeda et al., 2000), (Olivares et al., 2009), (Cui et al., 2012,

1. It is possible to find devices with 3-axis gyroscopes and accelerometers from 40 USD by October 2012, *e.g.* , https://www.sparkfun.com/products/10251

and references therein), has been of intensive research. It is hard to model all nonlinear characteristics from these instruments and how they interact. Therefore, authors often employ first or second order approximations, *i.e.*

$$\mu(t) = b_\mu + M_1\mu(t) + \mu^T M_2\mu(t) + \dots,$$

in order to model the dynamics of a μ, *e.g.* the accelerometers or gyroscopes. A plausible model defines $b_\mu \in \mathbb{R}^3$ as the measurement bias, $M_1 \in \mathbb{M}(3)$ is non-singular matrix that represents scale factors and alignment errors due to the mechanical mount, and, sometimes, the higher order parameters $M_2 \in \mathbb{M}_3$ can model effects due to changes in temperature. Calibration procedures usually involve specific movements executed by robotic arms or rotating turn-tables that must "excite" all of the system modes in order to yield an observable model to obtain all of the pre-defined parameters.

The measurement bias is an important effect that is characterized by an additive offset measurement. In this thesis, we consider the following measurement model:

$$^{\mathcal{B}}\omega_y(t) = {}^{\mathcal{B}}\omega(t) + b_\omega\,, \tag{1.14}$$

$$^{\mathcal{B}}a_y(t) = {}^{\mathcal{B}}a_S(t) + b_a = {}^{\mathcal{B}}a(t) + b_a - {}^{\mathcal{B}}g(t)\,, \tag{1.15}$$

where b_ω, b_a denote gyro and accelerometers bias. The constant dynamics model is employed for additive bias:

$$\dot{b}_\omega = 0_{3\times1}, \qquad \dot{b}_a = 0_{3\times1}\,. \tag{1.16}$$

for the angular rate gyroscopes and accelerometers, respectively. That constant model may seem incomplete at a first glance, since these parameters can vary due to different factors such as temperature and battery level. The dynamics of the biases, however, varies slowly through relatively long periods of time. Hence, a constant model for these biases can indeed represent their governing dynamics. Furthermore, this constant model seldom changes the observability conditions of the system, and it is commonly employed in practice, *e.g.* , (Lefferts et al., 1982) and references therein.

Naturally, for the model (1.14)–(1.16) be valid, we assume higher order sensor characteristics, *i.e.* gain scales, alignment matrix and higher order nonlinear characteristics are either previously calibrated and compensated in a pre-treatment phase of the sensors or negligible. We make this simplification since the mechanical mount of the sensors generally present a "more constant" behavior than the additive term. Moreover, computing such multiplicative terms together with the bias and the pose variables increase the complexity of the system's observability conditions. Hence, we are unlikely to improve the results for these complex configurations without a controlled environment often accessible to manufacturers only.

1.3 POSE ESTIMATION WITH COMPUTER VISION

The sense of vision is very important for animal interaction with the environment. The eyes can perceive a spectrum of the waves reflected by objects inside the field of view, and, afterwards, the brain processes the information present in the observed scene. This whole process is a result of thousands of years of evolution, and animals seem to use vision very intuitively. Humans have employed cameras to play the role of the eyes, while images captured by these cameras carry the information about the environment. It is left to the computer vision algorithms to interpret what information can be extracted from each image, and how to exploit the information provided at its most.

Computer vision can be very effective for pose estimation. First, the cameras are non-intrusive and passive sensors. Secondly, multiple images can provide information of different

points of view of the scene, such that one can compute the trajectory of the camera from the changes shown in these images. The pose estimation problem can be treated similarly to the visual tracking. Visual tracking is a classic application of computer vision whose objective is to estimate the displacement of an object in a sequence of images. Therefore, one crucial task in visual tracking is to constantly identify the object.

We can define two techniques to model an object: feature-based and direct techniques. Although pose estimation can be treated as a visual tracking problem, this task is also generalized by a localization task, crucial in the *simultaneous localization and mapping* (SLAM), *c.f.* , for instance, (Klein and Murray, 2007), (Silveira et al., 2008), (Newcombe et al., 2011). Nevertheless, SLAM using visual techniques usually present one layer that involves either a feature-based or direct visual tracking method. This thesis therefore focuses on the visual tracking problem and the data fusion problem that is independent of the employed pose estimation technique. Next, we present the basics of image formation and the geometry of multiple views needed by Chapter 2.

FEATURE-BASED TECHNIQUES are build upon the extraction and matching of a sparse set of geometric characteristics (Shi and Tomasi, 1994). These characteristics can be described, for instance, by points of interest (Harris and Stephens, 1988), line segments (Hager and Toyama, 1998), (Marchand, 1999), etc. The representation of features is improved with the use of descriptors, such as, *e.g.* , SIFT (Lowe, 2004), SURF (Bay et al., 2006), and FERNS (Ozuysal et al., 2007). An interesting summary of several feature based techniques is shown in (Roth and Winter, 2008). Furthermore, feature-based techniques rely on a data association procedure, also known as feature matching. This procedure implies further computational burden and a new source of errors, as the cost of exhaustively comparing every feature is prohibitive and some approximations are employed in order to guarantee (close to) real-time execution.

One advantage of feature-based methods is that the ensemble of information provided by the images is completely represented by the features, which may, in turn, yield a representation that requires less memory space. Moreover, these methods are robust to large displacements, as the displacement of the object can be computed explicitly from feature matches.

The quality of the solution, however, depends on the number of observed features, which makes these methods prone to data association errors, partial occlusions and highly dependent on the density of features.

DIRECT TECHNIQUES, also known as intensity-based methods, exploit the brightness intensity of each individual pixel in order to solve the visual tracking. In contrast to feature-based methods, intensity-based methods can exploit the ensemble of information given by the image, thus these techniques can explore even areas where no features exist. Direct visual tracking methods have shown to be more accurate than feature-based techniques. However, the solution for pose estimation is not explicit and it is often obtained via iterative optimization of a similarity function.

The first solution to direct visual tracking was built upon the sum of squared differences (SSD) (Lucas and Kanade, 1981). The solution of the SSD is closely related to the least squares problem and the solution via SSD has proven to be very efficient, mainly because the optimization can be much simplified due to numerous solutions to nonlinear least squares, *e.g.* (Baker and Matthews, 2001) and (Benhimane and Malis, 2007). SSD tracking, however, is severely impaired when brightness constancy is violated, since motion and photometric variations are dealt in the same way by the similarity function. Different works improve the visual tracking using SSD by estimating online (Bartoli, 2008), (Silveira and Malis, 2010) or offline (Hager and Belhumeur, 1998) photometric parameters. Moreover, the problem of partial occlusion is usually treated by a robustly weighted SSD (Hager and Belhumeur, 1998). Nevertheless, as

we show, the SSD does not perform well under concurrent illumination changes and partial occlusions.

The normalized cross correlation (NCC) is a similarity function invariant to affine illumination changes, with radius of convergence comparable to the SSD. Typically, gradient-based solutions for the NCC resort to computationally expensive Newton's method (Irani and Anandan, 1998) or other first other simplifications such as (Evangelidis and Psarakis, 2008) or (Brooks and Arbel, 2010). More general examples of similarity measures include the mutual information (MI), *c.f.* (Viola and Wells, 1997), (Dame and Marchand, 2010), the sum of conditional variance (Pickering et al., 2009), (Richa et al., 2011), the correlation ratio (Roche et al., 1998b) and the cross cumulative residual entropy (Wang and Vemuri, 2007).

We dedicate the whole of Chapter 2 for a deeper discussion about direct visual tracking methods and pose estimation techniques, where we also present a new direct visual tracking method based on the NCC.

1.3.1 *Image formation and photometric model*

Image formation depend on multiple factors: type of visual sensor being employed, *e.g.* pinhole camera with thin lenses, and illumination properties from the sources and the scene. These factors result in geometric and photometric models for camera and the scene representation. We further describe the models of interest for this thesis, and these topics are covered with more details in computer vision books, *e.g.* (Ma et al., 2003), (Hartley and Zisserman, 2004) and (Szeliski, 2012).

We define an image by a two-dimensional brightness array that takes positive values for the brightness of each point. More specifically, the image $\boldsymbol{I}$ is a map defined on a compact set Ξ of a two dimensional surface of pixels that takes value in the positive real numbers, *i.e.*

$$\boldsymbol{I}\colon \Xi \subset \mathbb{P}^2 \to \mathbb{R}, \qquad \boldsymbol{p} \mapsto \boldsymbol{I}(\boldsymbol{p})$$

where $\boldsymbol{p} = [\boldsymbol{u},\, \boldsymbol{v},\, 1]^T \in \Xi$ defines the coordinates of a pixel, the origin $\boldsymbol{p}_0 = [0,\, 0,\, 1]$ is conveniently associated to the top-left pixel of the image. This thesis considers digital images, and, for that case, the domain Ξ and the range $\mathbb{R}$ are discretized, *e.g.* the image surface denotes $\Xi = [0, 799] \times [0, 599] \times 1 \subset \mathbb{N}^3$, and the brightness $[0, 255] \subset \mathbb{N}$.

The values of $\boldsymbol{I}$ depend on physical properties of the scene, *i.e.* reflectance of the material and shape of the object and light sources. According to experimental (Blinn, 1977) and physically-based (Cook and Torrance, 1982) models, the intensity measured at a particular pixel $\boldsymbol{p}$ depends on specular, diffuse and ambient reflections. The complexity of the scene's photometric model increases as we model these effects taking parameters such as ambient light or viewpoint into account. However, we can simplify the model by considering the scene composed by Lambertian surfaces, *i.e.* objects that maintain their appearance independent of the viewing direction, and we can use a photometric model reduced to an affine transformation for an image $\boldsymbol{I}$ considering a reference $\boldsymbol{I}^*$, *i.e.*

$$\boldsymbol{I}(\boldsymbol{p}) = \alpha(t)\boldsymbol{I}^*(\boldsymbol{p}) + \beta(t)\,, \tag{1.17}$$

with $\alpha(t), \beta(t) \in \mathbb{R}$. This approximation holds, at least locally, for most applications.

1.3.2 *Sub-pixel approach*

Despite the fact that brightness values are only available on discrete surface, high precision algorithms often need to compute the intensity values at a non-integer pixel position. Therefore, we must often resort to image interpolation techniques, for instance:

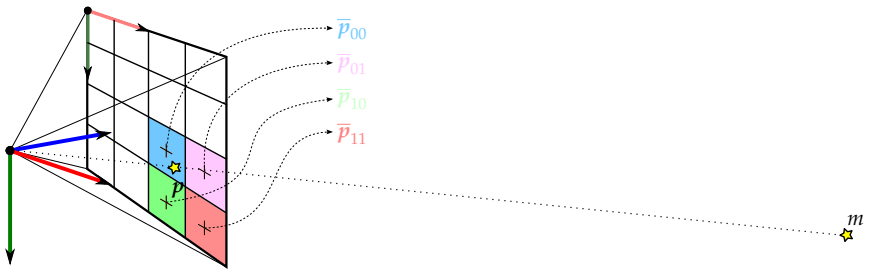


Figure 1.4: Bilinear interpolation

– *Nearest-neighbor interpolation* computes the nearest integers for the u and v elements of the pixel $\boldsymbol{p}$, *i.e.*

$$\overline{\boldsymbol{p}} = \begin{bmatrix} \text{int}(u), & \text{int}(v), & 1 \end{bmatrix}^T,$$

where int: $\mathbb{R} \to \mathbb{Z}$, $u \mapsto \text{int}(u)$ defines the function returning the closest integer to a real number. The resulting intensity is given by $\mathfrak{I}_n(\boldsymbol{p}) = \boldsymbol{I}(\overline{\boldsymbol{p}})$.
This technique is simple and fast, however, the resulting interpolation has low quality and presents discontinuities, *e.g.* , rounded shapes or textured surfaces.

– *Bilinear interpolation* computes the resulting intensity using 4 neighboring pixels as depicted in Figure 1.4. Denote the nearest-neighbor $\overline{\boldsymbol{p}}_{00} = \overline{\boldsymbol{p}}$, and let $\overline{\boldsymbol{p}}_{ij} = \overline{\boldsymbol{p}}_{00} + [i\ j\ 0]^T$, and $\alpha = \boldsymbol{p} - \overline{\boldsymbol{p}}_{00}$. We write the bilinear interpolation as:

$$\mathfrak{I}_l(\boldsymbol{p}) = \begin{bmatrix} 1-\alpha_u \\ \alpha_u \end{bmatrix}^T \begin{bmatrix} \boldsymbol{I}(\overline{p}_{00}) & \boldsymbol{I}(\overline{p}_{01}) \\ \boldsymbol{I}(\overline{p}_{10}) & \boldsymbol{I}(\overline{p}_{11}) \end{bmatrix} \begin{bmatrix} 1-\alpha_v \\ \alpha_v \end{bmatrix}. \tag{1.18}$$

This technique is slower than the nearest-neighbor, however, the contours of the resulting image are smoother.

– *Bicubic interpolation* uses the same approach as the bilinear interpolation, however the resulting intensity is computed over 16 neighboring pixels. We can compute this interpolation as:

$$\mathfrak{I}_c(\boldsymbol{p}) = \begin{bmatrix} f(1+\alpha_u) \\ f(\alpha_u) \\ f(1-\alpha_u) \\ f(2-\alpha_u) \end{bmatrix}^T \begin{bmatrix} \boldsymbol{I}(\overline{p}_{-1-1}) & \boldsymbol{I}(\overline{p}_{0-1}) & \boldsymbol{I}(\overline{p}_{1-1}) & \boldsymbol{I}(\overline{p}_{2-1}) \\ \boldsymbol{I}(\overline{p}_{-10}) & \boldsymbol{I}(\overline{p}_{20}) & \boldsymbol{I}(\overline{p}_{10}) & \boldsymbol{I}(\overline{p}_{20}) \\ \boldsymbol{I}(\overline{p}_{-11}) & \boldsymbol{I}(\overline{p}_{21}) & \boldsymbol{I}(\overline{p}_{11}) & \boldsymbol{I}(\overline{p}_{21}) \\ \boldsymbol{I}(\overline{p}_{-12}) & \boldsymbol{I}(\overline{p}_{22}) & \boldsymbol{I}(\overline{p}_{12}) & \boldsymbol{I}(\overline{p}_{22}) \end{bmatrix} \begin{bmatrix} f(1+\alpha_v) \\ f(\alpha_v) \\ f(1-\alpha_v) \\ f(2-\alpha_v) \end{bmatrix},$$

where f stands for the cardinal sinus function, *i.e.* $\text{sinc}(x) = 0$ for $x = 0$, and $\text{sinc}(x) = \sin(x)/x$ otherwise.
This technique is slower than both nearest-neighbor and bilinear interpolation, however the resulting is smooth and still maintains the contours, while the bilinear technique tends to smooth the gradients of the intensities.

Hence, the choice of the interpolation technique has to consider not only the quality of the result, but also the computational effort. In this thesis, we employ the bilinear interpolation due to the trade-off between computational effort and smoothness of the resulting image. Thus, for the sake of notation, every time we express $\boldsymbol{I}(\boldsymbol{p})$, we actually mean $\mathfrak{I}_l(\boldsymbol{p})$ using (1.18).

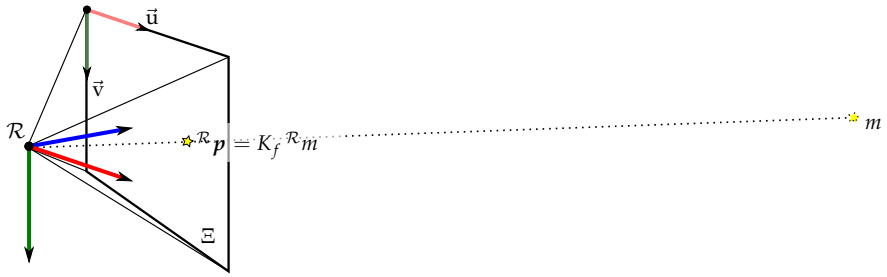


Figure 1.5: Perspective projection model

1.3.3 *Pinhole cameras*

Let us consider a frame $\mathcal{R}$ associated to the camera's optic center such that a point m in space has coordinates ${}^{\mathcal{R}}m$ in $\mathcal{R}$. Pinhole cameras are based on the perspective projection model, *i.e.* a point m with coordinates ${}^{\mathcal{R}}m$ is projected at

$$ {}^{\mathcal{R}}\boldsymbol{m} = {}^{\mathcal{R}}z^{-1}\,{}^{\mathcal{R}}m, $$

where ${}^{\mathcal{R}}z = e_3{}^T\,{}^{\mathcal{R}}m$ defines the depth of the point in $\mathcal{R}$ coordinates. In the above equation, the point ${}^{\mathcal{R}}\boldsymbol{m} = [{}^{\mathcal{R}}\boldsymbol{m}^1, {}^{\mathcal{R}}\boldsymbol{m}^2, 1]^T \in \mathbb{P}^2$ lies on a virtual plane perpendicular to the optical axis of the camera and distant one meter from the projection center. Pinhole cameras, however, change that virtual projection plane into the image plane Ξ such that the point m is associated to the pixel ${}^{\mathcal{R}}\boldsymbol{p}$, *c.f.* Figure 1.5, via

$$ {}^{\mathcal{R}}\boldsymbol{p} = K_f\,{}^{\mathcal{R}}\boldsymbol{m} = {}^{\mathcal{R}}z^{-1}K_f\,{}^{\mathcal{R}}m \in \mathbb{P}^2, \tag{1.19} $$

where the matrix $K_f \in \mathbb{M}(3)$ contains the intrinsic parameters of the camera

$$ K_f = \begin{bmatrix} f & fs & d_{\mathrm{u}} \\ 0 & fr & d_{\mathrm{v}} \\ 0 & 0 & 1 \end{bmatrix}, \tag{1.20} $$

with f as the focal length expressed in pixels, s the cosine between the image frame axes, r the aspect ratio and d_u, d_v the coordinates of the camera's principal point, w.r.t. the center of the image in pixels. The intrinsic parameters can be obtained by calibration procedures, *c.f.* (Tsai, 1987), (Zhang, 2000), using multiple images from different views of an object with known dimensions. We consider cameras with fixed lenses in this thesis, therefore K_f is constant and can be obtained using the aforementioned techniques.

The pinhole camera model (1.19) holds for many applications, however some lenses produce radian distortions on the projected image, *e.g.* the light is not propagated linearly causing straight lines to appear as curves, *c.f.* Figure 1.6. This phenomenon is important in wide-angle and fish-eye lenses. We can compensate the radial distortion towards the perspective-like model using a polynomial expression:

$$ \boldsymbol{p}^r = \boldsymbol{p}^d + (1 + k_1 r^2 + k_2 r^4 + \ldots + k_n r^{2n})(\boldsymbol{p} - \boldsymbol{p}^d), $$

where $\boldsymbol{p}^r$ denotes the rectified coordinates of $\boldsymbol{p}$, the parameters k_i denote the distortion coefficients, and $r = |K_f{}^{-1}(\boldsymbol{p} - \boldsymbol{p}^d)|$ is the distance from the current pixel to the optical center $\boldsymbol{p}^d = \begin{bmatrix} d^{\mathrm{u}} & d^{\mathrm{v}} & 1 \end{bmatrix}^T$. The radial distortion is constant, likewise the intrinsic parameters, since this

effect only depends on the pixel distance with respect to the center of projection. We can compute the distortion during the calibration procedure, then rectify the images in a pre-treatment step. We consider undistorted images in this thesis for the sake of simplicity.

1.3.4 *Geometry of two views*

We have discussed so far the basics on the formation of images in a single view. Concerning pose estimation, we are particularly interested in understanding how the projections of points in an image change between two views, and how we can identify the relative pose between these two views.

Let us consider the case described in Figure 1.7. The camera observes the point m initially in the pose associated to a frame $\mathcal{R}$, such that m is projected at the pixel

$$ {}^{\mathcal{R}}\boldsymbol{p} = {}^{\mathcal{R}}z^{-1} K_f \, {}^{\mathcal{R}}m \, . $$

Afterwards, the camera moves towards the pose described by the coordinate frame $\mathcal{C}$, and the same point m is projected at the pixel

$$ {}^{\mathcal{C}}\boldsymbol{p} = {}^{\mathcal{C}}z^{-1} K_f \, {}^{\mathcal{C}}m \, . $$

We can obtain the generic relation between the position of the pixels using (1.1)

$$ {}^{\mathcal{C}}\boldsymbol{p} \propto K_f \, {}^{\mathcal{C}}R_{\mathcal{R}} K_f{}^{-1} \, {}^{\mathcal{R}}\boldsymbol{p} + {}^{\mathcal{R}}z^{-1} K_f \, {}^{\mathcal{C}}p_{\mathcal{R}}, \tag{1.21} $$

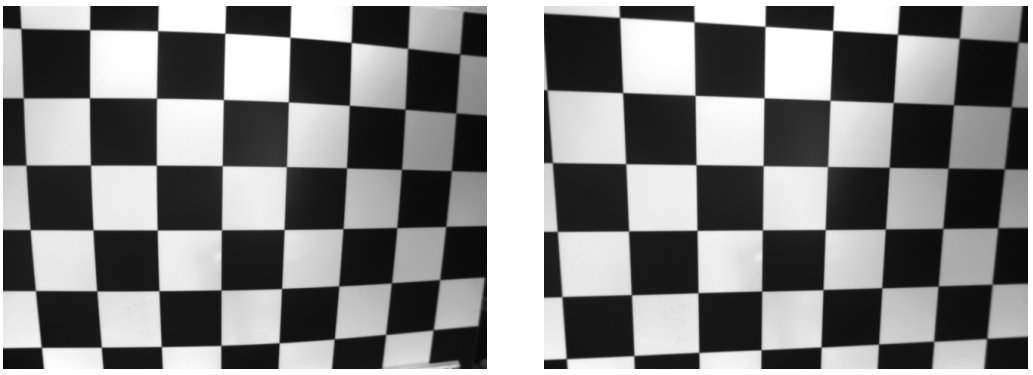
where the pair ${}^{\mathcal{C}}(p, R)_{\mathcal{R}}$ denotes the coordinates of frame $\mathcal{R}$ with respect to frame $\mathcal{C}$.

Essential matrix approach

One relation between the projection of a generic point and the relative pose using two images is given by the *Essential matrix* $E = \mathrm{S}({}^{\mathcal{C}}p_{\mathcal{R}}) \, {}^{\mathcal{C}}R_{\mathcal{R}}$. We obtain the Longuet-Higgins constraint multiplying both sides of (1.21) on the left by ${}^{\mathcal{C}}\boldsymbol{p}^T K_f{}^{-T} \mathrm{S}({}^{\mathcal{C}}p_{\mathcal{R}}) K_f{}^{-1}$, and using Eq. (1.19):

$$ {}^{\mathcal{C}}\boldsymbol{p}^T K_f{}^{-T} E K_f{}^{-1} \, {}^{\mathcal{R}}\boldsymbol{p} = {}^{\mathcal{C}}m^T \, E \, {}^{\mathcal{R}}m = 0, $$

since ${}^{\mathcal{C}}\boldsymbol{p}^T K_f{}^{-T} \mathrm{S}({}^{\mathcal{C}}p_{\mathcal{R}}) K_f{}^{-1} {}^{\mathcal{C}}\boldsymbol{p} = {}^{\mathcal{C}}\boldsymbol{p}^T K_f{}^{-T} \mathrm{S}({}^{\mathcal{C}}p_{\mathcal{R}}) \, {}^{\mathcal{C}}p_{\mathcal{R}} = 0$.



(a) (b)

Figure 1.6: Effects due to imperfect lenses: (a) captured; (b) rectified image

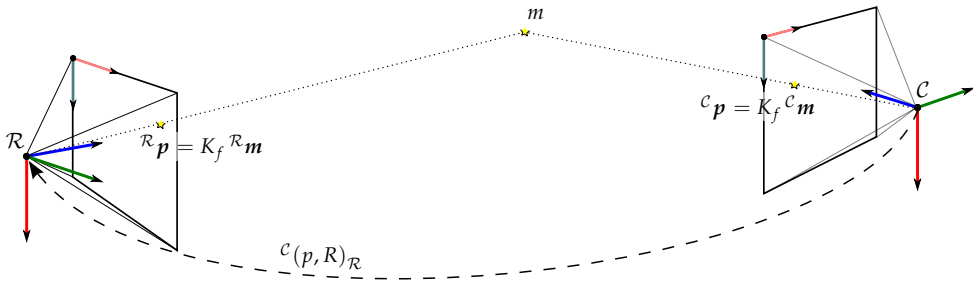


Figure 1.7: Projection of a point in two different views

The Essential matrix can be computed, for instance, from the Longuet-Higgins constraint using the 8 point method and E is further decomposed into $^{\mathcal{C}}(p,R)_{\mathcal{R}}$, *c.f.* (Ma et al., 2003, p. 121) for example. We need $N \geq 8$ pixel correspondences of different points m_i between two views, with $^{\mathcal{C}}p_i$ and $^{\mathcal{R}}p_i$ denoting the i-th correspondence. A first approximation E_* is computed via the SVD decomposition of

$$\boldsymbol{P} = \begin{bmatrix} ^{\mathcal{R}}\boldsymbol{u}_0\,^{\mathcal{C}}\boldsymbol{p}_0{}^T & ^{\mathcal{R}}\boldsymbol{v}_0\,^{\mathcal{C}}\boldsymbol{p}_0{}^T & ^{\mathcal{C}}\boldsymbol{p}_0{}^T \\ ^{\mathcal{R}}\boldsymbol{u}_1\,^{\mathcal{C}}\boldsymbol{p}_1{}^T & ^{\mathcal{R}}\boldsymbol{v}_1\,^{\mathcal{C}}\boldsymbol{p}_1{}^T & ^{\mathcal{C}}\boldsymbol{p}_1{}^T \\ \vdots & & \\ ^{\mathcal{R}}\boldsymbol{u}_N\,^{\mathcal{C}}\boldsymbol{p}_N{}^T & ^{\mathcal{R}}\boldsymbol{v}_N\,^{\mathcal{C}}\boldsymbol{p}_N{}^T & ^{\mathcal{C}}\boldsymbol{p}_N{}^T \end{bmatrix} = USV^T,$$

with $^{\mathcal{R}}\boldsymbol{p}_i^{\mathrm{u}}$ and $^{\mathcal{R}}\boldsymbol{p}_i^{\mathrm{v}}$ denoting the first and second components of the pixel $^{\mathcal{R}}p$, and the elements of E_* are given by the last column of V. The Essential matrix is obtained via

$$E = K_f{}^T U_*(I_3 - e_3e_3{}^T)V_*{}^T K_f$$

such that U_* and V_* are computed from the SVD decomposition $E_* = U_*S_*V_*{}^T$.

Some complications, however, appear due to this approach. First, we need sufficiently large parallax, translational displacement roughly speaking, to obtain a well conditioned Essential matrix, *i.e.* different from the nullity. Secondly, this approach needs general 3D points as correspondences, and the method fails namely when all of the correspondences lie in a planar surface.

The planar case

The planar case is depicted by Figure 1.8. We have that every point m over the planar surface Π verifies $^{\mathcal{R}}\boldsymbol{n}^T\,^{\mathcal{R}}m = {}^{\mathcal{R}}d$ in $\mathcal{R}$ coordinates, where $^{\mathcal{R}}\boldsymbol{n}$ denotes the unit normal vector and $^{\mathcal{R}}d \in \mathbb{R}$ the distance of the plane to the camera's optic center. Thus, we can write the depth of point m in $\mathcal{R}$ as

$$^{\mathcal{R}}z^{-1} = {}^{\mathcal{R}}d^{-1}\,^{\mathcal{R}}\boldsymbol{n}^T K_f{}^{-1}\,^{\mathcal{R}}\boldsymbol{p}\,,$$

such that, using (1.21), the projection $^{\mathcal{C}}\boldsymbol{p}$ of m at the pose corresponding to $\mathcal{C}$ writes

$$\begin{aligned} &^{\mathcal{C}}\boldsymbol{p} \propto {}^{\mathcal{C}}H_{\mathcal{R}}\,^{\mathcal{R}}\boldsymbol{p}\,, \\ &^{\mathcal{C}}\boldsymbol{p} \propto K_f\,^{\mathcal{C}}G_{\mathcal{R}}K_f{}^{-1\,\mathcal{R}}\boldsymbol{p}\,, \\ &^{\mathcal{C}}\boldsymbol{p} \propto K_f({}^{\mathcal{C}}R_{\mathcal{R}} + {}^{\mathcal{R}}d^{-1\,\mathcal{C}}p_{\mathcal{R}}\,^{\mathcal{R}}\boldsymbol{n}^T)K_f{}^{-1\,\mathcal{R}}\boldsymbol{p}\,. \end{aligned} \tag{1.22}$$

We define the Euclidean homography by $^{\mathcal{C}}G_{\mathcal{R}} = {}^{\mathcal{C}}R_{\mathcal{R}} + ({}^{\mathcal{R}}d^{-1})^{\mathcal{C}}p_{\mathcal{R}}\,^{\mathcal{R}}\boldsymbol{n}^T$, and the projective homography $^{\mathcal{C}}H_{\mathcal{R}} = K_f\,^{\mathcal{C}}G_{\mathcal{R}}K_f{}^{-1}$ of point in the plane Π for the $\mathcal{R}$ frame expressed in $\mathcal{C}$ coordinates.

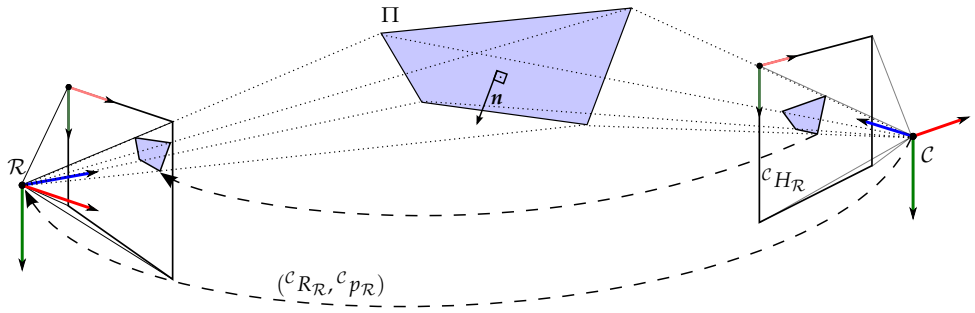


Figure 1.8: Projection of a planar surface in two different views

We can compute the Euclidean or projective homographies via different methods, *e.g.* , direct methods described in the next Chapter, or via the correspondence of $N \geq 4$ pixels. The technique based on correspondence of pixels is simple and similar to the technique employed by the Essential matrix. Let us consider the pixels ${}^{\mathcal{R}}p_i$ and ${}^{\mathcal{C}}p_i$ denoting the i-th pixel correspondence of a points m_i, then, the projective homography ${}^{\mathcal{C}}H_{\mathcal{R}}$ is computed via the SVD decomposition of

$$\boldsymbol{P} = \begin{bmatrix} {}^{\mathcal{R}}\boldsymbol{u}_0 \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_0) & {}^{\mathcal{R}}\boldsymbol{v}_0 \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_0) & \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_0) \\ {}^{\mathcal{R}}\boldsymbol{u}_1 \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_1) & {}^{\mathcal{R}}\boldsymbol{v}_1 \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_1) & \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_1) \\ \vdots & & \\ {}^{\mathcal{R}}\boldsymbol{u}_N \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_N) & {}^{\mathcal{R}}\boldsymbol{v}_N \, \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_N) & \mathrm{S}({}^{\mathcal{C}}\boldsymbol{p}_N) \end{bmatrix} = USV^T,$$

where the elements of ${}^{\mathcal{C}}H_{\mathcal{C}}$ are given by the last column of V. The Euclidean homography is computed directly via ${}^{\mathcal{C}}G_{\mathcal{R}} = {K_f}^{-1} \, {}^{\mathcal{C}}H_{\mathcal{R}} K_f^{-1}$.

After computing the Euclidean homography, we can recover a scaled pose, ${}^{\mathcal{C}}(\lambda p, R)_{\mathcal{R}}$, and the unit normal vector ${}^{\mathcal{R}}\boldsymbol{n}$ using, *e.g.* , SVD decomposition (Faugueras and Lustman, 1988), (Zhang and Hanson, 1995) or closed formulæ (Malis and Vargas, 2007). The number of possible solutions varies from eight, without any assumption on the scene, down to two, considering that the target is in front of the camera, *i.e.* $d_{\mathcal{R}} \in \mathbb{R}^+$. In order to obtain the full relative pose ${}^{\mathcal{C}}(p, R)_{\mathcal{R}}$, we must know the scale factor $\lambda = {}^{\mathcal{R}}d^{-1}$, and also be sure which extracted ${}^{\mathcal{R}}\boldsymbol{n}$ is the "correct" one.

When the plane is seen in front of the camera, *i.e.* nominal operation of pinhole camera, we can constrain the homographies up to $\det({}^{\mathcal{C}}G_{\mathcal{R}}) = \det({}^{\mathcal{C}}H_{\mathcal{R}}) = 1$, because (1.22) is a proportional relation. That constraint forces the matrices to belong to the special linear group $\mathbb{SL}(3)$ (Warner, 1987), *i.e.* the group of $\mathbb{M}(3)$ matrices with unitary determinant. We can replace (1.22) by the group action $\boldsymbol{w}\colon \mathbb{SL}(3) \times \mathbb{P}^2 \to \mathbb{P}^2$:

$$\boldsymbol{w}(H, \boldsymbol{p}) = \left[\frac{{e_1}^T H \boldsymbol{p}}{{e_3}^T H \boldsymbol{p}}, \quad \frac{{e_2}^T H \boldsymbol{p}}{{e_3}^T H \boldsymbol{p}}, \quad 1 \right]^T \tag{1.23}$$

such that for $\boldsymbol{p} \in \mathbb{P}^2$, and $H, H_1, H_2 \in \mathbb{SL}(3)$:

$$\begin{cases} \boldsymbol{w}(I_3, \boldsymbol{p}) = \boldsymbol{p}, \\ \boldsymbol{w}(H_1, \ w(H_2, \boldsymbol{p})) = \boldsymbol{w}(H_1 H_2, \boldsymbol{p}), \\ \boldsymbol{w}^{-1}(H, \boldsymbol{p})) = \boldsymbol{w}(H^{-1}, \boldsymbol{p}). \end{cases} \tag{1.24}$$

where $\boldsymbol{w}^{-1}$ denotes the inverse group action of $\boldsymbol{w}$.

1.4 STATE ESTIMATION

Control and estimation research communities use extensively the term state. In simple words, a state symbolizes the minimal amount of information to represent the internal condition of a target system at a given time instant. Notice that a different state representation can be defined as conveniently for a system, of course, whilst it satisfies the task's purpose. The term estimation derives from the Latin *æstimare* "determine", and the problem of state estimation can be defined as determining the internal conditions of a system at a certain time. The estimation problem is closely related to observability characteristics of a system (Kalman, 1960b), *i.e.* whether it is possible to reconstruct the state of a system from the information provided by trajectory of the measurements and known inputs of the system.

Generalizing the estimation problem, we can enumerate three situations:

1. determination of the current state using information up to the current time;
2. determination of a future state using information up to the current time;
3. determination of a past state using information up to the current time.

The first situation defines a filtering problem, while the second defines a prediction problem, and the third determines a smoothing problem (Jazwinski, 1970). This thesis addresses the filtering problem, and the words estimation and filtering are used interchangeably throughout the text.

The problem of state estimation has been studied for a long time, a history that dates up to, at least, 1800 when Gauss developed the technique known today as least squares to estimate planetary orbits. The theory used in state estimation can be divided in two ways of reasoning: deterministic and stochastic filtering. The leitmotif of stochastic filtering is that any system performs randomly. A stochastic system is described by probability density functions in the process dynamics and sensor models. Deterministic filtering, on the other hand, considers that the same system always behaves identically, *i.e.* a system will always exhibit the same outputs from the same starting condition and the same inputs. A deterministic system is described by differential equations.Whilst stochastic approaches aim at obtaining the full probability distribution of the system, deterministic solutions work toward the convergence of the estimates, *i.e.* the values of the estimates should coincide with their real values in the absence of noise.

The Kalman filter is unarguably the most successful result in stochastic linear filtering. This estimator is named after its author Rudolph Kalman, who first claimed the development of the optimal filter for linear discrete time systems with noisy inputs in (Kalman, 1960a), the extension for continuous systems was shown later in (Kalman and Bucy, 1961). In his first paper, Kalman proved the optimality using the principle of orthogonality, and since then his result has been "rediscovered" several times using different approaches. For example, the recursive least-squares formulation shows that the Kalman filter computes the estimates with the minimum squared-error in both deterministic and stochastic concepts. Furthermore, considering systems with Gaussian process and measurement noise, the Kalman filter yields the maximum likelihood, and also the minimum covariance filter, *c.f.* (Jazwinski, 1970, pp. 200-210, and references therein).

The Kalman filter is also shown as the Bayesian filter for Gaussian processes subjected to linear dynamics (Ho and Lee, 1964). Those results arrive although Kalman, in his seminal works, did not specify the quality of the probability distribution, as long as they are convex and symmetric with respect to the mean. Linear Kalman filtering has been an active research subject for the last 50 years and the literature is incredibly broad, some classical references include the seminal works of Kalman and Bucy, as well as the classic textbooks (Jazwinski, 1970),

(Sorenson, 1985), or (Simon, 2006) that discusses some newer results concerning robust filtering and estimation with state constraints. There are two other properties of Kalman filtering worth mentioning. Kalman's gain is computed from the uncertainties of the predicted state and the current measurements. Hence, in order to calculate the innovation gain, Kalman's filter must also compute the full covariance matrix of the system, which can be seen as a "state augmentation" of the system. For example, the estimation of a system with three states is accomplished computing the estimates together with other nine elements of the covariance matrix. That augmentation may not be a problem in system with few states. However, memory and computation power can be an issue for systems with larger dimensions. The optimality claim of the Kalman filter is strongly related to the knowledge of the model, *i.e.* parameters of the dynamics, covariance and correlation of process and measurement noises. Estimate divergence can be a problem if the filter operates using an erroneous model over a large amount of data. In such cases, the filter will "learn the wrong model too well", and eventually future observations will have small influence in future estimate calculations. There exist several techniques that retune the filter parameters in order to mitigate the divergence or recover the estimates from divergence (Jazwinski, 1970, pp. 305-323) and, at this level, Kalman filter tuning can become a mixture of rules of the thumb and art.

In many practical applications, a deterministic approach can be preferable due to the lack of suitable statistical models for the system dynamics and sensor measurements. State observers are an option to the filtering problem using a deterministic rationale. State observers for linear time-invariant systems are also called Luenberger observers, for the early work of David Luenberger on the design and characterization of properties for this approach, *i.e.* (Luenberger, 1964), (Luenberger, 1966). Classic control techniques require full state feedback, which is specially difficult for cases where each state could not be directly measured. Luenberger observers can provide a solution to that problem, and they are computationally simpler than the Kalman filter because the estimates are obtained using a constant matrix in the feedback dynamics of the estimates. Therefore, liner observer filtering is performed using only the estimates and measurements, *i.e.* without "state augmentation", and state observers need less memory and computational power. Luenberger proved that if the system is completely observable, then a state observer can be designed to provide an asymptotically convergent estimate with an arbitrary settling time for the transient response. He also remarked that both state observer and Kalman filter share the same dynamical structure. From this equivalence property, it is possible to compute a steady state Kalman gain and use it in a state observer form, thus resulting in what is often called steady state Kalman filter. The literature also covers really well the design and analysis of Luenberger observers, main references are Luenberger seminal papers as well as other classical linear system text books, *e.g.* (Kailath, 1979), and (Chen, 1984).

The solutions discussed so far deal with the state estimation problem in linear systems, where important results, such as optimality and asymptotic stability of the estimate, were obtained. However, most of real-world problems are nonlinear in nature, notably the classic rigid body motion employed for pose estimation. Unfortunately, solutions obtained for linear applications may not perform satisfactorily in nonlinear applications, and, for this reason, state estimation for nonlinear systems has been a trend research area in the past 50 years. For the state estimation of nonlinear systems, the differences between stochastic and deterministic filtering become more noticeable. Nonlinear versions derived after Kalman's linear solution are largely available and still under active research. The extended Kalman filter (EKF) is unarguably the simplest Kalman-based approach for nonlinear estimation, and has been present in countless applications since early applications. The EKF is computed after the first order Taylor expansion of the system equations using the current state estimate. Differently from the linear version, however, that approach provides only weak convergence properties, *e.g.* the es-

timation error is likely to converge to zero only for small initial errors, which can be seen from EKF's equivalence to a first-step computation of the Gauss-Newton method for the likelihood maximization (Bell and Cathey, 1993). The convergence properties of the filter in continuous time will depend on other properties as boundedness of the linearized model. Another recurring problem is the filter divergence when the linearized model is not computed close enough to the current state. In this case, as discussed for the linear case, the filter can "learn too well" the erroneous model, thus leading to the divergence of the estimates.

The literature shows numerous improvements to overcome difficulties faced by the EKF. For instance, the iterated EKF can reduce the effects due to nonlinearities of the measurement function (Jazwinski, 1970, pp. 279-280). The iterated EKF is equivalent to computing the nonlinear maximum likelihood solution using an iterative Gauss-Newton formulation (Bell and Cathey, 1993). Furthermore, the multiplicative EKF (Lefferts et al., 1982) addresses attitude estimation using unitary quaternions. Quaternions form a Lie Group (Warner, 1987), whose group operator is nonlinear, differently from the element-wise sum for $\mathbb{R}^n$ vectors. Leffets et al. noticed this problem and proposed an *ad-hoc* solution using reduced space and changing the update rule so as to respect the quaternion group operation. More recently, (Bonnabel et al., 2009b) proposed an invariant EKF that exploits system symmetries with a Kalman-based formulation. This invariant formulation aims at respecting the geometry of the state space when it corresponds to a manifold. As a result, the invariant EKF provides an autonomous linearized dynamics for a larger set of trajectories, where the invariant EKF is expected to perform better than the standard EKF. These three solutions are not but few examples that have been presented in this 50-year literature. There are still numerous other technical solutions, and most of these can be combined with other engineering tricks such as using other coordinates to describe the states, different procedures to tune the process noise. Therefore, it is difficult to claim such thing as "the EKF", but instead an EKF implementation, that relies on the employed techniques.

The unscented Kalman filter (UKF) is another Kalman-based solution that has earned some attention in the past 15 years. This technique is named after the unscented transform (Julier and Uhlmann, 2004; Julier et al., 2000), which is supposed to approximate the covariance of distributions in nonlinear transformations better than the approximation provided by first order Taylor series. The computation of the UKF assumes a Gaussian distribution, and the authors provide a set of rules to carefully choose the points to approximate better the nonlinear transformation of the probability distribution. The authors show large improvements using the unscented transform for polar to Cartesian coordinate transformation, and other works have presented the UKF as very effective alternative for attitude and position estimation, *e.g.* (Crassidis, 2006). Up to this author's knowledge, the problem of filter divergence is, more than often, unaddressed in the literature, most likely because the UKF does not depend on system linearization. However, the UKF seems to be as prone to modeling errors as the Kalman filter, of course, referring the UKF to the nonlinear and the Kalman filter to linear problem.

Concluding the review on stochastic nonlinear estimation, notice that Kalman-based filtering is not the only methodology available. In the past 20 years, Particle filters (Doucet et al., 2000b, and references therein) have been another recurring solution for the data fusion problem. Each particle represents a realization of the state, which is obtained, for instance, using sequential Monte Carlo sampling and techniques to avoid the degeneracy of the algorithm. These methods can cope with nonlinear dynamics and general probability distributions, instead of the common Gaussian assumption. Particle filters, however, suffer from the "curse of dimensionality" (Daum, 2005), since the number of samples must increase with the dimension of the system's states. Furthermore, the Rao-Blackwelisation technique (Doucet et al., 2000a) aims at decoupling states of the system in a hierarchical form to reduce the problem due to the

growth of state dimension. That technique is somewhat similar to the nonlinear design based on interconnected subsystems (Nijmeijer and van der Schaft, 1990, pp. 337-344). Due to its computational cost, particle filters remain restricted to problems where Kalman-based filtering is not efficient. An interesting and rather complete reference for discussion, pros-and-cons of stochastic filtering methods can be found in (Daum, 2005), including a set of methods that can solve the probability density function for some particular families of nonlinear dynamics.

As we have discussed, deterministic approaches do not focus on approximating the probability functions, but on providing filters that guarantee convergence of the estimates to the real state values in the absence of noise. Nonlinear observers are a class of deterministic filters employed when linear filtering, *i.e.* Luenberger observers, is unsuited to guarantee estimate convergence for nonlinear systems. The development of nonlinear observers is closely related to nonlinear control theory, and relies very often on rigorous proofs of stability. The literature presents several solutions for different nonlinear structures. For instance, systems with a state-affine form can be solved efficiently using Kalman-like observers (Hammouri and de Leon Morales, 1990). Furthermore, systems linearizable by the output allow a linear representation using a well defined variable change (Hammouri and Gauthier, 1992), afterwards methods from linear theory can be applied. High-gain observers are capable of providing exponentially stable estimates for nonlinear systems, as the technique's names suggests, under high gain assumptions (Esfandiari and Khalil, 1992), (Gauthier and Kupka, 1994). Also, we should not be surprised to see that the EKF yields a locally stable observer, whose basin of convergence can be large for systems with weak non-linearities (Song and Grizzle, 1995), equivalently to that Kalman filter that can be expressed as a Luenberger observer. (Busvelle and Gauthier, 2002) shows how to include a high-gain structure to increase the basin of convergence from the EKF. These filters have, of course, much more technical details than discussed here. Surveys such as (Besançon, 2007; Nijmeijer and Mareels, 1997, and references therein) discuss deeper properties needed and results provided by these nonlinear observers.

The deterministic solutions previously discussed often consider state spaces given by manifolds, however, they do not explore Lie group properties often present in system dynamics. There have been works in the literature discussing the estimation on Lie groups. The seminal (Salcudean, 1991) proposes a nonlinear observer for the attitude estimation of a rigid body using the unit-quaternion parametrization for the special orthogonal group. That work is followed by several others on attitude estimation that, either employ quaternion representation, *e.g.* (Vik and Fossen, 2001), (Thienel and Sanner, 2003), and (Martin and Salaun, 2008), or directly exploit the structure from special orthogonal group, *e.g.* (Mahony et al., 2005), (Campolo et al., 2006), and (Vasconcelos et al., 2008b). However, attitude estimation is not the sole domain where Lie group structure has been applied, other works exploit the special Euclidean group (Rehbinder and Ghosh, 2003), (Baldwin et al., 2007), (Vasconcelos et al., 2007) for pose estimation, and the special linear group (Mahony et al., 2012) for planar homography estimation. These works share the goal of exploring lie Group properties, however, each one is developed for their own specific application. On the other hand, other authors have exploited group properties for general system dynamics with a Lie group structure (Maithripala et al., 2005), (Bonnabel et al., 2008), (Bonnabel et al., 2009a), (Lageman et al., 2010). For instance, Maithipala et al. present a method to estimate the velocity of the system based on configuration measurements. Furthermore, Bonnabel et al. propose a similar structure to the invariant tracking (Martin and Rudolph, 1999) for the design of nonlinear observers for systems with state symmetries. The provided design is constructive, and, although their estimator can only guarantee local stability, the convergence properties are stronger than the ones provided by EKF as they are reinforced by symmetry properties present in the error dynamics. More recently, Lageman et al. proposed a method to write gradient-like observers for systems with

full measurements and proves almost global stability if the structure allows a Morse-Bott energy function.

This thesis treats the problem of pose estimation using information from inertial and visual sensors. Rigid body motion is the main dynamics considered, where the main part of the state is defined by body pose, body orientation and position, together with its time-derivatives, linear and angular velocities and accelerations. We also take other static parameters, *e.g.* sensory effects and frame-to-frame coordinates, into account to improve system's representation and the data fusion process. However, these static parameters may also increase the complexity of the dynamics. Hence, we exchange a simple but inaccurate model for more complete one with the compromise of having a system that is often observable under specific conditions.

1.5 ATTITUDE AND POSE ESTIMATION

We have seen that state estimation is a broad research domain by itself. Furthermore, the topic of attitude estimation has been of great interest to part of the research community since early works on state estimation. In essence, the attitude estimation problem consists of computing the rotation matrix of a frame $\mathcal{B}$ in $\mathcal{R}$ coordinates, *i.e.* ${}^{\mathcal{R}}R_{\mathcal{B}}$, from N correspondences of multiple vectors ${}^{\mathcal{R}}v_i$ with $i = 1, \cdots, N$, measured in $\mathcal{B}$ coordinates, *i.e.* ${}^{\mathcal{B}}v_i$. Furthermore, this estimation problem can be improved considering the information provided by angular rate gyroscopes. The works by (Shuster, 1993) and (Crassidis et al., 2007) present a deep review of early developments on attitude estimation, while the former has a deeper historical focus and the latter is more technical. (Hua, 2009b, Chapter 3) presents another very complete review of the developments shown in the literature until the referred year.

One reason why attitude estimation is such an active research area is that there exist several solutions that, in turn, can provide different properties and guarantees. The TRIAD algorithm (Kim et al., 1964), for instance, provides a closed form solution for the attitude estimation problem, *i.e.*

$$ {}^{\mathcal{R}}R_{\mathcal{B}} = \begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix}^T $$

with the orthonormal vector triads given by s_i and r_i

$$ s_1 \triangleq \frac{{}^{\mathcal{B}}v_1}{|{}^{\mathcal{B}}v_1|}, \quad s_2 \triangleq \frac{{}^{\mathcal{B}}v_1 \times {}^{\mathcal{B}}v_2}{|{}^{\mathcal{B}}v_1 \times {}^{\mathcal{B}}v_2|}, \quad s_3 \triangleq s_1 \times s_2, $$
$$ r_1 \triangleq \frac{{}^{\mathcal{R}}v_1}{|{}^{\mathcal{R}}v_1|}, \quad r_2 \triangleq \frac{{}^{\mathcal{R}}v_1 \times {}^{\mathcal{R}}v_2}{|{}^{\mathcal{R}}v_1 \times {}^{\mathcal{R}}v_2|}, \quad r_3 \triangleq r_1 \times r_2. $$

This method, however, presents some disadvantages. First, it is computed *exclusively* from two vector measurements. Secondly, although this method provides the exact orientation matrix for perfect measurements, it is severely impaired by noisy measurements. (Wahba, 1965) stated the least-squares criterion for the attitude using multiple measurements, *i.e.*

$$ {}^{\mathcal{R}}R_{\mathcal{B}} \triangleq \underset{R \in \mathbb{M}(3)}{\arg\min}\ J(R) = \sum_{i=1}^{N} a_i |{}^{\mathcal{R}}v - M\,{}^{\mathcal{B}}v|^2, $$

with weights $a_i > 0$, and the solution is subjected to $MM^T = I_3$ and $\det(M) = 1$. Several authors presented solutions in (Farrel et al., 1966), however, the solutions involved some type of decomposition and were not practical in time due to computational complexity. Even recently, other solutions based on decompositions have been proposed to solve that least-squares problem, *c.f.* (Sanyal, 2006), for example. This problem was considered a difficult task until the quaternion parametrization was effectively exploited by the q-method (Davenport, 1968), and

the further development of quaternion estimation (QUEST) algorithm (Shuster, 1978). Davenport's method computes the optimal solution using the unitary quaternion instead of the attitude matrix, and the solution is given by the unitary eigenvector associated to the largest eigenvalue of the matrix

$$D = \begin{bmatrix} \gamma & z^T \\ z & C - \gamma I_3 \end{bmatrix},$$

with $z = \sum_{i=1}^{N} a_i {}^{\mathcal{R}}v_i \times {}^{\mathcal{B}}v_i$, and $B = \sum_{i=1}^{N} a_i {}^{\mathcal{R}}v_i {}^{\mathcal{B}}v_i{}^T$, such that $\gamma = \mathrm{tr}(B)$ and $C = B + B^T$. Initially, the eigenvalue computation was the bottleneck of Davenport's method, but Shuster showed with the QUEST algorithm that the largest eigenvalue is obtained solving the 4-th polynomial in λ

$$\lambda^4 - (a+b)\lambda^2 - c\lambda = (ab + c\gamma - d) = 0,$$

with $a = \gamma^2 - \mathrm{tr}(\mathrm{adj}(C))$, $b = \gamma^2 + |z|^2$, $c = \det(C) + z^T C z$. One important simplification provided by QUEST is that λ is obtained after a few steps of a Newton-Raphson iterative solver initialized at $\lambda_0 = \sum_{i=1}^{N} a_i$, *c.f.* (Shuster, 1978).

Those techniques refer to the computation of an attitude matrix from vector measurements. Moreover, similarly to the general case of state estimation for nonlinear systems, the extended Kalman filter has been the incipient method for data problem with other sensors (Crassidis et al., 2007), *e.g.* angular rate gyroscopes. However, the attitude estimation problem is often over-parametrized, *i.e.* a rotation matrix can be represented by three parameters instead of the 9 elements of a rotation matrix, or the 4 elements of a quaternion, which can lead to instability of the filter. The multiplicative Kalman-filter addresses this issue by reducing the dimension of the covariance matrix, and such a solution has been studied, for instance, by (Lefferts et al., 1982) using an EKF approach, (Vandyke et al., 2004) via unscented, or (Crassidis, 2006) via sigma-point Kalman filtering. Although the latter results can improve significantly the propagation of the probability distribution, up to the authors knowledge, the achievement of *almost* global convergence using these methods remains to be achieved. More recently, the invariant Kalman filter (Bonnabel et al., 2009b) was also employed for attitude estimation. This result is more interesting with respect to the previous approaches, since the resulting error dynamics is autonomous for a larger set of trajectories. Additionally, the attitude estimation problem can be rewritten as estimation of two non-collinear and unconstrained vectors, instead of the attitude matrix or some parametrization (Batista et al., 2009). The resulting system becomes linear time-varying where the estimation problem is solved optimally via a standard Kalman-filter. The problem with this approach is that, since the estimated vectors are not necessarily orthonormal, one must still rely on techniques such as TRIAD or QUEST in order to recover the rotation matrix.

In the past years, some effort has been put in the development of nonlinear observers for attitude, attitude-heading reference (AHR) and pose estimation. These nonlinear observers are usually application-specific estimators that take advantage of structural properties of the models. Although some technical differences can be noticed between different nonlinear observers, this class of estimators commonly share the benefits of global, or at least semi-global, stability proofs. (Salcudean, 1991) is likely the seminal nonlinear observer for attitude estimation. This observer provides a unitary quaternion estimate of the orientation, and it is followed by numerous works exploiting either the unit quaternion, *e.g.* (Thienel and Sanner, 2003), (Tayebi et al., 2007), (Martin and Salaün, 2010), or the $\mathbb{SO}(3)$ Lie Group structure, *e.g.* (Campolo et al., 2006), (Vasconcelos et al., 2008b), (Mahony et al., 2008), (Hua, 2009a), (Zamani et al., 2011), (Hua et al., 2013). Most of these nonlinear observers consider the estimation of gyroscope bias, moreover, the AHR nonlinear observer shown in (Martin and Salaun, 2008) also accomplishes (partial) accelerometer bias estimation.

The full pose estimation problem has also drawn the attention of the research community. In order to solve the full pose estimation using nonlinear observers, the decoupling of the attitude and translational displacement is often considered. (Vik and Fossen, 2001) is one of the first works to consider nonlinear observers for pose estimation using a decoupled approach for attitude–position estimation. An observer for attitude and position using inertial data and visual line features, is presented in (Rehbinder and Ghosh, 2003), despite the fact that biases are considered for the presented simulation results, there is no procedure for gyroscope nor accelerometer bias estimation in the observer. The estimation using the $\mathbb{SE}(3)$ representation by means of IMU and bearing measurements is addressed in (Baldwin et al., 2009), but online IMU bias estimation is not performed. A cascaded nonlinear observer for attitude and position is presented in (Vasconcelos et al., 2008a) exploiting IMU measurements together with GPS measurements. Exponential convergence for attitude and position is achieved, together with gyroscope bias estimation. (Cheviron et al., 2007) presents a solution for full pose estimation, however due to misconceived hypothesis on the accelerometer measurement definition the convergence proof for position estimation is only valid for small angular velocities. (Barczyk and Lynch, 2012) present an implementation of invariant observers for pose estimation with gyroscope and accelerometer bias estimation, however, the resulting observer has but local stability properties.

1.6 VISUO-INERTIAL SYSTEMS

Visuo-inertial systems consist of a set of a camera together with inertial sensors, *i.e.* accelerometers and angular rate gyroscopes, of an IMU. (Viéville and Faugueras, 1989) are likely the first researches to propose the use of inertial information for improving pose estimation via computer vision. This data fusion application has shown to be very adequate since high frequency measurements, usually from 100 up to 1000 Hz, obtained by the IMU provide the incremental displacement of the body, while the camera can provide accurate relative pose measurements, however, at low frequencies, from 5 up to 40 Hz. The sensors present complementary characteristics as the IMU provides information about fast movements of the systems, however, the pose estimation via pure inertial data drifts after a few seconds. Moreover, we can obtain accurate relative pose estimates using information provided by the camera to bound that drift. The data fusion process is impaired by measurement errors and uncertainties on the system dynamics. A first source of difficulties comes from IMU measurement bias, which can be significant for most low-cost IMUs used in robotic applications. Another source of difficulties concerns various parameters related to the use of different coordinate frames, *e.g.* the camera and the IMU frames.

Researchers made extensive research in visuo-inertial fusion during the past decade. A large part of the effort was put towards either immediate applications of Kalman filters, *c.f.* , for example, (Azuma et al., 1999), (Lobo and Dias, 2003), (Armesto et al., 2004), and (Servant et al., 2010) or nonlinear observers, *e.g.* , (Rehbinder and Ghosh, 2003), (Brás et al., 2011), for the fusion of inertial data with pose estimates from visual tracking algorithms. The calibration of bias and scale factors of the inertial sensors was already discussed in (Viéville and Faugueras, 1989), and thenceforth most of visuo-inertial data fusion algorithms take these parameters into account. However, biases and scales factor are not the only parasite modeling errors, since other parameters related to the use of multiple coordinate frames can also impair the estimation process.

The calibration of parameters relating to multiple frames was a known problem in computer vision before the early years of visuo-inertial data fusion (Tsai and Lenz, 1989). Several works neglect the calibration topic, mostly because these parameters can be negligible, approximated via CAD model or calibrated in a previous phase. The calibration task is not simple tough. Up

to the author's knowledge, (Foxlin and Naimark, 2003) is the pioneer to discuss this problem with more details, where the authors presented a solution based on an EKF with multiple frame parameters, also bias and scales of inertial sensors. Other works followed by trying to understand the problem with more depth. (Lang and Pinz, 2005) presented a technique to estimate the rotation from the camera-frame to IMU-frame (c-to-IMU frame) using the difference of rotation from the sensors. (Lobo and Dias, 2007) later proposed a technique that requires multiple synchronized measurements of accelerometers and rotation of the camera in static positions in order to estimate the c-to-IMU rotation, while the motion of a turning table is employed to recover c-to-IMU translation. (Hol et al., 2010) present the problem as a hybrid solution of an EKF and a black-box optimization of the residual errors due to the parameters, and (Fleps et al., 2011) present a batch optimization technique based on the robust pseudo-Huber residual, *c.f.* Appendix B.

In common, those techniques conclude that one must obtain measurements in different positions to obtain satisfactory results for the calibration procedure. In fact, that conclusion is strictly related to the the observability properties of the system, which were not discussed until (Jones et al., 2007) and (Mirzaei and Roumeliotis, 2008). Later, (Kelly and Sukhatme, 2011), (Jones and Soatto, 2011) extended the observability analysis to other system configurations, *e.g.* in order to include measurements from monocular vision. For instance, every analysis performed so far characterized that the frame calibration and bias estimation of the system is observable under certain angular motion. However, there is not much that can be claimed on the properties of the movements along which observability is granted due to the analysis there performed. More recently, the works from (Martinelli, 2011) and (Martinelli, 2012) showed an observability analysis for different configurations of visuo-inertial systems. This study evaluates the conditions under which one obtains indistinguishable output trajectories from different movements of the sensor. That analysis include cases with specific motions and configurations under which resulting system dynamics is observable. The latter work provides also a closed-form solution for the estimation of pose and multiple parameters of the system.

The main objective of this thesis is to develop new techniques for concurrent pose estimation, IMU bias and c-to-IMU frame calibration. We first study the observability properties of the system in order not to simply state whether the system is observable, but also to define angular movements that can guarantee the complete observability of the system. We dedicate Chapter 3 to that observability analysis and to the development of the new estimation techniques. Next, we describe the variables considered in different system configurations with the respective dynamics.

System description

We reviewed the tools and theory to state the problem of visual inertial estimation in the previous sections. We have recalled the basic structure of image formation, *i.e.* how to compute the relative pose between two images and sensors that can be employed to obtain angular velocity and linear acceleration. A visuo-inertial sensor consists of a rigid mount of camera and inertial sensors. The objective is to combine pose measurements provided by a camera at a relatively low frequency with high frequency measurements of the angular velocity and proper acceleration provided by the gyros and accelerometers. Figure 1.9 depicts the basic structure, *i.e.* coordinate frames, employed in this thesis to model pose estimation using visuo-inertial sensors. We assume two frames $\mathcal{B}$ and $\mathcal{C}$ attached to the same rigid body. The objective of the problem is to estimate the pose (orientation and position) of $\mathcal{B}$ with respect to some (partially) known inertial reference frame $\mathcal{R}$. The rigid body is moving, therefore frames $\mathcal{B}$ and $\mathcal{C}$ depend on time. The representation of frame $\mathcal{B}$ in $\mathcal{C}$ frame is constant however. Hence,

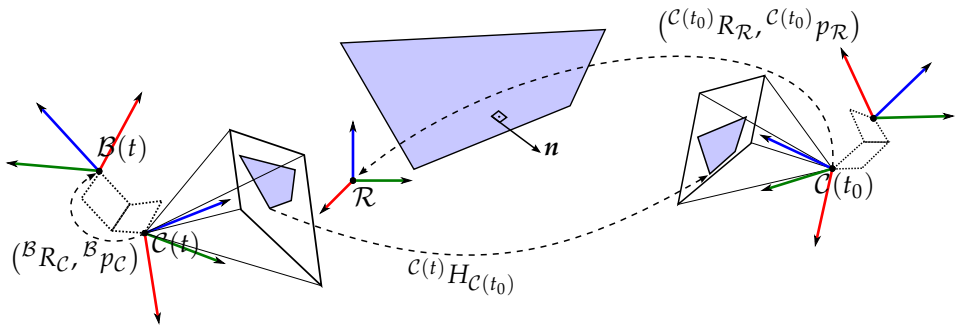


Figure 1.9: Important coordinate frames in visuo-inertial systems

we can define ${}^{\mathcal{B}}(R,p)_{\mathcal{C}}$ as the constant rotation and translational displacement of $\mathcal{B}$ in $\mathcal{C}$. Moreover, we have seen in Section 1.3.4 how we can exploit images from a planar surface with an associated normal vector $\boldsymbol{n}$ in order to measure relative pose. It is therefore reasonable to assign one frame, *e.g.* $\mathcal{C}$, to the optical center of the camera. In this way, we are able to obtain relative pose measurements from with respect to this frame.

In order to define the pose of a body, we must first assign a frame $\mathcal{B}$ to the rigid-body and an inertial reference frame $\mathcal{R}$ as appropriate. *E.g.* , we can employ the canonical frame for $\mathcal{R}$ such that the origin is placed in the initial position of the body, e_3 coincides with the direction of the gravitational acceleration, e_1 points to the magnetic north and e_2 is defined to obtain a right hand frame. In this situation, we are strictly related to the knowledge of ${}^{\mathcal{R}}\boldsymbol{n}$. In other situations, we can assume $\mathcal{R}$ as the initial configuration of the rigid body, however, we may deal with unknown variables, such as the value of gravitational acceleration in $\mathcal{R}$ frame.

We write the body pose by its orientation ${}^{\mathcal{R}}R_{\mathcal{B}}(t)$ and position ${}^{\mathcal{R}}p_{\mathcal{B}}(t)$, and, using Eqs. (1.4) and (1.6), we can write the dynamics of the body frame as:

$$ {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \quad {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \quad {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \tag{1.25} $$

where ω, v and a define, respectively, the angular velocity, linear velocity and linear acceleration of $\mathcal{B}$ with respect to $\mathcal{R}$, such that ${}^{\mathcal{B}}\omega$, ${}^{\mathcal{B}}a$ denote the coordinates of ω and a expressed in $\mathcal{B}$ and ${}^{\mathcal{R}}v$, denotes coordinates of v expressed in $\mathcal{R}$.

Strapped-down gyros and accelerometers measure respectively ${}^{\mathcal{B}}\omega_{\mathrm{y}}$ and ${}^{\mathcal{B}}a_{\mathrm{y}}$ according to the models (1.14) and (1.15). Recall that b_ω and b_a denote gyroscope and accelerometer biases, and the acceleration due to gravitational field in body coordinates writes ${}^{\mathcal{B}}g = {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g$, then, we write measurements and the respective bias dynamics as

$$ {}^{\mathcal{B}}\omega_{\mathrm{y}} = {}^{\mathcal{B}}\omega + b_\omega\,, \quad {}^{\mathcal{B}}a_{\mathrm{y}} = {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g\,, \quad \dot{b}_\omega = 0_{3\times 1}, \quad \dot{b}_a = 0_{3\times 1}\,. \tag{1.26} $$

We are not capable, in general, of making direct pose measurements ${}^{\mathcal{R}}(p,R)_{\mathcal{B}}$ from images. Instead, we employ the camera and obtain relative pose measurements. More specifically, let us denote by $\mathcal{C}$ a coordinate system attached to the optical center of the camera, with sensor-to-sensor pose ${}^{\mathcal{B}}(p,R)_{\mathcal{C}}$ and, for the sake of notation, we define the trajectory of the camera pose by $\mathcal{C}(t)$. Using the proposed notation, we thus have the relative pose between the current $\mathcal{C}(t)$ and initial $\mathcal{C}(t_0)$ given by the pair ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{C}(t_0)}$.

We obtain different dynamics for the estimation procedure depending on the configuration of the sensors and the previous knowledge about the scene. Next, we present 3 different situations with increased degree of complexity. The first two situations concern systems with calibrated relative pose between camera and inertial sensors, while the last discusses concurrent pose estimation and sensor-to-sensor calibration.

1.6.1 *Dynamics with known frames*

The first case considers previously known, or calibrated sensor-to-sensor pose and full knowledge of the scene. More specifically, we assume:
- knowledge of gravitational field expressed in $\mathcal{C}(t_0)$ coordinates;
- full pose measurement, *i.e.* ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{C}(t_0)}$;
- knowledge of sensor-to-sensor relative pose, *i.e.*

$$ {}^{\mathcal{B}}(p,R)_{\mathcal{C}} \triangleq {}^{\mathcal{B}(t)}(p,R)_{\mathcal{C}(t)} = {}^{\mathcal{B}(t_0)}(p,R)_{\mathcal{C}(t_0)}. $$

The first two items refer to the knowledge of the scene. The first item allows us to define the inertial reference frame $\mathcal{R} = \mathcal{C}(t_0)$ and specify ${}^{\mathcal{R}}g$. Furthermore, the second item allows us to compute ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{R}}$. The third item, associated to the second one, enables us to compute full pose ${}^{\mathcal{R}}(p,R)_{\mathcal{B}(t)}$. The full system model, *i.e.* body pose and sensor bias, is written after Eqs. (1.25) and (1.26):

$$ \begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \\ {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \\ \dot{b}_{\omega} = 0_{3\times 1}\,, \\ \dot{b}_{a} = 0_{3\times 1}\,, \end{cases} \tag{1.27} $$

with measurements

$$ (p_{\mathrm{y}}, R_{\mathrm{y}}, {}^{\mathcal{B}}a_{\mathrm{y}}, {}^{\mathcal{B}}\omega_{\mathrm{y}}) = \left({}^{\mathcal{R}}p_{\mathcal{B}}\,, {}^{\mathcal{R}}R_{\mathcal{B}}\,, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g\,, {}^{\mathcal{B}}\omega + b_{\omega}\right). \tag{1.28} $$

1.6.2 *Unknown gravitational field*

The second case concerns an extension of the previous one, *i.e.* the gravitational field ${}^{\mathcal{R}}g$ is unknown. In this situation, we write the inertial reference frame $\mathcal{R} = \mathcal{C}(t_0)$, however, we must include the gravitational field $g_{\mathcal{R}}$ in the dynamics. Since $\mathcal{R}$ is inertial, remark that $\dot{g}_{\mathcal{R}} = 0$. The configuration for this section considers
- relative full pose measurement, *i.e.* ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{C}(t_0)}$;
- knowledge of sensor-to-sensor pose, *i.e.* $({}^{\mathcal{B}}R_{\mathcal{C}}, {}^{\mathcal{B}}p_{\mathcal{C}})$.

Similarly to the previous system, we can determine the current camera pose ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{R}}$, thus, via the parameters of sensor-to-sensor pose, we compute ${}^{\mathcal{R}}(p,R)_{\mathcal{B}}$. The full system model, *i.e.* pose, sensor bias and gravitational field, is written using Eqs. (1.25) and (1.26):

$$ \begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \\ {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \\ \dot{b}_{\omega} = 0_{3\times 1}\,, \\ \dot{b}_{a} = 0_{3\times 1}\,, \\ {}^{\mathcal{R}}\dot{g} = 0_{3\times 1}, \end{cases} \tag{1.29} $$

with measurements

$$ (p_{\mathrm{y}}, R_{\mathrm{y}}, {}^{\mathcal{B}}a_{\mathrm{y}}, {}^{\mathcal{B}}\omega_{\mathrm{y}}) = \left({}^{\mathcal{R}}p_{\mathcal{B}}\,, {}^{\mathcal{R}}R_{\mathcal{B}}\,, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g\,, {}^{\mathcal{B}}\omega + b_{\omega}\right). \tag{1.30} $$

1.6.3 *Sensor-to-sensor self calibration*

The previous two systems considered a pre-calibration or knowledge of sensor-to-sensor relative pose. That assumption much simplifies the problem because we can assume to recover full body pose. The next cases deal with the problem of pose estimation and concurrent camera-to-inertial sensors pose estimation.

The third situation concerns the extension of Section 1.6.2 with sensor-to-sensor calibration. We assume for this section full relative pose measurement, *i.e.* ${}^{\mathcal{C}(t)}(p,R)_{\mathcal{C}(t_0)}$, and define the inertial reference frame $\mathcal{R}=\mathcal{C}(t_0)$. Full body pose, however, is obtained using the parameters ${}^{\mathcal{B}}(p,R)_{\mathcal{C}}$ included in the model. Since we have a rigid body mount for the camera and inertial sensors, sensor-to-sensor relative pose is constant. The full system model, *i.e.* pose, sensor bias, gravitational field and sensor-to-sensor relative pose, is given by:

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \\ {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \\ \dot{b}_{\omega} = 0_{3\times 1}\,, \\ \dot{b}_{a} = 0_{3\times 1}\,, \\ {}^{\mathcal{R}}\dot{g} = 0_{3\times 1}\,, \\ {}^{\mathcal{B}}\dot{R}_{\mathcal{C}} = 0_{3\times 3}\,, \\ {}^{\mathcal{B}}\dot{p}_{\mathcal{C}} = 0_{3\times 1}\,, \end{cases} \tag{1.31}$$

with measurements

$$(p_{\mathrm{y}}, R_{\mathrm{y}}, {}^{\mathcal{B}}a_{\mathrm{y}}, {}^{\mathcal{B}}\omega_{\mathrm{y}}) = \big({}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}\,, {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}\,, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g\,, {}^{\mathcal{B}}\omega + b_{\omega}\big). \tag{1.32}$$

2

DIRECT VISUAL TRACKING

Visual tracking can be an effective method to compute relative pose. First, the cameras are non-intrusive and passive sensors, and multiple images can provide information of different points of view of the scene. Therefore, we can can compute the trajectory of the camera from the changes shown in these images. The visual tracking problem can be defined as the problem of finding the transformation parameters that best align a reference image to the following frames in a video stream. In contrast to feature-based methods, which are built on the extraction and matching of a sparse set of characteristics from the image, direct visual tracking methods exploit each individual pixel's intensity to solve the visual tracking. This chapter concerns direct visual tracking methods.

The quality of the match between two images is measured by a similarity function. Typically, direct visual tracking methods are built upon the sum of squared differences (SSD). The SSD has proven to be very efficient, mainly because the optimization can be much simplified due to numerous solutions to nonlinear least squares (Baker and Matthews, 2001; Benhimane and Malis, 2007). SSD tracking, however, is severely impaired when brightness constancy is violated, since displacement of the camera and photometric variations are dealt in the same way by the similarity function. In order to overcome this problem, the photometric model of the scene can be estimated using off-line (Hager and Belhumeur, 1998) or on-line procedures (Bartoli, 2008), (Silveira and Malis, 2010). The normalized cross correlation (NCC) is another suitable choice for a similarity measure. The NCC has radius of convergence comparable to the SSD, additionally, this similarity is invariant to affine illumination changes, and its computation is simpler than probabilistic solutions. Other examples of similarity measures include sum of the conditional variance (Pickering et al., 2009), mutual information (Viola and Wells, 1997), the cross cumulative residual entropy (Wang and Vemuri, 2007) and the correlation ratio (Roche et al., 1998b). These similarities have also been applied to direct visual tracking, and they relax the brightness constancy to more complex photometric variations and multimodal images.

Given a function that provides the quality of the match between two images, we must further solve a nonlinear optimization problem that is not globally convex, or concave in the case of a maximization problem, in order to obtain the parameters of the best transformation. It is not an easy task to find the globally optimal solution to this problem. Global optimization techniques such as the simulated annealing, *c.f.* (Horst and Pardalos, 1995), are too cumbersome for real-time implementation. Therefore, direct visual tracking relies strongly on gradient-based optimization methods, since these strategies present a good trade off between region of convergence and computational cost.

We propose a novel solution using the NCC as similarity measure. This similarity is chosen because of its simplicity. We rely on the property that the similarity is intrinsically invariant to affine illumination changes, which is a powerful characteristic that allied with two techniques here presented can improve the robustness to nonlinear illumination and partial occlusion. These techniques are based on subregion partitioning, and weighting using a residue invariant

to affine illumination variations. We also propose a method to improve the gradient solution while having a well defined optimization problem. This method was introduced in (Scandaroli et al., 2012).

We test several direct tracking methods using different optimization techniques with SSD, SCV and MI similarities. These techniques are evaluated using synthetic images, the planar based visual tracking benchmark dataset from (Lieberknecht et al., 2009) and challenging real-world video sequences. The results obtained show that the choice of the similarity is important, however, the optimization approach also plays an determinant role in a visual tracking method. The method prosed in this thesis is suited for visual tracking under complex illumination variation, and tracking can still be performed for partially occluded targets under extreme illumination settings.

2.1 PROBLEM DESCRIPTION

Direct visual tracking exploits the intensity of the pixels in order to define the best alignment between two images $\boldsymbol{I}_{\mathrm{R}}$ and $\boldsymbol{I}_{\mathrm{C}}$. Hence, the problem corresponds to finding the parameters x associated with a warp function $\boldsymbol{w}(x,\cdot)$ that maps the intensities of the pixels from an image $\boldsymbol{I}_{\mathrm{C}}$ to intensities of $\boldsymbol{I}_{\mathrm{R}}$. Similarity functions $\boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{I}_{\mathrm{C}}) \in \mathbb{R}$ represent scores of how good is the matching between two images, thus the direct visual tracking problem can be represented by the following optimization:

$$\widehat{x} = \arg\operatorname*{opt}_{x} \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big)\,. \tag{2.1}$$

If $\boldsymbol{S}$ is maximized when the images are best matched, then (2.1) is defined as a maximization problem. Conversely, (2.1) is defined as a minimization problem if $\boldsymbol{S}$ is minimized when the images are best matched.

We can point out three main components that make a direct visual tracking method:
- the similarity function,
- the optimization approach,
- the warp function.

The following sections discuss these aspects.

Recalls on notation

Let us recall from Section 1.3 that an image with m rows and n columns is defined by $\mathfrak{I}$. An image $\mathfrak{I}$ with bit depth $B_{\boldsymbol{I}}$ stores the intensities $\boldsymbol{I} \in [0, B_{\boldsymbol{I}} - 1]$ of pixels with coordinates $\boldsymbol{p} \in \Xi \subset \mathbb{P}^2$, and

$$\mathcal{H} = \{\boldsymbol{I} \colon \Xi \subset \mathbb{P}^2 \to \mathbb{R}\}, \quad p \mapsto \boldsymbol{I}(\boldsymbol{p})$$

where $\mathcal{H}$ defines the set of functions that map the coordinates $\boldsymbol{p}$ of a pixel to its respective intensity in some $\mathfrak{I}$.

The coordinates of a pixel can be changed via parameters $X \in \mathbb{X}$ and

$$\mathcal{W} = \{w \colon \mathbb{X} \times \mathbb{P}^2 \to \mathbb{P}^2\}, \qquad (X, \boldsymbol{p}) \mapsto w(X, \boldsymbol{p})$$

where $\mathcal{W}$ defines the set of warp functions w. Let us consider specifically the set of warp functions that generate group actions, *i.e.* , functions that satisfy for X_1, $X_2 \in \mathbb{X}$

$$\begin{cases} w\big(X_1, w(X_2, p)\big) = w(X_1 \circ X_2, p), \\ w\big(X_1, w({X_1}^{-1}, p)\big) = w\big(X_1 \circ {X_1}^{-1}, p)\big) = p. \end{cases} \tag{2.2}$$

In general, $\mathbb{X}$ is a Lie Group of finite dimension and w is differentiable with respect to $X \in \mathbb{X}$. For the sake of simplicity, we can also write

$$w\colon \mathbb{R}^{\dim(\mathbb{X})} \times \mathbb{P}^2 \to \mathbb{P}^2, \qquad (x, \boldsymbol{p}) \mapsto w(x, \boldsymbol{p}), \tag{2.3}$$

since the group element $X \in \mathbb{X}$ can be identified with $x \in \mathbb{R}^n$ such that $X = \exp\{\phi(x)\}$. In that case, we have that ϕ is a minimal parametrization sending the identity element in $\mathbb{X}$ to the null vector, *e.g.* , with $\phi = \log$. Furthermore, we can define the resulting intensity map obtained from warp of *every* pixel from an image

$$\boldsymbol{w}\colon \mathbb{X} \times \mathcal{H} \mapsto \mathcal{H}, \qquad (x, \boldsymbol{I}) \mapsto \boldsymbol{w}(x, \boldsymbol{I}).$$

For simplicity of notation, we define the Jacobian of the warp of an image

$$\boldsymbol{J}\colon \mathcal{W} \times \mathbb{X} \times \mathcal{H} \to \mathbb{M}(mn, dim(\mathbb{X})), \qquad (\boldsymbol{w}, x, \boldsymbol{I}) \mapsto \boldsymbol{J}(\boldsymbol{w}, x, \boldsymbol{I})$$

with each row given by

$$J\colon \mathcal{W} \times \mathbb{X} \times \mathcal{H} \times \mathbb{P}^2 \to \mathbb{M}(1, \dim(\mathbb{X})), \qquad (\boldsymbol{w}, x, \boldsymbol{I}, \boldsymbol{p}) \mapsto J(\boldsymbol{w}, x, \boldsymbol{I}, \boldsymbol{p}).$$

More specifically $J(w, x, \boldsymbol{I}, \boldsymbol{p}_i) \triangleq \partial_x \boldsymbol{I}(w(x, \boldsymbol{p}_i))$.We address the computation of the Jacobian with more details in Section 2.6.

2.2 SIMILARITY FUNCTIONS FOR DIRECT VISUAL TRACKING

2.2.1 *Sum of squared differences*

The sum of squared differences (SSD) writes

$$\boldsymbol{S}_\Sigma\big(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C}\big) = \frac{1}{2}\sum_{i=1}^{mn} \mu_i \big(\boldsymbol{I}_\mathrm{C}(\boldsymbol{p}_i) - \boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i)\big)^2, \tag{2.4}$$

with $\mu_i > 0$. This problem is also known as the weighted nonlinear-least squares The solution to the linear version, together with possible choices for the weights μ_i, is recalled in Appendix B. This similarity measure has been applied in direct visual tracking for many years and it has proven itself to be a good solution. Nevertheless, the SSD is prone to problems due to illumination changes or partial occlusion.

2.2.2 *Sum of the conditional variance*

The sum of the conditional variance (SCV) is a similarity that was conceived for multi-model medical image alignment (Pickering et al., 2009). This similarity builds upon the SSD, however, instead of employing $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$ explicitly, the SCV considers the SSD between the expectation of the $\boldsymbol{I}_\mathrm{C}$ given $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$. More specifically, the joint intensity distribution provides the probability of co-occurrence of two intensities r and s in the images $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$, *i.e.*

$$\boldsymbol{P}(r, \boldsymbol{I}_\mathrm{R}, s, \boldsymbol{I}_\mathrm{C}) = \frac{1}{mn}\sum_{i=0}^{mn} \phi(\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i) - r)\phi(\boldsymbol{I}_\mathrm{C}(\boldsymbol{p}_i) - s)$$

with ϕ a Parzen density function. Notice that the computation joint intensity distribution can be associated to a matrix $\boldsymbol{P} \in \mathbb{M}(B_{\boldsymbol{I}}, B_{\boldsymbol{I}})$ with each element (r, s) given by $\boldsymbol{P}(r, \boldsymbol{I}_\mathrm{R}, s, \boldsymbol{I}_\mathrm{C})$. Therefore, the expected intensity $\mathcal{E}\big(\boldsymbol{I}_\mathrm{C}(\boldsymbol{p}_i)\big|\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i)\big)$ can be computed as

$$\mathcal{E}\big(\boldsymbol{I}_\mathrm{C}(\boldsymbol{p}_i)\big|\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i)\big) = \left(\sum_{s_i=0}^{Nb-1} \boldsymbol{P}\big(\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i), s_i\big)\right)^{-1} \left(\sum_{s_i=0}^{Nb-1} s_i\, \boldsymbol{P}\big(\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i), s_i\big)\right).$$

The SCV for two images $\boldsymbol{I}_{\mathrm{R}}$ and $\boldsymbol{I}_{\mathrm{C}}$ writes

$$\boldsymbol{S}_{\mathcal{E}}\left(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{I}_{\mathrm{C}}\right) = \frac{1}{2}\sum_{i=1}^{mn}\Big(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i) - \mathcal{E}\left(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i)\middle|\boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i)\right)\Big)^2,$$

and the above equation is equivalent to the SSD if $\mathcal{E}\left(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i)\middle|\boldsymbol{I}_{\mathrm{R}}\right) = \boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i)$. The SCV is invariant to illumination changes given by injective functions (Richa et al., 2011). From this point of view, this similarity encapsulates affine illumination model from (1.17), and is able to further cope with a larger set of illumination variations.

2.2.3 *Normalized cross-correlation*

The normalized cross-correlation (NCC) provides a correlation coefficient $\boldsymbol{S}_{\times} \in [-1, 1]$ for two images. The NCC of $\boldsymbol{I}_{\mathrm{R}}$ and $\boldsymbol{I}_{\mathrm{C}}$ is given by

$$\boldsymbol{S}_{\times}\left(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{I}_{\mathrm{C}}\right) = \frac{\displaystyle\sum_{i=1}^{mn}\Big(\boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_{j=1}^{mn}\boldsymbol{I}_{\mathrm{R}}(p_j)\Big)\Big(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_{j=1}^{mn}\boldsymbol{I}_{\mathrm{C}}(p_j)\Big)}{\sqrt{\displaystyle\sum_{i=1}^{mn}\Big(\boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_{i=1}^{mn}\boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i)\Big)^2}\sqrt{\displaystyle\sum_{i=1}^{mn}\Big(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_{i=1}^{mn}\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i)\Big)^2}}, \tag{2.5}$$

Notice that we can also write the NCC using vector notation, defining two vectors $\boldsymbol{i}_{\mathrm{R}}$, and $\boldsymbol{i}_{\mathrm{C}}$ obtained stacking the intensities of $\boldsymbol{I}_{\mathrm{R}}$ and $\boldsymbol{I}_{\mathrm{C}}$, respectively, such that the i-th element of each vector writes $\boldsymbol{i}_{i,\mathrm{R}} = \boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_j \boldsymbol{I}_{\mathrm{R}}(p_j)$, and $\boldsymbol{i}_{i,\mathrm{C}} = \boldsymbol{I}_{\mathrm{C}}(\boldsymbol{p}_i) - \frac{1}{mn}\sum_j \boldsymbol{I}_{\mathrm{C}}(p_j)$. Hence, the NCC similarity can be written in the more compact form

$$\boldsymbol{S}_{\times}\left(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{I}_{\mathrm{C}}\right) = \frac{\boldsymbol{i}_{\mathrm{R}}{}^{T}\boldsymbol{i}_{\mathrm{C}}}{|\boldsymbol{i}_{\mathrm{R}}||\boldsymbol{i}_{\mathrm{C}}|}. \tag{2.6}$$

Let us discuss a few remarks for the NCC using this vector interpretation. A NCC coefficient $\boldsymbol{S}_{\times}\left(\boldsymbol{I}_{\mathrm{R}},\ \boldsymbol{I}_{\mathrm{C}}\right) = 0$ implies that the vectors $\boldsymbol{i}_{\mathrm{R}}$ and $\boldsymbol{i}_{\mathrm{C}}$ are *orthogonal*, thus the images share no information. Furthermore, a coefficient $\boldsymbol{S}_{\times}\left(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{I}_{\mathrm{C}}\right) = 1$ implies that the vectors are *parallel*, therefore the images are perfectly aligned. Recalling inner product properties, we have that the correlation remains unaffected after any shift and/or (positive) scale. Note that the absolute value of NCC remains the same after a negative scaling, however, the sign of the resulting correlation coefficient is inverted. Scales and shifts on $\boldsymbol{i}_{\mathrm{C}}$ are directly related to illumination variations, *i.e.* α and β of photometric model (1.17), and invariance to such effects is indeed a good property for a similarity function.

2.2.4 *Mutual information*

The mutual information (MI) is deduced from the entropy of a discrete random variable z with probability $\boldsymbol{P}(z)$:

$$\boldsymbol{h}(z) \triangleq -\sum_{z_i}\boldsymbol{P}(z_i)\ln\left(\boldsymbol{P}(z_i)\right). \tag{2.7}$$

We can further expand the concept of entropy of a variable towards the joint entropy of two discrete random variables z and y with joint probability $\boldsymbol{P}(z, y)$:

$$\boldsymbol{h}(z, y) \triangleq -\sum_{y_i}\sum_{z_i}\boldsymbol{P}(z_i, y_i)\ln\left(\boldsymbol{P}(z_i, y_i)\right). \tag{2.8}$$

The mutual information is given by three components: the entropy of z, the entropy of y and the negative joint entropy of z and y (Viola and Wells, 1997), *i.e.*

$$\boldsymbol{H}(z, y) = \boldsymbol{h}(z) + \boldsymbol{h}(y) - \boldsymbol{h}(z, y). \tag{2.9}$$

The joint intensity distribution provides the probability of co-occurrence of two intensities r and s in $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$, *i.e.*

$$\boldsymbol{P}(r, \boldsymbol{I}_\mathrm{R}, s, \boldsymbol{I}_\mathrm{C}) = \frac{1}{mn} \sum_{i=0}^{mn} \phi(\boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i) - r)\phi(\boldsymbol{I}_\mathrm{C}(\boldsymbol{p}_i) - s)$$

where ϕ is a Parzen density function. The joint intensity distribution can be associated to a matrix $\boldsymbol{P} \in \mathbb{M}(B_{\boldsymbol{I}}, B_{\boldsymbol{I}})$ with each element (r, s) given by $\boldsymbol{P}(r, \boldsymbol{I}_\mathrm{R}, s, \boldsymbol{I}_\mathrm{C})$, and the probability of the occurrence of an certain intensity r in $\boldsymbol{I}_\mathrm{R}$ or s in $\boldsymbol{I}_\mathrm{C}$ can be computed from the marginals of $\boldsymbol{P}(r, s)$ with respect to s and r respectively, *i.e.*

$$\boldsymbol{P}(r, s) \triangleq \boldsymbol{P}(r, \boldsymbol{I}_\mathrm{R}, s, \boldsymbol{I}_\mathrm{C}), \qquad \boldsymbol{P}_{\boldsymbol{I}_\mathrm{R}}(r) = \sum_{s_j=0}^{B_{\boldsymbol{I}}} \boldsymbol{P}(r, s_j), \qquad \boldsymbol{P}_{\boldsymbol{I}_\mathrm{C}}(s) = \sum_{r_j=0}^{B_{\boldsymbol{I}}} \boldsymbol{P}(r_j, s)\,.$$

Using (2.7), (2.8), (2.9) and (2.10), the MI for two images $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$ writes:

$$\boldsymbol{S}_\mathrm{MI}(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C}) = \sum_{r_i=0}^{B_{\boldsymbol{I}}-1} \sum_{s_i=0}^{B_{\boldsymbol{I}}-1} \boldsymbol{P}(r_i, s_i) \log \left(\frac{\boldsymbol{P}(r_i, s_i)}{\boldsymbol{P}_{\boldsymbol{I}_\mathrm{R}}(r_i)\boldsymbol{P}_{\boldsymbol{I}_\mathrm{C}}(s_i)} \right). \tag{2.10}$$

Despite the fact that the MI is not invariant with respect to illumination changes, this similarity shows superior robustness to multi-modal transformations. The robustness is improved as the MI is high only for sparse joint distributions $\boldsymbol{P}(r_i, s_i)$, which occurs only when the images are well aligned. However, the MI can be affected by local artifacts that induce multiple local maxima close to the optimal solution (Dame and Marchand, 2010). These other maxima can be suppressed smoothing the joint intensity distribution by reducing the number of bits $B_{\boldsymbol{I}}$ for which the joint distribution is computed, *c.f.* (Dame and Marchand, 2010).

2.2.5 *Other examples from medical imaging*

The aforementioned similarity functions have been applied in direct visual tracking with success. These are not the only similarity functions however, and other examples have been applied to multi-modal medical imaging. *E.g.* , the correlation ratio (CR) (Roche et al., 1998b), and the cross cumulative residual entropy (CCRE) (Wang and Vemuri, 2007). These similarities appear to be as robust as the MI for multi-modal image registration, however, the solution of $\boldsymbol{S}(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C})$ is not necessarily equivalent to $\boldsymbol{S}(\boldsymbol{I}_\mathrm{C}, \boldsymbol{I}_\mathrm{R})$. In fact, for the CR similarity, $\boldsymbol{S}(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C}) = \boldsymbol{S}(\boldsymbol{I}_\mathrm{C}, \boldsymbol{I}_\mathrm{R})$ only for affine transformations of the intensities (Roche et al., 1998a). That equivalence is an interesting property for the optimization techniques treated in the next section, as we obtain the same solution for the same inputs independently of the order analyzed in the similarity. Otherwise, the solution may be always prone to the order of $\boldsymbol{I}_\mathrm{R}$ or $\boldsymbol{I}_\mathrm{C}$, and we choose to avoid these kind of technical detail in the implementation of tracking methods.

2.3 GRADIENT BASED OPTIMIZATION

In general, the optimization problem proposed in the Eq. (2.1) using functions from Section 2.2 is not globally convex (or concave in the case of a maximization problem). Hence,

it is not an easy task to find the optimal solution to this problem. Global optimization techniques (Horst and Pardalos, 1995) such as the simulated annealing are too cumbersome for real-time implementation. Therefore, gradient-based optimization methods have been widely employed to solve the problem, as these techniques present a good trade off between region of convergence and computational cost.

2.3.1 *Steepest-descent*

The steepest-descent is one of the simplest solutions to gradient-based optimization (Nocedal and Wright, 2000). We can express the similarity function from the problem (2.1) as a first order Taylor expansion around $\widehat{x}_0$

$$\boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big) = \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\mathrm{C}})\big) + \partial_x \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\mathrm{C}})\big)\widetilde{x} + \boldsymbol{o}(\widetilde{x}, 2),$$

with $\widetilde{x}$ an increment of parameters given by $\widetilde{x} = {\widehat{x}_0}^{-1} \circ x$, and $\boldsymbol{o}(\widetilde{x}, n)$ a polynomial of $\widetilde{x}$ composed by elements of n-th and higher order.

The solution for (2.1) can be computed iteratively by an increment using the gradient descent direction $g(x) = \partial_x \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big)$, *i.e.*

$$\widetilde{x}^* = -Mg(\widehat{x}_k), \tag{2.11}$$

with $M \in \mathbb{M}(\dim(\mathbb{X}))$ a positive definite matrix for a minimization problem or negative definite for a maximization. The solution $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*$ is computed until the obtained increment $\widetilde{x}^*$ is conveniently small, $|\widetilde{x}^*| < \varepsilon$. Defining $M = |g(\widehat{x}_k)|^{-1} I_{\dim(\mathbb{X})}$ is the trivial choice for the steepest-descent, however, this option is not always satisfactory, as can even lead to instability of the method, or need too many iterations to converge to the optimal solution. It depends strongly on the function being optimized. Other line-search methods can improve the convergence rate of the solution.

2.3.2 *Newton-based solution and the forward compositional*

We can also solve (2.1) using Newton's method (Nocedal and Wright, 2000), *i.e.* , the similarity function $\boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big)$ is expressed by the second order Taylor expansion around $\widehat{x}_0$:

$$\begin{aligned}\boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big) = \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\mathrm{C}})\big) + \partial_x \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\mathrm{C}})\big)\widetilde{x} \\ + \frac{1}{2}\widetilde{x}^T \Big(\partial_x{}^2 \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\mathrm{C}})\big)\Big)\widetilde{x} + \boldsymbol{o}(\widetilde{x}, 3),\end{aligned}$$

and defining

$$g_{\mathrm{FC}}(x) \triangleq \partial_x \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big), \qquad M_{\mathrm{FC}}(x) \triangleq \partial_x{}^2 \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big) \tag{2.12}$$

respectively the gradient and the Hessian of the similarity, we obtain the optimal solution the maximum for this approximation of $\boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big)$ at $\partial_{\widetilde{x}} \boldsymbol{S}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})\big) = 0$:

$$\widetilde{x}^* = -M_{\mathrm{FC}}(\widehat{x}_k)^{-1} g_{\mathrm{FC}}(\widehat{x}_k)^T. \tag{2.13}$$

Remark that the problem is solved for an increment $\widetilde{x}$, and the solution for the parameters is then given by $\widehat{x} = \widehat{x}_0 \circ \widetilde{x}^*$. Newton's methods converge in one iteration for quadratic functions, but, for non-quadratic functions, the solution is obtained computing (2.13) iteratively for $\widehat{x}_n$, with $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*$, until the obtained increment $\widetilde{x}$ is conveniently small, *i.e.* , $|\widetilde{x}| < \varepsilon$. According to the classification of (Baker and Matthews, 2001), Eq. (2.13) provides a forward compositional method.

Newton's method has a second order convergence rate, therefore it converges to the practical solution in fewer steps than the steepest-descent. The radius of convergence, however, is smaller and directly related to the Hessian. For instance, for a minimization problem, we verify that an increment obtained in (2.13) is a stable solution for (2.1) if M_{FC} is positive definite. Conversely, the Hessian obtained for a maximization problem must be negative definite. Computational effort of Newton's method is another relevant matter. The computation of the Hessian is seldom an easy task and its computational effort may favor other first order methods.

2.3.3 *Inverse compositional*

In some cases, the optimization of the similarity function can be solved more efficiently by inverting the roles of reference and current images. The inverse compositional technique rewrites the problem (2.1) to "virtually" warp $\boldsymbol{I}_{\text{R}}$ towards $\boldsymbol{I}_{\text{C}}$ using an incremental $\widetilde{x} = \widehat{x}_0{}^{-1} \circ x$:

$$\begin{aligned}
\widehat{x} = \arg\max_x \boldsymbol{S}\big(\boldsymbol{I}_{\text{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\text{C}})\big) &= \arg\max_x \boldsymbol{S}\big(\boldsymbol{w}(0, \boldsymbol{I}_{\text{R}}), \boldsymbol{w}(x, \boldsymbol{I}_{\text{C}})\big) \\
&= \arg\max_{\widetilde{x}:\, x=\widehat{x}_0\circ\widetilde{x}} \boldsymbol{S}\Big(\boldsymbol{w}\big(0, \boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{R}})\big), \boldsymbol{w}\big(x, \boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{C}})\big)\Big) \\
&= \arg\max_{\widetilde{x}:\, x=\widehat{x}_0\circ\widetilde{x}} \boldsymbol{S}\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{R}}), \boldsymbol{w}(x \circ \widetilde{x}^{-1}, \boldsymbol{I}_{\text{C}})\big) \\
&= \arg\max_{\widetilde{x}:\, x=\widehat{x}_0\circ\widetilde{x}} \boldsymbol{S}\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{R}}), \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big). \qquad (2.14)
\end{aligned}$$

Analogously to the forward compositional technique, we write the second order Taylor expansion for $\boldsymbol{S}\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{R}}), \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big)$ and, recalling that $\widetilde{x}^{-1} \approx -\widetilde{x}$ close to $\widetilde{x} = 0$,

$$\begin{aligned}
\boldsymbol{S}\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\text{R}}), \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big) \approx \boldsymbol{S}\big(\boldsymbol{I}_{\text{R}}, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big) &- \partial_{\widetilde{x}}\boldsymbol{S}\big(\boldsymbol{w}(\boldsymbol{I}_{\text{R}}, 0), \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big)\widetilde{x} \\
&+ \frac{1}{2}\widetilde{x}^T\Big(\partial_{\widetilde{x}}{}^2\boldsymbol{S}\big(\boldsymbol{w}(\boldsymbol{I}_{\text{R}}, 0), \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_{\text{C}})\big)\Big)\widetilde{x} + \boldsymbol{o}(\widetilde{x}, 3).
\end{aligned}$$

We can define

$$g_{\text{IC}}(\widehat{x}) \triangleq \partial_x \boldsymbol{S}\big(\boldsymbol{w}(\boldsymbol{I}_{\text{R}}, 0), \boldsymbol{w}(\widehat{x}, \boldsymbol{I}_{\text{C}})\big), \quad M_{\text{IC}}(\widehat{x}) \triangleq \partial_x{}^2 \boldsymbol{S}\big(\boldsymbol{w}(\boldsymbol{I}_{\text{R}}, 0), \boldsymbol{w}(\widehat{x}, \boldsymbol{I}_{\text{C}})\big) \qquad (2.15)$$

respectively as the gradient and the Hessian of the similarity, and the local maximum is obtained at $\partial_{\widetilde{x}}\boldsymbol{S}\big(\boldsymbol{I}_{\text{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\text{C}})\big) = 0$, *i.e.*

$$\widetilde{x}^* = M_{\text{IC}}(\widehat{x}_k)^{-1} g_{\text{IC}}(\widehat{x}_k)^T. \qquad (2.16)$$

Notice that g_{IC}, M_{IC} are equivalent to g_{FC}, M_{FC} in 2.12, literally inverting the roles of the $\boldsymbol{I}_{\text{R}}$ and $\boldsymbol{I}_{\text{C}}$, *i.e.* $\boldsymbol{w}(x, \boldsymbol{I}_{\text{C}})$ becomes the reference image, and $\boldsymbol{I}_{\text{R}}$ is the current. Of course, that property is valid only if $\boldsymbol{S}\big(\boldsymbol{I}_{\text{R}}, \boldsymbol{I}_{\text{C}}\big) = \boldsymbol{S}\big(\boldsymbol{I}_{\text{C}}, \boldsymbol{I}_{\text{R}}\big)$. The seminal inverse solution of the visual tracking was proposed by (Hager and Belhumeur, 1998), that considered transformations with additive parameters. This technique was latter extended by (Baker and Matthews, 2001) to consider compositional transformations.

2.4 GRADIENT OPTIMIZATION FOR SPECIFIC SIMILARITIES

2.4.1 *Inverse compositional and the SSD*

The inverse compositional technique is particularly efficient for the SSD similarity because the value of the Hessian is constant for a first order, and therefore needs to be computed only

once. This property much simplifies the computation of the iterations. For instance, let us consider the similarity (2.4) applied to the inverse compositional problem (2.14)

$$\boldsymbol{S}_\Sigma\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_\mathrm{R}),\, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_\mathrm{C})\big) = \frac{1}{2}\sum_{i=1}^{mn} \mu_i \Big(\boldsymbol{I}_\mathrm{C}\big(w(\widehat{x}_0, \boldsymbol{p}_i)\big) - \boldsymbol{I}_\mathrm{R}\big(w(\widetilde{x}^{-1}, \boldsymbol{p}_i)\big)\Big)^2, \tag{2.17}$$

and let the image Jacobian $J(\boldsymbol{w}, x, \boldsymbol{I}, \boldsymbol{p}_i) \triangleq \partial_x \boldsymbol{I}(\boldsymbol{w}(x, \boldsymbol{p}_i))$ and recall that $x \approx -x^{-1}$ close to the identity, we can thus approximate the warp of the image

$$\boldsymbol{I}_\mathrm{R}\big(w(\widetilde{x}^{-1}, \boldsymbol{p}_i)\big) = \boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i) - J(w, 0, \boldsymbol{I}_\mathrm{R}, \boldsymbol{p}_i)\widetilde{x} + \boldsymbol{o}(\widetilde{x}, 2) \tag{2.18}$$

hence (2.17) is approximated up to the first order by

$$\boldsymbol{S}_\Sigma\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_\mathrm{R}),\, \boldsymbol{w}(\widehat{x}_0, \boldsymbol{I}_\mathrm{C})\big) \approx \frac{1}{2}\sum_{i=1}^{mn} \mu_i \Big(\boldsymbol{I}_\mathrm{C}\big(w(\widehat{x}_0, \boldsymbol{p}_i)\big) - \boldsymbol{I}_\mathrm{R}(\boldsymbol{p}_i) - J(w, 0, \boldsymbol{I}_\mathrm{R}, \boldsymbol{p}_i))\widetilde{x}\Big)^2, \tag{2.19}$$

The above problem is similar to the robust least squares, *c.f.* Appendix B. The optimal increment $\widetilde{x}^*$ is obtained by (B.4), and the solution is obtained computing (2.16) iteratively with $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*$, until the obtained increment $\widetilde{x}^*$ is conveniently small.

This solution is equivalent to the well-known Gauss-Newton technique. Since the Hessian approximation is always definite positive, the convergence radius is larger than the original formulation described in Section 2.3.3. However, this method has a first order convergence rate, and it is often necessary to compute more iterations than Newton's method given by the Hessian of (2.15) in order to reach the solution. Nevertheless, since $\boldsymbol{J}(w, 0, \boldsymbol{I}_\mathrm{R})$ can be evaluated only once, the computational burden of this method is mostly given by the warped image $\boldsymbol{I}_\mathrm{C}\big(w(\widehat{x}_0, \boldsymbol{p}_i)\big)$ and μ_i, which much reduces the total effort of the solution.

2.4.2 *Efficient second order optimization for the SSD*

The approximation in (2.19) provides a simplified solution for the inverse compositional problem in terms of computational effort at expense, however, of reduced convergence rate. The ESM is an efficient second order optimization method proposed in (Malis, 2004) and (Benhimane and Malis, 2007). This method considers a second order approximation of the warped image, instead of analyzing the derivatives of the similarity function. More specifically, we can evaluate the forward compositional (2.13) for the SSD (2.4)

$$\boldsymbol{S}_\Sigma\big(\boldsymbol{I}_\mathrm{R},\, \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big) = \frac{1}{2}\sum_{i=1}^{mn} \mu_i \big(\boldsymbol{I}_\mathrm{C}\big(w(x, \boldsymbol{p}_i)\big) - \boldsymbol{I}_\mathrm{R}\big)^2, \tag{2.20}$$

and consider following approximation for the warp of the warped image

$$\boldsymbol{I}_\mathrm{C}(\boldsymbol{w}(x, p_i)) = \boldsymbol{I}_\mathrm{C}(\boldsymbol{w}(\widehat{x}, p_i)) + J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i)\widetilde{x} + \frac{1}{2}\widetilde{x}^T\big(\partial_x J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i)\big)\widetilde{x} + \boldsymbol{o}(\widetilde{x}, 3), \tag{2.21}$$

since w is a group action, we can further approximate the Jacobian by

$$\begin{aligned} J(w, 0, \boldsymbol{I}_\mathrm{R}, p_i) &= J(w, 0, \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C}), p_i) \\ &= J(w, x, \boldsymbol{I}_\mathrm{C}, p_i) \\ &= J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i) + \widetilde{x}^T\big(\partial_x J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i)\big) + \boldsymbol{o}(\widetilde{x}, 2), \end{aligned}$$

so that

$$\widetilde{x}^T\big(\partial_x J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i)\big) = J(w, 0, \boldsymbol{I}_\mathrm{R}, p_i) - J(w, \widehat{x}, \boldsymbol{I}_\mathrm{C}, p_i) - \boldsymbol{o}(\widetilde{x}, 2), \tag{2.22}$$

We obtain replacing (2.22) in (2.21)

$$\begin{aligned}\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{w}(x,p_i)) &= \boldsymbol{I}_{\mathrm{C}}(\boldsymbol{w}(\widehat{x},p_i))+J(w,\widehat{x},\boldsymbol{I}_{\mathrm{C}},p_i)\widetilde{x}+\frac{1}{2}\Big(J(w,0,\boldsymbol{I}_{\mathrm{R}},p_i)-J(w,\widehat{x},\boldsymbol{I}_{\mathrm{C}},p_i)\Big)\widetilde{x}+\boldsymbol{o}(\widetilde{x},3)\\ &= \boldsymbol{I}_{\mathrm{C}}(\boldsymbol{w}(\widehat{x},p_i))+\frac{1}{2}\Big(J(w,0,\boldsymbol{I}_{\mathrm{R}},p_i)+J(w,\widehat{x},\boldsymbol{I}_{\mathrm{C}},p_i)\Big)\widetilde{x}+\boldsymbol{o}(\widetilde{x},3)\\ &= \boldsymbol{I}_{\mathrm{C}}(\boldsymbol{w}(\widehat{x},p_i))+\frac{1}{2}J_{\mathrm{ESM}}(w,\widehat{x},\boldsymbol{I}_{\mathrm{R}},\boldsymbol{I}_{\mathrm{C}},p_i)\widetilde{x}+\boldsymbol{o}(\widetilde{x},3)\end{aligned}$$

so that (2.20) can be approximated up to the second order by

$$\boldsymbol{S}_{\Sigma}\big(\boldsymbol{I}_{\mathrm{R}},\,\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)\approx\frac{1}{2}\sum_{i=1}^{mn}\mu_i\bigg(\boldsymbol{I}_{\mathrm{C}}(\boldsymbol{w}(\widehat{x},p_i))-\boldsymbol{I}_{\mathrm{R}}+\frac{1}{2}J_{\mathrm{ESM}}(w,0,\boldsymbol{I}_{\mathrm{R}},\boldsymbol{I}_{\mathrm{C}},p_i)\widetilde{x}\bigg)^2. \tag{2.23}$$

The above problem is similar to the robust least squares, *c.f.* Appendix B. The optimal increment $\widetilde{x}^*$ is obtained by (B.4), and the solution is obtained computing (2.16) iteratively with $\widehat{x}_{k+1}=\widehat{x}_k\circ\widetilde{x}^*$, until the obtained increment $\widetilde{x}^*$ is conveniently small.

This technique is very efficient since the solution has a second order convergence rate without the explicit evaluation of the onerous Hessian. This efficient framework was extended to the SCV in (Richa et al., 2011).

2.4.3 *NCC-based direct visual tracking*

We describe with more details the forward and inverse compositional solutions for the NCC-based direct visual tracking in this section. This approach is discussed originally in (Scandaroli et al., 2012). Let us consider the NCC of two images $\boldsymbol{I}_{\mathrm{R}}$, $\boldsymbol{I}_{\mathrm{C}}$ with warp function $\boldsymbol{w}$ and parameters x given by $\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)$ in Eq. 2.5.

Forward compositional

Let us start with the forward compositional approach given by Eq. (2.13) and recall for $u\in\mathbb{R}^n$ and $f\colon\mathbb{R}^n\to\mathbb{R}^n$ that $|u|=\sqrt{u^Tu}$, and $\partial_u\frac{1}{|f(u)|}=-\frac{f(u)^T}{|f(u)|^3}\partial_v f(u)$. We can write the gradient of $\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)$ using the compact form (2.6) as

$$\begin{aligned}g_{\mathrm{FC}}(x)\triangleq\partial_x\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big) &= \frac{\boldsymbol{i}_{\mathrm{R}}{}^T\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})}{|\boldsymbol{i}_{\mathrm{R}}||\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}-\frac{\boldsymbol{i}_{\mathrm{R}}{}^T\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})}{|\boldsymbol{i}_{\mathrm{R}}||\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|^3}\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})^T\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}}),\\ &= \left(\frac{\boldsymbol{i}_{\mathrm{R}}}{|\boldsymbol{i}_{\mathrm{R}}|}-\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)\frac{\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\right)^T\frac{\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|},\end{aligned} \tag{2.24}$$

with the Jacobian $\boldsymbol{J}_{\times}(w,x,\boldsymbol{I})$ obtained by the stacking the mn rows computed with the i-th element $J_{\times}(\boldsymbol{w},x,\boldsymbol{I},\boldsymbol{p}_i)=J\big(w,x,\boldsymbol{I},\boldsymbol{p}_i\big)-\frac{1}{mn}\sum_j J\big(w,x,\boldsymbol{I},\boldsymbol{p}_j\big)$.

Evaluating the partial derivative w.r.t. x^T of Eq. (2.24), we obtain the full expression for the Hessian as:

$$\begin{aligned}\partial_x{}^2\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big) = &-\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)\frac{\boldsymbol{J}_{\times}{}^T\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|^2}\\ &-\frac{\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})^T}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\left(\frac{\boldsymbol{i}_{\mathrm{R}}\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})^T}{|\boldsymbol{i}_{\mathrm{R}}||\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}+\frac{\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})\boldsymbol{i}_{\mathrm{R}}{}^T}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})||\boldsymbol{i}_{\mathrm{R}}|}\right)\frac{\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\\ &+3\frac{\boldsymbol{J}_{\times}{}^T\big(\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\frac{\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})^T}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|^2}\frac{\boldsymbol{J}_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}})}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\\ &+\sum_{i=0}^{mn}\frac{H_{\times}(w,x,\boldsymbol{I}_{\mathrm{C}},\boldsymbol{p}_i)}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\left(\frac{\boldsymbol{i}_{i,\mathrm{R}}}{|\boldsymbol{i}_{\mathrm{R}}|}-\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}},\boldsymbol{w}(x,\boldsymbol{I}_{\mathrm{C}})\big)\frac{\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})_i}{|\boldsymbol{w}(x,\boldsymbol{i}_{\mathrm{C}})|}\right),\end{aligned} \tag{2.25}$$

where $H_\times(w, x, \boldsymbol{I}_\mathrm{C}, \boldsymbol{p}_i) = \partial_x J_\times(w, x, \boldsymbol{I}_\mathrm{C}, \boldsymbol{p}_i)$ denotes the Hessian of the image (not to be mistaken for the Hessian of the similarity function, which is given by (2.25)). Although (Irani and Anandan, 1998) compute the solution given by the forward compositional problem using (2.13) with (2.24), (2.25), the calculation of Hessian at each iteration is not an easy task, specially as the computation of $\boldsymbol{H}_\times(\boldsymbol{w}, x, \boldsymbol{I}_\mathrm{C})$ involves more complex computations and is not as stable as the computation of the Jacobian.

An approximation that simplifies the evaluation of (2.25) is certainly welcome, there are some complications however. Concerning the optimization of the mutual information, (Dame and Marchand, 2010) state that the Hessian should not be approximated by neglecting only the term that involves the image Laplacian. We can relate this statement to the Hessian of the NCC, more specifically, the last term in (2.25). It can be shown that an approximation obtained simply neglecting the last term of Eq. 2.25 is not necessarily negative definite, which disagrees indeed with the maximization problem posed by (2.1). (Brooks and Arbel, 2010) suggest a conflicting approximation, that might lead to an unexpected behavior caused by an unstable Hessian. Nevertheless, there are practical Newton methods (Nocedal and Wright, 2000) that can improve the conditioning of the Hessian, but these modifications increase the computational effort without a guarantee of increased speed nor basin of convergence.

(Dame and Marchand, 2010) suggest an interesting approximation for the mutual information, where the Hessian is evaluated with reference image computed at the solution of the problem, *i.e.* with $\boldsymbol{I}_\mathrm{R} = \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})$. Translating this approximation to the NCC corresponds to evaluating the Hessian around some x, *i.e.* at $\boldsymbol{S}_\times\big(\boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C}), \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big)$. Hence, let $\sigma_\times$ denote the sign of the NCC coefficient $\boldsymbol{S}_\times\big(\boldsymbol{I}_\mathrm{R}, \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big)$, we obtain the approximation

$$\begin{aligned} M_\mathrm{FC}(x) &\triangleq \partial_x{}^2 \boldsymbol{S}_\times\big(\boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C}), \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big) \\ &\approx -\sigma_\times \frac{\boldsymbol{J}_\times{}^T \boldsymbol{J}_\times(w, x, \boldsymbol{I}_\mathrm{C})}{|\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})|^2} + \frac{\boldsymbol{J}_\times(w, x, \boldsymbol{I}_\mathrm{C})^T}{|\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})|} \frac{\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C}) \boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})^T}{|\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})|^2} \frac{\boldsymbol{J}_\times(w, x, \boldsymbol{I}_\mathrm{C})}{|\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})|}, \end{aligned} \tag{2.26}$$

yielding a definite negative matrix. Technically, however, $M_\mathrm{FC}(x)$ can be semi-definite negative. It can be verified from the eigenvectors of $I_{mn} - \boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C}) \boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})^T$ that semi-definiteness happens iff the gradients of $\boldsymbol{I}_\mathrm{C}(w(x, \boldsymbol{p}_i)) = 0$ for every pixel $\boldsymbol{p}_i$, *i.e.* the warped image is not textured. Notice that the situation provoking this degenerate case is not relevant for visual tracking. The solution for the DVT is obtained after computing the increment (2.13) iteratively using Eqs. (2.24) and (2.26):

$$\widetilde{x}^*_\mathrm{FC} = -M_\mathrm{FC}(\widehat{x}_k)^{-1} g_\mathrm{FC}(\widehat{x}_k)^T, \tag{2.27}$$

where $\widetilde{x}^*_\mathrm{FC}$ is the increment for $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*_\mathrm{FC}$ computed until $|\widetilde{x}^*_\mathrm{FC}| < \varepsilon$.

Inverse compositional

The same procedure can be used to compute an inverse compositional solution for the NCC. Recall that locally and close to the identify of the group $x^{-1} \approx -x$, with the gradient

$$\begin{aligned} g_\mathrm{IC}(x) &\triangleq \partial_x \boldsymbol{S}_\times\big(\boldsymbol{I}_\mathrm{R}, \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big) \\ &= -\left(\boldsymbol{S}_\times\big(\boldsymbol{I}_\mathrm{R}, \boldsymbol{w}(x, \boldsymbol{I}_\mathrm{C})\big) \frac{\boldsymbol{i}_\mathrm{R}}{|\boldsymbol{i}_\mathrm{R}|} - \frac{\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})}{|\boldsymbol{w}(x, \boldsymbol{i}_\mathrm{C})|}\right)^T \frac{\boldsymbol{J}_\times\big(\boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R})\big)}{|\boldsymbol{i}_\mathrm{R}|}, \end{aligned} \tag{2.28}$$

and the Hessian approximation

$$\begin{aligned} M_\mathrm{IC} &\triangleq \partial_x{}^2 \boldsymbol{S}_\times\big(\boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R}), \boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R})\big) \\ &\approx -\sigma_\times \frac{\boldsymbol{J}_\times{}^T \boldsymbol{J}_\times\big(\boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R})\big)}{|\boldsymbol{i}_\mathrm{R}|^2} + \frac{\boldsymbol{J}_\times\big(\boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R})\big)^T}{|\boldsymbol{i}_\mathrm{R}|} \frac{\boldsymbol{i}_\mathrm{R} \boldsymbol{i}_\mathrm{R}{}^T}{|\boldsymbol{i}_\mathrm{R}|^2} \frac{\boldsymbol{J}_\times\big(\boldsymbol{w}(0, \boldsymbol{I}_\mathrm{R})\big)}{|\boldsymbol{i}_\mathrm{R}|}, \end{aligned} \tag{2.29}$$

the solution for the DVT is obtained after computing the increment (2.16) iteratively using the gradient (2.28) and approximated Hessian (2.29):

$$\widetilde{x}^*_{\mathrm{IC}} = {M_{\mathrm{IC}}}^{-1} g_{\mathrm{IC}}(\widehat{x}_k)^T \,, \tag{2.30}$$

where $\widetilde{x}^*_{\mathrm{IC}}$ is the increment for $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*_{\mathrm{IC}}$ computed until we obtain a convenient $\widetilde{x}^*_{\mathrm{IC}}$ such that $|\widetilde{x}^*_{\mathrm{IC}}| < \varepsilon$. This approach can be considered as an improved version for the steepest descent, and this solution is very interesting as M_{IC} and its inverse can be computed only once, thus reducing the computational cost of each iteration. Nevertheless, the basin of convergence can be small compared to the forward compositional. Despite these techniques for the NCC-DVT can be quite adequate for some applications, it is still possible to improve the solution of the problem.

A similar solution is provided in (Evangelidis and Psarakis, 2008). The authors, however, derive their solution from the first order approximation of image, *c.f.* Eq. (2.18), showing that maximization of the approximated NCC is still well posed. The approximated Hessian is simpler than the one provided by Newton's method, however, it still depends on $\boldsymbol{I}_{\mathrm{R}}$ and $\boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}})$ at must be computed at each iteration.

2.5 IMPROVING THE NCC-BASED DIRECT VISUAL TRACKING

The NCC is intrinsically robust against affine illumination changes, but there is not a simple and transparent approach to reject occlusion and unmodeled illumination, *e.g.* specular reflections. We propose to redefine the NCC as

$$\boldsymbol{S}^W_{\times}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{I}_{\mathrm{C}}) = \frac{{\boldsymbol{i}_{\mathrm{R}}}^T W \boldsymbol{i}_{\mathrm{C}}}{|\boldsymbol{i}_{\mathrm{R}}|_W |\boldsymbol{i}_{\mathrm{C}}|_W} \,, \tag{2.31}$$

where W is symmetric positive definite weighting matrix and $|v|_P = \sqrt{v^T P v}$. A simpler option suggests W be written as a diagonal matrix with elements $\mu_i > 0$. The remainder of section addresses two techniques to define the μ_i and an approach to improve the gradient optimization. These results were presented originally in (Scandaroli et al., 2012).

2.5.1 *Local illumination changes*

Maximizing the NCC of the whole reference image makes the implicit assumption that the same affine illumination parameters in (1.17) are shared by every pixel. This hypothesis is but seldom satisfied due to reflective properties of the target and local illumination sources. Instead of assuming that the reference image represents a target with constant reflective properties, we split $\boldsymbol{I}_{\mathrm{R}}$ in a grid $\mathbf{G}$ composed by several tiles (subregions) $\mathbf{G}_i$. A similar grid approach is proposed in (Irani and Anandan, 1998) to improve the robustness adding two simple steps:

- only the "concave tiles" are taken into account, *i.e.* tiles whose Hessian has only negative eigenvalues;
- each pixel from a subregion is weighted by the determinant computed from the Hessian of the respective subregion.

This technique can be helpful whilst using Newton's method and the forward compositional approach, but weighting by the determinant may not be very robust with the approximations presented in Section 2.4.3.

We propose a technique to improve the optimization based on Hessian approximations. The *k-means* algorithm (Lloyd, 1982) is employed to partition the tiles into two clusters. We classify the cluster with absolute NCC closer to the unit as *good* cluster $\mathbf{G}^+$, and the other as *bad* cluster $\mathbf{G}^-$. Afterwards, we assign a weight to every subregion. For values lower than

$\mathbf{G}^+$'s centroid, we assign weights μ_i^g from the current distance to the centroid using Huber's influence function (Huber, 1981), *c.f.* Appendix B. The other tiles have a weight assigned to $\mu_i^g = 1$. Figure 2.1 presents a detailed scheme for representing the grid-weighting approach.

We show an example of the state of the art and the proposed grid-weighting approaches in Figure 2.2, where Figure 2.2 (a) shows a reference image and (b) the same image corrupted by non-uniform illumination. We illustrate the effects of the determinant weighting in Figure 2.2 (c) and (d). Remark that the corrupted region in Figure 2.2 (b) is not well identified using any of the constant Hessians. Figure 2.2 (e) displays the grid weighting using the proposed technique. We can verify that the proposed method is able to identify the degraded portion of the image, and reduce their corresponding influence in the optimization.

2.5.2 *Specular reflections and occlusion*

Other types of unmodeled changes in the current image can impair direct visual tracking applications, *e.g.* specular reflections and partial occlusions. These effects can indeed be treated by the technique proposed in Section 2.5.1. Moreover, if the reference image is already small, the local approach it not very recommended.

(Arya et al., 2007) treats such local variations by weighting each pixel from $\boldsymbol{I}_\mathrm{R}$ and $\boldsymbol{I}_\mathrm{C}$ using their histograms and Huber's influence function. This technique tries to approximate the images by mono-modal distributions. Nevertheless, this weighting might not present the desired effect depending on the degradation level. As an illustration, Figure 2.3 (a) represents a reference image and (b) the same image corrupted by specular reflection. Figure 2.3 (c) represents the weights using the method (Arya et al., 2007). Remark that the specular reflection was not detected, and only the pixels with larger gradients are affected.

Our approach is directly connected to the gradients of the NCC similarity, *c.f.* (2.24) and (2.28). It is natural to define the residues

$$\boldsymbol{r}(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C}) \triangleq \frac{\boldsymbol{i}_\mathrm{R}}{|\boldsymbol{i}_\mathrm{R}|} - \boldsymbol{S}_\times(\boldsymbol{I}_\mathrm{R}, \boldsymbol{I}_\mathrm{C}) \frac{\boldsymbol{i}_\mathrm{C}}{|\boldsymbol{i}_\mathrm{C}|}.$$

This residue $\boldsymbol{r}$ defines a new distribution. We compute the weights μ_i^p using Huber's influence function together with the median and the median absolute deviation of $\boldsymbol{r}$. This weighting approach is similar to the one employed by robust least-squares, however, the NCC defines a distribution invariant to affine illumination changes. Figure 2.3 (d) displays the weights computed using the proposed approach. Despite weighting the strong gradients from the right side, the specular reflection is well identified.

This method can be combined with Section 2.5.1. We can compute the median and the median absolute deviation of the residue $\bar{\boldsymbol{r}}$ defined by the good subregions $\mathbf{G}^+$. Afterwards, we can compute the weights μ_i^p for every pixel.

2.5.3 *Improving the gradient solution*

It is well known in the literature that robust estimators reduce convergence speed of the optimization in favor of the robustness against outliers. Furthermore, using solely the inverse or forward solution neglects all the gradient information that could be provided either by the current or reference images. We propose a method to improve the solution. First, we

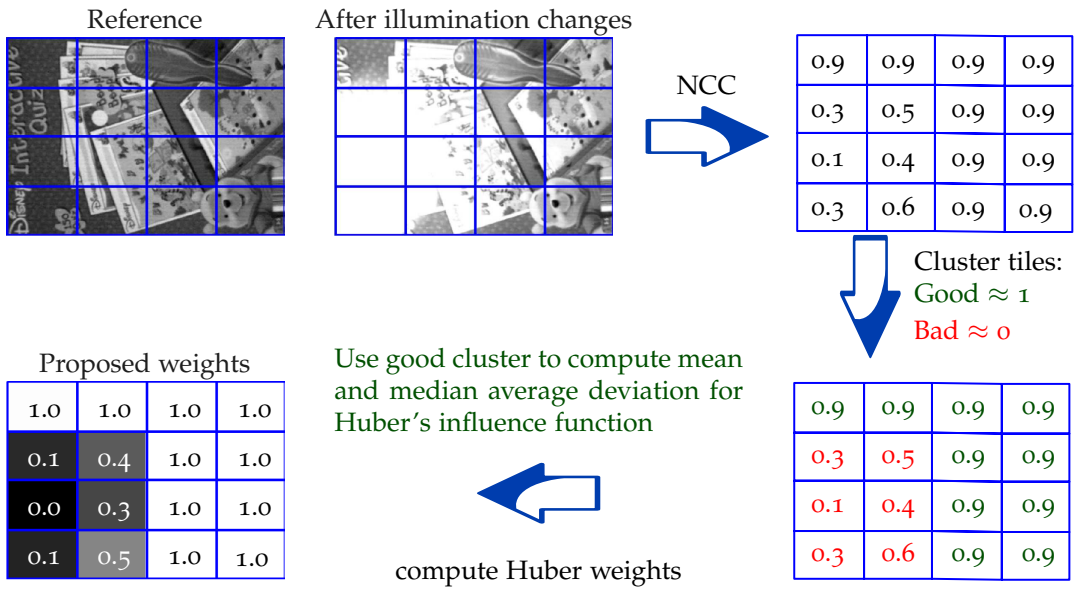
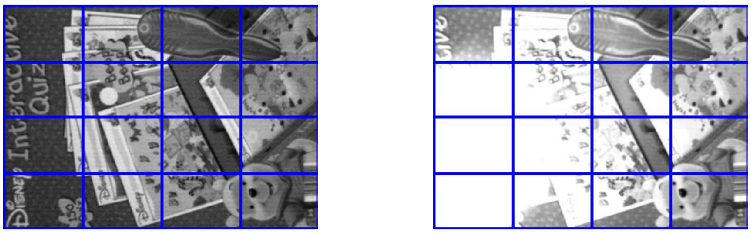
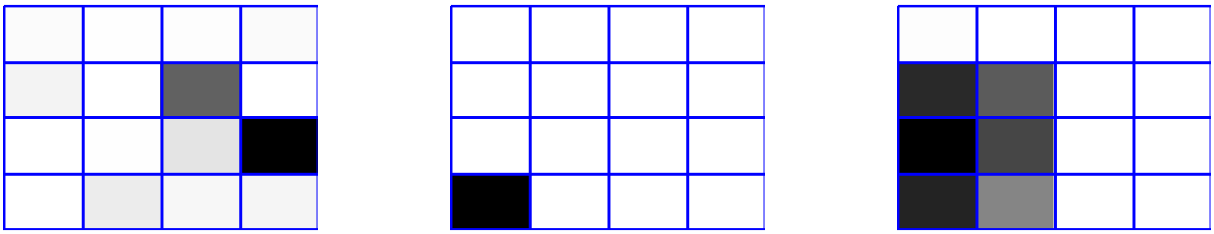


Figure 2.1: Coping with local illumination changes



(a) I_R. (b) I_C.



(c) Inverse compositional (d) Forward compositional (e) Proposed weighting.

Figure 2.2: Grid weighting procedure; (a) reference image; (b) current image; (c) weights from (Irani and Anandan, 1998) using H_{IC}; (d) weights from (Irani and Anandan, 1998) using H_{FC}; (e) proposed weights. All weights vary from black: $\mu_i = 0$ to white: $\mu_i = 1$.

(a) $\boldsymbol{I}_{\mathrm{R}}$ (b) $\boldsymbol{I}_{\mathrm{C}}$ (c) (Arya et al., 2007) (d) Proposed weighting

Figure 2.3: Pixel weighting procedure; (a) reference image; (b) current image; (c) weights from (Arya et al., 2007) vary from black: $\mu_i^p = 0.8$; to white: $\mu_i^p = 1$ (d) proposed weights vary from black: $\mu_i^p = 0.25$ to white: $\mu_i^p = 1$.

heuristically approximate the parabolas for the forward and inverse compositional methods using constant Hessians

$$\begin{cases} \boldsymbol{S}_{\times}^{\mathrm{FC}}(\widetilde{x}) \triangleq \boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_k \circ \widetilde{x}, \boldsymbol{I}_{\mathrm{C}})\big) \approx \boldsymbol{S}_{\times}^{\mathrm{FC}}(\widehat{x}_k) + g_{\mathrm{FC}}(\widehat{x}_k)\widetilde{x} + \dfrac{1}{2}\widetilde{x}^T H_{\mathrm{FC}}\widetilde{x}\,, \\ \boldsymbol{S}_{\times}^{\mathrm{IC}}(\widetilde{x}) \triangleq \boldsymbol{S}_{\times}\big(\boldsymbol{w}(\widetilde{x}^{-1}, \boldsymbol{I}_{\mathrm{R}}), \boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})\big) \approx \boldsymbol{S}_{\times}^{\mathrm{IC}}(\widehat{x}_k) - g_{\mathrm{IC}}(\widehat{x}_k)\widetilde{x} + \dfrac{1}{2}\widetilde{x}^T H_{\mathrm{IC}}(\widehat{x}_k)\widetilde{x}\,. \end{cases} \tag{2.32}$$

Thus, to obtain the maximum we compute the partial derivative with respect to $\widetilde{x}$ for (2.32) that equals to zero and obtain:

$$\begin{cases} 0 = g_{\mathrm{FC}}(\widehat{x}_k)^T + H_{\mathrm{FC}}\widetilde{x}^*\,, \\ 0 = -g_{\mathrm{IC}}(\widehat{x}_k)^T + H_{\mathrm{IC}}(\widehat{x}_k)\widetilde{x}^*\,. \end{cases} \tag{2.33}$$

Ideally, under the assumption that the similarity function is quadratic, the inverse and the forward solutions are the same. Nevertheless, in practice, these solutions give complementary information that we propose to exploit. Adding both right hand sides of (2.33), we obtain the optimal increment:

$$\widetilde{x}^* = \big(H_{\mathrm{FC}} + H_{\mathrm{IC}}(\widehat{x}_k)\big)^{-1}\big(g_{\mathrm{FC}}(\widehat{x}_k) - g_{\mathrm{IC}}(\widehat{x}_k)\big)^T\,. \tag{2.34}$$

The computational effort of the proposed solution is increased comparing to the inverse compositional, however, we double the information employed to solve the optimization. This is at the expense of recomputing $H_{\mathrm{FC}}(\widehat{x}_k)$ and $g_{\mathrm{FC}}(\widehat{x}_k)$ at each iteration.

Solution (2.34) is inspired by the ESM (Benhimane and Malis, 2007), *c.f.* Section 2.4.2, which achieves a second order convergence rate for the SSD without computing the Hessian explicitly. The ESM uses the information of both reference and current image Jacobians, nevertheless, the estimation of the photometric parameters is a necessary task to accomplish a similar result with the NCC. (Brooks and Arbel, 2010) propose an ESM extension for other similarities than the SSD. They proceed by directly adding reference and current image Jacobians, as proposed by the ESM. The primal ESM solution by (Benhimane and Malis, 2007) assumes brightness constancy, therefore it is sound to directly sum the Jacobians, but one must be careful under illumination changes.

The trivial sum of the image Jacobian is contradicted by considering the, theoretically possible, case where the photometric gradients have inverse signals whilst the proposed solution (2.34) is still valid in this ill-conditioned case. Moreover, (Keller and Averbuch, 2004) and (Mégret et al., 2008) follow the same heuristics, however, the rationale is entirely based on the Jacobians of the image, where our approach considers directly the gradient and the Hessian of the similarity. Remark that result from (2.34) could be employed independently of the similarity. We have verified in experiments outside the scope of this thesis, however, that such an approach does not necessarily improve histogram based approaches such as the MI.

Algorithm 1 Visual tracking using proposed method

Require: warp $\boldsymbol{w}$, initial parameter guess $\widehat{x}_0$, threshold ε, maximum number of iterations $\bar{k}$, grid $\mathbf{G}$ with tiles $\mathbf{G}_i \in \mathbf{G}$.

1: **for all** $\mathbf{G}_i \in \mathbf{G}$, and pixel $\boldsymbol{p}$: $\boldsymbol{p}_n \in \mathbf{G}_i$ **do**
2: $\quad \boldsymbol{i}_{\mathrm{R}}^{\mathbf{G}_i}$: $\boldsymbol{i}_{n,\mathrm{R}}^{\mathbf{G}_i} = \boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_n) - \frac{1}{mn}\sum_{\boldsymbol{p}_j}^{\boldsymbol{p}_j \in \mathbf{G}_i} \boldsymbol{I}_{\mathrm{R}}(\boldsymbol{p}_j)$,
3: $\quad J_{\times}^{\mathbf{G}_i}\big(\boldsymbol{w}(0, \boldsymbol{I}_{\mathrm{R}})\big)$: $J_{\times}^{\mathbf{G}_i}\big(\boldsymbol{w}(0, \boldsymbol{I}_{\mathrm{R}})\big)_n = J\big(\boldsymbol{w}(0, \boldsymbol{I}_{\mathrm{R}})\big)_n - \frac{1}{mn}\sum_{\boldsymbol{p}_j}^{\boldsymbol{p}_j \in \mathbf{G}_i} J\big(\boldsymbol{w}(0, \boldsymbol{I}_{\mathrm{R}})\big)_j$.
4: **end for**
5: **for all** new image $\boldsymbol{I}_{\mathrm{C}}$ **do**
6: $\quad$ **repeat**
7: $\quad\quad$ **for all** $\mathbf{G}_i \in \mathbf{G}$, and pixel $\boldsymbol{p}$: $\boldsymbol{p}_n \in \mathbf{G}_i$ **do**
8: $\quad\quad\quad \boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})^{\mathbf{G}_i}$: $\boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})_n^{\mathbf{G}_i} = \boldsymbol{I}_{\mathrm{C}}\big(\boldsymbol{w}(\widehat{x}_k, \boldsymbol{p}_n)\big) - \frac{1}{mn}\sum_{\boldsymbol{p}_j}^{\boldsymbol{p}_j \in \mathbf{G}_i} \boldsymbol{I}_{\mathrm{C}}\big(\boldsymbol{w}(\widehat{x}_k, \boldsymbol{p}_j)\big)$,
9: $\quad\quad\quad J_{\times}^{\mathbf{G}_i}\big(\boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})\big)$: $J_{\times}^{\mathbf{G}_i}\big(\boldsymbol{w}(0, \boldsymbol{I}_{\mathrm{C}})\big)_n = J\big(\boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})\big)_n - \frac{1}{mn}\sum_{\boldsymbol{p}_j}^{\boldsymbol{p}_j \in \mathbf{G}_i} J\big(\boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})\big)_j$.
10: $\quad\quad$ **end for**
11: $\quad\quad$ Cluster good $\mathbf{G}^+$ and bad $\mathbf{G}^-$ subregions by $\boldsymbol{S}_{\times}\big(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})\big) = \frac{\boldsymbol{i}_{\mathrm{R}}{}^T \boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})}{|\boldsymbol{i}_{\mathrm{R}}||\boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})|}$, and compute weights $\mu_i^{\mathbf{G}}$ of every tile $\mathbf{G}_i$, *c.f.* Section 2.5.1.
12: $\quad\quad$ Compute median $\bar{\boldsymbol{r}}$ and median of average distances $\sigma_{\boldsymbol{r}}$ of residues
$$\boldsymbol{r} = \frac{\boldsymbol{i}_{\mathrm{R}}^{\mathbf{G}_i}}{|\boldsymbol{i}_{\mathrm{R}}^{\mathbf{G}_i}|} - \boldsymbol{S}_{\times}(\boldsymbol{I}_{\mathrm{R}}^{\mathbf{G}_i}, \boldsymbol{w}(\widehat{x}_k, \boldsymbol{I}_{\mathrm{C}})^{\mathbf{G}_i}) \frac{\boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})^{\mathbf{G}_i}}{|\boldsymbol{w}(\widehat{x}_k, \boldsymbol{i}_{\mathrm{C}})^{\mathbf{G}_i}|},$$
$\quad\quad$ for all $\mathbf{G_j} \in \mathbf{G}^+$.
13: $\quad\quad$ Compute weights $\mu_i^{\boldsymbol{p}}$ of every pixel $\boldsymbol{p}_n$ using Huber's influence function and the distribution defined by, *c.f.* Section 2.5.2.
14: $\quad\quad$ Compute W with i-th diagonal element given by $W_i = \mu_i^g \mu_i^{\boldsymbol{p}}$.
15: $\quad\quad$ Compute $\widetilde{x}^*$ via (2.35), and $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*$.
16: $\quad$ **until** $|\widetilde{x}^*| < \varepsilon$, or $k > \bar{k}$
17: **end for**

Summary of the proposed algorithm

We define the weighting matrix W in Eq. 2.31 as a diagonal matrix with the i-th diagonal element W_i obtained by the multiplication of the weights μ_i^g from Section 2.5.1 and $\mu_i^{\boldsymbol{p}}$ from Section 2.5.2. Remark that the use of the weighting is optional, and grid or pixel weighting are neglected by setting $\mu_i^g = 1$ or $\mu_i^{\boldsymbol{p}} = 1$, respectively. Using the same procedure as Section 2.4.3, it is direct to obtain the explicit forms of $M_{\mathrm{FC}}^W(x)$, M_{IC}^W, $g_{\mathrm{IC}}^W(x)$, and $g_{\mathrm{IC}}^W(x)$ needed by the optimization of the revisited NCC (2.31). The solution is given by:

$$\widetilde{x}^* = -\big(M_{\mathrm{FC}}^W + M_{\mathrm{IC}}^W(\widehat{x}_k)\big)^{-1}\big(g_{\mathrm{FC}}^W(\widehat{x}_k) - g_{\mathrm{IC}}^W(\widehat{x}_k)\big)\,, \tag{2.35}$$

where $\widetilde{x}^*$ is the increment that composes $\widehat{x}_{k+1} = \widehat{x}_k \circ \widetilde{x}^*$.

Algorithm 1 describes the full proposed technique.

2.6 VISUAL TRACKING OF PLANAR SURFACES

We have presented the problem of direct visual tracking in Section 2.1 divided in three main components: the similarity function, the optimization approach and the warp function. We have covered the two first topics in Sections 2.2 and 2.3. Briefly, recall from (2.3) that $\boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}}))$ actually simplifies

$$\boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(X, \boldsymbol{I}_{\mathrm{C}})) = \boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}\big(\phi(x), \boldsymbol{I}_{\mathrm{C}}\big)).$$

We have verified in Section 2.3 that gradient-based methods must compute, at least, the gradient of the similarities

$$\begin{aligned}\partial_x \boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(X, \boldsymbol{I}_{\mathrm{C}})) &= \partial_{\boldsymbol{I}_{\mathrm{C}}} \boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\phi(x), \boldsymbol{I}_{\mathrm{C}})) \cdot \boldsymbol{J}(\boldsymbol{w}, x, \boldsymbol{I}_{\mathrm{C}}) \\ &= \partial_{\boldsymbol{I}_{\mathrm{C}}} \boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(\phi(x), \boldsymbol{I}_{\mathrm{C}})) \cdot \partial_{\boldsymbol{p}} \boldsymbol{I}_{\mathrm{C}}\big(\boldsymbol{w}(\phi(x), \boldsymbol{p}_i)\big) \cdot \partial_X \boldsymbol{w}(\phi(x), \boldsymbol{p}_i) \cdot \partial_x \phi(x).\end{aligned}$$

Considering the SSD and the SCV similarities, for instance, the computation of the rightmost term $\partial_{\boldsymbol{I}_{\mathrm{C}}} \boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x, \boldsymbol{I}_{\mathrm{C}}))$ is trivial, but this derivative demands a little more caution for the NCC and the MI. The image Jacobian is given by the other terms,

$$\boldsymbol{J}(\boldsymbol{w}, x, \boldsymbol{I}_{\mathrm{C}}) \triangleq \partial_{\boldsymbol{p}} \boldsymbol{I}_{\mathrm{C}}\big(\boldsymbol{w}(x, \boldsymbol{p}_i)\big) \cdot \partial_X \boldsymbol{w}(\phi(x), \boldsymbol{p}_i) \cdot \partial_x \phi(x).$$

We can divide $\boldsymbol{J}$ considering the photometric and geometric components, *i.e.* the components due to the illumination $\partial_{\boldsymbol{p}} \boldsymbol{I}_{\mathrm{C}}$ and warp $\partial_X \boldsymbol{w}(I_3, \boldsymbol{p}_i) \cdot \partial_x \exp(\phi(x))$. The prior term is easily computed using (Ma et al., 2003, pp. ?) and depends only on the texture of the image. The latter term, however, depends on the transformation that we consider for the pixels. Although we deal mostly with planar surfaces and monocular pinhole cameras in this thesis, there are other interesting applications using deformable surfaces (Gay-bellile et al., 2010), or multi-camera systems that employ, for instance, a spherical representation for the warp (Meilland et al., 2010).

We have seen in Section 1.3.4 that the geometry of two views from a planar target presents some interesting properties that can be exploited in visual tracking. The planar properties have been well explored in direct visual tracking using the SSD similarity with ESM optimization in (Benhimane and Malis, 2007). Afterwards, (Silveira and Malis, 2010) extend the approach to consider to illumination changes and non-planar surfaces. These techniques consider the use of Homography as pixel transformation. We refer the reader to these works for more details.

One choice for the warp function is given by (1.23)

$$w_{\mathrm{P}}(H, \boldsymbol{p}) = \left[\frac{{e_1}^T H \boldsymbol{p}}{{e_3}^T H \boldsymbol{p}}, \quad \frac{{e_2}^T H \boldsymbol{p}}{{e_3}^T H \boldsymbol{p}}, \quad 1\right]^T.$$

We can verify that $w = w_{\mathrm{P}}$ is indeed a group action directly via the definition (2.2) and properties (1.24). The choice of a transformation using parameters from Lie group is very interesting as, *c.f.* , for instance, (Warner, 1987),

$$\partial_x w(\phi(x), \boldsymbol{p}_i) \cdot \partial_x \phi(x) = \partial_x w(I_3, \boldsymbol{p}_i) \cdot \partial_x \phi(0).$$

The Special Linear group is 8 dimensional, and its lie algebra $\mathfrak{sl}(3)$ is given by the set of matrices with trace zero that can be associated, for instance, by $x \in \mathbb{R}^8$ as

$$X = \phi(x) = \exp\left(\begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & -x_1 - x_5 \end{bmatrix}\right).$$

Therefore, considering the case given by (1.23), we obtain that

$$\partial_H w_{\mathrm{P}}(I_3, \boldsymbol{p}_i) = \begin{bmatrix} {\boldsymbol{p}_i}^T & 0_{1\times 3} & ({e_1}^T \boldsymbol{p}) \boldsymbol{p}^T \\ 0_{1\times 3} & {\boldsymbol{p}_i}^T & ({e_2}^T \boldsymbol{p}) \boldsymbol{p}^T \end{bmatrix}, \tag{2.36}$$

and

$$\partial_x \phi(0) = \begin{bmatrix} I_8 \\ \mathbf{w} \end{bmatrix}, \quad \mathbf{w} = -\begin{bmatrix} {e_1}^T & {e_2}^T & 0_{1\times 2} \end{bmatrix}.$$

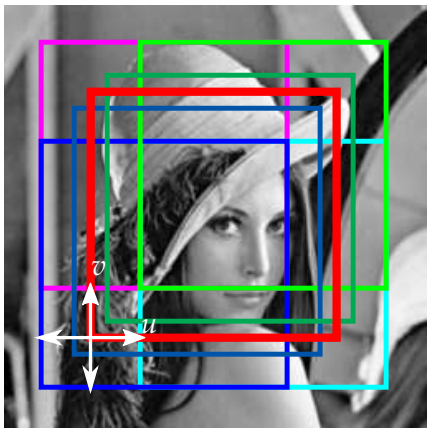


Figure 2.4: Experimental settings for evaluation of similarities.

2.7 COMPARATIVE RESULTS FOR PLANAR TRACKING

We evaluate the tracking accuracy and the robustness with respect to illumination variations and partial occlusions using a benchmark dataset and challenging video sequences. Our objective is to show that the proposed techniques can improve the NCC tracking, in terms of computational effort and speed, to the same level as other state of the art methods whilst improving the robustness to concurrent illumination changes and partial occlusions. We first compare the convergence radius of different optimization strategies for the similarity functions described in Section 2.2. Next, we present the results of the visual tracking techniques using a benchmark dataset, and compare to results obtained by other authors using the same dataset. At last, we evaluate the techniques using several sequences with challenging illumination changes. All of the experiments consider warps using the homography structure, *c.f.* Section 2.6.

2.7.1 *Convergence radii*

We compare the convergence radii of different similarity functions and optimization techniques in this Section. We present two experiments. The first result concerns an evaluation of the similarities for 2D displacements, the setup of this experiment is depicted in Figure 2.4. We first define $\boldsymbol{I}_{\mathrm{R}}$ as a subregion with 150 by 150 pixels at center of a larger image, *i.e.* the $\boldsymbol{I}_{\mathrm{R}}$ is given by the red square in Figure 2.4, and multiple $\boldsymbol{I}_{\mathrm{R}}$ obtained for $x_i = (u_i, v_i)$ displacements. We cover the range from -30 to +30 pixels in each direction, and for each x_i we associate $\boldsymbol{S}(\boldsymbol{I}_{\mathrm{R}}, \boldsymbol{w}(x_i, \boldsymbol{I}_{\mathrm{C}}))$ for the SSD, SCV, NCC and MI similarities. Using these values, we are able to evaluate level curves for each similarity with the respective gradients at each x_i, as depicted in Figure 2.5. Figure 2.5 (a) presents the results for the SSD, (b) for the SCV considering a bit-depth of 256, (c) for the NCC and (d) for the MI considering a bit-depth of 8. The regions in blue refer to the worst scores of the similarities (low results for similarities that are maximized, *i.e.* NCC and MI, and highest results for similarities that are minimized, *i.e.* SSD and SCV), and each arrow represent the gradient evaluated numerically. These results allow the analysis of the each similarity regardless of the optimization technique, under the brightness constancy assumption. We can verify that the SSD and the NCC have indeed the largest basin of convergence for this situation. We can clearly spot that the gradients are stronger within a radius of around 20 pixels from the center of the image, this situation happens only for 10 and 15 pixels for the MI and SCV, respectively. We can conclude that the performance of histogram-based techniques is very inferior to the others considering, of course, this hypothesis of constant il-

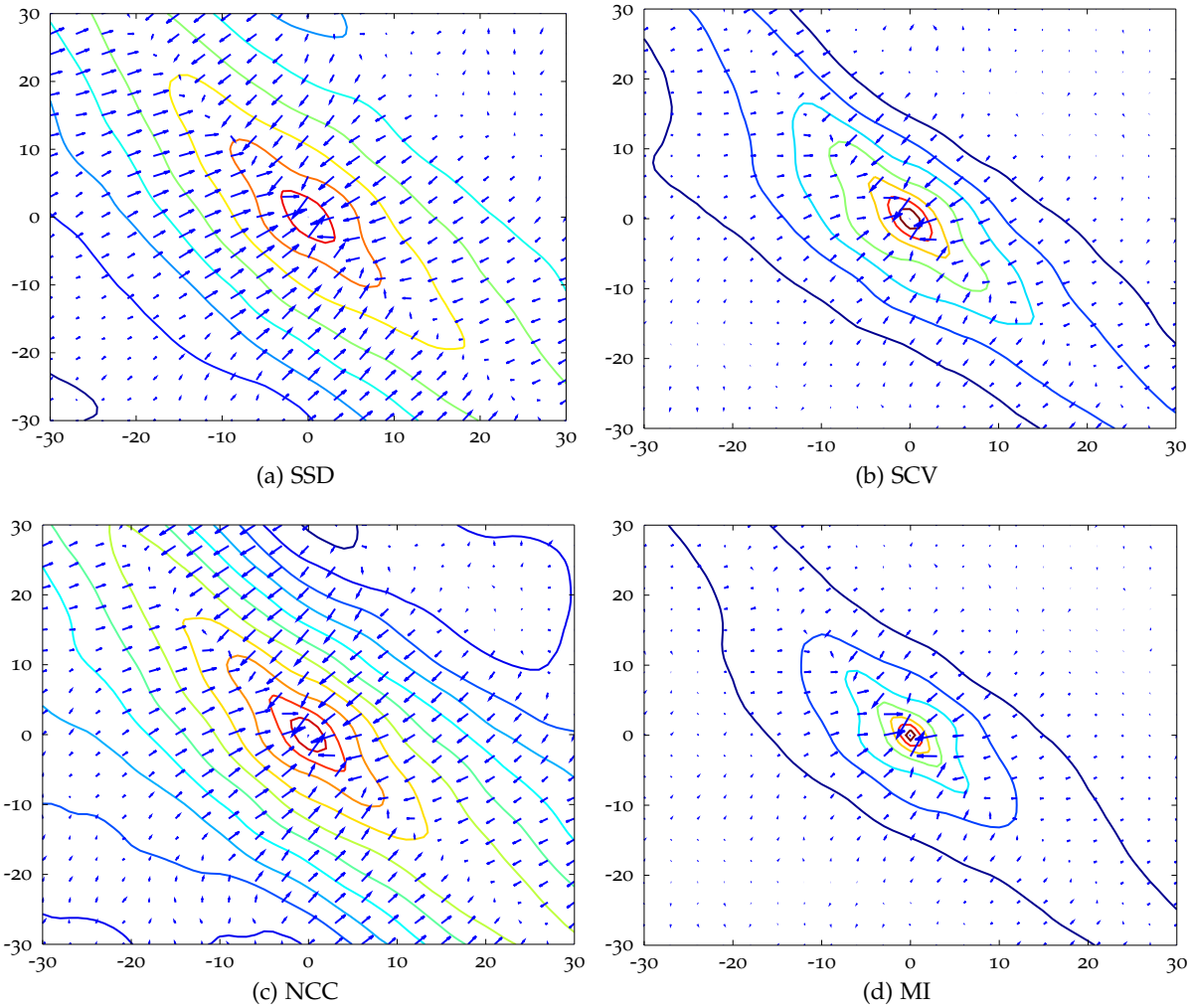


Figure 2.5: Level curves and experimental region of convergence for different similarities.

lumination settings. This is indeed expected as the SCV and MI are designed for multi-modal image alignment, as the best performance of SSD is obtained for the hypothesis used in this experiment.

The second experiment evaluates the convergence rate of the tracking methods for the similarities using different optimization techniques. We use a similar configuration as the experiment from (Baker and Matthews, 2001), and compare six implementations:

- SSD with inverse compositional (*SSD+IC*) from (Baker and Matthews, 2001);
- SSD with ESM (*SSD+ESM*) from (Benhimane and Malis, 2007);
- SCV with ESM (*SCV+ESM*) from (Richa et al., 2011),;
- NCC with inverse compositional *NCC+IC* from Section 2.4.3;
- NCC with inverse and forward compositional as discussed in Algorithm 1;
- MI with inverse compositional from (Dame and Marchand, 2010);

We consider a reference image with 150 by 150 pixels, and generate 20000 random homographies. These random samples are obtained by adding a random offset to each of the four corners of the original image. The bounds of the noise added to the corners increase from 1 up to 20 pixels. We consider that the frame was solved with success if the root mean squared error the coordinates of the corners is below 0.25 pixels. Initially, we set 300 iterations for the methods, and the optimization stops when an increment $|\widetilde{x}| \leq \varepsilon = 10^{-3}$ is obtained. The

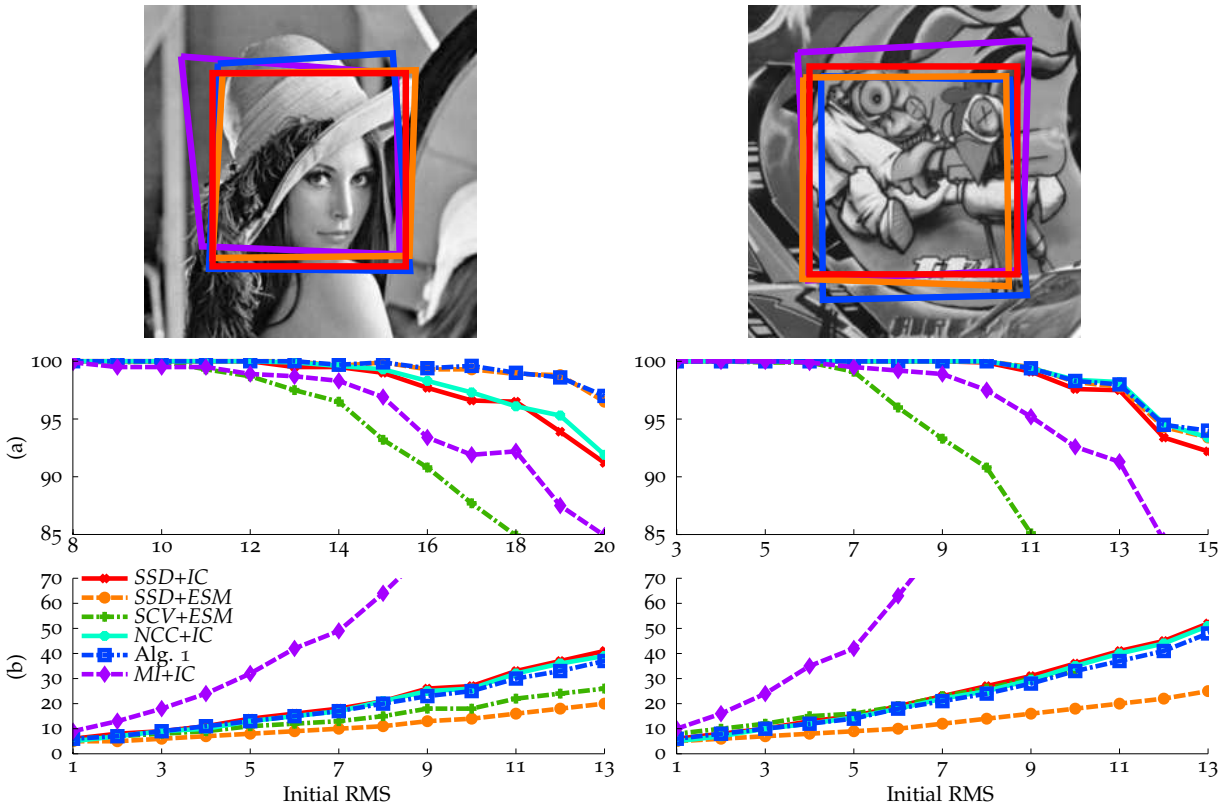
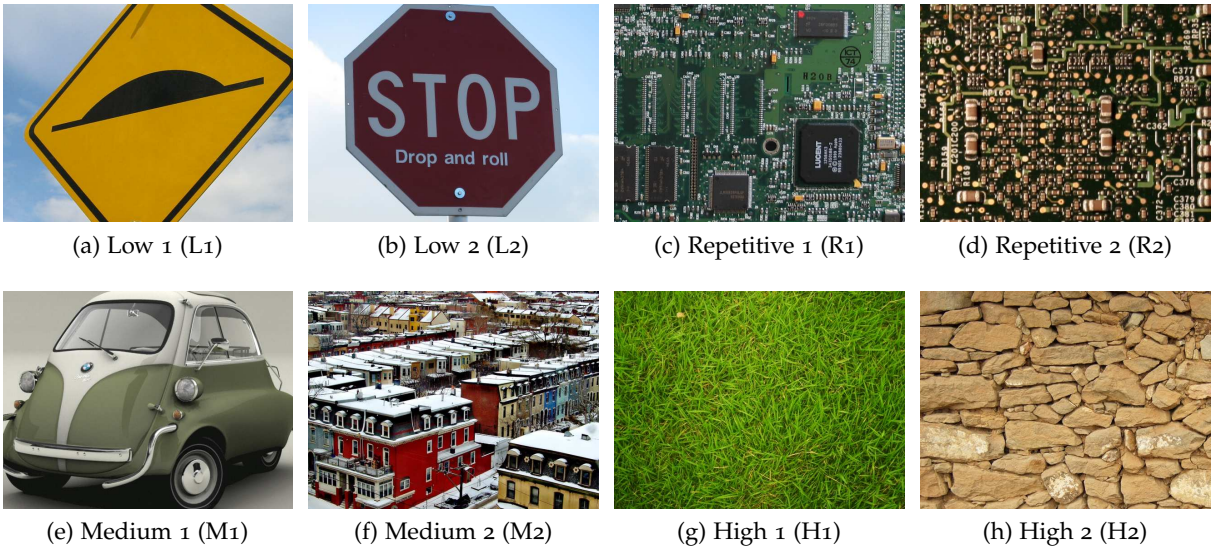


Figure 2.6: Convergence rate for the tracking of synthetic images. (a) percentage of correct attempts; (b) Average number of iterations.

methods are allowed to run at most 500 iterations. We evaluated the techniques on two different images, and the results are displayed in Figure 2.6. The results on the left correspond to the experiments performed using the Lena image, while the rightmost results correspond to the experiments using the Graffiti image. We can verify that the *SSD+ESM* has the best results among the methods, this method presents a high convergence rate over a larger range initial errors with the lowest number of iterations. Considering the Lena image, the method from Algorithm 1 shows the second best convergence rate, however, the solution needs twice as much iterations as the *SSD+ESM*. The inverse compositional based methods of NCC and SSD show a similar result, and at last we have the histogram based techniques *MI+IC* and *SCV+ESM*. The results follow the analysis from the first experiment, *i.e.* , histogram based methods show a smaller region of convergence under brightness constancy assumption. However, even though the SCV shows lower convergence rate than the MI, notice that the number of iterations differ considerably as the SCV needs to compute, in average, the same number of iterations as the NCC and SSD, the number of iterations needed by the MI increases almost exponentially.

2.7.2 *Metaio Benchmark dataset*

We evaluate different similarities and optimization techniques using the planar-based visual tracking benchmark presented in (Lieberknecht et al., 2009). This benchmark consists of 8 different reference images classified among low, repetitive, normal and high texture, *c.f.* Figure 2.7. There are 5 sequences of 1200 frames contemplating different motion types and illumination settings for each target: high angles, distance range, fast far motion, fast close motion



(a) Low 1 (L1) (b) Low 2 (L2) (c) Repetitive 1 (R1) (d) Repetitive 2 (R2)

(e) Medium 1 (M1) (f) Medium 2 (M2) (g) High 1 (H1) (h) High 2 (H2)

Figure 2.7: Reference images of planar tracking benchmark (Lieberknecht et al., 2009).

and, at last, illumination changes. The estimated position of the corners obtained from the visual tracking are compared to a ground truth database, and the tracking is considered successful if the sum of the squared errors of the 4 corners is lower than 10 pixels. The results are given as rate of successfully tracked frames.

We compare the results obtained for this dataset using our implementation of several methods, and comparing to other results found in the Literature. We contemplate three implementations of the SSD, the first version is based on a robust weighted-least squares with Huber's influence function, the second is an implementation of the SSD with local illumination changes (*SSD+i*) from (Silveira and Malis, 2010) and the third is an inverse compositional implementation from (Baker and Matthews, 2001), these results are compared to the ESM results presented in (Lieberknecht et al., 2009). We follow with an implementation of the SCV using the ESM optimization as presented in (Richa et al., 2011). Concerning the NCC similarity, we evaluate an implementation of the NCC with inverse compositional-like steepest descent optimization (*NCC+ICS*), *c.f.* Section 2.4.3 and Eq. (2.30), and also two implementations of the proposed method. First, without the pixel-wise robust weighting (Proposed-I), *i.e.* Algorithm 1 without computing line 13; and secondly, with pixel-wise robust weighting (Proposed-II), *i.e.* Algorithm 1 fully implemented. The last results concern an implementation of the MI using the inverse compositional approach presented in (Dame and Marchand, 2010), we compare the obtained scores with the original result.

In order to provide the same basis for comparison, we use the same parameters to every method when possible. More specifically, we compute all of the methods using three layers of Gaussian-pyramid. Moreover, we stop the optimization when the increment norm is below $\varepsilon = 10^{-3}$, and each method is allowed to run at most 500 iterations. We employ such a large number of iterations to provide similar opportunity to each method independently of the convergence rate at the expense of not reproducing a real-time application. All of the warps are computed using bilinear interpolation, and we use a downsized template to 320×240 pixels instead of the original 640×480 pixels in order to avoid oversampling in most of the sequences. The only difference between the settings of the methods refers to the number of tiles of the grids and number of bins of the histogram, when applicable. The *SSD+i* uses a grid of 5×5 tiles, where the proposed NCC methods use a grid of 3×3 tiles. The SCV

implementation considers a bit depth $B_I = 256$ for the histogram computation, whilst the MI employs $B_I = 8$.

SSD

Table 2.1 presents the scores obtained by four techniques using the SSD similarity, *(a)* refers to an implementation of the SSD with weights given by Huber's influence function and ESM optimization, *(b)* refers to an implementation of the SSD with compensation of the affine illumination *SSD+I* changes using ESM optimization proposed in (Silveira and Malis, 2010), *(c)* refers to the result of the SSD with ESM optimization presented in (Lieberknecht et al., 2009) and *(d)* refers to the result of an implementation of the inverse compositional method from (Baker and Matthews, 2001). Each row of the table presents the result for the targets from Figure 2.7. Moreover, each column displays the results for sequences angle, range, fast and far, fast and close and illumination changes. The scores in bold refer to the best result obtained for the dataset. For the sake of comparison, we consider any score difference below 5% is irrelevant. The last row of each table displays the mean of scores obtained by all targets for each type of sequence.

The difference between scores of the implemented algorithms *(a)* and *(b)* and the results from the benchmark paper *(c)* is notorious. The result of the benchmark paper outperformed the others in a single sequence with an improvement around 15% of tracked frames. However, we do not have access to the experimental settings employed by the authors. Therefore, we cannot conjecture whether the weight of the pixels or the illumination parameters improved the other technique. Furthermore, we can verify that results with ESM optimization scored better than the inverse compositional *(d)*. We thus conclude that the second order optimization considering both reference and current images can improve the results over considering the reference image only. We can also remark that the SSD with Huber weights outperformed the *SSD+i* in 40% of the sequences with an average improvement of 29.8% successfully tracked framed, *i.e.* around 357 frames per sequence, while the converse is shown only once with and improvement of 22.7% of tracked frames, around 272 frames. Therefore, for these benchmark sequences, it is evident that ignoring the pixels that do not belong to the model is more important than considering affine illumination changes.

SCV

Table 2.2 displays the scores obtained by an implementation of the SCV similarity with ESM optimization. Each row of the table presents the result for the targets from Figure 2.7, and the columns display the results for sequences angle, range, fast and far, fast and close and illumination changes. The last row of each table displays the mean obtained by all targets for each type of sequence. The scores obtained for this similarity are comparable to the *SSD+i* technique previously discussed. The SCV outperforms the *SSD+i* technique in 12.5% of the sequences, while the converse happens also 12.5% of the sequences. The differences between the scores, however, are not very large.

NCC

Table 2.3 shows the scores obtained by three techniques using the NCC similarity, *(a)* refers to an implementation of Algorithm 1 without the pixel-wise robust weighting, which we denote as PROPOSED-I, *i.e.* Algorithm 1 without computing line 13, *(b)* refers to the implementation of Algorithm 1 fully implemented, which we denote as PROPOSED-II, and *(c)* refers to an implementation of the NCC with inverse compositional-like steepest descent optimization (*NCC+IC*), *c.f.* Section 2.4.3 and Eq. (2.30). Each row of the table presents the result for the

Table 2.1: Benchmark scores for SSD. *(a)* ESM optimization using Huber weights, *(b)* ESM optimization considering illumination changes, based on (Silveira and Malis, 2010), *(c)* ESM optimization, result from (Lieberknecht et al., 2009), *(d)* Inverse compositional based on (Baker and Matthews, 2001).

(a)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	76.2%	**73.4%**	39.7%	97.7%
L2	100%	99.3%	55.3%	49.5%	68.3%
R1	100%	**100%**	**52.9%**	**87.3%**	100%
R2	91.2%	**99.4%**	15.7%	70%	**96.1%**
M1	100%	99.6%	**79.3%**	**85.6%**	99.6%
M2	99.8%	99.9%	14.8%	84.4%	100%
H1	**99.3%**	**80.9%**	**29.3%**	15%	**90%**
H2	100%	**79.8%**	27.1%	**72.1%**	**100%**
	98.8%	91.9%	43.5%	63%	94%

(b)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	99.8%	77.3%	62.1%	39.6%	97.8%
L2	100%	98.9%	54.9%	51%	**91.1%**
R1	100%	87.6%	26.4%	72.6%	100%
R2	91.2%	64.9%	12.7%	72.6%	61.4%
M1	99.8%	98.9%	38.8%	50.1%	99.6%
M2	100%	99.9%	14.7%	84.5%	100%
H1	66.3%	28.6%	7%	11.2%	34.8%
H2	100%	51.8%	13.2%	36.2%	90.2%
	94.6%	76%	28.7%	52.2%	84.4%

(c)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	**92.3%**	35.0%	21.5%	71.1%
L2	100%	64.2%	10.5%	26.8%	56.2%
R1	61.9%	50.4%	22.5%	50.2%	34.5%
R2	2.92%	11.3%	6.8%	35.8%	11.3%
M1	95.4%	77.7%	7.5%	67.1%	90.7%
M2	100%	99.9%	14.7%	84.5%	100%
H1	0%	0%	0%	0%	0%
H2	100%	61.4%	22.8%	45.5%	79.6%
	70%	57.2%	15%	41.4%	55.4%

(d)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	74.1%	27.7%	35%	96.6%
L2	100%	92.9%	21.9%	31.8%	66.3%
R1	17.6%	22.3%	3.9%	18.5%	31.5%
R2	23.3%	13.5%	7.4%	35.8%	24.4%
M1	100%	81.1%	78.8%	21.2%	96.1%
M2	55.7%	44.8%	6.2%	37.1%	49.8%
H1	8.8%	4.8%	2.9%	3.6%	5%
H2	59.9%	25.6%	7.1%	12.9%	39.3%
	58.2%	44.9%	19.5%	24.5%	51.1%

Table 2.2: Benchmark scores for SCV based on (Richa et al., 2011).

	Angle	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	99.8%	78.1%	60.1%	33.1%	98.5%
L2	100%	98.8%	62.5%	70.6%	94%
R1	99.8%	69.4%	24%	69.3%	100%
R2	81.1%	64.8%	12.8%	50%	61.8%
M1	100%	99.1%	52.4%	88.8%	100%
M2	99.8%	99.9%	14.9%	85.2%	100%
H1	64%	24.2%	6.9%	9.2%	34.5%
H2	92.8%	68.2%	17.9%	52.2%	100%
	92.2%	75.3%	31.4%	57.3%	86.1%

targets from Figure 2.7, and each column display the results for sequences angle, range, fast and far, fast and close and illumination changes. The scores in bold refer to the best result obtained for the dataset. For the sake of comparison, we consider any score difference below 5% to be irrelevant. The last row of each table displays the mean of scores obtained by all targets for each type of sequence.

We can notice from the scores that the *NCC+IC* presents the worst results, as it outperforms the other methods in only one sequence. This scored shows a 7.2% improvement of successfully tracked frames. Additionally, the *NCC+IC* could only achieve scores similar to proposed methods in 25% of the sequences, *i.e.* the scores are at least 10% worse than the other techniques for half of the sequences. The two proposed methods obtained similar results for 80% of sequences. We remark the outstanding performance of PROPOSED-I for sequences with illumination changes, where the method was able to track more than 99.8% of all images. This method performed better in 2 out of 8 of illumination sequences probably because the robust weighting reduces the influence of pixels with strong gradients, and these are specially responsible for the method's accuracy.

MI

Table 2.4 presents the scores obtained by two implementations of the MI similarity, *(a)* refers to our implementation of the inverse compositional algorithm of (Dame and Marchand, 2010), *(b)* refers to results of the same method presented in the original paper. Each row of the table presents the result for the targets from Figure 2.7. Furthermore, each column displays the results for sequences angle, range, fast and far, fast and close and illumination changes. The bold scores refer to the best result between the methods, *i.e.* if a method has a performance at least 5% better than the other. The last row of each table displays the mean of scores obtained by all targets for each type of sequence.

Noticeably, our implementation of the MI scores less than the original algorithm in 62.5% of the sequences, with an average difference of 36% of successfully tracked frames, *i.e.* 423 frames per sequence. However, an explicit comparison of the accuracy of the methods is unreasonable, as we do not have access to the experimental settings from (Dame and Marchand, 2010). There are several technical factors that may inflate the results. For instance, (Dame and Marchand, 2010) addresses that smoothing the image after the bi-linear interpolation can improve the tracking results. We do not consider that fine tuning technique, since they would alter the analysis of the similarities themselves. In fact, while some similarities can be improved via the smoothing of the reference image, others can be damaged. Moreover, the parametrization employed in each layer of the image pyramid can also improve the results. For instance, instead of computing three layers with the $\mathbb{SL}(3)$, one could compute the last layer using a simpler parametrization, *e.g.* an affine transformation on the pixel position, and finish the computation of the other layers using the original parametrization.

At last, we verify that our implementation of the inverse compositional approach for the MI shows better scores comparing to both SSD and NCC similarities using the same optimization technique for the sequences with angle and illumination changes. We can conclude from these results that the MI is indeed a good similarity function, at the expense of computational effort, of course. However, we can also conclude that other factors, *e.g.* optimization technique, also play an important in direct visual tracking, and should be taken more into account in direct visual tracking.

Table 2.3: Benchmark scores for NCC similarity. *(a)* PROPOSED-I, *(b)* PROPOSED-II, *(c) NCC+IC*.

(a)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	99.7%	76.8%	52.7%	27.6%	100%
L2	100%	99.9%	21.6%	**66%**	**100%**
R1	100%	57.7%	22.2%	68.2%	100%
R2	100%	81.3%	12.2%	53.6%	100%
M1	100%	96.8%	58.2%	**90.5%**	100%
M2	99.9%	99.9%	20.1%	80.5%	100%
H1	**93.6%**	**52.3%**	9.2%	14%	**98.9%**
H2	100%	51.5%	22%	**75%**	100%
	99.1%	77.0%	27.3%	59.4%	99.9%

(b)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	99.8%	92.2%	51.8%	31.6%	100%
L2	100%	95.8%	13.5%	42.1%	85.2%
R1	100%	59.1%	22.3%	68.1%	100%
R2	100%	81.1%	10.5%	**69.1%**	100%
M1	100%	96.1%	58.5%	86.1%	100%
M2	99.8%	99.9%	20.5%	85.3%	100%
H1	76.4%	16.9%	7.2%	9.7%	59.8%
H2	100%	**69.7%**	19.7%	42.8%	100%
	97%	76.4%	25.5%	54.4%	93.1%

(c)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	96.4%	16.4%	**38.8%**	97.3%
L2	100%	96.3%	21.8%	42.2%	80.9%
R1	17.6%	21.8%	3.9%	25.4%	31.6%
R2	12.2%	7.3%	6.8%	29.2%	20.3%
M1	100%	99.2%	38.4%	71.5%	99.4%
M2	55.3%	26.7%	6.2%	34.8%	40.8%
H1	10.3%	8.3%	5.2%	3.7%	6.5%
H2	42.8%	21.8%	6.8%	13.4%	18.3%
	54.8%	47.2%	13.2%	32.4%	49.4%

Table 2.4: Benchmark scores for MI similarity. *(a)* our implementation using an approximation of inverse compositional optimization, *(b)* result from (Dame and Marchand, 2010).

(a)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	97%	49.2%	41.9%	99.8%
L2	100%	99.9%	21.1%	42.8%	93.8%
R1	39.5%	23%	14.2%	44.5%	81.4%
R2	22%	9.1%	7%	29%	25%
M1	85.4%	83.2%	20.4%	**91.8%**	99.7%
M2	59.4%	13.5%	6.5%	25.6%	94.0%
H1	25.7%	22.4%	7.1%	9.6%	24.6%
H2	71.8%	31.2%	8.2%	34.1%	76.2%
	63%	47.4%	16.7%	39.9%	74.3%

(b)	*Angle*	*Range*	*F. F.*	*F. C.*	*Illum.*
L1	100%	94.1%	**75.2%**	56.5%	99.5%
L2	100%	98.1%	**69.9%**	43.7%	93%
R1	**76.9%**	**67.9%**	**22.8%**	63.6%	**100%**
R2	**91.3%**	**67.1%**	10.4%	70.5%	**96.2%**
M1	**99.2%**	**99.3%**	**43.9%**	86.7%	99.6%
M2	**100%**	**100%**	14.8%	84.5%	**100%**
H1	**47.1%**	**23.2%**	7.2%	10%	**50.6%**
H2	**100%**	**69.8%**	20.8%	**83.8%**	**100%**
	89.3%	77.4%	33.1%	62.4%	92.4%

Table 2.5: Comparative results under challenging illumination.

BEAR

Method	#Img.	#Its.	#Parms.	RMS	$S_\times$
SSD+i	1573	10	8G+36P	20	0.807
SCV	1573	9	8G	0.02	0.874
NCC, Alg. 1	1573	12	8G	–	0.807

BOOK

Method	#Img.	#Its.	#Parms.	RMS	$S_\times$
SSD+i	163	11	8G+49P	10.1	0.962
SCV	203	10	8G	0.04	0.970
SCV, $\varepsilon = 10^{-4}$	283	13	8G	0.03	0.970
NCC, Alg. 1	203	14	8G	–	0.957
NCC, ε=10^{-4}	283	19	8G	–	0.959

ROBOT

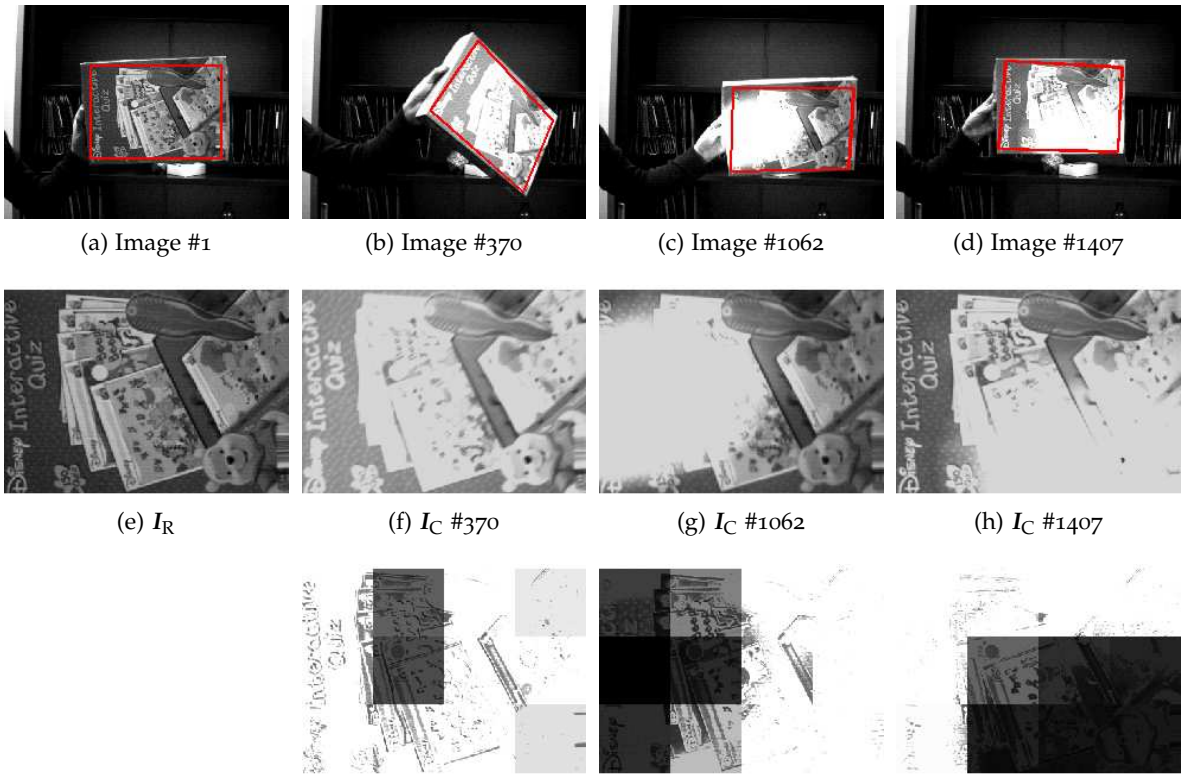
Method	#Img.	#Its.	#Parms.	RMS	$S_\times$
SSD+i	723	10	8G+36P	21	0.864
SCV	723	10	8G	0.02	0.882
NCC, Alg. 1	723	12	8G	–	0.864

2.7.3 *Evaluation under challenging illumination*

Next, we evaluate the techniques in sequences with challenging illumination settings. We selected the following implementations based on the results obtained in the Metaio benchmark:

- *SSD+i* from (Silveira and Malis, 2010);
- *SCV* from (Richa et al., 2011);
- *NCC* similarly from Algorithm 1.

We neglect the SSD using Huber weights and ESM optimization because this results obtained for this technique are inferior to the others in the sequences analyzed. We evaluate the three techniques on sequences BEAR, and BOOK from (Silveira and Malis, 2010) and ROBOT from (Dame and Marchand, 2010). These sequences represent extreme real-world situations with challenging illumination and targets from different materials and sizes. Remark that, besides 8 geometric parameters from the $\mathbb{SL}(3)$, the *SSD+i* must estimate other photometric parameters that increase with the number of grids. We evaluate the three methods using the same reference image, minimum step size $\varepsilon = 10^{-3}$, maximum of 50 iterations. These are reasonable parameters for most real-time applications. Furthermore, we found the BOOK to be more complex because none of the methods was able to complete the tracking using the default parameters. We thus reevaluate this sequence using the SCV and the proposed method with $\varepsilon = 10^{-4}$ and 500 iterations. Table 2.5 presents the comparative result in terms of total tracked images, median of iterations per image, number of estimated parameters, and root mean squared (RMS) error of intensities, and the median NCC of the resulting $\boldsymbol{I}_\mathrm{C}$ and $\boldsymbol{I}_\mathrm{R}$. The root mean squared error of intensities and the NCC for the SCV technique are computed using the expected image. Figs. 2.8, 2.9, 2.10 present key samples obtained for the proposed method in sequences BEAR, BOOK and, ROBOT respectively.

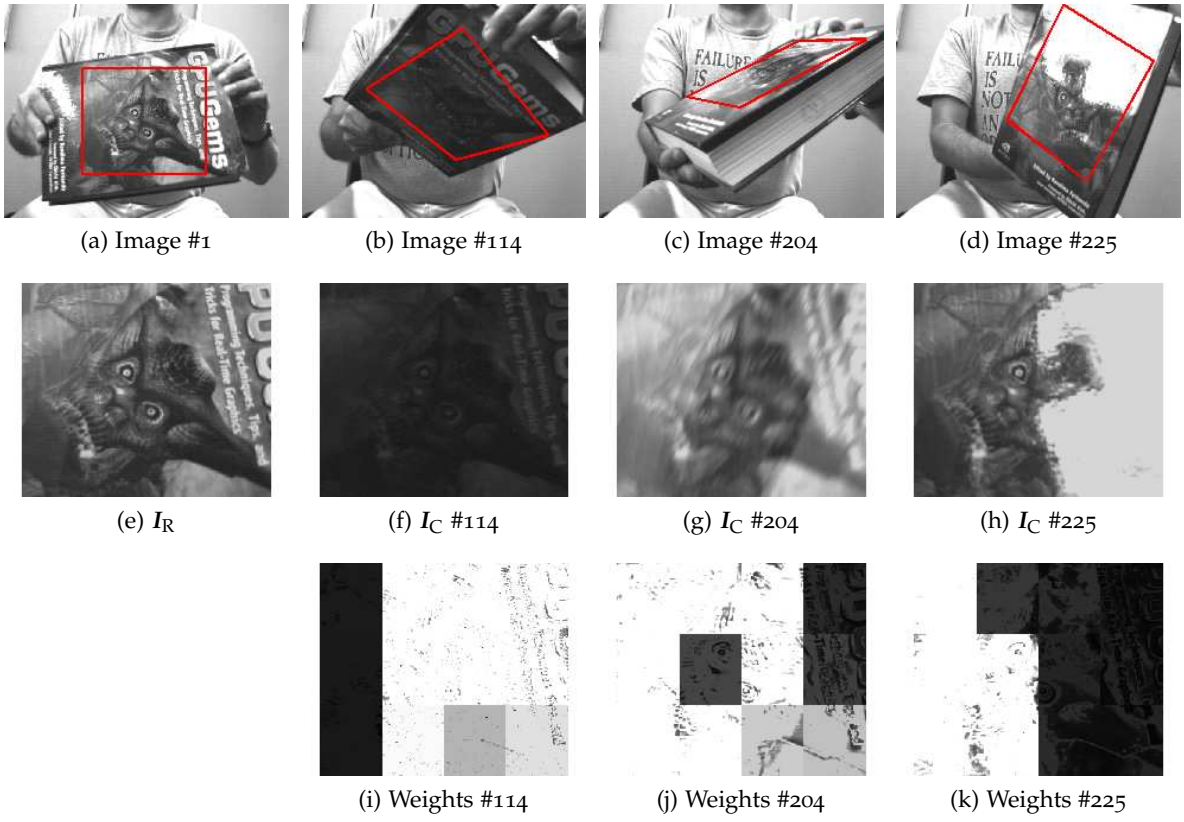


(a) Image #1 (b) Image #370 (c) Image #1062 (d) Image #1407

(e) I_R (f) I_C #370 (g) I_C #1062 (h) I_C #1407

(i) Weights #370 (j) Weights #1062 (k) Weights #1407

Figure 2.8: Samples from BEAR sequence – Total of 1573 images.



(a) Image #1 (b) Image #114 (c) Image #204 (d) Image #225

(e) I_R (f) I_C #114 (g) I_C #204 (h) I_C #225

(i) Weights #114 (j) Weights #204 (k) Weights #225

Figure 2.9: Samples from BOOK sequence – Total of 283 images.

Table 2.6: Comparative results under challenging illumination and partial occlusion.

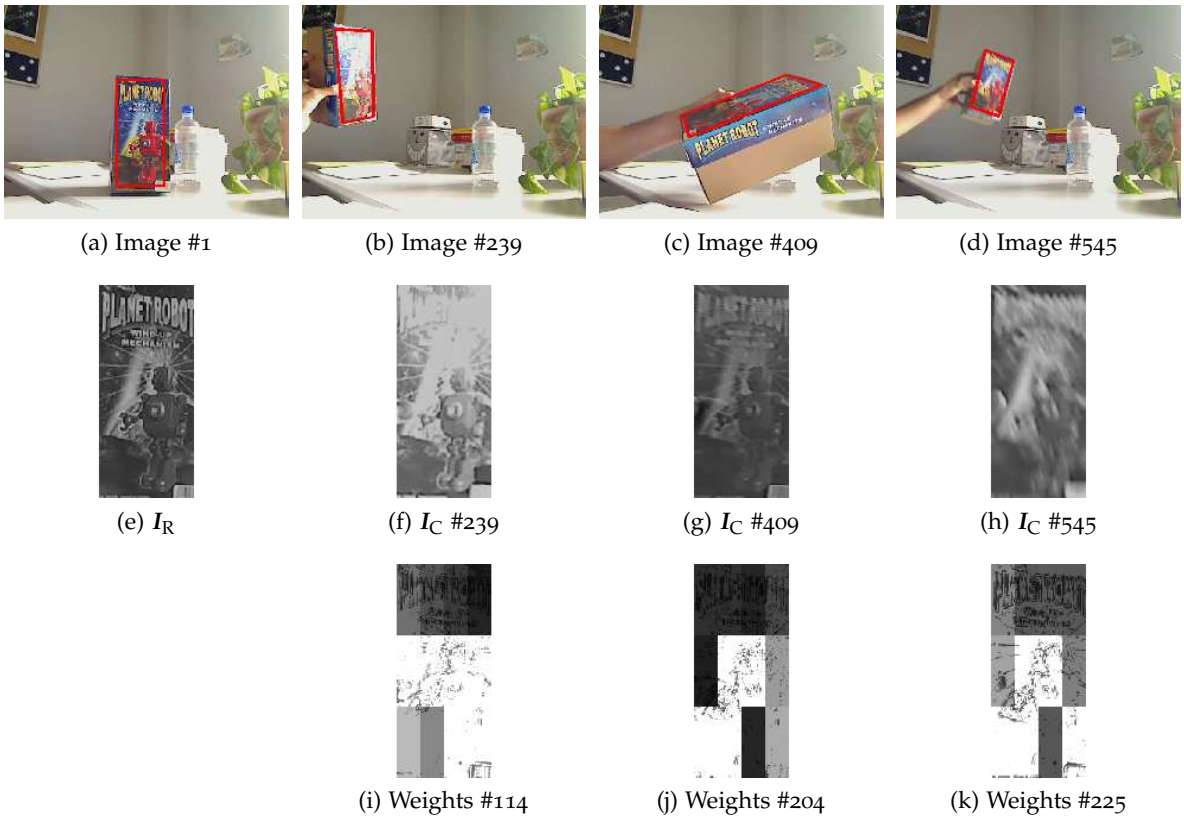
Starry-Night

Method	#Img.	#Its.	#Parms.	RMS	$S_\times$
SSD+i	400	12	8G+43P	13.3	0.939
SSD+i/Huber	680	21	8G+43P	21.4	0.876
SCV	536	11	8G	0.05	0.943
NCC, Alg. 1	1600	22	8G	–	0.849

We can verify that the proposed method performed at least similarly to the *SSD+i*. Note that the proposed method presents a slight increase in the number of iterations, however, we obtained a median of 23 iterations for the *NCC+ICS* (using the robust techniques). This result highlights the importance of the improvement proposed in Section 2.35, since, using the information from *inverse* and *forward* solutions, the NCC compares to a second order method in terms of iterations. The SCV displays a better score for both RMS and the NCC $S_\times$, however, the result takes into account the information computed in joint intensity distribution. Remark that the proposed method and the SCV performed better than the *SSD+i* for the BOOK sequence. The *SSD+i* gets stuck in a local minimum at frame 163, however, the other methods are able to continue until 203 using the same parameters. The decrease of the median NCC and increase on the iteration numbers is directly related to the sequences that the *SSD+i* was unable to track. We can consider frame 204 to be the most difficult from this sequence. Our method and the SCV were only capable of completing the sequence without the real-time constraint imposed by the iterations. The *SSD+i* still failed in the same local minimum (frame 163). Concluding, the ROBOT sequence shows similar results for the three techniques. We can verify, however, that the proposed technique is also suited for tracking smaller reference images as the one provided by this sequence.

2.7.4 *Evaluation under partial occlusion and illumination changes*

This section evaluates the same techniques discussed previously in a real-world situation where a region of the reference image is partially occluded and the illumination of the environment changes throughout the experiment. Again, we use the same parameters: reference image, minimum step and maximum iterations for every method. Table 2.6 presents the comparative results for all of the methods, but we also include and an implementation of the *SSD+i* solving a robust least-squares with Huber weights. Fig. 2.11 presents some key frames results obtained for the proposed method in the sequence STARRY NIGHT. The proposed method was the only capable of completing the full sequence. The *SSD+i* was unable to cope with the partial occlusion, and this is the main reason why it presents less median iterations per frame and a larger NCC than the other two methods. The *SSD+i* with M-estimator was capable of tracking the occluded patch as long as there were no illumination changes. We can infer that the *SSD+i* is unable to handle partial occlusion and illumination changes at the same time. The capability of facing illumination changes and occlusion supports the weighting techniques presented in Section 2.5. To the authors knowledge, this is the only approach capable of dealing with such extreme situations.

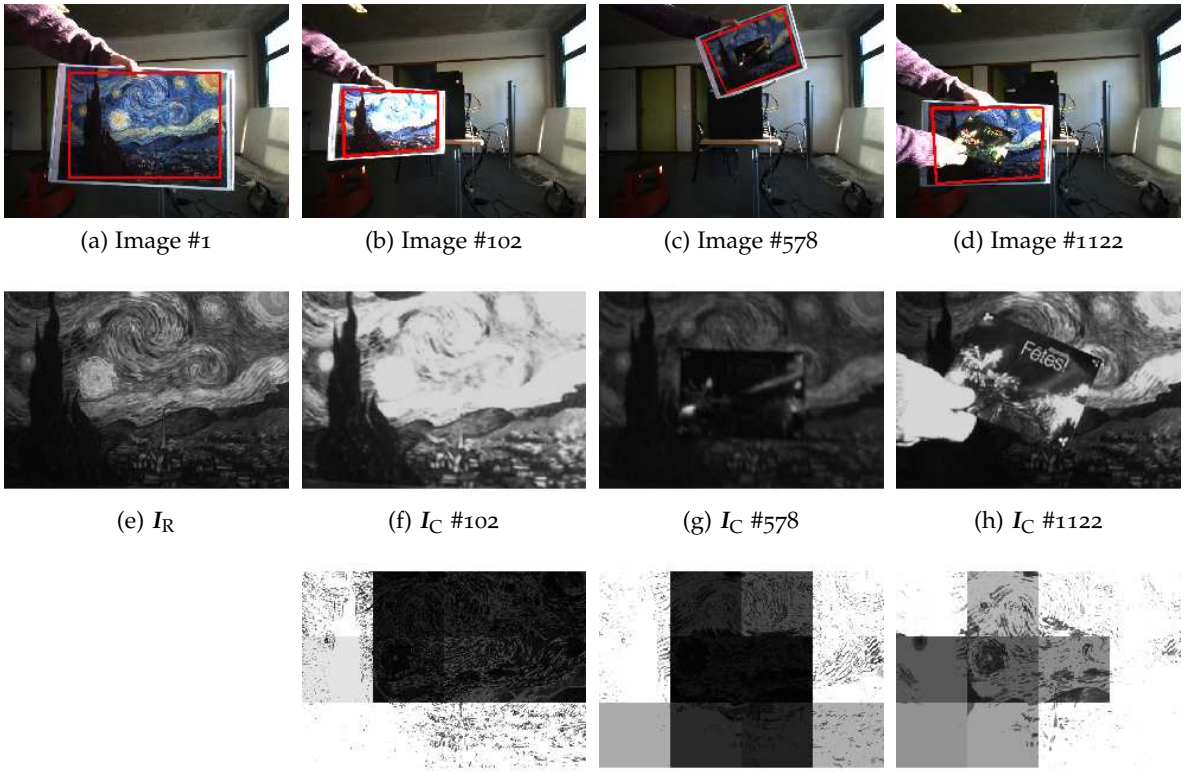


(a) Image #1 (b) Image #239 (c) Image #409 (d) Image #545

(e) I_R (f) I_C #239 (g) I_C #409 (h) I_C #545

(i) Weights #114 (j) Weights #204 (k) Weights #225

Figure 2.10: Samples from Robot sequence – Total of 723 images.



(a) Image #1 (b) Image #102 (c) Image #578 (d) Image #1122

(e) I_R (f) I_C #102 (g) I_C #578 (h) I_C #1122

(i) Weights #102 (j) Weights #578 (k) Weights #1122

Figure 2.11: Samples from Starry Night sequence – Total of 1600 images.

2.8 CONCLUSION

This chapter addressed different aspects of direct visual tracking methods. Many different direct visual tracking methods present in the literature can differ by the employed similarity measure, optimization strategy and geometric parameters. We first introduced different similarity measures and discussed several properties, such as invariance and robustness to illumination changes. Secondly, we presented different optimization techniques that can be employed to obtain the tracking solution with reasonable computational effort.

The main contribution of this chapter is a novel solution using the NCC as similarity measure. This solution is based on three main pillars. Two techniques are presented to increase the robustness against non-modeled effects, *e.g.* specular reflections and partial occlusion, and these techniques highly improve the rejection of degraded areas. We also address a Newton-like gradient solution using both *inverse* and *forward compositional* approaches.

The proposed method is exhaustively compared to other state of the art methods via the analysis of the basin of convergence, the scores obtained using a planar based visual tracking benchmark dataset, and challenging real-world video sequences. We verify that the proposed method is able to cope with tracking partially occluded objects even under severe illumination changes.

The downside of every gradient-based direct visual tracking method, however, is the need of an initialization close enough to the optimal solution. In practice, it is difficult to initialize the methods after large displacements, unless we either impose strong constraints on camera motion, or introduce other sensors capable of measuring incremental displacements in faster rates. Inertial sensors provide the latter property, and the next chapter discusses properties and techniques of pose estimation with these sensors.

3

NONLINEAR OBSERVERS FOR POSE ESTIMATION

This chapter concentrates on the data fusion process for pose estimation. The objective is to combine low-frequency pose measurements with high-frequency measurements of angular velocity and linear acceleration obtained by an IMU. In order to achieve this objective, we develop nonlinear observers that can provide filtered pose estimates, and various parameters of the sensors, *e.g.* IMU bias, gravitational acceleration, rotation and translation from camera-to-IMU (c-to-IMU) frames. Bias estimates of these parameters can severely impair the data fusion process. A first source of difficulties comes from IMU measurement bias. Although a constant dynamics represents a good model for the these parameters, *c.f.* Section 1.2.4, the biases may vary due to many factors (*e.g.* temperature variation, battery level, etc). Therefore, it is interesting to permanently estimate these parameters. Another source of difficulties concerns various parameters related to the use of multiple sensors and different coordinate frames, *e.g.* camera and IMU frames, or the computation of relative pose with respect to a certain reference frame. Usually, relative pose from c-to-IMU frames can be estimated in a preliminary step using the accelerometers as a measurement of gravity (Lobo and Dias, 2007), however, such that technique assumes that the IMU bias is known. An underlying difficulty of the concurrent estimation of frame parameters and IMU bias is that persistent motion conditions must be satisfied for the system to be completely observable.

The following sections discuss the design of nonlinear observers for the estimation of pose and multiple parameters of the system, as well as analyses of motion condition under which the system is observable. In order to obtain these condition, we initially recall some theoretical results on the observability of systems, state-of-the-art results on filter design. We then provide sufficient conditions for the uniform observability of linear time-varying systems, and one technique to obtain inputs that provide local-weak observability for nonlinear systems. The analysis of the observability is specially important for the design and stability analysis of the nonlinear observers next proposed. The most important discussions of this chapter concern the design of nonlinear observers for pose estimation with concurrent IMU bias, and c-to-IMU rotation. The respective nonlinear observers have been introduced in (Scandaroli and Morin, 2011), (Scandaroli et al., 2011). These results are complemented with the observability analysis of the system, together with an original nonlinear observer for attitude and gyro bias estimation, and new results on pose estimation with estimation of the gravitational acceleration and c-to-IMU translation. We conclude the chapter with simulation results, and compare the proposed nonlinear observers with other methods proposed in the Literature.

In order to discern the results presented in this thesis from other results provided by the Literature, we use a non-standard nomenclature for the results. The results presented in this thesis are stated as Propositions, whilst the results from the Literature are stated as Theorems. We also separate completely the presentation of the results from their proofs. The proofs for the propositions of the thesis are discussed in Appendix A.

3.1 THEORETICAL RECALLS

This section reviews some theoretical tools on nonlinear system analysis and nonlinear observers. Most of these results are found in control references, *e.g.* (Kalman, 1960b), (Chen, 1984), (Nijmeijer and van der Schaft, 1990), and tutorials on observability and observers (Besançon, 2007). Let us start from a generic nonlinear system:

$$\begin{cases} \dot{x} = f(x, u, t), \\ y = h(x, u, t), \end{cases} \tag{3.1}$$

where t denotes the time, the state $x \in \mathbb{X}$, the input $u \in \mathbb{U}$ and the output $y \in \mathbb{Y}$, such that $\mathbb{X}$, $\mathbb{U}$ and $\mathbb{Y}$ are connected manifolds of dimension n, m and p respectively, $f(x, u, t)$ and $h(x, u, t)$ are vector functions with proper dimensions.

The general problem of observability and observer design concerns the reconstruction of the state $x(t_0)$ of the dynamical system (3.1) from the knowledge of output trajectory $y(\cdot)$ and inputs $u(\cdot)$.

3.1.1 *Observability of systems*

Rudolph Kalman first introduced the concept of observability for the analysis of linear time-invariant systems (Kalman, 1960b), and thenceforward the observability property has been extensively studied and extended for other classes: linear time-varying, state-affine, and nonlinear systems. The observability is a prerequisite to derive exponentially stable observers. In this section, we review some results presented so far in the literature, complementing with a Proposition to identify sufficient conditions for the uniform observability of linear time-varying systems. This result delineates a simple method to evaluate universal inputs that yield a uniformly observable state-affine system.

Linear time-invariant systems

Linear time-invariant systems write

$$\begin{cases} \dot{x} = Ax + Bu, \\ y = Cx + Du, \end{cases} \tag{3.2}$$

where the state vector $x \in \mathbb{R}^n$, the input vector $u \in \mathbb{R}^m$ and the output vector $y \in \mathbb{R}^p$, such that A, B, C, and D are matrices with compatible dimension.

We can verify the observability of the system (3.2) via the observation space

$$\mathcal{O} \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

computed from A and C. The Kalman rank condition states that the system is observable iff $\operatorname{rank}(\mathcal{O}) = n$, see *e.g.* (Kailath, 1979). If the system is observable, the pair (A, C) is defined as observable. Remark that the observability of linear time invariant systems is independent of the inputs u and time t, which may not be true for other classes of systems.

Linear time-varying systems

Linear time-varying systems write

$$\begin{cases} \dot{x} = A(t)x + B(t)u, \\ y = C(t)x + D(t)u, \end{cases} \tag{3.3}$$

where $A(t)$, $B(t)$, $C(t)$, and $D(t)$ are matrix-valued functions of the time t with compatible dimension. The observability of linear time-varying systems can be categorized by different aspects and (Chen, 1984) classifies the observability of a linear time-varying system according to the following definitions.

Definition 3.1. A system is *differentially observable* if, $\forall\, t$, the state $x(t)$ can be computed from the inputs $u(\tau)$ and outputs $y(\tau)$ during $\tau \in [t, t+\tau]$ for $\tau > 0$ arbitrarily small.

Definition 3.2. A system is *instantaneously observable* if, $\forall\, t$, the states $x(t)$ can be computed from the inputs $u(t)$, outputs $y(t)$ and time derivatives $u^{(k)}(t)$, $y^{(k)}(t)$ with $k \leq n+1$.

Definition 3.3. A system is *uniformly observable* if $\exists\ \tau > 0$ such that, $\forall\, t$, the states $x(t)$ can be computed from the inputs $u(\tau)$ and outputs $y(\tau)$ during $[t, t+\tau]$.

The above definition of uniform observability is different from the one presented, *e.g.* (Gauthier and Kupka, 1994) or (Besançon, 2007), where uniformity is related to the inputs rather than time.

We are specifically interested in uniform observability, which ensures that the state estimation process is well-conditioned and can be solved via the design of exponentially stable observers. The following assumption is made to simplify the exposition.

Assumption 3.1. The matrix-valued functions A, B, C and D of the linear time varying system (3.3) are continuous and bounded on $[0, \infty)$.

The next theorem is a well known result on uniform observability (Chen, 1984, Ch. 5).

Theorem 3.1. A linear time-varying system (3.3) satisfying Assumption 3.1 is uniformly observable if there exist $\tau, \delta > 0$ such that

$$\forall t \geq 0, \quad 0 < \delta I_n \leq W(t, t+\tau), \tag{3.4}$$

where

$$W(t, t+\tau) \triangleq \int_t^{t+\tau} \Psi(s,t)^T C^T(s) C(s) \Psi(s,t)\, ds,$$

and $\Psi(s,t)$ is the state transition matrix of $\dot{x} = A(t)x$. The matrix W is called *observability Grammian* of System (3.3).

Remark that the observability of linear time varying systems is independent of the input u, but it is directly related to the system's Grammian which, in turn, only depends on the matrices $A(t)$ and $C(t)$. Hence, it should be clear from this definition that uniform observability is independent of B. We say without distinction that System (3.3) or the pair (A, C) is uniformly observable. Note also, as a consequence of Assumption 3.1, that $W(t, t+\tau)$ is upper bounded by some $\bar{\delta} I_n$ for any $t \geq 0$.

An important property of uniform observability is that it ensures the existence of exponentially stable linear time-varying observers with bounded gain matrix. Uniform observability of a given pair (A, C), however, is usually difficult to establish since calculation of the Grammian matrix requires explicit integration of the solutions of $\dot{x} = A(t)x$. It is well known that

observability properties of linear time-varying systems are strongly related with properties of the observation space of linear time varying systems (Chen, 1984, Ch. 5)

$$\mathcal{O}(t) \triangleq \begin{pmatrix} N_0(t) \\ N_1(t) \\ \vdots \end{pmatrix} \tag{3.5}$$

with

$$N_0 \triangleq C, \quad N_{k+1} \triangleq N_k A + \dot{N}_k \quad \text{for } k = 1, \cdots \tag{3.6}$$

For example, instantaneous observability can be guaranteed at t if $\text{rank}(\mathcal{O}_{n-1}(t)) = n$. However, uniform observability cannot be characterized in term of rank condition. We next propose a Lemma that provides a sufficient condition for uniform observability in terms of the observability space $\mathcal{O}$.

Lemma 3.1. Consider a linear time-varying system (3.3) satisfying Assumption 3.1 such that the following statements hold:

1. The k-th order derivative of A (resp. C) is well defined and bounded on $[0, +\infty)$ up to $k = K \geq 0$ (resp. up to $k = K + 1$).
2. There exist an $n \times m$ matrix M composed of row vectors of $N_0, \cdots, N_K$, and two scalars $\delta, \tau > 0$ such that

$$\forall t \geq 0, \quad 0 < \delta \leq \int_t^{t+\tau} |\det(M(s)^T M(s))| \, ds. \tag{3.7}$$

Then, System (3.3) is uniformly observable.

The proof for this lemma is discussed in Appendix A.1.

A similar criterion is proposed in (Bristeau et al., 2010). The observability condition therein, however, requires the evaluation of a similar inequality for every instant $t \in [t_0, t_0 + \tau]$, while Proposition 3.1 presents the observability condition based on the state trajectory along the same interval.

State-affine systems

State-affine systems write

$$\begin{cases} \dot{x} = A(u)x + b(u), \\ y = C(u)x + d(u), \end{cases} \tag{3.8}$$

where $A(u)$, $C(u)$ are matrix valued functions, and $b(u)$, $d(u)$ vector valued functions with proper dimensions. The observability of state-affine systems, differently from linear time invariant and time varying systems, depends on the inputs u, thus we are specially interested in identifying the inputs that yield the observability of the system.

Definition 3.4. An input u is *universal* for the system (3.8) if the resulting system is observable. An input u is *singular* if it is not universal (Sussmann, 1979).

We can see the observability analysis of state-affine systems likewise linear time varying ones. More specifically, we define $A_u(t) = A(u(t))$ and $C_u(t) = C(u(t))$ and the same tools from the previous section apply for the pair $\big(A_u(t), C_u(t)\big)$. Specially, Proposition 3.1 gives sufficient conditions on the dynamics of uniformly universal inputs for System (3.8), *i.e.* we can verify conditions on the trajectory of the inputs $u(t)$ and via its derivatives $\dot{u}(t)$, $\ddot{u}(t)$, $\cdots$, under which the state-affine system (3.8) is uniformly observable.

In the cases discussed in this thesis, we determine the dynamics of universal inputs, that must, in turn, be continuous and bounded on $[0, \infty)$ to satisfy Assumption 3.1.

Nonlinear systems

Nonlinear systems write

$$\begin{cases} \dot{x} = f(x, u, t), \\ y = h(x, u, t), \end{cases}$$

where the state $x \in \mathbb{X}$, the input $u \in \mathbb{U}$ and the output $y \in \mathbb{Y}$, such that $\mathbb{X}$, $\mathbb{U}$ and $\mathbb{Y}$ are connected manifolds of dimension n, m and p respectively, $f(x, u, t)$ and $h(x, u, t)$ are vector functions with proper dimensions.

The concept of observability in nonlinear systems is different from the previous systems, since some states can be *indistinguishable* due to the structural properties of certain nonlinear systems (Hermann and Krener, 1977).

Definition 3.5. A pair (x_0, x_1) is *indistinguishable* if they realize the same input-output map. A state x is indistinguishable from x_0 if the pair (x, x_0) is indistinguishable.

Nevertheless, we can relate the above definition to the original observability definition, *i.e.* we can reconstruct the state $x(t_0)$ of a system from its inputs and outputs iff there exist no indistinguishable pair (x, x_0). This concept of observability, however, is still too general for many nonlinear systems. Indeed, for such cases, one might be interested in distinguishing a state x from its neighbors instead of the whole space of states.

The main result for the observability of nonlinear systems (Hermann and Krener, 1977) defines a weaker notion of observability called *local weak observability*. More specifically, the system (3.1) is *locally weakly observable* if the state $x(t_0)$ can be distinguished from its neighbors without "going too far" (Besançon, 2007). Another advantage of the weaker observability is that a simple algebraic test on the observation space $\mathcal{O}(h, x, u, t)$ characterizes that property.

A rank test is proposed in (Hermann and Krener, 1977) for the observability of nonlinear time-invariant systems

$$\begin{cases} \dot{x} = f(x, u), \\ y = h(x), \end{cases} \tag{3.9}$$

using the Lie derivate operator

$$L_f h(x) = \big(\partial_x h(x)\big) f(x, u),$$

and the k-th successive computation of a Lie derivative defined by

$$L_f^k h(x) = L_f\big(L_f^{k-1} h\big)(x) = \big(\partial_x L_f^{k-1} h(x)\big) f(x, u)$$

for $k = 2, \cdots$. Define the observability space

$$\mathcal{O}(h, x) \triangleq \begin{pmatrix} h(x) \\ L_f h(x) \\ L_f^2 h(x) \\ \vdots \end{pmatrix}. \tag{3.10}$$

computed with $f(x, u_i)$ for every possible constant value u_i.

Theorem 3.2. Let the observable space $\mathcal{O}$ given by (3.10) evaluated for some constant input $u_0 \in \mathbb{U}$. A nonlinear system (3.9) is locally weakly observable at $x \in \mathbb{X}$

$$\dim\big(\partial_x \mathcal{O}(h, x)\big) = \dim(\mathbb{X}). \tag{3.11}$$

The system is weakly observable if (3.11) holds $\forall\, x \in \mathbb{X}$.

This result is particularly interesting for time invariant systems and systems where the output is a sum of the initial state and a function of the inputs. The latter case even present the strong property that if one input distinguishes between two initial states, then every input does also distinguish between two states (Hermann and Krener, 1977). Remark that the Kalman rank condition in linear time invariant systems, for instance, is obtained directly from Theorem 3.2.

The rank condition provided by Theorem 3.2 may not be trivially verified when the observability depends on the inputs. Besides, the observability of nonlinear affine systems relies often on the inputs, for example, consider

$$\begin{cases} \dot{x} = f_0(x) + \sum_{i=1}^{m} f_i(x) u_i, \\ y = h(x), \end{cases} \tag{3.12}$$

with u_i the i-th element of the input $u \in \mathbb{R}^m$, may not be guaranteed with $u = 0$. Moreover, instead of tediously computing the observability space for "infinite" values of u, we can utilize another definition of observability space, *c.f.* (Nijmeijer and van der Schaft, 1990),

$$\mathcal{O}(h, x) \triangleq \begin{pmatrix} h(x) \\ L_{f_{\theta_0}}(h)(x) \\ L_{f_{\theta_1}} L_{f_{\theta_0}}(h)(x) \\ L_{f_{\theta_2}} L_{f_{\theta_1}} L_{f_{\theta_0}}(h)(x) \\ \vdots \end{pmatrix}, \tag{3.13}$$

with $\theta_i \in \{0, 1, \cdots, m\}$, and a similar rank test to verify the observability.

Corollary 3.1. Let observable space $\mathcal{O}$ given by (3.13). A nonlinear system (3.12) is locally weakly observable at $x \in \mathbb{X}$ if:

$$\dim\left(\partial_x \mathcal{O}(h, x)\right) = \dim(\mathbb{X}). \tag{3.14}$$

for some input $u \in \mathbb{U}$. The system is weakly observable if (3.11) holds $\forall\, x \in \mathbb{X}$.

The above corollary follows from Theorem 3.2 using (3.13). This result guarantees that there exist some trajectory of the input u such that (3.12) is locally weakly observable. It does not, however, specify which are the universal inputs.

3.1.2 *Definition of an observer*

Given a process and measurement model, *e.g.* dynamics (3.1), we are interested in reconstructing the state $x(t)$ from the information of its inputs $u(t)$ and outputs $y(t)$. This problem is trivial if the measurement function is invertible w.r.t. the state, *i.e.* $x = h^{-1}(y, u, t)$. That is not the case in most situations however.

Definition 3.6. An observer f^* for system (3.1) writes:

$$\dot{\hat{x}} = f(\hat{x}, u, t) + k(\hat{x}, u, y, t), \tag{3.15}$$

where $\hat{x}(t) \in \mathbb{X}$ is the estimate of the state $x(t)$ and k is an innovation term defined such that:

$$\hat{x}(0) = x(0) \;\Rightarrow\; \hat{x}(t) = x(t),\ \forall\, u \in \mathcal{U}, \text{and } t > 0.$$

Note that $k(x, u, y(x, u, t), t) \equiv 0$. The above definition concerns only the equivalence between the observer and system dynamics, thus, if the estimation is exact in a certain instant t it will continue being so afterwards. Most situations do not satisfy the premise $\widehat{x}(0) = x(0)$ however. The *convergence* properties of an observer will characterize the ability of obtaining the state $x(t)$ starting with some $\widehat{x}(0) \neq x(0)$. Let us first consider the particular case where $\mathbb{X} = \mathbb{R}^n$, and define the estimation error by

$$\widetilde{x}(t) \triangleq x(t) - \widehat{x}(t). \tag{3.16}$$

Definition 3.7. An observer (3.15) of system (3.1) is asymptotically stable if:
- $\widetilde{x} = 0$ is stable in the sense of Lyapunov;
- $\widetilde{x}(t) \to 0_{n\times 1}$ as $t \to \infty$ for $\widehat{x}(0) \in \mathbb{V} \subset \mathbb{X}$.

The observer is globally stable if $\mathbb{V} = \mathbb{X}$. Moreover, the observer is exponentially stable with convergence rate $a > 0$ if there exists some $c_a > 0$ such that $|\widetilde{x}(t)| \leq c_a e^{-a(t-t_0)}|\widetilde{x}(t_0)|$ for any $t \geq t_0$.

The estimates of an asymptotically (resp. exponentially) stable observer are stable (resp. exponentially stable). The difference among various observer-based techniques lies on the innovation terms $k(\widehat{x}, u, y, t)$ designed to obtain the stability of the estimates. The remainder of this section covers some results already shown in the literature.

3.1.3 *Existence of observers*

The estimation problem is directly related to the observability properties of the system. The following Lemma summarizes two properties of uniformly observable systems. The first property follows from (Anderson and Moore, 1969, Lemma 3) and the duality principle, *c.f.* (Chen, 1984, Th. 5-10). This principle, together with (Ikeda et al., 1972, Th. 3), imply the second property.

Lemma 3.2. The following properties hold for linear time varying systems (3.3) satisfying Assumption 3.1.

1. The pair (A, C) is uniformly observable iff the pair $(A - KC, C)$ is uniformly observable, with $K(\cdot)$ any bounded matrix-valued time-function.
2. If the pair (A, C) is uniformly observable, then for any $a > 0$ there exists a bounded matrix $K(t)$ such that the linear observer

$$\dot{\widehat{x}} = A(t)\widehat{x} + B(t)u + K(t)\big(y - (C(t)\widehat{x} + D(t)u)\big) \tag{3.17}$$

is uniformly globally exponentially stable with convergence rate a.

3.1.4 *Observers for linear systems*

Luenberguer observers

Luenberguer observers are defined for linear time invariant systems (3.2) exploiting directly the general form (3.17), with K defined such $A - KC$ is Hurwitz stable. This observer is globally exponentially stable (Luenberger, 1966). The direct use of Luenberguer observers for linear time varying systems does not imply necessarily a stable observer, *i.e.* choosing some $K(t)$ such that $A(t) - K(t)C(t)$ is Hurwitz stable at each $t \geq 0$ for system (3.3) does not necessarily imply that the observer is stable, *c.f.* (Reinhard, 1989, p. 131). Considering linear time varying systems, we must employ other tools, such as Lyapunov's direct method, to verify the stability of the system.

Kalman-Bucy filter and variants

The Kalman-Bucy filter is defined for linear time varying systems (3.3) exploiting directly the general form (3.17), with K given by

$$\begin{aligned} K(t) &= P(t)C^T(t)W^{-1}(t), \\ \dot{P} &= A(t)P + PA^T(t) - PC^T(t)W^{-1}(t)C(t)P + Q(t). \end{aligned} \tag{3.18}$$

such that $P(0) > 0$, $Q(t) \geq 0$, and $W(t) > 0$.

Using a stochastic rationale, $Q(t)$ corresponds to the covariance matrix associated to the process model and $W(t)$ the covariance associated to the measurement and the Kalman-Bucy filter obtains the optimal estimate for Gaussian probability distributions if the system is uniformly observable (Kalman and Bucy, 1961). Using the deterministic concept, the resulting observer is uniformly globally exponentially stable iff the linear time varying system (3.3) is uniformly observable.

Kalman-like filters are characterized by the dependence of the gain on an auxiliary matrix $P(t)$ computed from a Riccati equation. Remark that this matrix must be computed in addition to the states x. Defining $Q(t) = Q$ and $W(t) = W$ for linear time invariant systems, however, the matrix $P(t)$ converges to a constant P^* if the pair (A, C) is observable (Kalman and Bucy, 1961). An observer employing $K(t) = P^*C^TW^{-1}$ is known as "steady-state" Kalman-filter.

It is also possible to employ the Kalman-Bucy filter for the estimation of state-affine systems (3.8) by changing the observer structure into

$$\dot{\hat{x}} = A(u,t)\hat{x} + b(u,t) - K(t)\big(y - (C(u,t)\hat{x} + d(u,t))\big)$$

and compute K as in (3.18) with $A(t) = A(u(t),t)$, $C(t) = C(u(t),t)$, *c.f.* (Bornard et al., 1989). Furthermore, another Kalman-like gain can be employed:

$$\begin{aligned} K(t) &= P^{-1}(t)C^T(u,t), \\ \dot{P} &= -kP + A(u,t)P + PA^T(u,t) - C^T(u,t)C(u,t). \end{aligned} \tag{3.19}$$

such that $P(0) > 0$ and $k > 0$, *c.f.* (Hammouri and de Leon Morales, 1990). In the latter formulation, there are less parameters to tune, thus a simpler Kalman-like filter is obtained. The stability of observers for state affine systems is directly related to the inputs, *i.e.* the estimates are uniformly globally exponentially stable if system (3.8) satisfies Assumption 3.1 and the input $u(t)$ yields a uniformly observable system.

3.1.5 *Observers for nonlinear systems*

Extended Kalman filter

The extended Kalman filter is a standard linearization method for approximate nonlinear filtering. The following observer is defined for a generic nonlinear system (3.1)

$$\dot{\hat{x}} = f(\hat{x},u,t) + K(t)\big(y - h(\hat{x},u,t)\big)$$

with gain $K(t)$ is given by (3.18) such that $P(0) > 0$, $Q(t) \geq 0$, $W(t) > 0$, and $A(t)$, $C(t)$ given by:

$$\begin{aligned} A(t) &= \partial_x f(\hat{x}(t),u(t),t), \\ C(t) &= \partial_x h(\hat{x}(t),u(t),t). \end{aligned}$$

Assuming that
- $A(t)$ and $C(t)$ are continuous and bounded in $t \in [0, \infty]$,
- the pair (A, C) is uniformly observable using,

one can verify, *e.g.* , using Lyapunov's indirect method (Khalil, 2002, p. 161), that the extended Kalman filter is locally exponentially stable. The nonlinear form (3.1) is quite universal, however one must be aware that the observability of the linearized system may also depend on the inputs. Furthermore, remark that these assumptions are often difficult to be ensured, since, for instance, the boundedness of A and C require that $\widehat{x}$ is also bounded. That becomes, in a certain level, a circular argument where the hypothesis merge with the result, as we can ensure that $\widehat{x}$ is bounded only if the filter is stable.

Nonlinear observer design via output injection

There are some nonlinear structures that admit a Luenberguer-like observer. For instance, let us consider the following system:

$$\begin{cases} \dot{x} = Ax + Bu + \varphi(Cx + Du, u, t), \\ y = Cx + Du. \end{cases}$$

The above system presents an interesting characteristic: the nonlinear contribution is given by an additive nonlinear function of the outputs. Therefore, we can asses this filtering problem with the following observer form:

$$\dot{\widehat{x}} = A\widehat{x} + Bu + \varphi(y, u, t) + K(t)\big(y - (C\widehat{x} + Du)\big). \tag{3.20}$$

This technique is known as linearization by output injection (Krener and Isidori, 1983). We can obtain directly a Luenberguer-like observer defining K to obtain globally exponentially stable estimates if the pair (A, C) is observable and $A - KC$ is stable. The nonlinear observer form (3.20) is also extensible to the time-varying and state-affine cases via a Kalman-like approach, *i.e.* the aforementioned gains (3.18) and (3.19). Notice that the latter solution is uniformly globally exponentially stable if the pair (A, C) is uniformly observable.

Luenberguer-like observers

Although the literature provides several filtering linear and nonlinear techniques, most time-varying and nonlinear solutions must often compute Kalman-like gains to guarantee the stability estimates on uniformly observable systems. In specific situations, however, we can define filters with constant K and still obtain stable estimates. In order to do so, one must often resort to nonlinear innovation terms, instead of the aforementioned ones. We define Luenberguer-like observers as the estimators satisfying the latter assertion, and rigorous proofs demonstrate the stability properties of the resulting observer. These filters are written more often for specific applications using tools from nonlinear control theory.

3.1.6 *Observer definition for Lie groups*

The stability of observers, as stated on Definition 3.7, depends explicitly on the estimation error. Although it is quite standard (Luenberger, 1966) and intuitive to define the estimation error with the Euclidean difference (3.16), that error may not preserve some geometric properties of the manifold $\mathbb{X}$. For instance, the Euclidean difference does not affect the structure of $\mathbb{R}^n$, but that is generally not the case for Lie groups (Warner, 1987). Instead, an error $\widetilde{x}\colon \mathbb{X} \times \mathbb{X} \to \mathbb{X}$ for a Lie group $\mathbb{X}$ can be defined using three important characteristics: operation, existence of inverse and identity elements. More specifically, the error can be written such that $\widetilde{x}(x_1, x_2) = \boldsymbol{e}$ iff $x_1 = x_2$, with $\boldsymbol{e}$ denoting the identity element of the group.

The error structure defined in this section encompasses (3.16), since $\mathbb{R}^n$ is a Lie group under the operation of addition with identity $0_{1\times n}$ and the inverse element of x is $-x$. Indeed, the error (3.16) does not affect the structure of $\mathbb{R}^n$ because it is defined within its structure. A counter example, however, is given by the special orthogonal group $\mathbb{SO}(3)$, *i.e.* the Lie group that represents rotation matrices. Considering the state given by $R(t) \in \mathbb{SO}(3)$ and an estimate $\widehat{R} \in \mathbb{SO}(3)$, it is trivial to verify that $R - \widehat{R} \notin \mathbb{SO}(3)$. The special orthogonal $\mathbb{SO}(3)$ is a group under the operation of multiplication with identity I_3 and the inverse element of $R \in \mathbb{SO}(3)$ is R^T. There exist, however, two different ways of writing the estimation error:

$$\widetilde{R} = \widehat{R}^T R, \qquad \text{or} \qquad \widetilde{R} = R\widehat{R}^T. \tag{3.21}$$

The errors provided by the above definition are not necessarily equal due to the non-abelian characteristic of $\mathbb{SO}(3)$. However, the errors are invariant, *i.e.* they preserve their value after the simultaneous multiplication by a same element, on the left of R and $\widehat{R}$ respectively for the leftmost definition. This is easily verified defining the invariant transformation for $G \in \mathbb{SO}(3)$,

$$\begin{aligned} \varphi\colon \mathbb{SO}(3) \times \mathbb{SO}(3) &\to \mathbb{SO}(3), \\ (G, R) &\mapsto \varphi_G(R) = GR \end{aligned}$$

thus computing

$$\widetilde{R} = \big(\varphi_G(\widehat{R})\big)^T \varphi_G(R) = \widehat{R}^T G^T G R = \widehat{R}^T R.$$

It is trivial to verify that the rightmost error definition in (3.21) is invariant to the simultaneous multiplication by a same element on the right of R and $\widehat{R}$ respectively.

The nonlinear observer design in this thesis considers the definition for the estimation error such that $\widetilde{x} \in \mathbb{X}$. In order to define the exponential stability of the system, let us recall that there also exists a local parametrization $\psi\colon \mathbb{R}^n \to \mathbb{X}$, such that $\psi(\widetilde{\xi}) = \boldsymbol{e}_3$ iff $\widetilde{\xi} = 0_{n\times 1}$ with inverse ψ^{-1}. We can further generalize the definition of nonlinear observer stability.

Definition 3.8. An observer (3.15) of system (3.1) is asymptotically stable if:
- the estimation error $\widetilde{x} = \boldsymbol{e}$ is an equilibrium of $\dot{\widetilde{x}}(t)$;
- the equilibrium $\boldsymbol{e}$ is stable in the sense of Lyapunov;
- $\widetilde{x}(t) \to \boldsymbol{e}$ as $t \to \infty$ for $\widehat{x}(0) \in \mathbb{V} \subset \mathbb{X}$.

The observer is globally stable if $\mathbb{V} = \mathbb{X}$. Furthermore, the observer is locally exponentially stable with convergence rate $a > 0$ if there exists some $c_a > 0$ such that $|\widetilde{\xi}(t)| \leq c_a e^{-a(t-t_0)} |\widetilde{\xi}(t_0)|$ for $\widetilde{\xi} = \psi^{-1}(\widetilde{x})$.

3.1.7 *Decoupling the dynamics*

In some cases, it is possible to partially address the observability analysis of a nonlinear system using tools from both linear and nonlinear system theory. To illustrate this point, let us consider the following system:

$$\begin{cases} \dot{x}_1 = f_1(x_1, u_1, t), \\ \dot{x}_2 = A_2(x_1, u_1, t)x_2 + B_2(x_1, u_1, t)u_2, \\ y_1 = h_1(x_1, u_1, t), \\ y_2 = C_2(x_1, u_1, t)x_2. \end{cases}$$

with $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$ such that $n = n_1 + n_2$, f_1 and h_1 are vector functions and A_2, B_2, and C_2 matrix functions of proper dimensions. We can view this system as either a nonlinear

system with state $x = (x_1, x_2)$, or as the cascade of a nonlinear system (in x_1) interconnected with a linear time-varying system (in x_2), *i.e.*

$$\dot{x}_2 = A_2^*(t)x_2 + B_2^*(t)u_2, \qquad y_2 = C_2^*(t)x_2,$$

with $M^*(t) = M(x_1(t), u_1(t), t)$ for $M = \{A_2, B_2, C_2\}$. The latter interpretation will prevail in this thesis due to the possibility to use techniques from linear-systems to (part of) the system, that can provide stronger observability and stability properties.

NONLINEAR OBSERVERS FOR POSE ESTIMATION

So far, we have made preliminary recalls in this chapter. Let us proceed next to the design of observers for pose estimation. We recall that Propositions describe the results developed in this thesis, and the results from other works are labeled as Theorems. We present the proofs of every proposition in Appendix A. Furthermore, we divide the results in two sections, the prior refers to the estimation of the rotational dynamics as the latter refers to the estimation of the translational dynamics. The following assumption is made for the sake of observability and stability analysis of the nonlinear observers presented next.

Assumption 3.2. There exist four positive constants $\bar{c}_\omega$, $\bar{c}_{\dot{\omega}}$, $\bar{c}_{\ddot{\omega}}$ and $\bar{c}_a$, such that $\forall t \in [0, \infty)$: $|\omega_{\mathcal{B}}(t)| \leq \bar{c}_\omega$, $|\dot{\omega}_{\mathcal{B}}(t)| \leq \bar{c}_{\dot{\omega}}$, $|\ddot{\omega}_{\mathcal{B}}(t)| \leq \bar{c}_{\ddot{\omega}}$, and $|\dot{v}_{\mathcal{R}}(t)| \leq \bar{c}_a$.

Clearly, the above hypothesis is satisfied for physical systems.

3.2 ESTIMATION OF THE ROTATIONAL DYNAMICS

Let us start with the analysis of the rotational dynamics. We can define two cases: calibrated and uncalibrated frames. The first case present the systems defined in Sections 1.6.1 and 1.6.2, these systems present equivalent dynamics, hence the same nonlinear observer can be employed for the orientation and gyro bias estimation. The main characteristic of this class of system is that pose and angular velocity measurements are made with respect to the same frame. The second case represent the system defined in Section 1.6.3. This system considers that the angular velocity and orientation measurements are made in with respect to different frames with an unknown rotation, and we obtain that such class of system is observable under certain motion conditions. The designed nonlinear observer is stable if the observability conditions are satisfied. We also show that the calibrated case can be seen as a special solution of the system with uncalibrated frames.

3.2.1 *Calibrated frames*

We have that the orientation dynamics for the case with calibrated frame described in Sections 1.6.1 and 1.6.2 writes

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \\ \dot{b}_\omega = 0_{3\times 1}\,. \end{cases}$$

with measurements given by $(R_y, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}R_{\mathcal{B}}, {}^{\mathcal{B}}\omega + b_\omega)$. However, it is common to define a system by its states, *known*-inputs and outputs. Even though ${}^{\mathcal{R}}R_{\mathcal{B}}$ and ${}^{\mathcal{B}}\omega_y$ are driven by the actual angular velocity ${}^{\mathcal{B}}\omega$, we only *know* effectively the value of the ${}^{\mathcal{B}}\omega + b_\omega$ however. Thus, the original system is equivalent to

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - b_\omega)\,, \\ \dot{b}_\omega = 0_{3\times 1}\,. \end{cases} \tag{3.22}$$

with measurements $R_y = {}^{\mathcal{R}}R_{\mathcal{B}}$.

System (3.22) is invariant in the sense of (Martin et al., 2004) and (Bonnabel et al., 2008). Let us consider a new reference frame $\mathcal{R}_1$ and body frame $\mathcal{B}_1$, then the system is invariant with respect to changes in $\mathcal{R}$ and $\mathcal{B}$ frames to $\mathcal{R}_1$ and $\mathcal{B}_1$ and additive gyro bias b_{ω_0}. For instance, let the states $x = ({}^{\mathcal{R}}R_{\mathcal{B}}, b_\omega)$, inputs $u = {}^{\mathcal{B}}\omega_y$ and outputs $y = R_y$. Furthermore, let $f(x,u)$ the right hand side of (3.22) and $h(x,u) = {}^{\mathcal{R}}R_{\mathcal{B}}$, we can define the group $\mathbb{G} = \mathbb{SO}(3) \times \mathbb{SO}(3) \times \mathbb{R}^3$ with elements $G = ({}^{\mathcal{R}_1}R_{\mathcal{R}}, {}^{\mathcal{B}}R_{\mathcal{B}_1}, b_{\omega_0}) \in \mathbb{G}$, and the invariant actions

$$
\begin{aligned}
\varphi_G(x) &\triangleq \left({}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{B}_1},\ {}^{\mathcal{B}_1}R_{\mathcal{B}}b_\omega + b_{\omega_0}\right),\\
\psi_G(u) &\triangleq {}^{\mathcal{B}_1}R_{\mathcal{B}}\,{}^{\mathcal{B}}\omega_y + b_{\omega_0},\\
\rho_G(y) &\triangleq {}^{\mathcal{R}_1}R_{\mathcal{R}}R_y\,{}^{\mathcal{B}}R_{\mathcal{B}_1}
\end{aligned}
$$

where $\dot{\overbrace{\varphi_G(x)}} = f(\varphi_G(x), \psi_G(u))$, *i.e.*

$$
\begin{aligned}
\dot{\overbrace{{}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{B}_1}}} &= {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - b_\omega)\,{}^{\mathcal{B}}R_{\mathcal{B}_1}\\
&= {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{B}_1}\mathrm{S}\left(({}^{\mathcal{B}_1}R_{\mathcal{B}}\,{}^{\mathcal{B}}\omega + b_{\omega_0}) - ({}^{\mathcal{B}_1}R_{\mathcal{B}}b_\omega + b_{\omega_0})\right),\\
\dot{\overbrace{{}^{\mathcal{B}_1}R_{\mathcal{B}}b_\omega + b_{\omega_0}}} &= 0_{3\times 1},
\end{aligned}
$$

and $\rho_G(h(x,u)) = h(\varphi_G(x), \psi_G(u))$, *i.e.* ${}^{\mathcal{R}_1}R_{\mathcal{R}}R_y\,{}^{\mathcal{B}}R_{\mathcal{B}_1} = {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{B}_1}$.

Furthermore, System (3.22) with the proposed measurements is uniformly observable. We can thus obtain asymptotically stable nonlinear observers independently of the angular motion. Let us define the following nonlinear observer:

$$
\begin{cases}
{}^{\mathcal{R}}\dot{\widehat{R}}_{\mathcal{B}} = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega + \alpha_{R_{\mathcal{B}}}),\\
\dot{\widehat{b}}_\omega = \alpha_\omega .
\end{cases}
\tag{3.23}
$$

Notice that the observer (3.23) satisfies indeed Definition 3.6, leaving us the design of the innovation $\alpha_{R_{\mathcal{B}}}$ and α_ω to provide an asymptotically stable observer. Additionally, this observer form preserves the invariance properties of the original system.

In order to establish the innovation terms and stability properties of the nonlinear observer, let us recall that that there exists two forms to denote the error in $\mathbb{SO}(3)$, *c.f.* (3.21), which yield errors in $\mathcal{B}$ and $\mathcal{R}$ frames, respectively. One can obtain similar results with both error forms with slight modifications on the innovation terms. We proceed the analysis using the error definitions in the $\mathcal{R}$ frame, *i.e.*

$$
\widetilde{R} = {}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}, \qquad \widetilde{b}_\omega = b_\omega - \widehat{b}_\omega,
$$

which results in the following error dynamics

$$
\begin{cases}
\dot{\widetilde{R}} = -\widetilde{R}\mathrm{S}({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}(\widetilde{b}_\omega + \alpha_{R_{\mathcal{B}}})),\\
\dot{\widetilde{b}}_\omega = -\alpha_\omega .
\end{cases}
\tag{3.24}
$$

The objective of the nonlinear observer, according to Definition 3.8, is to define $\alpha_{R_{\mathcal{B}}}$ and α_ω so that the point $(\widetilde{R}, \widetilde{b}_\omega) = (I_3, 0_{3\times 1})$ defines an asymptotically stable equilibrium of the above dynamics. Even though we would be eager to define the innovations in order to obtain a globally stable equilibrium, there exists a topological obstruction on $\mathbb{SO}(3)$ that limits the every result computed using the group to semi-global stability (Sontag, 1998, p. 250-252). Several works in the literature have discussed solutions for attitude and gyro bias estimation with semi-global stability properties, and the passive complementary filter (Mahony et al., 2008) states the following result.

Theorem 3.3 (Passive complementary filter for $\mathbb{SO}(3)$)**.** Let

$$\begin{cases} \alpha_{R_{\mathcal{B}}} = k_{R_{\mathcal{B}}} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big), \\ \alpha_{\omega} = -k_{\omega} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big). \end{cases} \tag{3.25}$$

with $k_{R_{\mathcal{B}}}, k_{\omega} > 0$. The following statements hold for the error dynamics (3.24):

- Every solution converges to $\mathbb{E}_s \cup \mathbb{E}_u$, with $\mathbb{E}_s = (I_3, 0_{3\times 1})$, and $\mathbb{E}_u = \{(\widetilde{R}, \widetilde{b}_{\omega}) \in \mathbb{SO}(3) \times \mathbb{R}^3 \,|\, \mathrm{tr}(\widetilde{R}) = -1\}$.
- The equilibrium point $(\widetilde{R}, \widetilde{b}_{\omega}) = (I_3, 0_{3\times 1})$ is locally exponentially stable.
- The elements of the equilibrium set $\mathbb{E}_u^* = \{(\widetilde{R}, \widetilde{b}_{\omega}) \in \mathbb{SO}(3) \times \mathbb{R}^3 \colon \mathrm{tr}(\widetilde{R}) = -1,\ \widetilde{b}_{\omega} = 0_{3\times 1}\} \subset \mathbb{E}_u$ are unstable.
- The domain of convergence of $\mathbb{E}_s$ increases as $k_2 \to \infty$.

Theorem 3.3 yields a nonlinear observer invariant to coordinate changes in the $\mathcal{B}$ frame. This observer has an interesting semi-global convergence property, and *almost* every solution converges to the desired equilibrium. In order to increase the domain domain of attraction, however, one must increase the gain k_{ω}, as, for instance, shown in (Vasconcelos et al., 2008b). The values of the gains are related to the dynamics of the estimates, and, as we show in Section 3.4, a larger value for k_{ω} implies a larger bandwidth for the bias estimate. Actually, this is not a good property for the filter dynamics, as we are interested in the lower frequencies of error, as we modeled the gyroscope bias by a constant. A simple change in the innovation terms provide stronger stability properties, as shown in the following proposition.

Proposition 3.1. Let

$$\begin{cases} \alpha_{R_{\mathcal{B}}} = k_{R_{\mathcal{B}}} \dfrac{{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big)}{\big(1 + \mathrm{tr}(\widetilde{R})\big)^2}, \\ \alpha_{\omega} = -k_{\omega} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big). \end{cases} \tag{3.26}$$

with $k_{R_{\mathcal{B}}}, k_{\omega} > 0$. Then, the following statements hold for the error dynamics (3.24):

- The equilibrium $(\widetilde{R}, \widetilde{b}_{\omega}) = (I_3, 0_{3\times 1})$ is locally exponentially stable.
- Every solution starting from $\big(\widetilde{R}(0), \widetilde{b}_{\omega}(0)\big) \notin \mathbb{E}_u$ converges to $\mathbb{E}_s$, with $\mathbb{E}_s = (I_3, 0_{3\times 1})$, and $\mathbb{E}_u = \{(\widetilde{R}, \widetilde{b}_{\omega}) \in \mathbb{SO}(3) \times \mathbb{R}^3 \,|\, \mathrm{tr}(\widetilde{R}) = -1\}$.

The proof of this proposition is discussed in Appendix A.4.1. Proposition 3.1 yields a nonlinear observer invariant to coordinate changes in the $\mathcal{B}$ frame with stability domain independent of the innovation gains. This result is inspired by Theorem 3.3 and almost-global stabilizers for mobile robot trajectories, *c.f.* , *e.g.* , (Morin and Samson, 2008). Remark that, locally, the resulting error dynamics by Theorem 3.3 and Proposition 3.1 are similar. The proposed nonlinear observer, however, provides stronger stability properties with known domain of attraction. Furthermore we can tune the gains of the nonlinear observer in order to obtain a low bandwidth for the bias dynamics without reducing the claims on the domain of convergence.

The improvement on the stability property has a counterpart, because the innovation $\alpha_{R_{\mathcal{B}}}$ is singular for every point from $\mathbb{E}_s$. Although that singularity is more likely to occur theoretically than in practical situations, it is still important to ensure that we do not divide by zero in $\alpha_{R_{\mathcal{B}}}$. This protection can be done multiplying the innovation term by a function that shows the same properties as the denominator at the bad set, *i.e.* let

$$\mu(\varepsilon, R) = \begin{cases} \dfrac{\big(1 + \mathrm{tr}(R)\big)^2}{\varepsilon^2}, & \text{if } |\mathrm{tr}(R) + 1| < \varepsilon, \\ 1, & \text{otherwise,} \end{cases}$$

thus $\alpha_{R_\mathcal{B}} \triangleq k_{R_\mathcal{B}} \mu(\varepsilon, \widetilde{R}) \frac{{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \operatorname{vex}\left(\mathrm{P_a}(\widetilde{R})\right)}{\left(1+\operatorname{tr}(\widetilde{R})\right)^2}$ for $\varepsilon > 0$ very small in Proposition 3.1. In that case, we have that the equilibrium set $\mathbb{E}_{\mathrm{u}}^*$ is unstable, and the domain of convergence is enlarged as $\varepsilon \to 0$.

These nonlinear observers rely in a full reconstructed orientation matrix. Some systems, however, provide measurements of vectors in the $\mathcal{B}$ frame with known coordinates in the $\mathcal{R}$ frame. For these cases, a straightforward solution can compute the orientation matrix computed via an intermediate step using the TRIAD or QUEST methods, *c.f.* Section 1.5, and afterwards apply the afore-discussed techniques. Nevertheless, we can avoid this intermediate step by changing the innovation terms to consider directly vector measurements. (Vasconcelos et al., 2008b) present the following nonlinear observer that yields similar stability properties as Theorem 3.3.

Corollary 3.2. Consider $N > 3$ vectors with known-coordinates ${}^{\mathcal{R}}\beta_i$ in the $\mathcal{R}$ frame, with vector associated to a gain $k_{\beta_i} > 0$ such that $V = \sum_{i=1}^{N} k_{\beta_i} {}^{\mathcal{R}}\beta_i {}^{\mathcal{R}}\beta_i{}^T$ is nonsingular. Assume that we measure these vectors as ${}^{\mathcal{B}}\beta_i$ in $\mathcal{B}$ coordinates. Let:

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \sum_{i=1}^{N} k_i \left((V^{-1}){}^{\mathcal{R}}\beta_i\right) \times ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} {}^{\mathcal{B}}\beta_i), \\ \alpha_{\omega} = -k_{\omega} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \sum_{i=1}^{N} k_i \left((V^{-1}){}^{\mathcal{R}}\beta_i\right) \times ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} {}^{\mathcal{B}}\beta_i), \end{cases} \tag{3.27}$$

with $k_{R_\mathcal{B}}, k_\omega > 0$. The error dynamics (3.24) is equivalent to Theorem 3.3.

The observer provided by this Corollary is also invariant to changes of the $\mathcal{B}$ frame likewise Theorem 3.3. Notice that the first hypothesis requires at least three vectors as measurements. At a first glance, this requirement can sound too restrictive, because the system comprising orientation and gyro bias is observable using only *two* non-parallel vectors. However, we can use the previous observer by considering a pseudo-measurement ${}^{\mathcal{B}}\beta_3 = {}^{\mathcal{B}}\beta_1 \times {}^{\mathcal{B}}\beta_2$ of the known vector ${}^{\mathcal{R}}\beta_3 = {}^{\mathcal{R}}\beta_1 \times {}^{\mathcal{R}}\beta_2$. In a same manner that Proposition 3.2 extends the stability properties of Theorem 3.3, we can define a new nonlinear observer that extends the stability of Corollary 3.2.

Corollary 3.3. Consider $N > 3$ vectors with known-coordinates ${}^{\mathcal{R}}\beta_i$ in the $\mathcal{R}$ frame, with vector associated to a gain $k_{\beta_i} > 0$ such that $V = \sum_{i=1}^{N} k_{\beta_i} {}^{\mathcal{R}}\beta_i {}^{\mathcal{R}}\beta_i{}^T$ is nonsingular. Assume that we measure these vectors as ${}^{\mathcal{B}}\beta_i$ in $\mathcal{B}$ coordinates. Let:

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} \dfrac{{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \sum_{i=1}^{N} k_i ((V^{-1}){}^{\mathcal{R}}\beta_i) \times ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} {}^{\mathcal{B}}\beta_i)}{\left(1 + \sum_{i=1}^{N} k_{\beta_i} ((V^{-1}){}^{\mathcal{R}}\beta_i)^T ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} {}^{\mathcal{B}}\beta_i)\right)^2}, \\ \alpha_{\omega} = -k_{\omega} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \sum_{i=1}^{N} k_i ((V^{-1}){}^{\mathcal{R}}\beta_i) \times ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} {}^{\mathcal{B}}\beta_i), \end{cases} \tag{3.28}$$

with $k_{R_\mathcal{B}}, k_\omega > 0$. The error dynamics (3.24) is equivalent to Proposition 3.1.

The proof of this Corollary is presented in Appendix A.4.2 These are interesting solutions to avoid the explicit reconstruction of the orientation matrix, however these implementation present two main problems. The first problem refers to the initial three-vector hypothesis. Even though we can easily avoid the singularity of the auxiliary matrix V by including a pseudo-measurement, in practice, that pseudo-measurement is noisier than the two original vectors. The literature provides other nonlinear observers that do not consider the explicit computation of the attitude nor consider this hypothesis, *c.f.* (Mahony et al., 2008, Theorem 5.1), (Martin and Salaun, 2008) and (Hua et al., 2013). The second issue arrives due to the coupling of the directions measured by the vectors. This problem is particularly bad when we have one accurate and one inaccurate measurement that refer to non-orthogonal directions. In this case, the measurements provide overlapping information about the same direction, and the information provided by the good measurement can be impaired if the other measurement is strongly corrupted by noise or external direction.

3.2.2 *Uncalibrated frames*

The previous section considered the case where the angular velocity and orientation measurements are made with respect to the same $\mathcal{B}$ frame. This section considers the problem when the measurements are made from two different frames, *i.e.* the angular velocity is measured in IMU $\mathcal{B}$-frame and the orientation in the camera $\mathcal{C}$-frame, which refers to the configuration described in Section 1.6.3. The techniques discussed in the previous section can be employed then frame-to-frame rotation matrix, *e.g.* c-to-IMU rotation ${}^{\mathcal{B}}R_{\mathcal{C}}$, is negligible or previously known. However, it is fundamental to have a good estimation of ${}^{\mathcal{R}}R_{\mathcal{B}}$ when these conditions are not satisfied. We propose hereafter a solution for this problem.

We can write the rotational dynamics for the system described in Section 1.31 as

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\,, \\ \dot{b}_{\omega} = 0_{3\times 1}\,, \\ {}^{\mathcal{B}}\dot{R}_{\mathcal{C}} = 0_{3\times 3}\,. \end{cases}$$

with measurements given by $(R_y, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}R_{\mathcal{C}}, {}^{\mathcal{B}}\omega + b_{\omega}) = ({}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}, {}^{\mathcal{B}}\omega + b_{\omega})$. Moreover, due to the same arguments from Section 3.2.1, we rewrite the above system as

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - b_{\omega})\,, \\ \dot{b}_{\omega} = 0_{3\times 1}\,, \\ {}^{\mathcal{B}}\dot{R}_{\mathcal{C}} = 0_{3\times 3}\,. \end{cases} \tag{3.29}$$

with measurements $R_y = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}$.

The invariance properties (3.29) system are obtained analogously to (3.22), *i.e.* considering a new reference frame $\mathcal{R}_1$, body frame $\mathcal{B}_1$, and camera frame $\mathcal{C}_1$, then the system is invariant with respect to changes in $\mathcal{R}$, $\mathcal{B}$ and $\mathcal{C}$ frames to $\mathcal{R}_1$, $\mathcal{B}_1$ and $\mathcal{C}_1$, and an additive gyro bias b_{ω_0}. For instance, let the states $x = ({}^{\mathcal{R}}R_{\mathcal{B}}, b_{\omega}, {}^{\mathcal{B}}R_{\mathcal{C}})$, inputs $u = {}^{\mathcal{B}}\omega_y$, outputs $y = R_y$. Furthermore, let $f(x,u)$ the right hand side of (3.29), and $h(x,u) = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{R}}R_{\mathcal{C}}$, then we can define the group $\mathbb{G} = \mathbb{SO}(3) \times \mathbb{SO}(3) \times \mathbb{SO}(3) \times \mathbb{R}^3$ with elements $G = ({}^{\mathcal{R}_1}R_{\mathcal{R}}, {}^{\mathcal{B}}R_{\mathcal{B}_1}, {}^{\mathcal{C}}R_{\mathcal{C}_1}, b_{\omega_0})$, and the invariant actions

$$\begin{aligned} \varphi_G(x) &\triangleq \left({}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1},\ {}^{\mathcal{B}_1}R_{\mathcal{B}}b_{\omega} + b_{\omega_0},\ {}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}{}^{\mathcal{C}}R_{\mathcal{C}_1}\right), \\ \psi_G(u) &\triangleq {}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}\omega_y + b_{\omega_0}, \\ \rho_G(y) &\triangleq {}^{\mathcal{R}_1}R_{\mathcal{R}}R_y{}^{\mathcal{C}}R_{\mathcal{C}_1}\,. \end{aligned}$$

where $\dot{\overbrace{\varphi_G(x)}} = f(\varphi_G(x), \psi_G(u))$, *i.e.*

$$\begin{aligned} \dot{\overbrace{{}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1}}} &= {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - b_{\omega}){}^{\mathcal{B}}R_{\mathcal{B}_1} \\ &= {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1}\mathrm{S}\big(({}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}\omega_y + b_{\omega_0}) - ({}^{\mathcal{B}_1}R_{\mathcal{B}}b_{\omega} + b_{\omega_0})\big)\,, \\ \dot{\overbrace{{}^{\mathcal{B}_1}R_{\mathcal{B}}b_{\omega} + b_{\omega_0}}} &= 0_{3\times 1}\,, \\ \dot{\overbrace{{}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}{}^{\mathcal{C}}R_{\mathcal{C}_1}}} &= 0_{3\times 3}\,. \end{aligned}$$

and $\rho_G(h(x,y)) = h(\varphi_G(x), \psi_G(u))$, *i.e.*

$${}^{\mathcal{R}_1}R_{\mathcal{R}}R_y{}^{\mathcal{C}}R_{\mathcal{C}_1} = {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}{}^{\mathcal{C}}R_{\mathcal{C}_1} = ({}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1})({}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}{}^{\mathcal{C}}R_{\mathcal{C}_1}).$$

The observability properties of System (3.24) and (3.29) differ, because the latter class is observable under certain motion conditions. The following result refers to a condition under which the System (3.29) is observable.

Proposition 3.2. Let the angular velocity represented by a function ${}^{\mathcal{B}}\omega(t)$ that satisfies Assumption 3.2, and

$$\forall\, t \geq 0, \qquad |{}^{\mathcal{B}}\dot{\omega}(t) \times {}^{\mathcal{B}}\ddot{\omega}(t)|^2 > 0\,. \tag{3.30}$$

Then, System (3.29) represented by the states $x{=}({}^{\mathcal{R}}R_{\mathcal{B}},\, b_\omega,\, {}^{\mathcal{B}}R_{\mathcal{C}})$, input $u = {}^{\mathcal{B}}\omega(t)$ and measurements $y = (R_y,\, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}\ {}^{\mathcal{B}}\omega + b_\omega)$ is instantaneously observable $\forall t \geq 0$. [1]

The proof of this proposition is given in Appendix A.3.1. The above proposition describes a universal input of system (3.29), which, in words, claims that the system is observable if the angular acceleration is not parallel to the angular jerk.

The observability of c-to-IMU calibration was already analyzed in the Literature, *e.g.* , see (Jones et al., 2007), (Mirzaei and Roumeliotis, 2008) and (Kelly and Sukhatme, 2011). Although the three previous results arrive to the same conclusion, *i.e.* the system is observable, their interpretation of the observability conditions is obscure. For instance, (Mirzaei and Roumeliotis, 2008) and (Kelly and Sukhatme, 2011) claim that the system is observable if the body *undergoes rotation along two axis*. The observability condition of the system, however, is related to the trajectory of the angular velocity instead of its instantaneous value. Moreover, (Jones et al., 2007) claims that the system is observable if the motion of ${}^{\mathcal{B}}\omega$ *spans the entirety of the state space*. Notice that this latter condition is also sufficient, it is far from necessary, as comparing to the expression presented in Proposition 3.2.

Let us continue to the estimation design, defining the nonlinear observer scheme:

$$\begin{cases} {}^{\mathcal{R}}\dot{\widehat{R}}_{\mathcal{B}} = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega + \alpha_{R_{\mathcal{B}}})\,, \\ \dot{\widehat{b}}_\omega = \alpha_\omega\,, \\ {}^{\mathcal{B}}\dot{\widehat{R}}_{\mathcal{C}} = {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}\mathrm{S}(\alpha_{R_{\mathcal{C}}})\,. \end{cases} \tag{3.31}$$

We can continue the design by defining innovation terms $\alpha_{R_{\mathcal{B}}}$, α_ω and $\alpha_{R_{\mathcal{C}}}$ that provide an asymptotically stable observer.

Likewise the case with calibrated-frames, there exist two possible forms to define the error in $\mathrm{SO}(3)$. We can denote for the orientation error errors in $\mathcal{B}$ and $\mathcal{R}$ frames, and c-to-IMU rotation in $\mathcal{C}$ and $\mathcal{B}$ frames. We proceed the analysis using the errors in $\mathcal{R}$ frame and $\mathcal{B}$ frame for the attitude and c-to-IMU rotation respectively, *i.e.*

$$\widetilde{R} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\,, \quad \widetilde{b}_\omega = b_\omega - \widehat{b}_\omega\,, \quad \widetilde{Q} = {}^{\mathcal{B}}R_{\mathcal{C}}{}^{\mathcal{C}}\widehat{R}_{\mathcal{B}}\,.$$

which yields the following error dynamics

$$\begin{cases} \dot{\widetilde{R}} \;=\; -\,\widetilde{R}\mathrm{S}\big({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}(\widetilde{b}_\omega + \alpha_{R_{\mathcal{B}}})\big)\,, \\ \dot{\widetilde{b}}_\omega \;=\; -\,\alpha_\omega\,, \\ \dot{\widetilde{Q}} \;=\; -\,\widetilde{Q}\mathrm{S}({}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}\,\alpha_{R_{\mathcal{C}}})\,. \end{cases} \tag{3.32}$$

The objective, according to Definition 3.8, is to define the innovation terms $\alpha_{R_{\mathcal{B}}}$, α_ω and $\alpha_{R_{\mathcal{C}}}$ so that the point $(\widetilde{R},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ defines an asymptotically stable equilibrium of dynamics (3.32).

Let ${}^{\mathcal{R}}\widehat{R}_{\mathcal{C}} = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}$ denote the estimate of ${}^{\mathcal{R}}\widehat{R}_{\mathcal{C}}$, as deduced from the estimates ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}$ and ${}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}$, and $\widetilde{R}_{\mathcal{C}} = {}^{\mathcal{R}}R_{\mathcal{C}}{}^{\mathcal{C}}\widehat{R}_{\mathcal{R}}$. The following result states a new observer for orientation, gyro bias and c-to-IMU rotation estimation, and it is originally introduced in (Scandaroli et al., 2011).

1. Notice that even though Definition 3.2 concerns the instantaneous observability linear systems, we can extend it directly to nonlinear systems.

Proposition 3.3. Let

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_\mathcal{C})\big) - {}^{\mathcal{B}}\widehat{R}_\mathcal{C} \alpha_{R_\mathcal{C}}, \\ \alpha_\omega \;\; = \; - k_\omega {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_\mathcal{C})\big), \\ \alpha_{R_\mathcal{C}} = k_{R_\mathcal{C}} {}^{\mathcal{C}}R_{\mathcal{R}} \mathrm{P_a}(\widetilde{R}_\mathcal{C}) {}^{\mathcal{R}}\widehat{R}_\mathcal{B} ({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega), \end{cases} \tag{3.33}$$

with $k_{R_\mathcal{B}}, k_\omega, k_\mathcal{C} > 0$. Suppose that Assumption 3.2 holds and that the following condition is satisfied

$$\exists\, \tau, \delta > 0 : \; \forall\, t \geq 0, \quad \int_t^{t+\tau} |{}^{\mathcal{B}}\dot{\omega}(s) \times {}^{\mathcal{B}}\ddot{\omega}(s)|^2 ds > \delta. \tag{3.34}$$

Then, $(\widetilde{R}, \widetilde{b}_\omega, \widetilde{Q}) = (I_3, 0_{3\times 1}, I_3)$ is locally exponentially stable equilibrium point of the error dynamics (3.32).

The proof of this Proposition is discussed in Appendix A.4.3. Notice that the resulting nonlinear observer preserves the invariance properties of the original system. Moreover, relation (3.34) is a "persistent excitation" condition related to the observability properties of the system. Indeed, as stated in Proposition 3.2, System (3.32) with $({}^{\mathcal{R}}R_\mathcal{C}, {}^{\mathcal{B}}\omega + b_\omega)$ as measurements is not observable for every input ${}^{\mathcal{B}}\omega$. Trivially, it is not observable when ${}^{\mathcal{B}}\omega(t) = 0, \; \forall\, t \geq 0$. Furthermore, the relation can be can be seen as a uniform extension of Proposition 3.2, since the instantaneous observability does not take time intervals into account.

The proposed observer can be viewed as an extension of the passive complementary filter on $\mathbb{SO}(3)$. More precisely, setting $k_{R_\mathcal{C}} = 0$ in (3.33) and assuming that ${}^{\mathcal{B}}R_\mathcal{C} = {}^{\mathcal{B}}\widehat{R}_\mathcal{C} = I_3$, the observer reduces to an attitude and rate gyro bias estimator. In this special case, it has been shown in Theorem 3.3 that this estimator is semi-globally exponentially stable, independently of ${}^{\mathcal{B}}\omega$. Despite the fact that semi-global exponential stability seems more difficult to prove for this observer than the seminal case, we verify via simulation results that the domain of attraction is also large. Additionally, this nonlinear observer could be extended using the same procedure as Proposition 3.1, however, the local properties obtained by the proof are unchanged.

In practice, Condition (3.34) will not always be satisfied and one must be careful with the implementation of the proposed observer. A possibility consists in first using the full observer in a preliminary calibration step with persistent motion, thus obtaining a good estimate for ${}^{\mathcal{B}}R_\mathcal{C}$, and then setting $k_{R_\mathcal{C}}$ in order to use, as explained above, the observer as an attitude and gyro bias estimator. A second possibility consists in using the observability–stability condition (3.34) so as to tune the gain $k_{R_\mathcal{C}}$ in function of the level of "motion excitation". Basically, this gain associated with the estimation of the c-to-IMU rotation should be non-zero only when the quantity $|{}^{\mathcal{B}}\dot{\omega}(s) \times {}^{\mathcal{B}}\ddot{\omega}(s)|^2$ is significantly larger than zero, so as to avoid possible drift of ${}^{\mathcal{B}}\widehat{R}_\mathcal{C}$ in case of weak motion excitation due to the integration of measurement noise instead of the actual error. Although that quantity is not directly measured, it can be estimated from angular rate gyro measurements ω_y. This procedure is discussed with more details in Section 4.3.1.

3.3 ESTIMATION OF THE TRANSLATIONAL DYNAMICS

We continue with the analysis for the translational dynamics of the pose estimation. The estimation part respective to the translational dynamics can be divided in three cases: calibrated frames, unknown gravitational acceleration field and uncalibrated case. The first case is defined in Section 1.6.1, and it represents the situation where both acceleration and position measurements are made with respect to the same frame, and the actual value of gravitational

acceleration field in $\mathcal{R}$ coordinates is known. The second case is defined in Section 1.6.2 as an extension of the previous one. The second class of systems considers that the position and acceleration measurements are made with respect to the same frame, however, the actual value of the gravitational acceleration field in $\mathcal{R}$ coordinates is unknown. The third and last system is defined in Section 1.6.3, this class relaxes the hypothesis on the frames of position and acceleration measurements, while also it estimates the acceleration due to local gravity. The two previous configurations are observable under certain motion conditions, and the obtained nonlinear observers are stable if the observability conditions are satisfied.

Notice that, without loss of generality, we can consider that the problem of orientation estimation is already solved independently of the translational dynamics. This hypothesis much simplifies the following analysis, as we can assume that both the orientation ${}^{\mathcal{R}}R_{\mathcal{B}}$ and angular velocity ${}^{\mathcal{B}}\omega$ are directly measured. Notice that the hypothesis on the knowledge of the frames in which measurements are taken is then reduced to the knowledge of the position in which the measurements are taken.

3.3.1 *Calibrated frames*

We can write the translational dynamics for the case with calibrated frames described in Section 1.6.1 as

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v \\ {}^{\mathcal{R}}\dot{v} \;\; = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a \\ \dot{b}_a \;\; = 0_{3\times 1} \end{cases}$$

with measurements $(p_y, R_y, {}^{\mathcal{B}}a_y, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}R_{\mathcal{B}}, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g, {}^{\mathcal{B}}\omega)$.

This system is driven by the linear acceleration of the body. However, the accelerometers measure the specific linear acceleration (sum of body linear acceleration and the gravitational acceleration) with an additive bias. We thus suppose that ${}^{\mathcal{B}}a_y$ is a known input of the system, and in this section exclusively, ${}^{\mathcal{R}}g$ is known. We can also estimate the orientation dynamics using the nonlinear observers from Section 3.2, and we can consider initially ${}^{\mathcal{R}}R_{\mathcal{B}}$ as another known-input. The original system can be rewritten as

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} \;\; = {}^{\mathcal{R}}R_{\mathcal{B}}({}^{\mathcal{B}}a_y - b_a) + {}^{\mathcal{R}}g\,, \\ \dot{b}_a \;\; = 0_{3\times 1}\,, \end{cases} \tag{3.35}$$

with measurements $p_y = {}^{\mathcal{R}}p_{\mathcal{B}}$.

System (3.35) is invariant in the sense of (Martin et al., 2004) and (Bonnabel et al., 2008). Let us consider a new reference frame $\mathcal{R}_1$ and body frame $\mathcal{B}_1$, then the system is invariant with respect to changes of translation and rotation from the $\mathcal{R}$ frame to $\mathcal{R}_1$, rotations from the $\mathcal{B}$ frame to $\mathcal{B}_1$ and additive accelerometer biases b_{a_0}. For instance, recall the equations for frame changes (1.3) and define the states $x = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}v, b_a)$, inputs $u = ({}^{\mathcal{B}}a_y, {}^{\mathcal{R}}R_{\mathcal{B}}, {}^{\mathcal{R}}g)$, output $y = p_y$. Furthermore, let $f(x,u)$ the right hand side of (3.35) and $h(x,u) = {}^{\mathcal{R}}p_{\mathcal{B}}$, we can define the group $\mathbb{G} = \mathbb{R}^3 \times \mathbb{SO}(3) \times \mathbb{SO}(3) \times \mathbb{R}^3$ with elements $G = ({}^{\mathcal{R}_1}p_{\mathcal{R}}, {}^{\mathcal{R}_1}R_{\mathcal{R}}, {}^{\mathcal{B}}R_{\mathcal{B}_1}, b_{a_0}) \in \mathbb{G}$, and the invariant actions

$$\begin{aligned} \varphi_G(x) &\triangleq \left({}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}_1}p_{\mathcal{R}},\ {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}v,\ {}^{\mathcal{B}_1}R_{\mathcal{B}}b_a + b_{a_0}\right), \\ \psi_G(u) &\triangleq \left({}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}a_y + b_{a_0},\ {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1},\ {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}g\right), \\ \rho_G(y) &\triangleq {}^{\mathcal{R}_1}R_{\mathcal{R}}p_y + {}^{\mathcal{R}_1}p_{\mathcal{R}} \end{aligned}$$

where $\dot{\overbrace{\varphi_G(x)}} = f(\varphi_G(x), \psi_G(u))$, *i.e.*

$$\begin{aligned}
\dot{\overbrace{{}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}_1}p_{\mathcal{R}}}} &= {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}v\,,\\
\dot{\overbrace{{}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}v}} &= {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}({}^{\mathcal{B}}a_y - b_a) + {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}g\\
&= {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}R_{\mathcal{B}}\,{}^{\mathcal{B}}R_{\mathcal{B}_1}\big(({}^{\mathcal{B}_1}R_{\mathcal{B}}\,{}^{\mathcal{B}}a_y + b_{a_0}) - ({}^{\mathcal{B}_1}R_{\mathcal{B}}b_a + b_{a_0})\big) + {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}g,\\
\dot{\overbrace{{}^{\mathcal{B}_1}R_{\mathcal{B}}b_a + b_{a_0}}} &= 0_{3\times 1}\,,
\end{aligned}$$

and $\rho_G(h(x,u)) = h(\varphi_G(x), \psi_G(u))$, *i.e.*

$${}^{\mathcal{R}_1}R_{\mathcal{R}}p_y + {}^{\mathcal{R}_1}p_{\mathcal{R}} = {}^{\mathcal{R}_1}R_{\mathcal{R}}\,{}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}_1}p_{\mathcal{R}}.$$

Notice that the dynamics is not invariant to a translation in the $\mathcal{B}$ frame using the proposed configuration.

It is straightforward to verify that this system is uniformly observable and the observability does not depend on body motion. We define the following nonlinear observer:

$$\begin{cases}
{}^{\mathcal{R}}\dot{\widehat{p}}_{\mathcal{B}} = {}^{\mathcal{R}}\widehat{v} + \alpha_{p_{\mathcal{B}}}\,,\\
{}^{\mathcal{R}}\dot{\widehat{v}} \;= {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}a_y - \widehat{b}_a) + {}^{\mathcal{R}}g + \alpha_v\,,\\
\dot{\widehat{b}}_a \;= \alpha_a\,,
\end{cases} \tag{3.36}$$

The goal is to design the innovation terms $\alpha_{p_{\mathcal{B}}}$, α_v and α_a to obtain an asymptotically stable observer, and we define the estimation errors as

$$\widetilde{p} = {}^{\mathcal{R}}p_{\mathcal{B}} - {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}\,, \qquad \widetilde{v} = {}^{\mathcal{R}}v - {}^{\mathcal{R}}\widehat{v}\,, \qquad \widetilde{b}_a = b_a - \widehat{b}_a$$

that yield the following error dynamics:

$$\begin{cases}
\dot{\widetilde{p}} = \widetilde{v} - \alpha_{p_{\mathcal{B}}}\,,\\
\dot{\widetilde{v}} = \; - {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_a - (I_3 - \widetilde{R})\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}({}^{\mathcal{B}}a_y - \widehat{b}_a) - \alpha_v\,,\\
\dot{\widetilde{b}}_a = \; - \alpha_a\,.
\end{cases} \tag{3.37}$$

More specifically, the objective of the observer is to design $\alpha_{p_{\mathcal{B}}}$, α_v and α_a so that the origin $(0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is an asymptotically stable equilibrium of the above dynamics. The variables considered in this section do not present topological problems, therefore we can provide conditions for which the proposed filter has global exponential stability.

Initially, let us consider to measure ${}^{\mathcal{R}}R_{\mathcal{B}}$ and ${}^{\mathcal{B}}\omega$ explicitly. As a consequence, we can consider the dynamics of the nonlinear observer (3.36) with ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}$. The following result provides an observer for position and linear velocity of the body in $\mathcal{R}$ and accelerometer bias. This result is originally introduced in (Scandaroli and Morin, 2011).

Proposition 3.4. Let

$$\begin{cases}
\alpha_{p_{\mathcal{B}}} = k_{p_{\mathcal{B}}}\widetilde{p}\,,\\
\alpha_v = k_v\widetilde{p}\,,\\
\alpha_a = \; - k_a\big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega)\big)\,{}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{p}\,,
\end{cases} \tag{3.38}$$

with $k_{p_{\mathcal{B}}}$, k_v, $k_a > 0$ such that $k_a < k_{p_{\mathcal{B}}}k_v$. Then, $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a) = (0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is a globally exponentially stable equilibrium point of the estimation error dynamics (3.37).

The proof for this proposition is stated in Appendix A.4.4 Remark that the resulting nonlinear observer satisfies Definition 3.6 when ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}$, and also preserves the invariance properties of the original system. This nonlinear observer provides a globally exponentially stable estimator for position, velocity and accelerometer bias, independently of the rotational dynamics. Note that we must satisfy a stability condition on the parameters relating to the gains. However, this condition is due to the order of the system and is not induced by the observer. More specifically, let us consider the case when ${}^{\mathcal{B}}\omega = 0$, then the dynamics of the estimation error is linear and autonomous. Remark that the gain conditions presented by proposition 3.4 correspond exactly to the stability conditions of the linear autonomous system.

It is implicitly assumed in (3.38) that ${}^{\mathcal{B}}\omega$ is available as a measurement. In practice, this term should be replaced by ${}^{\mathcal{B}}\omega_y - \widehat{b}_\omega$, with $\widehat{b}_\omega$ being the output of an attitude observer from Section 3.2. The next result, easily derived from Proposition 3.4, shows that this can be done without further consequences on the stability of the observer.

Corollary 3.4. Let

$$\begin{cases} \alpha_{p_{\mathcal{B}}} = k_{p_{\mathcal{B}}}\widetilde{p}\,, \\ \alpha_v = k_v\widetilde{p}\,, \\ \alpha_a = -k_a\big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)\big){}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{p}\,, \end{cases} \tag{3.39}$$

with $k_{p_{\mathcal{B}}}, k_v, k_a > 0$ such that $k_a < k_{p_{\mathcal{B}}}k_v$. If $\widetilde{b}_\omega$ converges asymptotically to zero, then $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a)$ converges asymptotically to $(0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ along the solutions of the error dynamics (3.37).

The proof for this Corollary is given in Appendix A.4.5. Finally, the measured rotation matrix hypothesis that ${}^{\mathcal{R}}R_{\mathcal{B}}$ is explicitly measured can be replaced by an estimate ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}$. However, the global asymptotic convergence cannot be achieved anymore, since observers on $\mathbb{SO}(3)$ are not globally asymptotically stable.

Corollary 3.5. Let

$$\begin{cases} \alpha_{p_{\mathcal{B}}} = k_{p_{\mathcal{B}}}\widetilde{p}\,, \\ \alpha_v = k_v\widetilde{p}\,, \\ \alpha_a = -k_a\big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)\big){}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\widetilde{p}\,, \end{cases} \tag{3.40}$$

with $k_{p_{\mathcal{B}}}, k_v, k_a > 0$ such that $k_a < k_{p_{\mathcal{B}}}k_v$. If $\widetilde{R}$ and $\widetilde{b}_\omega$ converge asymptotically to I_3 and zero, then $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a)$ converges asymptotically to $(0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ along the solutions of the error dynamics (3.37).

The proof for this Corollary is very similar to Corollary 3.4 and therefore omitted.

3.3.2 *Estimation of the gravitational field*

The previous section discussed the case where the gravitational acceleration ${}^{\mathcal{R}}g$ was known. That case represent, for instance, applications of visual tracking where the pose of the reference image is known with respect to the gravity. However, that assumption is not satisfied very often. This section considers the situations where the gravitational acceleration ${}^{\mathcal{R}}g$ is unknown. The dynamics for the case with unknown gravitational field discussed in Section 1.6.2 is given by

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \\ \dot{b}_a = 0_{3\times1}\,, \\ {}^{\mathcal{R}}\dot{g} = 0_{3\times1}\,. \end{cases}$$

with measurements $(p_y, R_y, {}^{\mathcal{B}}a_y, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}R_{\mathcal{B}}, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g, {}^{\mathcal{B}}\omega)$.

Likewise the previous section, the linear acceleration is not available directly and we must incorporate the accelerometer measurements as a known-input of the system. We thus assume ${}^{\mathcal{B}}a_y$, ${}^{\mathcal{B}}\omega$, ${}^{\mathcal{R}}R_{\mathcal{B}}$ as inputs. However, we estimate the variable ${}^{\mathcal{R}}g$ in this section additionally to position, linear velocity and accelerometer bias. The system can be rewritten as

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} \;\; = {}^{\mathcal{R}}R_{\mathcal{B}}({}^{\mathcal{B}}a_y - b_a) + {}^{\mathcal{R}}g\,, \\ \dot{b}_a \;\; = 0_{3\times 1}\,, \\ {}^{\mathcal{R}}\dot{g} \;\; = 0_{3\times 1}\,. \end{cases} \tag{3.41}$$

with measurements $y = {}^{\mathcal{R}}p_{\mathcal{B}}$.

System (3.41) is also invariant with respect to changes of translation and rotation from the $\mathcal{R}$ frame to $\mathcal{R}_1$, rotation from the $\mathcal{B}$ frame to $\mathcal{B}_1$ and additive accelerometer biases b_{a_0}. This analysis is very similar to the previous Section, defining however the states $x = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}v, b_a, {}^{\mathcal{R}}g_{)}$, inputs $u = ({}^{\mathcal{B}}a_y), {}^{\mathcal{R}}R_{\mathcal{B}})$ and output $y = p_y$.

One strong difference from the system discussed in Section 3.3.1 is the observability of (3.41) *depends* on body motion. This is easily verified for constant values of ${}^{\mathcal{R}}R_{\mathcal{B}}$. In such case, of course, the expression ${}^{\mathcal{R}}g + {}^{\mathcal{R}}R_{\mathcal{B}}b_a$ is constant and only the sum of these parameters is observable. The following proposition refers to a condition under which System (3.41) is uniformly observable.

Proposition 3.5. Let the angular velocity represented by a function ${}^{\mathcal{B}}\omega(t)$ satisfying Assumption 3.2, and

$$\exists\, \tau, \delta > 0 :\ \forall\, t \geq 0, \quad \int_t^{t+\tau} |\omega_{\mathcal{B}}(s) \times \dot{\omega}_{\mathcal{B}}(s)| ds > \delta. \tag{3.42}$$

Then, System (3.41) given by the states $x{=}({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}v, b_a, {}^{\mathcal{R}}g)$ with inputs $u = ({}^{\mathcal{B}}a_y\ , {}^{\mathcal{B}}\omega)$ and measurements $y = (p_y, R_y) = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}R_{\mathcal{B}})$ is *uniformly observable*.

The proof for this proposition is provided in Appendix A.3.2. The above proposition asserts that the system comprising position, linear velocity, accelerometer bias and the gravitational acceleration is observable if the angular velocity is not parallel to the angular acceleration. Differently from Section 3.2.2, we now have a state-affine system, and we can thus obtain a uniform condition for the system observability. Analogously, the observability of this system was discussed in the Literature, *c.f.* (Jones et al., 2007), (Mirzaei and Roumeliotis, 2008) and (Kelly and Sukhatme, 2011). These works, however, address only the weak observability of the system, and therefore do not present an expression of the sufficient motion in order ensure the observability.

We continue the estimator design defining the following observer form:

$$\begin{cases} {}^{\mathcal{R}}\dot{\widehat{p}}_{\mathcal{B}} = {}^{\mathcal{R}}\widehat{v} + \alpha_{p_{\mathcal{B}}} \\ {}^{\mathcal{R}}\dot{\widehat{v}} \;\; = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}a_y - \widehat{b}_a) + {}^{\mathcal{R}}\widehat{g} + \alpha_v \\ \dot{\widehat{b}}_a \;\; = \alpha_a \\ {}^{\mathcal{R}}\dot{\widehat{g}} \;\; = \alpha_g \end{cases}$$

We also define the following estimation errors

$$\widetilde{p} = {}^{\mathcal{R}}p_{\mathcal{B}} - {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}, \qquad \widetilde{v} = {}^{\mathcal{R}}v - {}^{\mathcal{R}}\widehat{v}, \qquad \widetilde{b}_a = b_a - \widehat{b}_a, \qquad \widetilde{g} = {}^{\mathcal{R}}g - {}^{\mathcal{R}}\widehat{g},$$

then the objective of the filter design is to define innovation terms $\alpha_{p_\mathcal{B}}$, α_v, α_a and α_g that makes the origin of

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - \alpha_{p_\mathcal{B}}\,, \\ \dot{\widetilde{v}} = -{}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_a + \widetilde{g} - (I_3 - \widetilde{R})\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}({}^{\mathcal{B}}a_y - \widehat{b}_a) - \alpha_v\,, \\ \dot{\widetilde{b}}_a = -\alpha_a\,. \\ \dot{\widetilde{g}} = -\alpha_g \end{cases} \tag{3.43}$$

an asymptotically stable equilibrium. Since the variables considered in this section do not present topological complications, we can provide conditions for which the nonlinear observer has global exponential stability.

Similarly to the previous Section, we initially consider to measure ${}^{\mathcal{R}}R_{\mathcal{B}}$ and ${}^{\mathcal{B}}\omega$ explicitly. The following result provides an observer for the estimation of position, linear velocity, accelerometer bias and gravitational acceleration.

Proposition 3.6. Let

$$\begin{cases} \alpha_{p_\mathcal{B}} = k_{p_\mathcal{B}}\widetilde{p}\,, \\ \alpha_v = k_v\widetilde{p}\,, \\ \alpha_a = -k_a\big(I_3 + \frac{1}{k_{p_\mathcal{B}}}\mathrm{S}({}^{\mathcal{B}}\omega)\big)\,{}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{p}\,, \\ \alpha_g = k_g\widetilde{p}. \end{cases} \tag{3.44}$$

with $k_{p_\mathcal{B}}$, k_v, k_a, $k_g > 0$ such that $(k_a + k_g) < k_{p_\mathcal{B}}k_v$. Assume that

$$\exists\ \tau, \delta > 0 :\ \forall t \geq 0, \quad \int_t^{t+\tau} |\omega_\mathcal{B}(s) \times \dot{\omega}_\mathcal{B}(s)|ds > \delta. \tag{3.45}$$

Then, $(\widetilde{p},\, \widetilde{v},\, \widetilde{b}_a,\, \widetilde{g}) = (0_{3\times1},\, 0_{3\times1},\, 0_{3\times1},\, 0_{3\times1})$ is a globally exponentially stable equilibrium point of the (3.43).

The proof of the previous Proposition is presented in Appendix A.4.6. Remark that this observer satisfies Definition 3.6 when ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}$, and preserves the invariance properties of the original system. Notice that even though we have four variables, the condition on the gains resembles a third order system. Indeed, we can verify for constant ${}^{\mathcal{R}}R_{\mathcal{B}}$ that system is originally of third order. Moreover, we can separate the estimate of the accelerometer bias and gravitational acceleration only if the system is observable, *i.e.* if the condition provided by (3.42) is satisfied. That condition can be seen as the second "gain constraint" for a fourth order system. This observer provides a globally exponentially stable estimator for position, linear velocity, accelerometer bias and gravitational acceleration. This result has some properties similar to Propositions 3.3 and 3.4. For instance, (3.45) is a "persistent excitation" condition related to the observability properties of the system, *c.f.* Proposition 3.5.

We can see this observer as an extension of Proposition 3.4 to the case where the local gravitational field is unknown, where a condition on the gains is imposed additionally to the excitation term. More precisely, setting $k_g = 0$ in (3.44) and assuming ${}^{\mathcal{R}}\widehat{g} = {}^{\mathcal{R}}g$, the observer reduces to the position, linear velocity and accelerometer bias estimator. Since the structure of this observer is basically the same to the observers discussed in Section 3.3.1, we can extend the previous filter to include angular rate gyro measurements, and estimates of gyro bias and orientation provided by any filter of Section 3.2 as Corollaries 3.4 and 3.5. In these cases, of course, the global exponential guarantees of the original filter will become asymptotic, or local exponential stability properties depending on the orientation variables.

3.3.3 *Uncalibrated frames*

The previous sections treated the design of nonlinear observer for position, linear velocity and accelerometer bias with known or unknown gravitational acceleration. Those results consider that measurements of position and acceleration given by the accelerometers are made with respect to the same frame $\mathcal{B}$. This section considers the problem when the measurements are made from two different frames, *i.e.* the position is made in the camera $\mathcal{C}$ frame and the accelerometers in a different $\mathcal{B}$ frame. This task refers to the similar application discussed in Section 3.2.2.

We can write the dynamics for the system described in Section 1.6.3 as

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} \;\; = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a\,, \\ \dot{b}_a \;\; = 0_{3\times 1}\,, \\ {}^{\mathcal{R}}\dot{g} \;\; = 0_{3\times 1}\,, \\ {}^{\mathcal{B}}\dot{p}_{\mathcal{C}} \;\; = 0_{3\times 1} \end{cases}$$

with measurements $(p_y,\, R_y,\, {}^{\mathcal{B}}a_y,\, {}^{\mathcal{B}}\omega_y) = ({}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}},\, {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{R}}R_{\mathcal{C}},\, {}^{\mathcal{B}}a + b_a - {}^{\mathcal{B}}R_{\mathcal{R}}{}^{\mathcal{R}}g,\, {}^{\mathcal{B}}\omega)$. We have discussed the problems caused by using the linear acceleration as a *known* input of the system and its relations to the actual accelerometer measurements in the previous sections. Likewise, we consider that the c-to-IMU rotation is either known or given by the orientation observer discussed in Section 3.2.2. Therefore, we rewrite the original dynamics to encompass ${}^{\mathcal{B}}a_y$, ${}^{\mathcal{B}}\omega$ and ${}^{\mathcal{R}}R_{\mathcal{B}}$ as inputs, *i.e.*

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} \;\; = {}^{\mathcal{R}}R_{\mathcal{B}}({}^{\mathcal{B}}a_y - b_a) + {}^{\mathcal{R}}g\,, \\ \dot{b}_a \;\; = 0_{3\times 1}\,, \\ {}^{\mathcal{R}}\dot{g} \;\; = 0_{3\times 1}\,, \\ {}^{\mathcal{B}}\dot{p}_{\mathcal{C}} \;\; = 0_{3\times 1} \end{cases} \tag{3.46}$$

with measurements $p_y = {}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}$.

The invariance properties of the system can be analyzed as in Sections 3.3.1 and Section 3.3.2. There is, however, one additional degree of freedom, *i.e.* the system is invariant with respect to change in orientation and position for a new reference frame $\mathcal{R}_1$, changes in orientation of $\mathcal{B}_1$ frame and additive accelerometer bias. The invariance can be verified defining the states $x = ({}^{\mathcal{R}}p_{\mathcal{B}},\, {}^{\mathcal{R}}v,\, b_a,\, {}^{\mathcal{R}}g,\, {}^{\mathcal{B}}p_{\mathcal{C}})$, inputs $u = ({}^{\mathcal{B}}a_y,\, {}^{\mathcal{R}}R_{\mathcal{B}})$ and output $y = p_y$. Furthermore, let $f(x,y)$ the right hand side of (3.46) and $h(x,u) = {}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}$, we can define the group $\mathbb{G} = \mathbb{R}^3 \times \mathbb{SO}(3) \times \mathbb{SO}(3) \times \mathbb{R}^3 \times \mathbb{SO}(3) \times \mathbb{R}^3$ with elements $G = ({}^{\mathcal{R}_1}p_{\mathcal{R}},\, {}^{\mathcal{R}_1}R_{\mathcal{R}},\, {}^{\mathcal{B}_1}R_{\mathcal{B}},\, {}^{\mathcal{C}_1}p_{\mathcal{C}},\, {}^{\mathcal{C}_1}R_{\mathcal{C}},\, b_a)$ and the invariant actions

$$\begin{aligned} \varphi_G(x) &\triangleq \left({}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}_1}p_{\mathcal{R}},\, {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}v,\, {}^{\mathcal{B}_1}R_{\mathcal{B}}b_a + b_{a_0},\, {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}g,\, {}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}\right), \\ \psi_G(u) &\triangleq \left({}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}a_y + b_{a_0},\, {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1}\right), \\ \rho_G(y) &\triangleq {}^{\mathcal{R}_1}R_{\mathcal{R}}p_y + {}^{\mathcal{R}_1}p_{\mathcal{R}} \end{aligned}$$

where $\dot{\widetilde{\varphi_G(x)}} = f(\varphi_G(x), \psi_G(u))$ can be verified similarly to Section 3.3.1, and we can verify $\rho_G(h(x,u)) = h(\varphi_G(x), \psi_G(u))$ as

$$\begin{aligned} {}^{\mathcal{R}_1}R_{\mathcal{R}}p_y + {}^{\mathcal{R}_1}p_{\mathcal{R}} &= {}^{\mathcal{R}_1}R_{\mathcal{R}}({}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}) + {}^{\mathcal{R}_1}p_{\mathcal{R}} \\ &= ({}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}p_{\mathcal{B}} + {}^{\mathcal{R}_1}p_{\mathcal{R}}) + {}^{\mathcal{R}_1}R_{\mathcal{R}}{}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{B}_1}({}^{\mathcal{B}_1}R_{\mathcal{B}}{}^{\mathcal{B}}p_{\mathcal{C}}). \end{aligned}$$

Remark that even though the system is invariant to changes in the rotation of the $\mathcal{C}$ frame, *c.f.* Section 3.2.2, the final position ${}^{\mathcal{R}}p_{\mathcal{B}}$ is unchanged by the action, which indeed agrees with the invariance properties of obtained in Section 3.3.1.

Similarly to the observability properties discussed in Sections 3.2.2 and 3.3.2, the observability of System (3.46) depends on body motion. More specifically, we can verify that the observability of the system depends only on the angular velocity. It seems, however, more difficult to establish a universal input for this system than for the cases presented in Sections 3.2.2 and 3.3.2. Moreover, if the linear acceleration is known, *i.e.* $\widehat{b} - a = b_a$ and ${}^{\mathcal{R}}\widehat{g} = {}^{\mathcal{R}}g$), it is not very difficult to verify that the system comprising position, linear velocity and c-to-IMU translation is observable for angular velocities that satisfy Proposition 3.5.

We continue the design of the filter with the following observer form:

$$\begin{cases} {}^{\mathcal{R}}\dot{\widehat{p}}_{\mathcal{B}} = {}^{\mathcal{R}}\widehat{v} + \alpha_{p_{\mathcal{B}}}\,, \\ {}^{\mathcal{R}}\dot{\widehat{v}} \;\; = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}a_y - \widehat{b}_a) + {}^{\mathcal{R}}\widehat{g} + \alpha_v\,, \\ \dot{\widehat{b}}_a \;\; = \alpha_a\,, \\ {}^{\mathcal{R}}\dot{\widehat{g}} \;\; = \alpha_g\,, \\ {}^{\mathcal{B}}\dot{\widehat{p}}_{\mathcal{C}} \;\; = \alpha_{p_{\mathcal{C}}}\,, \end{cases}$$

Let us define the following estimation errors

$$\widetilde{p} = {}^{\mathcal{R}}p_{\mathcal{B}} - {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}, \quad \widetilde{v} = {}^{\mathcal{R}}v - {}^{\mathcal{R}}\widehat{v}, \quad \widetilde{b}_a = b_a - \widehat{b}_a,$$
$$\widetilde{g} = {}^{\mathcal{R}}g - {}^{\mathcal{R}}\widehat{g}, \quad \widetilde{q} = {}^{\mathcal{B}}p_{\mathcal{C}} - {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}},$$

then the goal of the design is to define innovation terms $\alpha_{p_{\mathcal{B}}}$, α_v, α_a and α_g and $\alpha_{p_{\mathcal{C}}}$ such that the origin of

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - \alpha_{p_{\mathcal{B}}}\,, \\ \dot{\widetilde{v}} = \; - {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_a + \widetilde{g} - (I_3 - \widetilde{R}){}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}({}^{\mathcal{B}}a_y - \widehat{b}_a) - \alpha_v\,, \\ \dot{\widetilde{b}}_a = \; - \alpha_a\,, \\ \dot{\widetilde{g}} = \; - \alpha_g\,, \\ \dot{\widetilde{q}} = \; - \alpha_{p_{\mathcal{C}}} \end{cases} \tag{3.47}$$

is an asymptotically stable equilibrium.

Let ${}^{\mathcal{R}}\widehat{p}_{\mathcal{C}} = {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}} + {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}{}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}$, and consider to measure ${}^{\mathcal{R}}R_{\mathcal{B}}$ and ${}^{\mathcal{B}}\omega$, similarly to the previous Sections, however, in order to prove stability for (part of) the system, we assume additionally that $\widehat{b}_a = b_a$, ${}^{\mathcal{R}}\widehat{g} = {}^{\mathcal{R}}g$. The next result concerns an observer for position, linear velocity and c-to-IMU translational displacement.5

Proposition 3.7. Assume that $\widehat{b}_a = b_a$, ${}^{\mathcal{R}}\widehat{g} = {}^{\mathcal{R}}g$ and $\alpha_a = \alpha_g = 0$. Let

$$\begin{cases} \alpha_{p_{\mathcal{B}}} = k_{p_{\mathcal{B}}}\widetilde{p} - {}^{\mathcal{R}}R_{\mathcal{B}}\alpha_{p_{\mathcal{C}}}\,, \\ \alpha_v = k_v\widetilde{p}\,, \\ \alpha_p = \; - k_{p_{\mathcal{C}}}\mathrm{S}({}^{\mathcal{B}}\omega){}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{p}\,, \end{cases} \tag{3.48}$$

with $k_{p_{\mathcal{B}}}$, k_v, $k_{p_{\mathcal{C}}} > 0$. Assume that

$$\exists\, \tau, \delta > 0 : \; \forall t \geq 0 \quad \int_t^{t+\tau} |\omega_{\mathcal{B}}(s) \times \dot{\omega}_{\mathcal{B}}(s)| ds > \delta. \tag{3.49}$$

is satisfied. Then, $(\widetilde{p}, \widetilde{v}, \widetilde{q}) = (0_{3\times 1}, 0_{3\times 1}, 0_{3\times 1})$ is a globally exponentially stable equilibrium point of the (3.43).

The proof of the above Proposition is similar to 3.6 and therefore it is omitted in the Appendix. The above observer satisfies definition 3.6 when ${}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}$, $\widehat{b}_a = b_a$, ${}^{\mathcal{R}}\widehat{g} = {}^{\mathcal{R}}g$ and $\alpha_a = \alpha_g = 0$ and preserves the invariance properties of the original system. This result provides a globally exponentially stable observer for a reduced set of System 3.46 under "persistent motion" conditions, however, we are not capable of extending a similar result to the full system. This result can be extended likewise Corollaries 3.4 and 3.5, however the result does not provide a full estimator for the system. Finally, we conjecture an observer for the full system, *i.e.* position, linear velocity, accelerometer bias, gravitational acceleration and c-to-IMU translational displacement.

Conjecture 3.1. Let

$$\begin{cases} \alpha_{p_{\mathcal{B}}} = k_{p_{\mathcal{B}}}\widetilde{p}_{\mathcal{C}} - {}^{\mathcal{R}}R_{\mathcal{B}}\alpha_{p_{\mathcal{C}}}\,, \\ \alpha_v = k_v\widetilde{p}_{\mathcal{C}}\,, \\ \alpha_a = -k_a\big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)\big){}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\widetilde{p}_{\mathcal{C}} + \frac{k_a}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)\alpha_{p_{\mathcal{C}}}\,, \\ \alpha_g = k_g\widetilde{p}_{\mathcal{C}} - \frac{k_g}{k_{p_{\mathcal{B}}}}{}^{\mathcal{R}}R_{\mathcal{B}}\alpha_{p_{\mathcal{C}}}\,, \\ \alpha_p = -k_{p_{\mathcal{C}}}\mathrm{S}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega){}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\widetilde{p}_{\mathcal{C}}\,, \end{cases} \tag{3.50}$$

with $k_{p_{\mathcal{B}}}, k_v, k_a, k_g, k_{p_{\mathcal{C}}} > 0$ and $k_a + k_g \leq k_{p_{\mathcal{B}}}k_v$. Assume that enough "excitation" of ${}^{\mathcal{B}}\omega$. Then, $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a, \widetilde{g}, \widetilde{q}) = (0_{3\times1}, 0_{3\times1}, 0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is a stable equilibrium point of the error dynamics (3.47).

Notice that the previous observer satisfies definition 3.6 and preserves the invariance properties of the original system.This proposition merges directly the results of Propositions 3.6 and 3.7, however we are unable to provide proofs of the stability properties or observability conditions for the system.

3.4 GAIN TUNING

We have discussed the design of nonlinear observers based on their stability analysis. A large basin of convergence is a prerequisite to guarantee unbiased estimates for pose, linear velocity, and the additive biased for angular rate gyroscopes and accelerometers. Moreover, understanding the effects on the estimation dynamics of the gains corresponding to each innovation term is also vital for obtaining high quality estimates under fast dynamics. The stability properties given by the nonlinear innovation terms, however, only provides necessary conditions for the convergence of the estimates. A solid and well-state tuning procedure for the innovation gains enables to ensure a good response to estimation errors as well as to respect the characteristics of the employed sensors.

For instance, the gains in stochastic filtering are directly related to noise and model uncertainty characteristics, such as sensor, process covariance matrices, as well as sensor to sensors stochastic correlation. The proposed filter design and respective analysis are based on the dynamics of the estimation errors, it is more sound to perform a gain tuning strategy in terms of time-response. Considering the IMU-based pose estimation discussed so far, we can distinguish two dynamics:

- *Fast* dynamics for pose variables and derivatives: non-modeled effects that corrupt these estimates, measurement noise itself for instance, usually present fast dynamics. Another interpretation for the desired behavior is that the nonlinear observers can also be seen as *low-pass filters*, thus a larger bandwidth is needed to track fast dynamics.
- *Slow* dynamics for angular rate gyroscopes and accelerometer biases: non-modeled effects in these variables, effects due to temperature changes for instance, also present slow

dynamics. The same low-pass filter interpretation applies. The variables should present a very low bandwidth in order to neglect the fast dynamics.

In the same way as the observer design, we can analyze the gain tuning as two independent procedures. First, let us consider the attitude observer from Proposition 3.1.

3.4.1 *Orientation estimation*

The gain tuning is made in function of two parameters $\tau_{R_\mathcal{B}}, \tau_\omega > 0$, that denote settling times such that $\tau_{R_\mathcal{B}}$ is much smaller than τ_ω.

The parameter $\tau_{R_\mathcal{B}}$ denotes a target settling time for the orientation estimation dynamics. However, the implementation of the filter in discrete time is constrained by the sampling frequency f_p of the pose measurements. Furthermore, the choice of a small settling time may yield to instability of the filter. A typical value considers $\tau_{R_\mathcal{B}} \in [\frac{1}{f_m}, 1]$. As for τ_ω, this parameter denotes the desired settling time for gyroscope bias estimation error. Since this bias varies slowly, a relatively large value of τ_ω can be considered, $\tau_\omega \geq 20$ for example.

We define the gains $k_{R_\mathcal{B}}$ and k_ω for the attitude and gyro bias innovation terms of Proposition 3.1 as

$$k_{R_\mathcal{B}} = 48\frac{\tau_{R_\mathcal{B}} + \tau_\omega}{\tau_{R_\mathcal{B}}\tau_\omega}, \quad k_\omega = 9\frac{1}{\tau_{R_\mathcal{B}}\tau_\omega}. \tag{3.51}$$

We draw these gains starting with dynamics of (3.22)

$$\begin{cases} \dot{\widetilde{R}} = -\widetilde{R}\,\mathrm{S}\left({}^{\mathcal{R}}\widehat{R}_\mathcal{B}\widetilde{b}_\omega + k_{R_\mathcal{B}}\dfrac{\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big)}{\big(1+\mathrm{tr}(\widetilde{R})\big)^2}\right), \\ \dot{\widetilde{b}}_\omega = k_\omega\,{}^{\mathcal{R}}\widehat{R}_\mathcal{B}\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big). \end{cases}$$

and further consider a parametrization of $\mathbb{SO}(3)$ such that $\widetilde{R} \approx I_3 + \mathrm{S}(\widetilde{\theta})$ around I_3, *e.g.* an element $\widetilde{\theta} \in \mathbb{R}^3$ of $\mathfrak{so}(3)$ writes $\widetilde{R} = \exp(\mathrm{S}(\theta)) \approx I_3 + \mathrm{S}(\widetilde{\theta})$ around I_3, and the variable change $\widetilde{z} = -{}^{\mathcal{R}}R_\mathcal{B}\widetilde{b}_\omega$. The linearized dynamics of the attitude estimation error dynamics around the equilibrium point $\widetilde{\theta} = 0_{3\times 1}$ yields

$$\overbrace{\begin{bmatrix}\widetilde{\theta}\\ \widetilde{z}\end{bmatrix}}^{\dot{}} = \begin{bmatrix}-(k_{R_\mathcal{B}}/16)I_3 & I_3\\ -k_\omega I_3 & 0_{3\times 3}\end{bmatrix}\begin{bmatrix}\widetilde{\theta}\\ \widetilde{z}\end{bmatrix}, \tag{3.52}$$

Note that the previous system can be decomposed into three independent autonomous linear time-invariant systems, where each characteristic polynomial is given by

$$\begin{aligned} p(s) &= s^2 + k_{R_\mathcal{B}}/16 s + k_\omega \\ &= s^2 + 3\frac{\tau_{R_\mathcal{B}} + \tau_\omega}{\tau_{R_\mathcal{B}}\tau_\omega}s + \frac{9}{\tau_{R_\mathcal{B}}\tau_\omega} = \left(s + \frac{3}{\tau_{R_\mathcal{B}}}\right)\left(s + \frac{3}{\tau_\omega}\right). \end{aligned}$$

Hence, the gain choice (3.51) yields two real eigenvalues $\lambda_{R_\mathcal{B}} = -3/\tau_{R_\mathcal{B}}, \lambda_\omega = -3/\tau_\omega$. Moreover, assuming $\tau_{R_\mathcal{B}} \neq \tau_\omega$ and using the following variable change

$$\begin{bmatrix}x_{R_\mathcal{B}}(t)\\ x_\omega(t)\end{bmatrix} = \begin{bmatrix}\lambda_{R_\mathcal{B}} & 1\\ \lambda_\omega & 1\end{bmatrix}\begin{bmatrix}\widetilde{\theta}_i(t)\\ \widetilde{z}_i(t)\end{bmatrix}, \tag{3.53}$$

then Eq. (3.52) writes

$$\overbrace{\begin{bmatrix}x_{R_\mathcal{B}}\\ x_\omega\end{bmatrix}}^{\dot{}} = \begin{bmatrix}\lambda_{R_\mathcal{B}} & 0\\ 0 & \lambda_\omega\end{bmatrix}\begin{bmatrix}x_{R_\mathcal{B}}\\ x_\omega\end{bmatrix}. \tag{3.54}$$

By using (3.53) and (3.54), it is not difficult to obtain the following expression for the solutions of System (3.52):

$$\widetilde{\theta}_i(t) = e^{\lambda_{R_\mathcal{B}} t}\widetilde{\theta}_i(0) + \lambda_\omega \frac{e^{\lambda_{R_\mathcal{B}} t} - e^{\lambda_\omega t}}{\lambda_{R_\mathcal{B}} - \lambda_\omega}\widetilde{\theta}_i(0) + \frac{e^{\lambda_{R_\mathcal{B}} t} - e^{\lambda_\omega t}}{\lambda_{R_\mathcal{B}} - \lambda_\omega}\widetilde{z}_i(0)$$
$$\widetilde{z}_i(t) = e^{\lambda_\omega t}\widetilde{z}_i(0) + \lambda_{R_\mathcal{B}}\lambda_\omega \frac{e^{\lambda_{R_\mathcal{B}} t} - e^{\lambda_\omega t}}{\lambda_{R_\mathcal{B}} - \lambda_\omega}\widetilde{\theta}_i(0) + \lambda_\omega \frac{e^{\lambda_{R_\mathcal{B}} t} - e^{\lambda_\omega t}}{\lambda_{R_\mathcal{B}} - \lambda_\omega}\widetilde{z}_i(0)$$

Therefore the following (partial) dynamics decoupling is obtained:

- Fast exponential decrease of $\widetilde{\theta}_i(t)$ to zero, corrupted by slowly decreasing terms with small amplitude: $\frac{\lambda_\omega}{\lambda_{R_\mathcal{B}} - \lambda_\omega}$ and $\frac{1}{\lambda_{R_\mathcal{B}} - \lambda_\omega}$ tend to zero as $\tau_{R_\mathcal{B}} \to 0$ and $\tau_\omega \to \infty$.
- Slow exponential decrease of $\widetilde{z}_i(t)$ to zero corrupted by slowly decreasing terms with small amplitude.

3.4.2 *Position estimation*

The same rationale leads to the following definition of the estimation of translational variables. The gain tuning is made in function of three settling times $\tau_{p_\mathcal{B}}$, τ_v, $\tau_a > 0$ such that $\tau_{p_\mathcal{B}}$, τ_v are much smaller than τ_a. Likewise, $\tau_{R_\mathcal{B}}$, the values for $\tau_{p_\mathcal{B}}$, $\tau_v \in [\frac{1}{f_m}, 1]$ and $\tau_a \geq 20$ often present a good response for the error dynamics.

We define the gains $k_{p_\mathcal{B}}$ and k_v and k_a for the position, linear velocity and accelerometer bias innovation terms of Proposition 3.1 as

$$k_{p_\mathcal{B}} = 3\frac{\tau_{p_\mathcal{B}}\tau_v + \tau_{p_\mathcal{B}}\tau_a + \tau_v\tau_a}{\tau_{p_\mathcal{B}}\tau_v\tau_a}, \quad k_v = 9\frac{\tau_{p_\mathcal{B}} + \tau_v + \tau_a}{\tau_{p_\mathcal{B}}\tau_v\tau_a}, \quad k_a = \frac{27}{\tau_{p_\mathcal{B}}\tau_v\tau_a}. \tag{3.55}$$

These gains satisfy the stability conditions of Proposition3.4. Choosing $\tau_{p_\mathcal{B}}, \tau_v \ll \tau_a$ leads to the same (partial) decoupling of the dynamics of $\widetilde{p}, \widetilde{v}$ on one hand, and $\widetilde{b}_a$ on the other hand.

3.4.3 *Gain tuning and observability conditions*

We determined in the previous sections a gain tuning technique for systems that are observable independently of the inputs. We have seen that other systems, however, are observable under certain movements, for example: the estimation of orientation, gyro bias and c-to-IMU rotation, *c.f.* Section 3.2.2, or position, linear velocity, accelerometer bias, with unknown gravitational acceleration, *c.f.* Section 3.3.2. For each of the prior cases, the resulting nonlinear observer is stable under certain conditions of the motion and the analysis neglecting the angular velocity will not be able to allocate correctly the poles of the system. We propose to employ the same gains to the other parameters as the ones obtained by the gyro and accelerometer biases. However, in practical situations, we can tune the gain of each variable depending on the each condition. We discuss such this approach in Section 4.3.1

3.5 SIMULATION RESULTS

The previous sections discussed different aspects of nonlinear observer design for pose estimation. The observers introduced have clear stability properties, which sometimes hold under certain observability conditions. In this section, we perform series of simulations to validate the convergence properties of the proposed filters, robustness to unmodeled parameters and sensory noise. Moreover, the performance of the proposed filters is compared to results obtained with implementation of other approaches from the Literature. Seven categories divide the experiments

- Orientation and gyro bias estimation;
- Orientation, gyro bias and c-to-IMU rotation estimation;
- Position and accelerometer bias estimation;
- Position, accelerometer bias and gravitational acceleration estimation;
- Position and c-to-IMU translation.
- Coupled estimators for orientation and position.

The first simulation compares Proposition 3.1 to the passive complementary filter, *i.e.* Theorem 3.3, and an implementation of the multiplicative extended Kalman filter (MEKF) (Lefferts et al., 1982). A brief introduction to the latter filter is given in Appendix C. The second simulation compares the nonlinear observer from Proposition 3.3 with an implementation of the MEKF that includes c-to-IMU rotation estimation. The third, and fourth simulations compare Propositions 3.4, 3.6 to an implementation of the Kalman filter (KF). Notice that we do consider full knowledge of orientation dynamics for this simulation, therefore it is possible to employ the KF instead of some extension. Those two simulations on translational displacement also evaluate the effects of unmodeled errors on gravitational acceleration and c-to-IMU translation. The fifth simulation concerns the performance of Corollary 3.7 and the results are compared to the KF. The last simulation validates the implementation of the filters for translation estimation using the orientation estimates, instead of the explicit variables.

The implementation of the KF and MEKF calls for a few remarks. First, the MEKF is based on quaternion parametrization, which uses four parameters to represent the three dimensional $\mathbb{SO}(3)$. Hence, the redundant degree may result in an ill conditioned covariance matrix (Lefferts et al., 1982) that also may impair the performance of the extended Kalman filter (EKF). Indeed, there are several MEKF techniques discussed in the seminal article. We refer to the specific formulation employing a *reduced representation of the covariance matrix*. This version is chosen because it reduces the consequences from ill-conditioning of the covariance matrix, also the diagonal values of the covariance matrix refer directly to the uncertainties of angular and gyro bias estimates. Nevertheless, as we have discussed in Section 1.4, there is a plethora of empirical techniques to improve the performance of EKF-based techniques. Therefore, it is inaccurate and restrictive to claim something as *the* EKF, and we would rather claim an implementation of the filter. Moreover, since the analyses employ simulated data, we have access to every uncertainty of the system, *i.e.* noise, initial errors, etc. We thus set the parameters of the MEKF using the nominal values of initial errors and sensory noises. The same procedure is made to tune the KF for estimation of translational dynamics and respective parameters. There are plenty of other Kalman-based techniques, as, *e.g.* , the unscented Kalman filter. We leave aside these other techniques since they present mostly heuristic improvements in the approximation of the covariance matrix of the estimates.

On the other hand, the setup of the proposed nonlinear observers requires fewer and rather simpler parameters. We rely on the gain tuning technique described in Section 3.4. The gains employed by Propositions 3.1 and 3.3 are computed using (3.51) with settling times $\tau_{R_{\mathcal{B}}} = 0.15$ [s], $\tau_{b_\omega} = 15$ [s]. The gains employed by Theorem 3.3 use the same settling times computed using a similar rationale as introduced originally in (Scandaroli and Morin, 2011). When required, the innovation c-to-IMU rotation employs the same gains as the gyro bias $\widehat{b}_\omega$. The observers for the translational dynamics consider settling times of $\tau_{p_{\mathcal{B}}} = 0.15$ [s], $\tau_v = 0.8$ [s], and $\tau_{b_a} = 15$ [s] and the gains are computed using (3.55). When required, the innovation terms of the gravitational acceleration ${}^{\mathcal{R}}\widehat{g}$ employs the gain as the accelerometer bias $\widehat{b}_a$ and c-to- IMU translation ${}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}$ employs gain with half of the value from the innovation of ${}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}$.

The simulations consider different aspects of rotational and translational body motion. More specifically, we designed a class of reference trajectories for which amplitude and frequency can be randomly changed for each execution of a simulation. Two types of angular motion are employed that we classify in *good* and *bad* angular motions. These adjectives refer

specifically to the observability properties provided by these inputs, and according to Assumption 3.2, the angular velocities must be continuous and have bounded derivatives. Good and bad angular motions are given in function of the amplitude γ_a, frequency γ_p and delay γ_d, with bad angular motion given by

$$^{\mathcal{B}}\vec{\omega}(t) = \left[\gamma_{a,1}\ \cos(\gamma_p t + \ \gamma_d) \quad \gamma_{a,2}\ \cos(\gamma_p t + \ \gamma_d) \quad \gamma_{a,3}\ \cos(\gamma_p t + \ \gamma_d)\right]^T,$$

and good motion given by

$$^{\mathcal{B}}\tilde{\omega}(t) = \left[\gamma_{a,1}\ \cos(\gamma_p t + \ \gamma_{d,1}) \quad \gamma_{a,2}\ \cos(\gamma_p t + \ \gamma_{d,2}) \quad \gamma_{a,3}\ \cos(\gamma_p t + \ \gamma_{d,3})\right]^T.$$

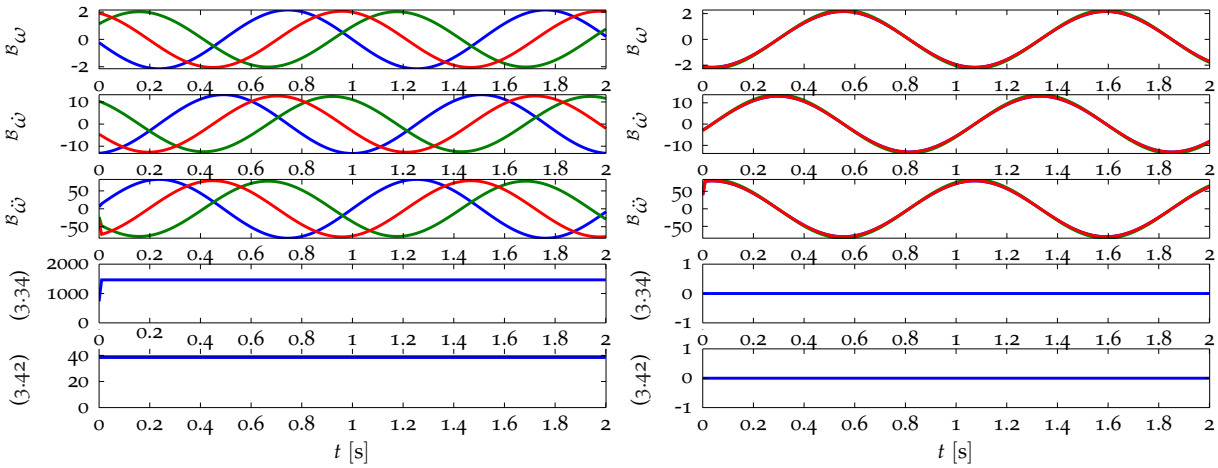
The parameters γ_a are drawn from a Gaussian distribution (GD) with mean 2.1 $[rad]$ and variance $0.5 \cdot 10^{-3}$ $[rad]^2$, this is equivalent to angular velocity with amplitude (maximum value minus minimum value of the angular velocity) of 240 degrees in average, with minimum 210 and maximum 270 degrees amplitude. The parameters referring to the angular period γ_p are drawn from a GD with mean 6.3 $[rad/s]$ and variance $1.1[rad/s]^2$, this is equivalent to 1 [Hz] frequencies in average, with minimum 0.5 and maximum 1.5 [Hz]. The parameters related to the delay are drawn from a GD with zero mean and variance 0.84 $[rad/s]^2$, this refers to delays up to 25 degrees in the phase of angular velocity components. The delay of each component of the angular velocity is the main difference between good and bad motions. Figure 3.1 depicts one example of good and bad angular velocities with the respective values given by conditions (3.34) and (3.45) computed at each instant.

The simulation is also designed to evaluate the effects due to the translational dynamics. A trivial solution could consider a constant acceleration in $\mathcal{B}$ frame for the trajectory, however, if the acceleration is not null, the resulting position becomes numerically unstable after a relatively short period. We designed a bounded trajectory divided in two main displacements:

$$^{\mathcal{R}}\vec{p}_{\mathcal{B}}(t) = \left[\ \gamma_p cos(\gamma_f t) \quad \tfrac{3}{4}\gamma_p cos(\tfrac{1}{2}\gamma_f t) \quad \tfrac{1}{3}\gamma_p cos(\tfrac{1}{2}\gamma_f t)\right],$$
$$^{\mathcal{R}}\overleftarrow{p}_{\mathcal{B}}(t) = -\left[\ \gamma_p cos(\gamma_f t) \quad \tfrac{3}{4}\gamma_p cos(\tfrac{1}{2}\gamma_f t) \quad \tfrac{1}{3}\gamma_p cos(\tfrac{1}{2}\gamma_f t)\right].$$

The constant γ_p is drawn from a GD with mean 1 [m] and variance $27.8 \cdot 10^{-3}$ $[\mathrm{m}]^2$, which provides a maximum and minimum amplitudes of 3 and 1 [m] respectively. Even though the observability of the system is independent of the acceleration, this variable must still satisfy Assumption 3.2, *i.e.* the linear accelerations and their first order time derivatives must be bounded, so that the observability analysis is valid. Remark that $^{\mathcal{R}}\vec{p}_{\mathcal{B}}(t)$ and $^{\mathcal{R}}\overleftarrow{p}_{\mathcal{B}}(t)$ yield linear accelerations that indeed satisfy Assumption 3.2, the transition between these two trajectories results a discontinuous acceleration however. Thus, the transition between $^{\mathcal{R}}\vec{p}_{\mathcal{B}}(t)$, and $^{\mathcal{R}}\overleftarrow{p}_{\mathcal{B}}(t)$ is given by a sixth-order polynomial to ensure the continuity of the accelerations. Moreover, the measurements of the accelerometers are given in $\mathcal{B}$ frame, while we compute analytically the acceleration in $\mathcal{R}$ frame. The simulation of the accelerometer measurements depends on the resulting angular velocity, and, of course, the gravitational acceleration in $\mathcal{R}$ frame. Figure 3.3 depicts the simulated specific acceleration using *good* and *bad* angular motion for the a reference trajectory.

Concluding the simulation setup, we consider that inertial and pose measurements are synchronous and sampled at a frequency of 200 [Hz]. We analyze different cases with and without measurement noise to compare different aspects of the filters. The measurement noises considered for the inertial sensors are similar to an IMU xSens MTi-G, *i.e.* gyroscopes are corrupted with noise drawn from a zero mean Gaussian distribution (ZMGD) with variance $33.8 \cdot 10^{-6}$ $[\mathrm{rad/s}]^2$, and the accelerometers from a ZMGD with variance $0.7 \cdot 10^{-6}$ $[\mathrm{m/s^2}]^2$.



(a) *Good* angular motion (b) *Bad* angular motion

Figure 3.1: Example of *good* and *bad* angular motions.

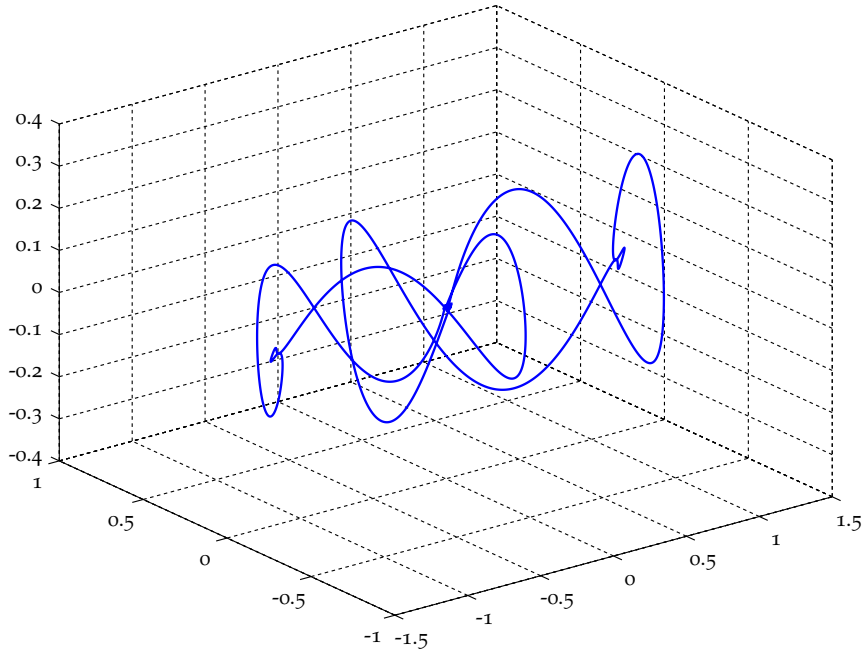
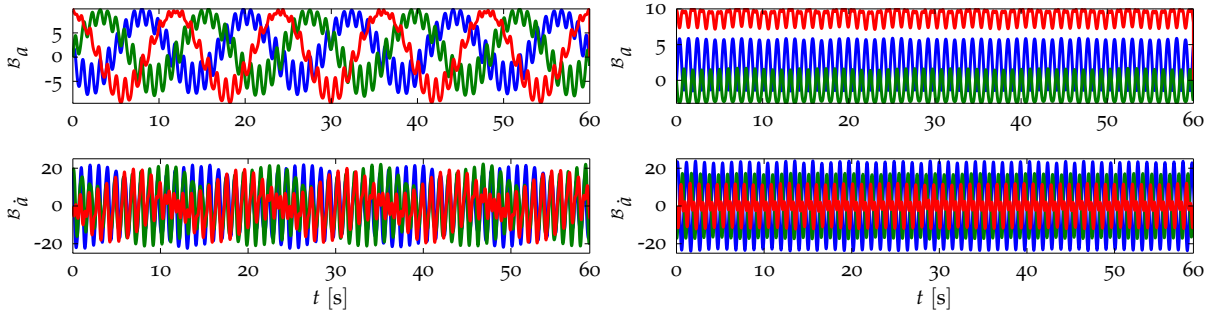


Figure 3.2: Sample of translational trajectory.



(a) Acceleration with *good* angular motion (b) Acceleration with *bad* angular motion

Figure 3.3: Specific acceleration of a reference trajectory with *good* and *bad* angular motions.

Furthermore, position measurements are corrupted by ZMGD with variance $2.8 \cdot 10^{-6}$ $[\mathrm{m}]^2$, that refer to measurements with an uncertainty of 0.5 [cm] roughly. In order to simulate noise on the rotation matrix, we resort to the angle axis representation, *i.e.* a rotation matrix $R \in \mathbb{SO}(3)$ can be represented by $(\theta, u) \in \{\mathbb{R} \times \mathbb{R}^3 \colon u^T u = 1\}$ with $R = \exp(\theta \mathbb{S}(u))$. We generate random matrices using angles θ drawn from ZMGD with variance $135.4 \cdot 10^{-6}$ $[\mathrm{rad}]^2$, which refer an uncertainty of 2 degrees. The rotation axis u is generated from three samples drawn out of a uniform distribution, and the resulting vector is normalized before the computation of the exponential matrix.

3.5.1 *Orientation and gyro bias*

The first simulation evaluates the performance of the filters for orientation and gyro bias estimation. This system is uniformly observable independently of the inputs. We compare three filters, Proposition 3.1, Theorem 3.3 and the MEKF in two situations: convergence from moderate initial errors and convergence from large initial errors. We can verify three aspects from these simulations

- The three filters have a large basin of convergence for this problem;
- Proposition 3.1 and Theorem 3.3 have similar responses locally;
- Proposition 3.1 improves the convergence properties from Theorem 3.3 for large errors, *i.e.* the estimates of the proposed nonlinear observer converge at the designed settling time.

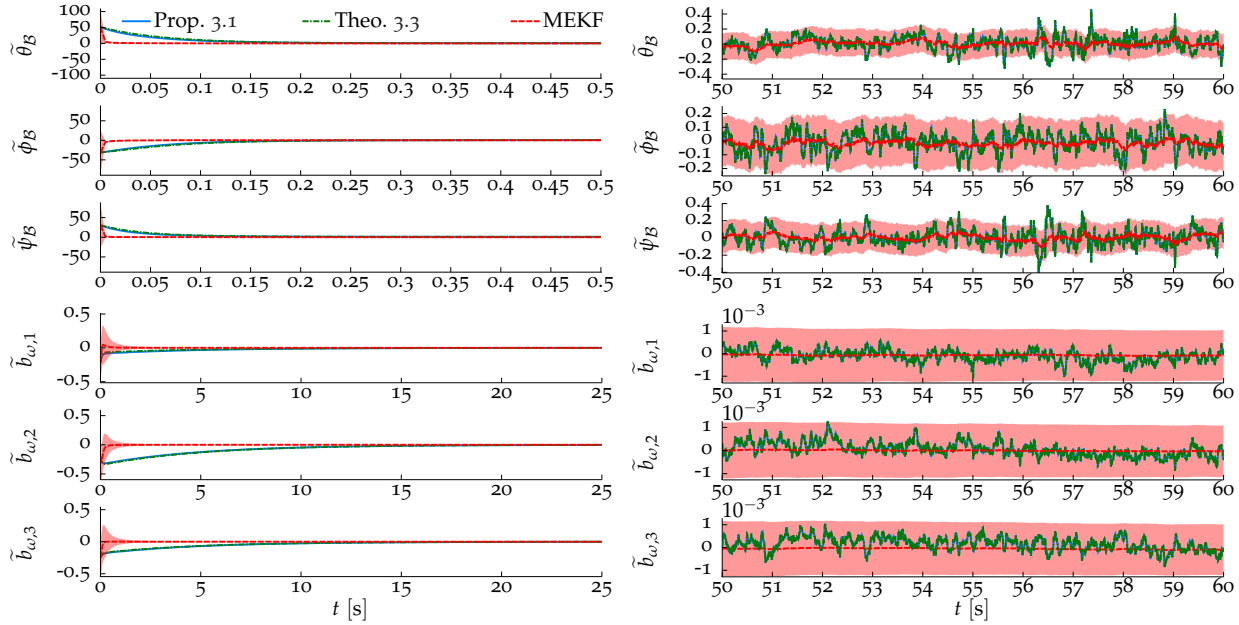
Convergence from moderate initial errors

This simulation tests the convergence of the filters starting from moderate initial errors. The estimates of orientation are initialized with random samples from a ZMGD with variance $121.8 \cdot 10^{-3}$ $[\mathrm{rad}]^2$, *i.e.* errors up to 60°. Gyro bias estimates are initialized with random samples from a ZMGD with variance $7.6 \cdot 10^{-3}$ $[\mathrm{rad}/s]^2$, this refers to biases up 15 [°/s]. Figure 3.4 shows a typical result obtained from repeated simulations. The curves in solid blue denote the responses of Proposition 3.1, in dashed green the responses of Theorem 3.3 and dashed red of the MEKF. Moreover, the light red areas represent the 3σ uncertainty regions provided by the MEKF for each variable. Figure 3.4 (a) displays the estimation error neglecting sensory noise, and Figure 3.4 (b) displays the steady state errors of the estimates considering sensor noises. From top to bottom, the results correspond to the estimation error for body orientation error in Euler angles in ° of roll $\widetilde{\theta}_{\mathcal{B}}$, pitch $\widetilde{\phi}_{\mathcal{B}}$, yaw $\widetilde{\psi}_{\mathcal{B}}$ angles, and gyro bias error for the first, second and third components of $\widetilde{b}_{\omega}$ in [rad/s].

The estimates provided by the three filters converge to the correct states. The orientation estimates from both nonlinear observers converge in about 0.15 [s] and the estimates of gyro bias converge in 15 [s], as predetermined by the gain tuning. This can be considered as a local convergence for both filters, we can also verify that the resulting dynamics is almost identical. Furthermore, we can verify that the estimates of the MEKF converges almost instantaneously to the correct states. This property is typical of a Kalman-like filter on systems that are observable regardless of the inputs. The steady-state responses of the estimates for noisy measurements allow us to verify that the uncertainties of nonlinear observers are of the same order of the MEKF, the observers do not compute the covariance matrix explicitly however.

Convergence from large initial errors

The next simulation analyzes the convergence of the estimates for large initial errors. The orientation estimates are initialized with errors from a GD with mean π [rad] and variance $3.4 \cdot$



(a) Convergence *without* measurement noise

(b) Steady-state *with* measurement noise

Figure 3.4: Orientation and gyro bias estimation with moderate initial errors.

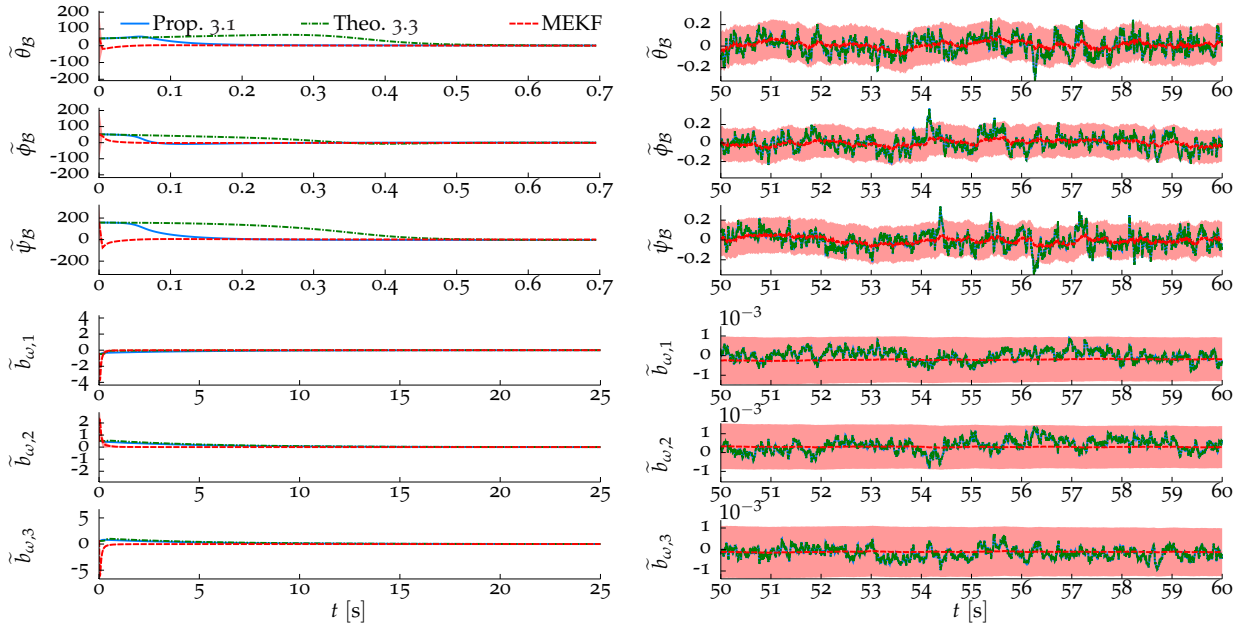
$10^{-9}[\text{rad}]^2$. This orientation refers points close to the unstable invariant set of Theorem 3.3 and the singular set of Proposition 3.1. Gyro bias estimates are initialized with errors from a ZMGD with variance $41.5 \cdot 10^{-3}$ $[\text{rad/s}]^2$, *i.e.* errors up to 35 [°/s]. Figure 3.5 shows a typical result obtained from repeated simulations and displacement of the results is the same from the previous simulation.

The estimates computed by the three filters converge to the correct states. The response of the MEKF is indeed similar to the previous one, *i.e.* the estimates converge to the real states immediately. Moreover, concerning the nonlinear observers, we can verify that their basin of convergence is very large, *i.e.* the estimates converge to the real states even very close to what can be called *bad* cases, *i.e.* the singular and unstable sets for Proposition 3.1 and Theorem 3.3. Remark, however, that the orientation estimates obtained using Proposition 3.1 still converge with the predefined settling time, *i.e.* 0.15 [s], whilst the estimates of the passive complementary filter converge only after 0.5 [s]. This improvement in the convergence of the estimates owes to the new innovation term that increases the innovation update for larger orientation errors.

The last remark concerns the exponential convergence of the estimates of the nonlinear observers. Let us consider the following energy function

$$\mathcal{V} = \text{tr}(I_3 - \widetilde{R}) + \frac{1}{k_{b_\omega}} |\widetilde{b}_\omega|^2.$$

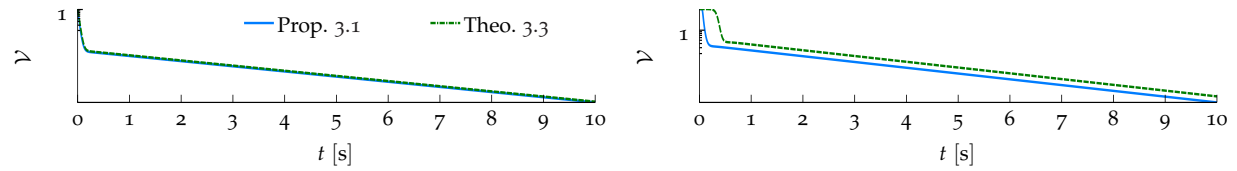
This energy function is strictly related to the stability proof for the estimates and it represents an error measure of the states. Figure 3.6 represents the response of this function in logarithmic scale for the experiments of Figures 3.4 and 3.5. The curves in solid blue represent the response of Proposition 3.1 and dashed green the response of Theorem 3.3. Notice in Figure 3.6 (a) that the curve is given by fast descent until 1 [s] followed by a straight decreasing line. This straight line represents the exponential decay, since the Figure presents the y-axis in logarithmic scale. We can remark a slightly different behavior in Figure 3.6 (b), where the response of Proposition 3.1 is similar to the the previous one, whilst the energy of the passive complementary filter remains constant until 0.5 [s], and after 0.75 [s] achieves the region



(a) Convergence *without* measurement noise

(b) Steady-state *with* measurement noise

Figure 3.5: Orientation and gyro bias estimation with large initial errors.



(a) Moderate initial errors

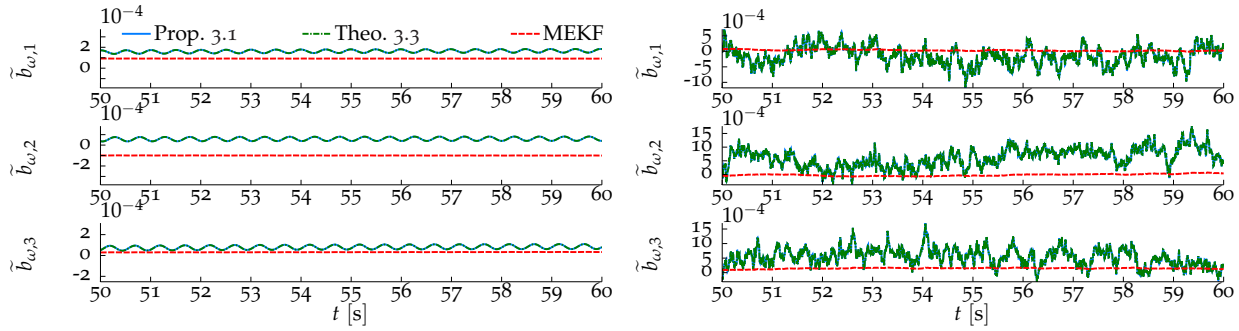
(b) High initial errors

Figure 3.6: Exponential convergence of the errors – logarithmic y-scale.

with exponential convergence. We thus verify the stronger stability properties of the proposed observer.

Identifying misaligned c-to-IMU rotation

The previous cases evaluated the case with known c-to-IMU rotation, and, in many applications, this variable can be roughly estimated either visually or using a CAD model. The value in practice, however, can differ from this rough estimate, and the resulting estimate of $\mathcal{B}$ orientation is biased. Let us analyze this case more carefully. Figure 3.7 presents the response of gyro-bias error considering a measurement $R_y = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}$ with c-to-IMU rotation ${}^{\mathcal{B}}R_{\mathcal{C}}$ obtained with a random axis and angle of 2°. Figure 3.7 (a) displays the estimation error without measurement noise and Figure 3.7 (b) shows the result with measurement noise. The effect due to the parasite c-to-IMU rotation is easily verified for the simulation without noise, since the bias estimated by the nonlinear observers oscillate in Figure 3.7 (a). The bias from the MEKF, however, shows only a small offset and practically no oscillation. The oscillation is hidden in Figure 3.7 (b) due to the measurement noise of the orientation. Therefore a bad estimate of the c-to-IMU orientation can be only identified using accurate orientation measurements, which is indeed the case explored in Chapter 4.



(a) Steady-state *without* noise

(b) Steady-state *with* noise

Figure 3.7: Estimated gyro bias with misaligned c-to-IMU rotation.

3.5.2 *Orientation, gyro bias and c-to-IMU rotation*

The second simulation concerns the estimation of orientation, gyro bias and c-to-IMU rotation. In this section, we analyze the convergence of estimates from Proposition 3.3 and an implementation of the MEKF for the same system. We can verify several aspects from these results:

- both filters are locally stable;
- moderate (and large) initial errors severely degrade the performance of the MEKF;
- the observer from Proposition 3.3 has very large basin of convergence;

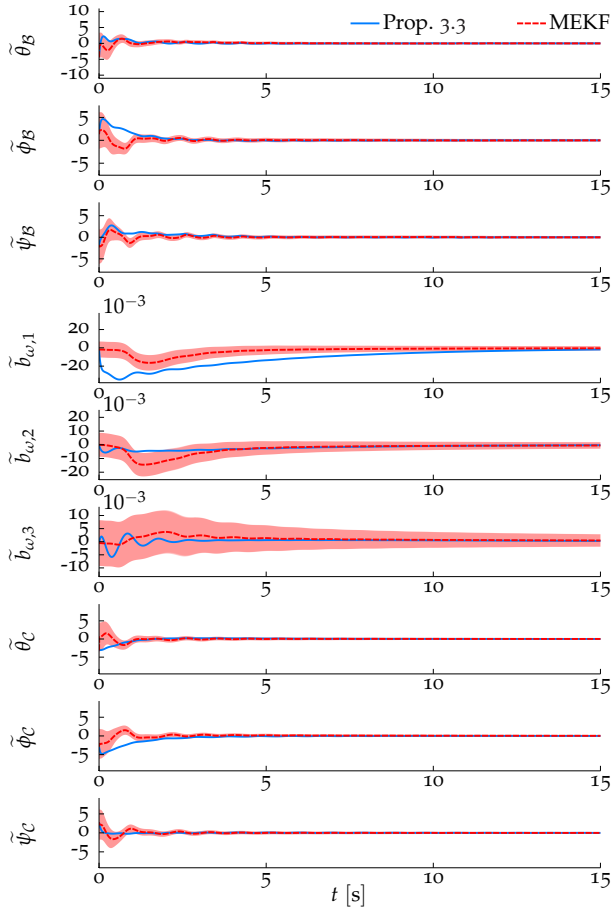
Convergence from small initial errors

This simulation analyzes the convergence of Proposition 3.3 and the MEKF for small initial errors and angular motions that satisfy condition (3.34). Orientation and c-to-IMU rotation estimates are initialized with errors from a ZMGD with variance $1.9 \cdot 10^{-3}$ $[\text{rad}]^2$, which to errors up to 7.5° of the rotation matrices. Gyro bias estimates are initialized from a ZMGD with variance $8.4 \cdot 10^{-6}$ $[\text{rad/s}]^2$, this distribution refers to errors up to 0.5[°/s]. We can see this simulation as a typical case with good initialization of the estimates. Figure 3.8 presents a typical solution obtained from repeated simulations. The curves in solid blue denote the response of Proposition 3.3 and dashed red the response for the MEKF. Moreover, the light red bounds represents the 3σ uncertainty regions computed for each variable by the MEKF. Figure 3.8 (a) displays the estimation error neglecting sensory noise and Figure 3.8 (b) displays the steady state errors with measurement noise. From top to bottom, the results correspond to body orientation error in Euler angles in ° of roll $\widetilde{\theta}_{\mathcal{B}}$, pitch $\widetilde{\phi}_{\mathcal{B}}$, yaw $\widetilde{\psi}_{\mathcal{B}}$ angles, the first, second and third components of gyro bias error $\widetilde{b}_{\omega}$ in [rad/s], and estimation errors of c-to-IMU rotation in Euler angles in ° of roll $\widetilde{\theta}_{\mathcal{C}}$, pitch $\widetilde{\phi}_{\mathcal{C}}$, and yaw $\widetilde{\psi}_{\mathcal{C}}$ angles. Figure 3.8 (c) depicts the response of the energy function

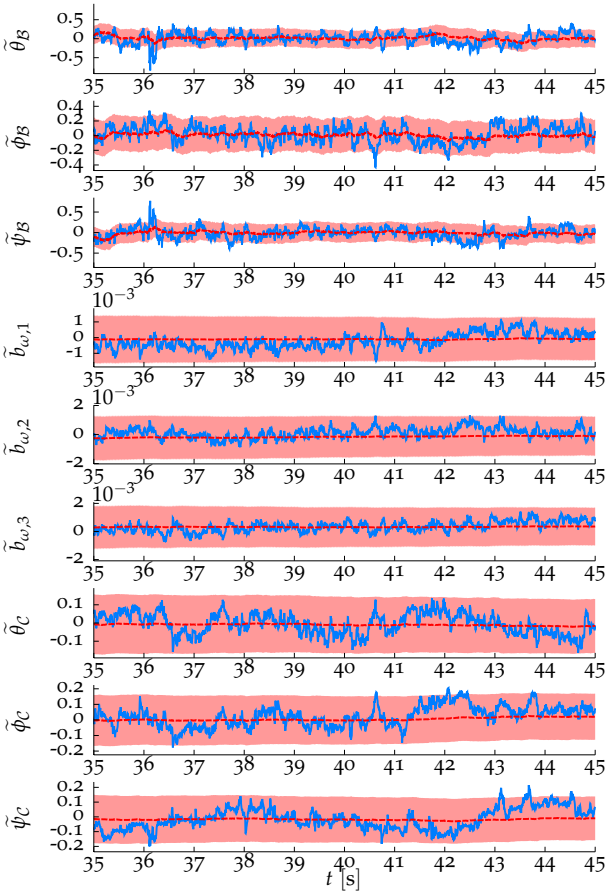
$$\mathcal{V} = \text{tr}(I_3 - \widetilde{R}_{\mathcal{C}}) + \frac{1}{k_{R_{\mathcal{B}}} k_{R_{\mathcal{C}}}} \text{tr}(I_3 - \widetilde{Q}) + \frac{1}{k_{R_{\mathcal{B}}} k_{b_\omega}} |\widetilde{b}_\omega|^2 \tag{3.56}$$

for the estimates obtained by the proposed nonlinear observer. Remark that the y-axis is shown in logarithmic scale. This energy function is strictly related to the stability proof for the estimates.

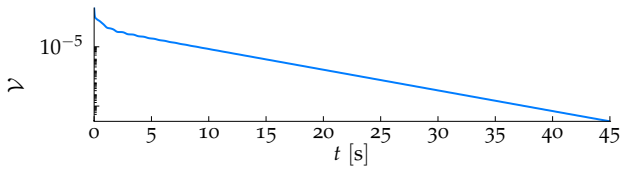
The proposed nonlinear observer and the MEKF solve this case trivially. First, let us focus on the result of Proposition 3.3. We can verify the convergence of the estimation errors to zero in Figure 3.8 (a) as predicted by the stability analysis of the observer. In this case, the rationale for the gain tuning using settling times do not hold as expected, mostly because



(a) Convergence *without* measurement noise

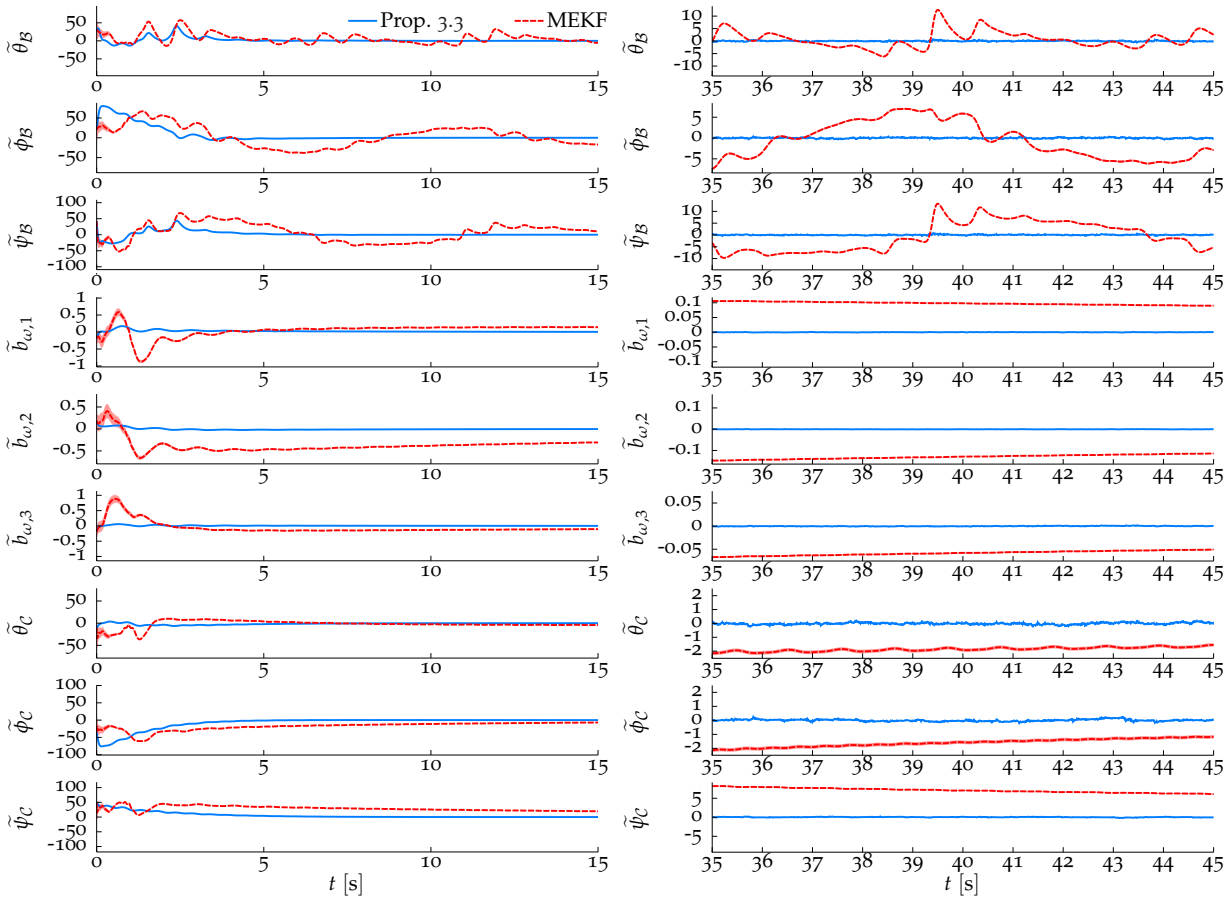


(b) Steady-state *with* measurement noise



(c) Energy function – logarithmic y-scale

Figure 3.8: Orientation, gyro bias and c-to-IMU rotation estimation with low initial errors.



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

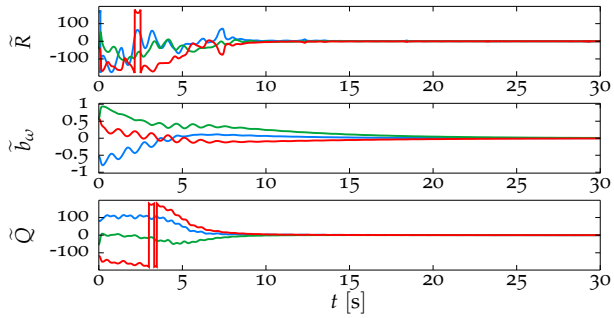
Figure 3.9: Orientation, gyro bias and c-to-IMU rotation estimation with moderate initial errors.

the observability of the system is related to the angular motion. Concerning the MEKF, differently from the previous simulation, the estimates do not converge immediately to the real solution. This effect is due to the observability conditions of the system, and the estimates do converge to the real states only after enough angular motion. We can notice likewise that the 3σ uncertainty bounds of the MEKF reduce in a slower rate than for the case where the observability is independent of the inputs. Finally, notice that uncertainties of the estimates given by Proposition 3.3 are of the same order of the uncertainty bound given by the MEKF.

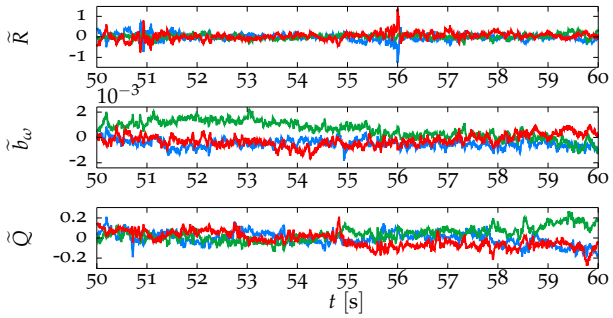
Convergence from moderate initial errors

This simulation analyzes the basin of convergence of both Proposition 3.3 and the MEKF with angular motion that satisfy condition (3.34). In this simulation, we initialize the orientation, c-to-IMU estimates and gyro bias with the same distributions as the moderate initial error simulation from Section 3.5.1. The orientation and c-to-IMU rotation estimates are initialized with an estimate drawn from a ZMGD with variance $121.8 \cdot 10^{-3}$ $[\text{rad}]^2$, *i.e.* errors from up 60°. Gyro bias estimates are initialized with random samples from a ZMGD with variance $7.6 \cdot 10^{-3}$ $[\text{rad}/s]^2$, which refers to biases up to 15 [°/s]. Figure 3.9 shows a typical result obtained from repeated simulations. Notation and displacement of the results are the same from the previous simulation.

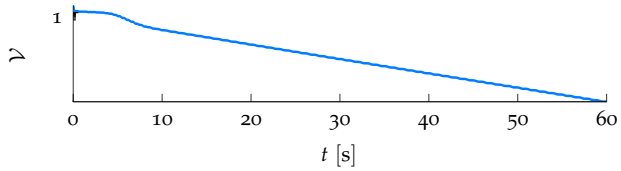
The trajectory of the estimates obtained using Proposition 3.3 perform similarly to the previous simulation, *i.e.* the estimates converge exponentially after some initial body motion. The



(a) Convergence *without* measurement noise



(b) Steady state *with* measurement noise



(c) Energy function

Figure 3.10: Orientation, gyro bias and c-to-IMU rotation estimation with large initial errors.

steady-state uncertainty of the estimates is also similar to the previous simulation, however, this behavior is masked by the poor response of the MEKF. The results for the MEKF are degraded severely by the model initial errors, since the estimates do not converge to the real states with similar response to the case with small errors. This situation may occur due to several factors. For instance, the MEKF relies strongly on the uncertainties given by the covariance matrix, since the Kalman gain is computed using this variable. Therefore, it is plausible that the covariance matrix becomes too confident before enough motion has been made in order to identify the real states. This process happens when a Kalman-like filter "learns" too well the wrong model. We can notice from Figure 3.9 (b) that the error in bias and c-to-IMU rotation are still decreasing towards zero, however, the MEKF estimates may take several minutes to reach the good solution. This is an example where the fine-tuning of a Kalman filter plays a role more important than using the real parameters of the system. On the other hand, the estimates obtained by the proposed filter are stable for moderate initial errors with the same parameters employed in the local stability analysis.

Convergence for very large initial errors

Previous simulation results support that MEKF performs well only for small errors, whilst the estimates given by Proposition 3.3 also converge for moderate errors. Clearly, if the angular motion satisfy (3.34). We can further verify that the estimates of the proposed filter are stable within an even larger domain of convergence. The next simulation initializes orientation and c-to-IMU rotation estimates are from a GD with mean of π [rad] and variance $3.4 \cdot 10^{-9}$ $[\text{rad}]^2$, *i.e.* errors close to the unstable set of passive complementary filter. Gyro bias estimates are initialized with random samples from a ZMGD with variance $41.5 \cdot 10^{-3}$ $[\text{rad}/s]^2$, *i.e.* errors up to 35 [°/s]. Figure 3.10 depicts a typical result obtained from repeated simulations, where Figure 3.10 (a) displays the error curves without measurement noise for the estimates, Figure 3.10 (b) shows the steady-state response of the estimate error considering measurement noise. From top to bottom, the results correspond to the three components components of the Euler angles of the orientation error $\widetilde{R}$, gyro bias error $\widetilde{b}_\omega$ in [rad/s] and the Euler angles in ° of c-to-IMU rotation error $\widetilde{Q}$. Figure 3.10 (c) depicts the response of the energy function $\mathcal{V}$ from (3.56). We can verify that the estimates clearly converge to the real states even for

large errors. Exponential stability, however, is obtained only after a long period, about 5 [s]. Unfortunately, we are not capable of defining accurately the domain of convergence for this case.

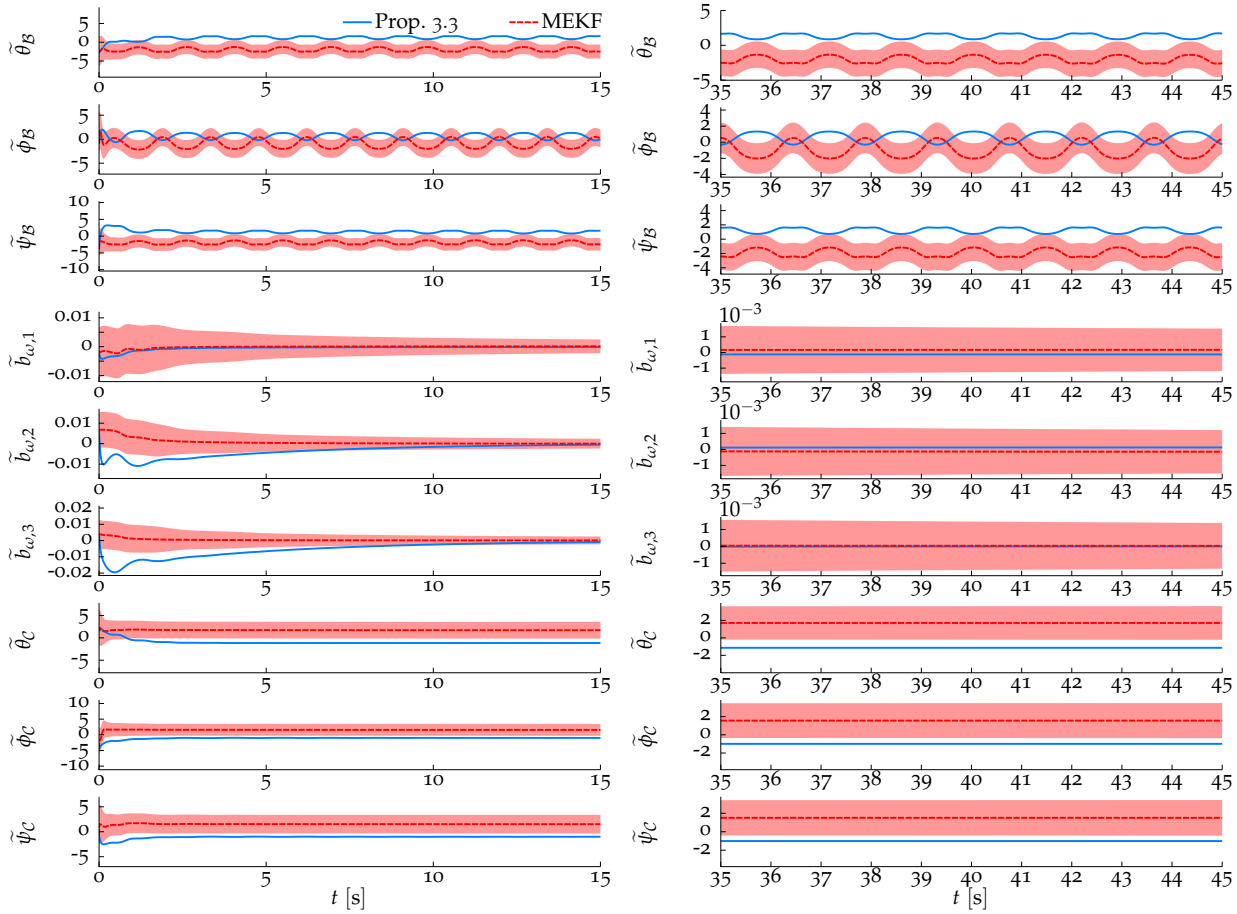
The case with unobservable motion

Convergence analysis for very large initialization errors may sound but a theoretical requirement for most systems. That property, however, is much important for systems that rely on specific observability conditions. Previous results analyzed several situations where the angular motion satisfy condition (3.34) indeed. However, the observability condition may not be always satisfied in practical situation, and filters are likely to behave unsatisfactorily under this ill-conditioned case. We now analyze the response of the estimates when the motion does not yield an observable system. Orientation and c-to-IMU rotation estimates are initialized with errors from a ZMGD with variance $1.9 \cdot 10^{-3}$ $[\text{rad}]^2$, this distribution refers to errors up to 7.5° in orientation. Gyro bias estimates are initialized from a ZMGD with variance $8.4 \cdot 10^{-6}$ $[\text{rad/s}]^2$, this distribution refers to errors up to 0.5 $[^\circ/\text{s}]$. Figure 3.11 represents a typical result from repeated simulations. Figure 3.11 (a) presents the error curves for the estimates of both filters regardless measurement noise, Figure 3.11 (b) presents the steady-state of the the same simulation with measurement noise. The orientation error given in Euler angles $[^\circ]$, gyros bias error in $[\text{rad/s}]$ and c-to-IMU rotation error in Euler angles $[^\circ]$ are displayed from top to bottom. Figure 3.11 (c) depicts the evolution of the energy function $\mathcal{V}$ from (3.56).

The estimates of both filters do not converge to the real states for this situation, as expected, since the observability condition is not satisfied. However, we can verify one interesting property for each filter. Analyzing Figure 3.11 (c), we can verify that the energy function decreases during the first second, and becomes constant afterwards. In practice, any movement of the body provides some information that allows us to identify a subset of the states. The main difference between persistently exciting and non-exciting motion is that the prior allows us to distinguish between every state, while the latter cannot distinguish the elements from a subset of the state-space. The constancy of the energy function is a direct result from the indistinguishability of a subset of the state-space. For the proposed filter, this is not a crucial problem however. Once the body performs any movement that yields instantaneous observability, the energy function will decrease as the estimates tend towards the real states. That property is valid due to the large basin of convergence of the filter.

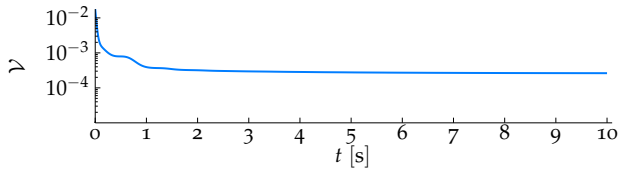
3.5.3 *Position and accelerometer bias*

We now analyze the results proposed for the estimation of the translational dynamics. Initially, let us consider to measure ${}^{\mathcal{R}}R_{\mathcal{B}}$ and ${}^{\mathcal{B}}\omega$ explicitly. This hypothesis is relaxed in Section 3.5.6. We first consider the estimation of position, linear velocity and accelerometer bias. This system is uniformly observable independently of the angular motion, therefore body motion does not play an effective role if gravitational acceleration and c-to-IMU translation are known. We compare the results obtained for the observer from Proposition 3.4 with a Kalman filter (KF) derived with the original state-affine system. The estimates of position and linear are initialized with samples from ZMGDs with variance of 11.1 $[\text{cm}]^2$ and 11.1 $[\text{cm/s}]^2$, *i.e.* position and velocity errors up to 10 [cm] and 10 [cm/s], respectively. The estimates for the accelerometer bias are initialized with samples from a ZMGD with variance $11.1 \cdot 10^{-6}$ $[\text{m/s}^2]^2$, *i.e.* biases up to 0.01 $[\text{m/s}^2]$. Figure 3.12 shows a typical result obtained from repeated simulations. The curves in solid blue refer to the response of Proposition 3.4 and dashed red the response of the KF. Furthermore, the light red areas represent the 3σ uncertainty region of the computed by the KF. Figure 3.12 (a) displays the estimates errors for a simulation without



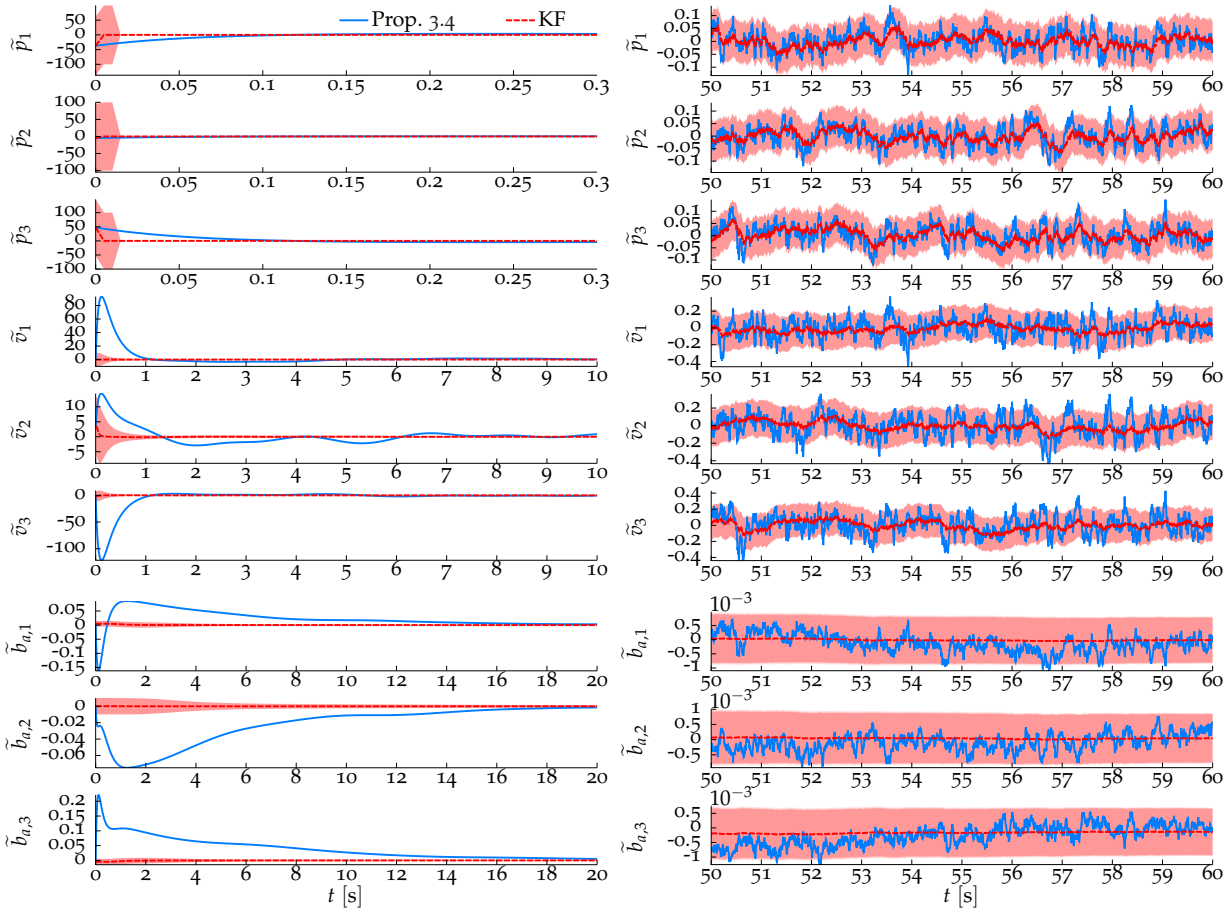
(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise



(c) Energy function - logarithmic y-scale

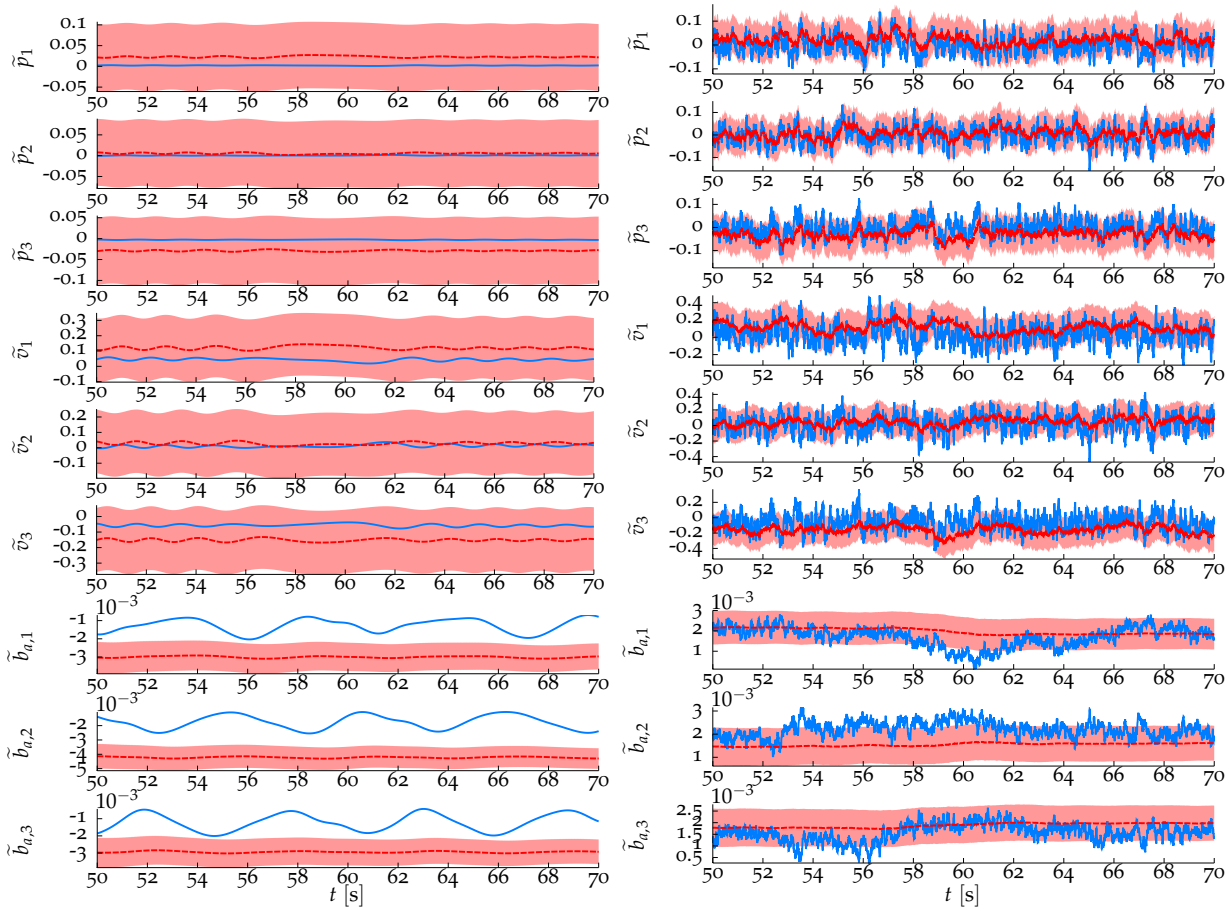
Figure 3.11: Orientation, gyro bias estimation and c-to-IMU orientation angular movements that do not satisfy condition (3.34).



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.12: Estimation of the translational dynamics – accelerometer bias.



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.13: Estimation of the translational dynamics – accelerometer bias with errors in local gravitational acceleration.

noise, while Figure 3.12 (b) shows the steady state results for the simulation with sensory noise. From top to bottom, the results correspond to the position estimation error $\widetilde{p}$ [cm], estimation error of linear velocity $\widetilde{v}$ [cm/s], and estimation error of accelerometer bias $\widetilde{b}_a$ [m/s^2].

We can verify that the estimates of both filters converge to real values of the states. As expected, the KF converges instantaneously without measurement noise. The proposed filter, on the other hand, converges in the settling times defined in Section 3.4. The uncertainties of the estimates given by the proposed observer are of the same order as the KF, however, Proposition 3.4 does not compute the covariance matrix of the system. We could obtain a bias with a steady-state "more constant" than Figure 3.12, however, this corresponds also to a slower settling time for the system.

Identifying misaligned local gravitational acceleration

In practice, the accelerometer bias is constant through long periods, other unmodeled effects can, however, mislead to an the interpretation of a time-varying bias. Let us consider, for instance, the estimation of pose, linear acceleration and accelerometer bias with an error of 2° in the direction of the gravitational acceleration in $\mathcal{R}$ frame. Figure 3.13 shows the steady state response of the estimate errors with and without measurements noise considering angular motion that satisfies Proposition 3.5 and the error in ${}^{\mathcal{R}}g$. The resulting estimates of the KF and Proposition 3.4 are biased, specifically accelerometers bias and velocity estimates. The

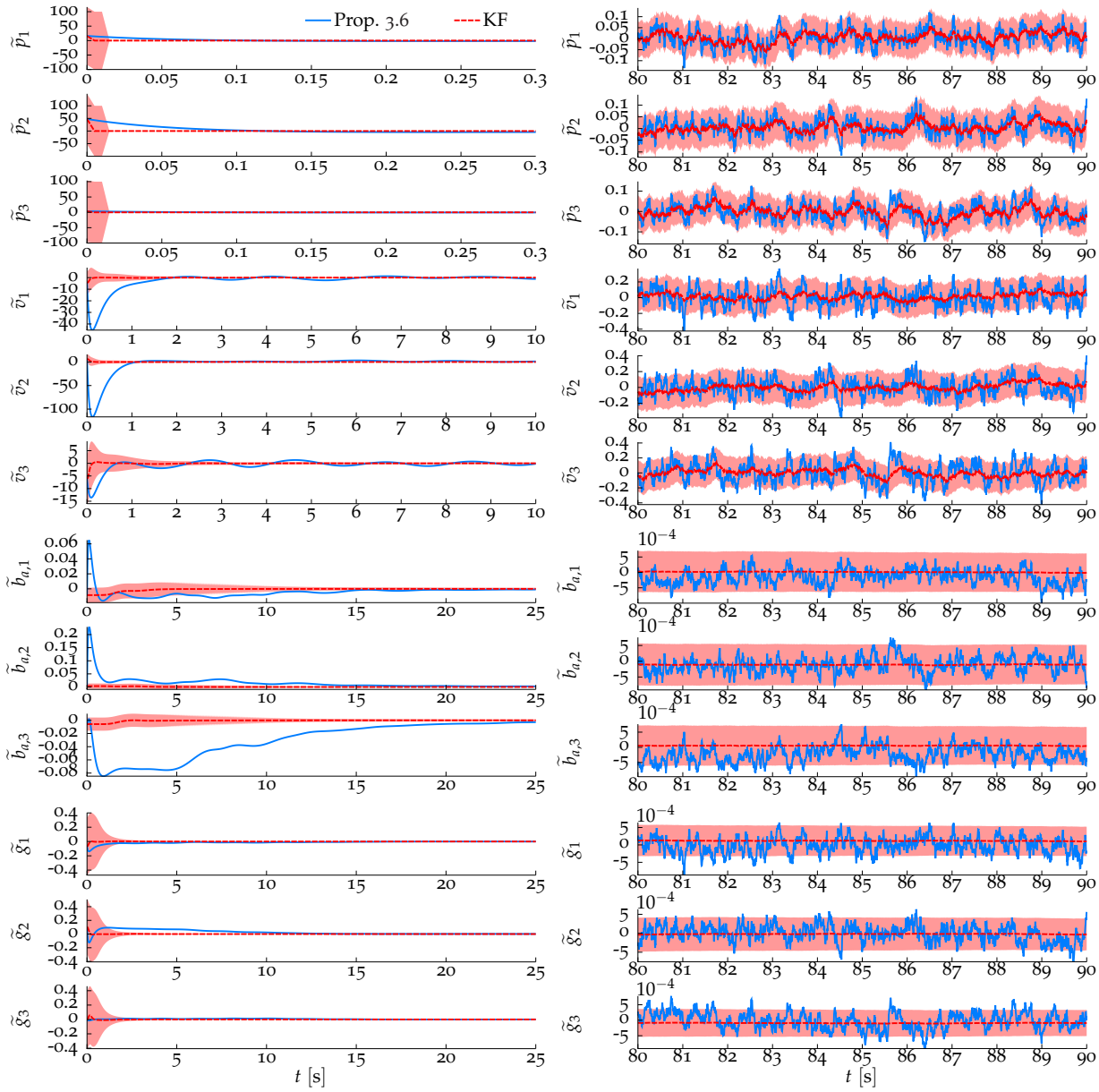
effects of the errors in each filter are slightly different however. The KF considers that the bias is constant, whilst the accelerometers add the uncertainties to the estimation of position. On the other hand, the rationale employed to tune the observer gains considers that the estimates of position estimates converge faster than the accelerometer bias. Notice that the effect due to the misaligned gravitational acceleration is easily verified for the estimates of the nonlinear filter in the case without measurement noise. Considering Gaussian noise, however, this effect is practically hidden. Therefore, a bad estimate of local gravitational acceleration can be only determined using accurate measurements, which is the case discussed in Chapter 4. We can compensate this bad effect by estimating the gravitational acceleration in $\mathcal{R}$ frame.

3.5.4 *Position, accelerometer bias and gravitational acceleration*

The next simulation considers the estimation of position, linear velocity, accelerometer bias and the local gravitational acceleration in $\mathcal{R}$ frame. This system is not uniformly observable for every angular velocity, and Proposition 3.5 enunciates an angular motion that provides uniform observability of this system. First, we verify the results obtained for the observer from Proposition 3.6 and a KF derived with the original state-affine system considering exciting motion. The estimates of position, linear velocity and accelerometer bias are drawn from the same distributions as Section 3.5.3, and the estimates of local acceleration are drawn considering a misalignment rotation R_g computed using angle-axis parametrization with the angle drawn from a ZMGD with variance $68.5 \cdot 10^{-3}[\text{rad}]^2$, *i.e.* the estimates ${}^{\mathcal{R}}\widehat{g}(0) = R_g \, {}^{\mathcal{R}}g$ are misaligned by an angle up to 45°. Figure 3.14 shows a typical result from repeated simulations. The curves in solid blue refer to the response of Proposition 3.6 and dashed red the response of the KF. Furthermore, the light red areas represent the 3σ uncertainty region of the computed by the KF. Figure 3.14 (a) displays the estimates errors for a simulation without noise, while Figure 3.14 (b) shows the steady state results for the simulation with sensory noise. From top to bottom, the results correspond to the estimation error of position $\widetilde{p}$ [cm], linear velocity $\widetilde{v}$ [cm/s], accelerometer bias $\widetilde{b}_a$ [m/s^2] and local gravitational acceleration $\widetilde{g}$ [m/s^2]. Similarly to the previous results, both filters converge to the real states in this simulation. Concerning the KF, we can remark that the estimate errors $\widetilde{b}_a$ and $\widetilde{g}$ do not converge immediately to zero for the noiseless case, instead, there is a short transient of about 2 seconds. This effect is due to the observability condition imposed by the system. The estimates of Proposition 3.6 also show these effects. Once again, notice that the proposed gain tuning using settling times present the similar uncertainty as the KF.

The case with unobservable motion

The previous simulation assumed that the angular angular motion satisfies Proposition 3.5, we now verify the effects of non-exciting inputs. We perform this simulation with the parameters initialized with the previous distributions, however, the angular motion does not satisfy the observability condition. Figure 3.15 shows a typical result from repeated simulations. The response of position and linear velocity estimation errors are presented on the left, accelerometer bias and local gravitational acceleration are shown on the right. We can remark some interesting points from this simulation. First, position and linear velocity estimation errors converge to zero independently of the angular motion. However, accelerometer bias and gravitational acceleration converge to a biased estimate. Although the non-exiting input does not directly influence the position and linear velocity error, the estimates will become erroneous as soon as the body perform an exiting motion.



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.14: Estimation of the translational dynamics – local gravitational acceleration.

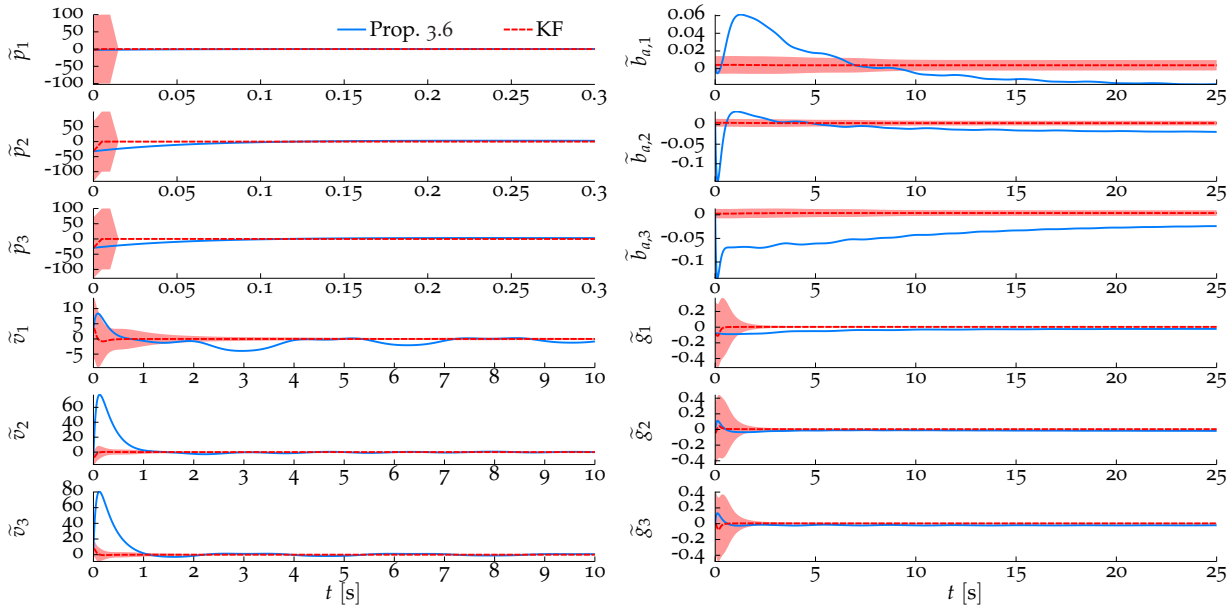


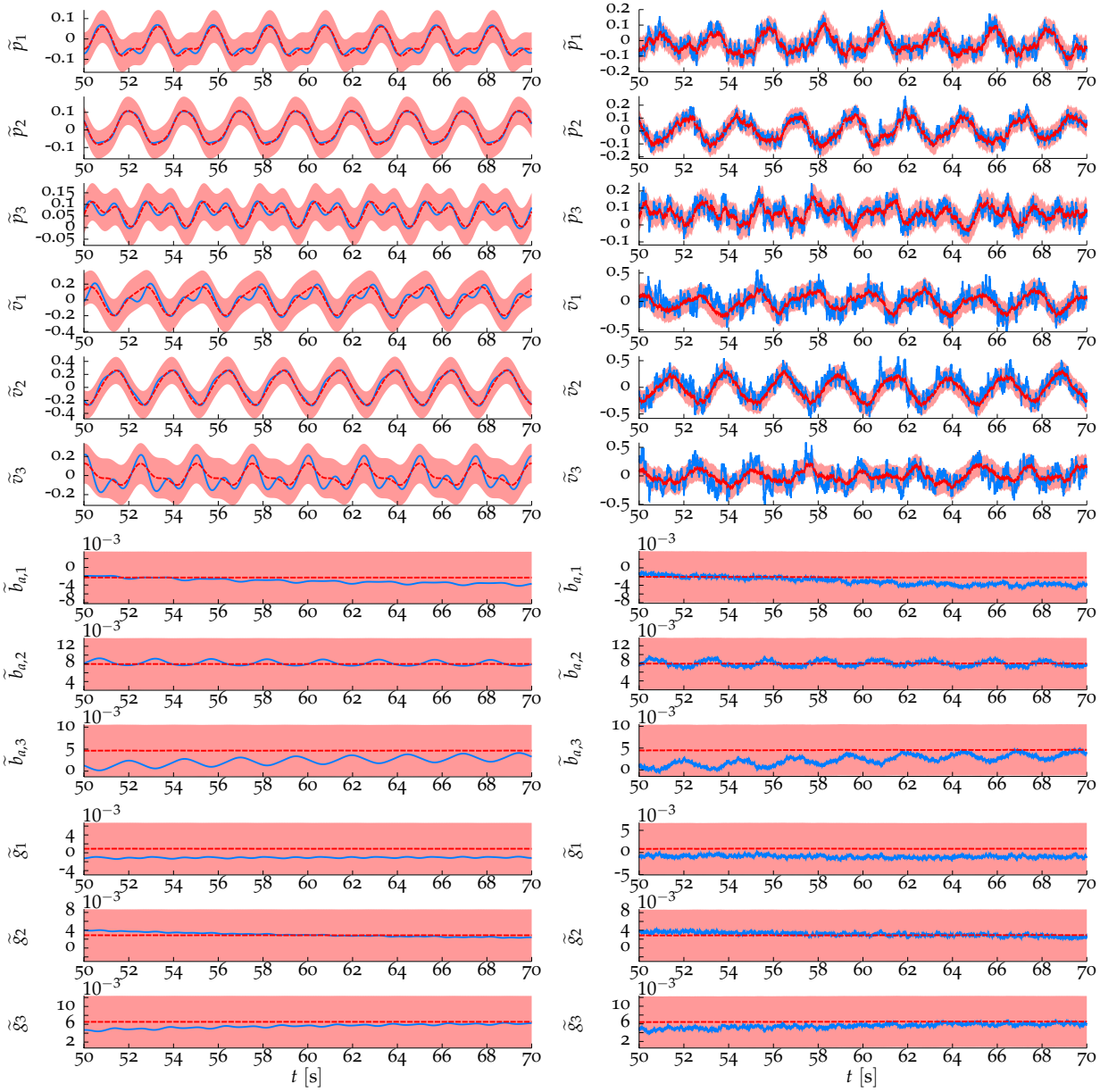
Figure 3.15: Estimation of the translational dynamics – local gravitational acceleration without persistent excitation (3.45).

Identifying misaligned c-to-IMU translation

We assume a constant model for accelerometer bias and local gravitational acceleration, additionally we consider that measurements of body position do coincide with $\mathcal{B}$ frame. In practice, the latter hypothesis is false. It can be quite difficult to measure directly and accurately the relative position of optic center of the lenses with respect to the IMU sensors in a visual inertial system. Of course, we may resort to the CAD model or rough measures using a ruler or a caliper, however we can loose in part the precision of the position measurements. Let us analyze the effect caused by errors in the estimate of c-to-IMU translation. Figure 3.16 shows the steady estimation error with and without measurement measurement for a situation considering a maximum error of 0.1 [cm] in the frame-to-frame translational displacement. The estimation errors are displaced similarly to the previous simulations. We can clearly verify that neither the estimates of the KF, nor the estimates from Proposition 3.6 converge to the actual states. Again, the effects of this unmodeled parameters impact each filter slightly differently. The KF, on one hand, considers that accelerometer bias and gravitational acceleration are constant and the errors propagate mostly on linear position and velocity. It can be quite difficult to identify this problem without a ground truth. The observer from Proposition 3.6, on the other hand, assumes that position and velocity estimates converge faster than accelerometer bias and local gravitational acceleration with the proposed gain tuning procedure. We can clearly verify that the accelerometer bias estimates oscillate in steady state, which clearly violates the constant model assumed previously. The effects from a bad estimate of c-to-IMU translation can be compensated by adding that variable to the estimator.

3.5.5 *Position and c-to-IMU translation*

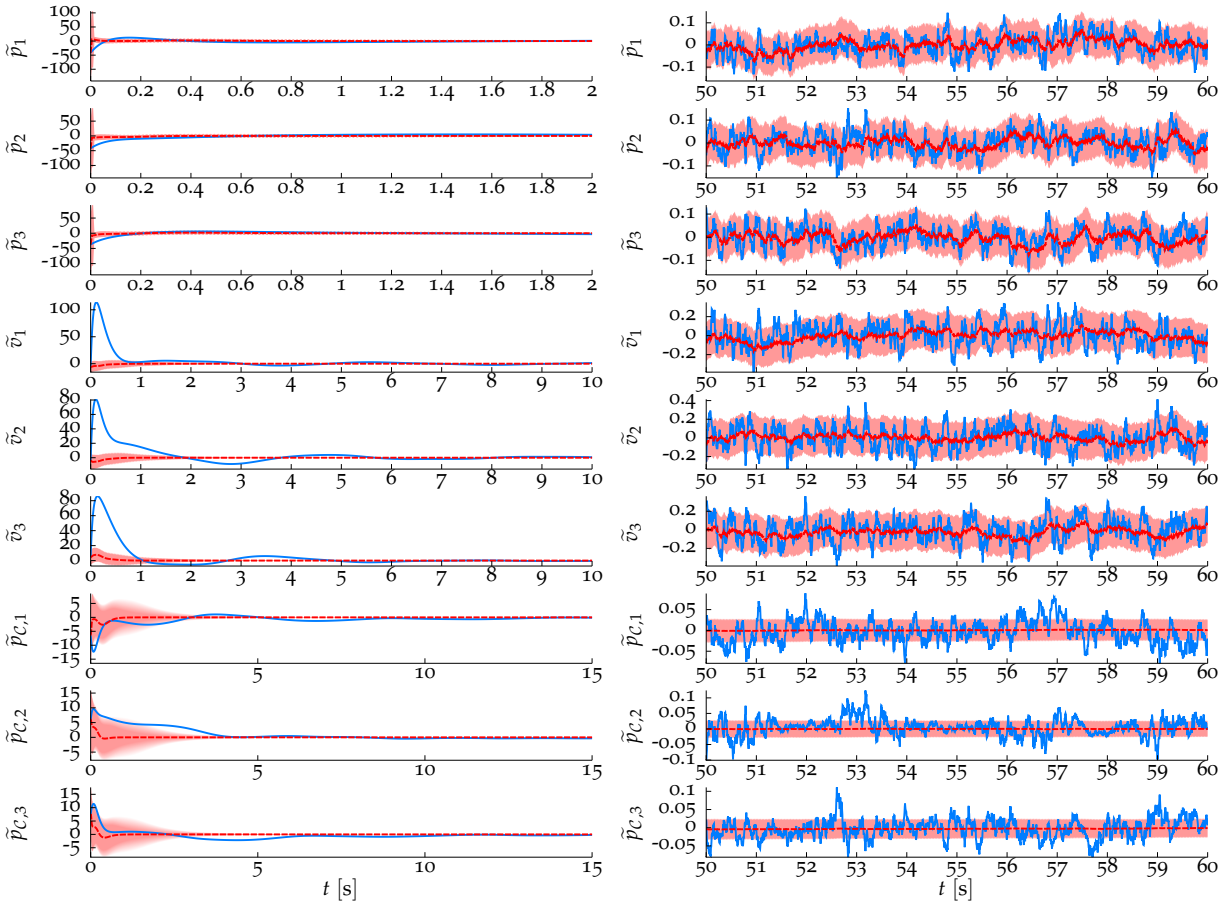
The next simulation considers the estimation of position, linear velocity and c-to-IMU translation. Proposition 3.7 assumes that accelerometer bias and the local gravitational acceleration are known, and the estimator is exponentially stable under condition 3.49. This is equivalent to the uniform condition of Proposition 3.5, that refers to the concurrent estimation of accelerometer bias and local gravitational acceleration. We verify the results obtained for the



(a) Steady-state *with* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.16: Estimation of the translational dynamics – accelerometer bias with error in c-to-IMU translation.



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.17: Position, linear velocity, and c-to-IMU translation.

observer introduced in Proposition 3.7 and a KF derived from the original state-affine system. We assume that position and linear velocity estimates are initialized from the same distribution as Section 3.5.3 and 3.5.4 and c-to-IMU translation estimates are drawn from a ZMGD with variance 11.1 $[\text{cm}]^3$, *i.e.* with a maximum error of 10 [cm]. Figure 3.17 presents a typical result from repeated simulations. The curves in solid blue denote the responses of Proposition 3.1, in dashed green the responses of Theorem 3.3, dashed red of the MEKF, and the light red areas represent the 3σ uncertainty regions provided by the MEKF for each variable. Figure 3.17 (a) presents the estimation errors for the estimates of both filters without measurement noise, while Figure 3.17 (b) displays the steady state considering measurement noises. From top to bottom, the results correspond to the estimation errors of position $\tilde{p}$ [cm], linear velocity $\tilde{v}$ [cm/s] and c-to-IMU translation $\tilde{q}$. Similarly to the previous results, the estimates of both filters converge to the values of the states. Furthermore, the system is observable under certain angular motion, which yields the estimates and uncertainties of the KF to be reduced within a transient response, instead of instantaneously. The analysis of Conjecture 3.1 is left out of the simulations, since we were unable to provide an input such that the filter is stable. We verify via actual experimental data in Chapter 4, however, that the result is comparable to the referring KF, as the observers previously discussed.

3.5.6 *Coupled estimators for orientation and position*

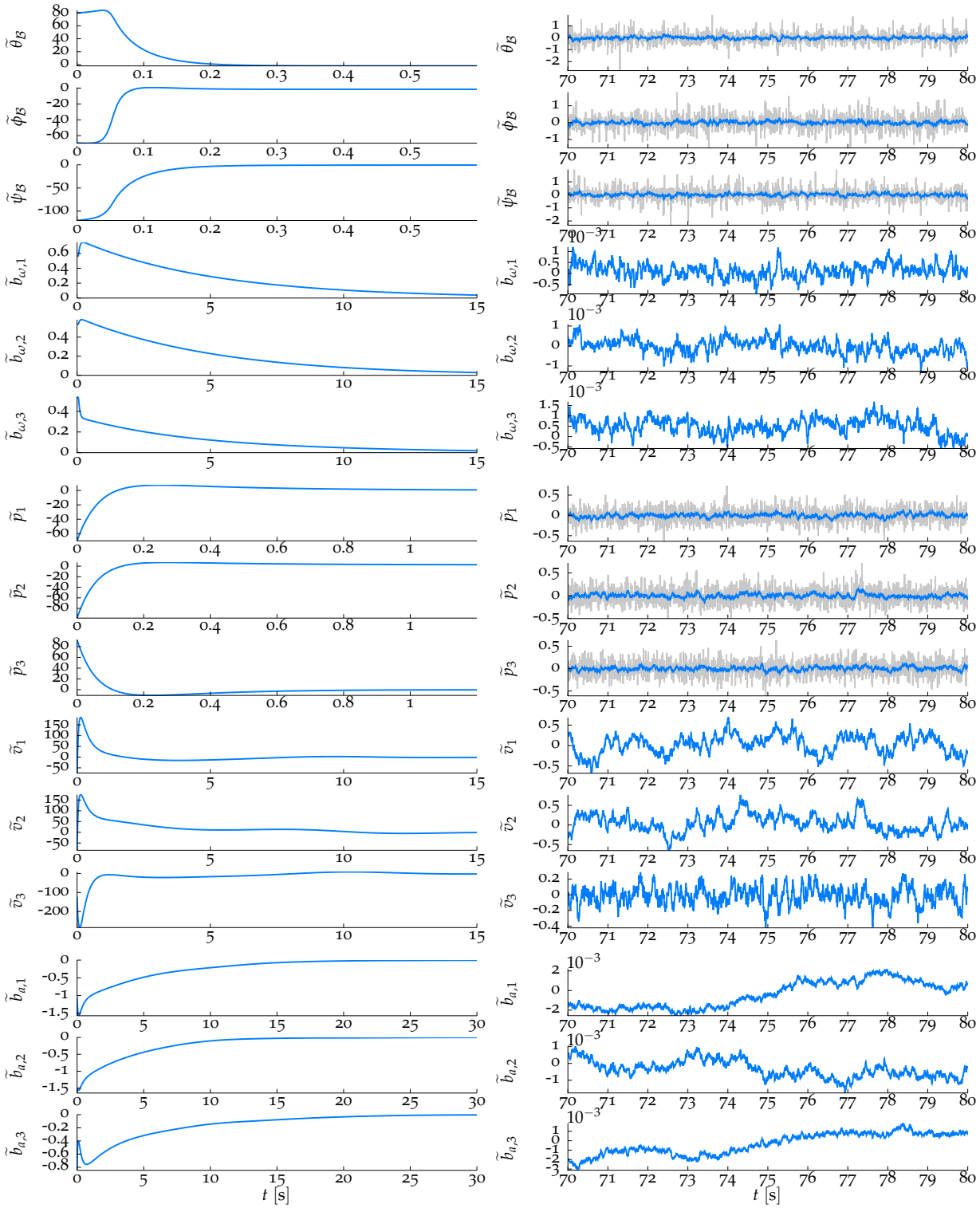
We have considered in Sections 3.5.3, 3.5.4 and 3.5.5 to measure body orientation and gyro bias explicitly. The observers introduced in Propositions 3.4, 3.6, 3.7 assume that this hypothesis holds indeed, but we have shown in respective corollaries that the estimates of a coupled pose estimator are also asymptotically stable, with domain of convergence given by the orientation observer. Let us investigate the effects of the coupled estimation in next simulation. For the sake of simplicity, we consider the case of pose estimation (orientation and position) with linear velocity, gyro and accelerometer bias estimation. This system is uniformly observable independently of the the dynamics, we thus consider the nonlinear observer from Proposition 3.1 for orientation estimation, and Corollary 3.5 for the position estimation. The orientation estimates are initialized with large errors from Sections 3.5.1 and 3.5.3. Figure 3.18 shows a typical result obtained from repeated simulations. The curves in solid blue represent the estimation error for the coupled estimation, the the solid gray lines present the explicit pose measurements. Figure 3.18 (a) presents the convergence results without measurement noise and Figure 3.18 (b) the steady-state with measurement noise. From top to bottom, the results correspond to the estimation error for body orientation error in Euler angles in $^\circ$ for roll $\widetilde{\theta}_{\mathcal{B}}$, pitch $\widetilde{\phi}_{\mathcal{B}}$, yaw $\widetilde{\psi}_{\mathcal{B}}$ and gyro bias error for the first, second and third components of $\widetilde{b}_\omega$ in [rad/s], position estimation error $\widetilde{p}$ [cm], estimation error of linear velocity $\widetilde{v}$ [cm/s], and estimation error of accelerometer bias $\widetilde{b}_a$ [m/s^2]. We can clearly verify that the estimates converge to the real states. In this extreme situation, however, the convergence of linear velocity and accelerometer bias estimates is slightly slower than in Section 3.5.3. This result holds also with c-to-IMU rotation estimation, and the estimation of translational parameters such as local gravitational acceleration and c-to-IMU translation. We consider this coupled approach in the experiments of Chapter 4.

3.6 CONCLUSION

This chapter addressed the data fusion process for pose estimation. In order to improve the data fusion, we estimate several parameters of the system, *e.g.* gyroscope and accelerometer bias, local gravitational accelerometers and camera-to-IMU rotation and translation.

We discussed the estimation of orientation and translational dynamics independently. First, we employed tools from nonlinear control to address the orientation dynamics. We provided three novel nonlinear observers on the group of rotation matrices. The first two observers are extensions of the passive complementary filter that ensure (almost) global asymptotic stability with domain of convergence independent of the magnitude of the innovation gains. We study the estimation of orientation, gyro bias and c-to-IMU rotation afterwards. The observability of this system is strictly related to angular body motion, and our observability analysis provides an explicit expression of the movements that ensure observability. We further propose an extension for the passive complementary filter so as to estimate c-to-IMU rotation also. We can ensure that the estimation error exponentially stability under specific (and specified) conditions. The advantages of our filter against classical techniques are twofold. First, the computation of the proposed observers is simpler, since we can evaluate the innovation terms using instantaneous information only, *i.e.* there is no need to compute the integrals to obtain the Kalman gains. Secondly, even though classical techniques are also locally exponential stable, we could verify via simulation results that the domain of convergence of the proposed method is larger than the one provided by Kalman-based techniques.

The translational dynamics is analyzed as a linear time-varying system. Initially, we studied the estimation of position, linear velocity and accelerometer bias. This system is uniformly observable independently of the body motion. We provided a globally exponentially stable



(a) Convergence *without* measurement noise

(b) Steady-state *without* measurement noise

Figure 3.18: Coupled estimation of orientation and position.

nonlinear observer computed using instantaneous information, and studied the problem of concurrent estimation of position, linear velocity, accelerometer bias with local gravitational acceleration and c-to-IMU translation. The observability of these systems is related to the angular motion, and our observability analysis provided explicit motion for (part of) the systems. The resulting Kalman filter for these systems is also globally exponentially stable, however, we provided observers that use innovation terms that use only instantaneous information. In this way, the resulting filters are simpler than Kalman-based techniques that must compute the covariance matrices additionally to the states themselves.

We also conjectured an observer for the estimation of the position, linear velocity and the three parameters that interact with these variables. This conjecture, however, is not endowed with a proof of stability. Concluding the chapter, we propose a procedure for tuning the gains of the nonlinear observers based on the settling times of the estimate errors. Furthermore, several simulation results endorse the simplicity and high performance of the proposed techniques compared to Kalman-based filtering.

The proposed nonlinear observers can improve the performance of direct visual tracking by providing accurate initialization after large displacements. The next Chapter presents the results obtained for a visuo-inertial sensor employing the visual methods introduced in Chapter 2 with the nonlinear observers previously introduced.

4

RESULTS ON VISUO-INERTIAL POSE ESTIMATION

This chapter addresses the validation of visuo-inertial pose estimation using the tools presented in this thesis. We discussed different direct visual tracking methods in Chapter 2, and verified that these methods are indeed accurate solutions to compute relative pose of a moving camera. The downside of all gradient-based direct visual tracking methods, however, is the need of an initialization close enough to the optimal solution. Inertial sensors can measure incremental displacements in faster rates than the camera. We introduced multiple nonlinear observers in Chapter 3. These observers take the inertial and pose data into account for pose estimation with concurrent identification of several parameters of the system.

4.1 VISUO-INERTIAL SENSOR

The experimental data analyzed in this chapter was obtained using the sensor depicted in Figure 4.1. This sensor consists of a xSens MTi–G IMU, with an AVT Stingray 125B camera. The IMU consists of accelerometers and angular rate gyroscopes that are capable of providing specific acceleration and angular velocity at 200 [Hz]. The camera provides a video stream of 40 images per second with resolution of 800×600 [pixel]. The samples obtained from the camera and IMU are synchronized. We can define two frames, $\mathcal{B}$ and $\mathcal{C}$ associated to the the IMU and camera respectively. Notice that we can easily determine an estimate of the sensor-to-sensor orientation by inspection

$$ {}^{\mathcal{B}}R_{\mathcal{C}} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, $$

and the translational displacement depends on the size of the lenses. We can either neglect this displacement, assuming ${}^{\mathcal{B}}p_{\mathcal{R}} = 0_{3\times 1}$, or try to measure it using a caliper. However, these are

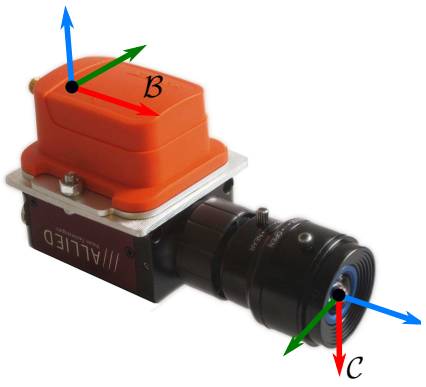


Figure 4.1: Inertial-visual sensor used in the experiments.

only rough estimates of the frame-to-frame calibration since we are unsure about the actual location of the optical center.

The visual-inertial system considered in the experiments employed Theia SY125M lenses. These are ultra wide-angle lenses that present practically no distortion at the the resolution we employed. A previous calibration procedure to obtain the intrinsic parameters yielded the matrix

$$K_f = \begin{bmatrix} 448.85088 & 0 & 394.30650 \\ 0 & 450.26420 & 292.82383 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.1}$$

4.2 POSE ESTIMATION AND DIRECT VISUAL TRACKING

We discussed in Section 2.6 how to employ the homography warp in direct visual tracking. That solution is interesting when we know previously the texture of a surface. More specifically, let us suppose that the reference image $\boldsymbol{I}_{\mathrm{R}}$ is obtained with the optic center of the camera coincident to a frame $\mathcal{R}$, and $\boldsymbol{I}_{\mathrm{C}}$ is obtained with the optic center of the camera coincident to another frame $\mathcal{C}$ at time t. We can express relative the pose ${}^{\mathcal{R}}(R,p)_{\mathcal{C}}$ by the Euclidean homography (1.22),

$${}^{\mathcal{C}}G_{\mathcal{R}} = {}^{\mathcal{C}}R_{\mathcal{R}} + ({}^{\mathcal{R}}z^{-1}){}^{\mathcal{C}}p_{\mathcal{R}}{}^{\mathcal{R}}\boldsymbol{n}^T,$$

where ${}^{\mathcal{R}}\boldsymbol{n}^T$ and ${}^{\mathcal{R}}z^{-1}$ denote the scaled normal vector and distance of the plane in $\mathcal{R}$ frame. Moreover, the projective Homography, *c.f.* Section 1.3.4,

$${}^{\mathcal{C}}H_{\mathcal{R}} \propto K_f{}^{\mathcal{C}}G_{\mathcal{R}}{K_f}^{-1}$$

is given as a function of the pose assuming that the intrinsic parameters K_f of the camera, and the normal vector ${}^{\mathcal{R}}n = ({}^{\mathcal{R}}z^{-1}){}^{\mathcal{R}}\boldsymbol{n}$. The relation to the Euclidean homography ${}^{\mathcal{R}}G_{\mathcal{C}}$ is straightforward, and we can decompose the Euclidean homography in components of orientation, scaled position, and unitary normal vector, *c.f.* Section 1.3.4. Such decomposition yields two possible solutions, and we can identify the correct one if the normal vector the scene is known. Remark, however, that, in this case, we solve an optimization in 8 dimensions in order to obtain a 6 dimensional pose. The extra two degrees of freedom are a drawback that increases the likelihood of being trapped in local minima of the similarity functions.

Assuming that we know the actual normal vector and intrinsic parameters K_f, we can reduce the dimension of the optimization by redefining the warp function in order to employ the actual pose instead of the projective Homography. For instance, we can define an application ζ for $\mathbb{X} = \mathbb{SO}(3) \times \mathbb{R}^3$, *i.e.* $X = {}^{\mathcal{C}}(R,p)_{\mathcal{R}}$,

$$\begin{aligned} \zeta \colon \mathbb{X} \times \mathbb{R}^3 \times \mathbb{M}(3) &\to \mathbb{SL}(3), \\ (X, {}^{\mathcal{R}}n, K_f) &\mapsto \zeta(X, {}^{\mathcal{R}}n, K_f), \end{aligned}$$

directly from (4.2), *i.e.* ,

$$\zeta(X, {}^{\mathcal{R}}n, K_f) = K_f \frac{{}^{\mathcal{R}}R_{\mathcal{C}} + {}^{\mathcal{C}}p_{\mathcal{R}}{}^{\mathcal{R}}n^T}{\sqrt[3]{1 + {}^{\mathcal{R}}n^T{}^{\mathcal{R}}R_{\mathcal{C}}{}^{\mathcal{C}}p_{\mathcal{R}}}} {K_f}^{-1}.$$

Consequently, we can define a warp $w_{\mathrm{E}}(X, \boldsymbol{p}) = w_{\mathrm{P}}\big(\zeta(X, {}^{\mathcal{R}}n, K_f), \boldsymbol{p}\big)$. Notice, however, that this is not actually a group action, since $w_{\mathrm{E}}(X_1X_2, p) \neq w_{\mathrm{E}}(X_1, w_{\mathrm{E}}(X_2, p))$, $\forall\ X_1, X_2 \in \mathbb{X}$.

Algorithm 2 Integrate estimates

Require: IMU sampling period Δ_I, gyro ${}^{\mathcal{B}}\omega_y[n]$ and accelerometer ${}^{\mathcal{B}}a_y[n]$ measurements.
Auxiliary variables:

$$\widehat{\omega} = {}^{\mathcal{B}}\omega_y[n] - \widehat{b}_\omega[n-1], \quad \widehat{a} = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n-1]\big({}^{\mathcal{B}}a_y[n] - \widehat{b}_a[n-1]\big) + \widehat{g}[n-1],$$
$$R_\omega = I_3 + \sin(|\widehat{\omega}|\Delta_I)\mathrm{S}\Big(\tfrac{\widehat{\omega}}{|\widehat{\omega}|}\Big) + \big(1 - \cos(|\widehat{\omega}|\Delta_I)\big)\mathrm{S}\Big(\tfrac{\widehat{\omega}}{|\widehat{\omega}|}\Big)^2$$

Integrate pose estimates:

$$\begin{aligned}
{}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n] &= {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n-1]R_\omega, \quad \widehat{b}_\omega[n] = \widehat{b}_\omega[n-1], \quad {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n] = {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n-1], \\
{}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}[n] &= {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}[n-1] + {}^{\mathcal{R}}\widehat{v}[n-1]\Delta_I + \widehat{a}\tfrac{\Delta_I{}^2}{2}, \qquad {}^{\mathcal{R}}\widehat{v}[n] = {}^{\mathcal{R}}\widehat{v}[n-1] + \widehat{a}\Delta_I, \\
\widehat{b}_a[n] &= \widehat{b}_a[n-1], \qquad {}^{\mathcal{R}}\widehat{g}[n] = {}^{\mathcal{R}}\widehat{g}[n-1], \quad {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}[n] = {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}[n-1].
\end{aligned}$$

For the sake of simplicity, we only assume that the group action holds only locally, *i.e.* close enough to X_1, and the geometric Jacobian yields

$$\partial_x \boldsymbol{w}_{\mathrm{E}}(\phi(x), \boldsymbol{p}_i) = \partial_H w_{\mathrm{P}}(I_3, \boldsymbol{p}_i) \cdot \partial_X \zeta(\phi(x), {}^{\mathcal{R}}n, K_f) \cdot \partial_x \phi(0)$$

with $\partial_H w_{\mathrm{P}}(I_3, \boldsymbol{p}_i)$ given by (2.36), and $\partial_X \zeta(\phi(x), {}^{\mathcal{R}}n, K_f) \cdot \partial_x \phi(0)$ is computed as (Benhimane, 2006, pp. 75-76). Notice that invariance properties (2.36) for group actions do not hold anymore, therefore the geometric Jacobian is not constant and must be recomputed at each iteration.

4.3 MULTI-RATE DATA FUSION

The main results of Chapter 3 concern nonlinear observers with stability proof for pose estimation. We employ tools from nonlinear control theory in order to determine the conditions and stability properties, and the ensemble of results is obtained for continuous time. However, as we have introduced in this chapter, the IMU and camera provide samples of the data different frequencies. Hence, the proposed nonlinear observers should be adapted to handle other three requisites:

- Discrete time integration;
- Discrete time update;
- Forecast–update decoupling.

The first requisite correspond to the periods of time during which we obtain only inertial measurements. During these periods, there is no pose measurement in order to compute the innovation terms. Algorithm 2 describes the integration procedure:

Concerning the translational dynamics, we can compute the integrals using Euler's method directly, however, the same method is not valid for the rotational dynamics. Instead, we can compute the integral explicitly

$$\int_{t_n-\Delta_I}^{t_n} \dot{R}\, dt = \int_{t_n-\Delta_I}^{t_n} R(t)\mathrm{S}(\omega(t))\, dt,$$

where Δ_I denotes the sampling period of the IMU, and, assuming that the angular rate is sampled by a zero-order holder, then $\omega(t) = \omega$ during the time interval $[t_0 - \Delta, t_0)$ and we can compute

$$R[n] = R[n-1]\exp\{\mathrm{S}(\omega\Delta_I)\},$$

Algorithm 3 Update estimates

Require: IMU measurements: gyroscope ${}^{\mathcal{B}}\omega_y[n]$.
Require: Camera measurements: ${}^{\mathcal{R}}p_{\mathcal{C}}[n]$, ${}^{\mathcal{R}}R_{\mathcal{C}}[n]$.
Auxiliary variables:

$$\widehat{\omega} = {}^{\mathcal{B}}\omega_y[n] - \widehat{b}_\omega[n],\ {}^{\mathcal{R}}\widehat{R}_{\mathcal{C}}[n] = {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n]\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n],\ {}^{\mathcal{R}}\widehat{p}_{\mathcal{C}}[n] = {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}[n] + {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n]\,{}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}[n].$$

Compute errors:

$$\widetilde{R}_{\mathcal{C}} = {}^{\mathcal{R}}R_{\mathcal{C}}[n]\,{}^{\mathcal{C}}\widehat{R}_{\mathcal{R}}[n], \qquad \widetilde{p}_{\mathcal{C}} = {}^{\mathcal{R}}p_{\mathcal{C}}[n] - {}^{\mathcal{R}}\widehat{p}_{\mathcal{C}}[n].$$

Compute innovation terms:

$$\begin{aligned}
\alpha_{R_{\mathcal{B}}} &= k_{R_{\mathcal{B}}}\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}[n]\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\big) - k_{R_{\mathcal{C}}}\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}[n]\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\,{}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\widehat{\omega},\\
\alpha_{\omega} &= -k_{\omega}\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}[n]\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\big), \qquad \alpha_{R_{\mathcal{C}}} = k_{R_{\mathcal{C}}}\,{}^{\mathcal{C}}\widehat{R}_{\mathcal{R}}\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\,{}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\widehat{\omega},\\
\alpha_{p_{\mathcal{B}}} &= k_{p_{\mathcal{B}}}\widetilde{p}_{\mathcal{C}} + k_{p_{\mathcal{C}}}\mathrm{S}\big({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n]\widehat{\omega}\big)\widetilde{p}_{\mathcal{C}}, \qquad \alpha_{v} = k_{v}\widetilde{p}_{\mathcal{C}},\\
\alpha_{a} &= -k_{a}\big(I_3 + \tfrac{1+k_{p_{\mathcal{C}}}}{k_{p_{\mathcal{B}}}}\mathrm{S}(\widehat{\omega})\big)\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}[n]\widetilde{p}_{\mathcal{C}},\\
\alpha_{g} &= k_{g}\widetilde{p}_{\mathcal{C}} + k_{g}\tfrac{k_{p_{\mathcal{C}}}}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n]\widehat{\omega})\widetilde{p}_{\mathcal{C}}, \qquad \alpha_{p_{\mathcal{C}}} = -k_{p_{\mathcal{C}}}\mathrm{S}(\widehat{\omega})\,{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}[n]\widetilde{p}_{\mathcal{C}},
\end{aligned}$$

Compute innovation rotation matrices:

$$\begin{aligned}
R_{\alpha_{R_{\mathcal{B}}}} &= I_3 + \sin(|\alpha_{R_{\mathcal{B}}}|\Delta_C)\mathrm{S}\Big(\tfrac{\alpha_{R_{\mathcal{B}}}}{|\alpha_{R_{\mathcal{B}}}|}\Big) + \big(1-\cos(|\alpha_{R_{\mathcal{B}}}|\Delta_C)\big)\mathrm{S}\Big(\tfrac{\alpha_{R_{\mathcal{B}}}}{|\alpha_{R_{\mathcal{B}}}|}\Big)^2,\\
R_{\alpha_{R_{\mathcal{C}}}} &= I_3 + \sin(|\alpha_{R_{\mathcal{C}}}|\Delta_C)\mathrm{S}\Big(\tfrac{\alpha_{R_{\mathcal{C}}}}{|\alpha_{R_{\mathcal{C}}}|}\Big) + \big(1-\cos(|\alpha_{R_{\mathcal{C}}}|\Delta_C)\big)\mathrm{S}\Big(\tfrac{\alpha_{R_{\mathcal{C}}}}{|\alpha_{R_{\mathcal{C}}}|}\Big)^2,
\end{aligned}$$

Update pose:

$$\begin{aligned}
&{}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n] = {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}[n]R_{\alpha_{R_{\mathcal{B}}}}, \qquad \widehat{b}_\omega[n] = \widehat{b}_\omega[n] + \alpha_\omega\Delta_C,\ {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n] = {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}[n]R_{\alpha_{R_{\mathcal{B}}}},\\
&{}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}[n] = {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}[n] + \alpha_{p_{\mathcal{B}}}\Delta_C,\ {}^{\mathcal{R}}\widehat{v}[n] = {}^{\mathcal{R}}\widehat{v}[n] + \alpha_v\Delta_C,\\
&\widehat{b}_a[n] = \widehat{b}_a[n] + \alpha_a\Delta_C, \qquad {}^{\mathcal{R}}\widehat{g}[n] = {}^{\mathcal{R}}\widehat{g}[n] + \alpha_g\Delta_C,\ {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}[n] = {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}}[n] + \alpha_{p_{\mathcal{C}}}\Delta_C.
\end{aligned}$$

where $n = t/\Delta_I$. The exponential matrix of $\mathbb{SO}(3)$ is computed using Rodrigues rotation formula:

$$\exp\{\mathrm{S}(\omega(t)\Delta_I)\} = I_3 + \sin(|\omega(t)\Delta_I|)\mathrm{S}\Big(\tfrac{\omega(t)}{|\omega(t)|}\Big) + \big(1-\cos(|\omega(t)\Delta_I|)\big)\mathrm{S}\Big(\tfrac{\omega(t)}{|\omega(t)|}\Big)^2.$$

Remark that we must be careful with numerical precision when $|\omega(t)| \approx 0$.

The second requisite corresponds to the update on the estimates due to the innovation terms defined by the nonlinear observers. The integrals in this case, are computed with respect to the sampling period Δ_C between two camera frames. We summarize Proposition 3.3 and Conjecture 3.7 in Algorithm 3. Notice that changing the filter towards Proposition 3.1 or Corollary 3.3 is direct, *i.e.* we do not intend to estimate ${}^{\mathcal{B}}R_{\mathcal{C}}$, ${}^{\mathcal{B}}p_{\mathcal{C}}$, or ${}^{\mathcal{R}}g$, we need simply to update the corresponding gain $k_{R_{\mathcal{C}}}$, $k_{p_{\mathcal{C}}}$ or k_g to zero.

The third requisite corresponds to the different frequencies for the integration of IMU and visual data. We summarize a general algorithm for visuo-inertial fusion in Algorithm 4.

Algorithm 4 Visuo-inertial fusion

Require: initial state estimate $\widehat{x}[0] = ({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}, \widehat{b}_{\omega}, {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}}, {}^{\mathcal{R}}\widehat{p}_{\mathcal{B}}, {}^{\mathcal{R}}\widehat{v}, \widehat{b}_{a}, {}^{\mathcal{R}}\widehat{g}, {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}})$.
Thread – IMU
for all new measurement $u[n] = ({}^{\mathcal{B}}\omega_y, {}^{\mathcal{B}}a_y)$ **do**
$\left(\widehat{x}[n], u[n]\right)$: integrate estimates $\rightarrow \widehat{x}[n+1]$
$n = n + 1$.
end for

Thread – Camera
for all new image **do**
verify corresponding sample $n_{\mathcal{C}} < n$
$\widehat{x}[n_{\mathcal{C}}]$: compute pose from using visual tracker $\rightarrow x[n_{\mathcal{C}}]$
$\left(\widehat{x}[n_{\mathcal{C}}], u[n_{\mathcal{C}}], x[n_{\mathcal{C}}]\right)$: update estimates $\rightarrow \widehat{x}[n_{\mathcal{C}}]$
for all $n_i \in \{n_{\mathcal{C}}, \ldots, n\}$ **do**
$\left(\widehat{x}[n_i], u[n_i]\right)$: integrate estimates $\rightarrow \widehat{x}[n_i + 1]$
end for
end for

4.3.1 *Gain tuning and observability conditions*

Section 3.4 presented a technique to tune the gains of the nonlinear observer based using settling times of the estimates. However, this method holds only for the cases where the system is uniformly observable independently of the motion. That hypothesis is not valid, for instance, for the cases introduced in Sections 3.2.2, 3.3.2 and 3.3.3. These systems have observability conditions based on the angular motion of the body. Moreover, the respective observers, Proposition 3.3, Proposition 3.6 and Proposition 3.7, also require sufficient motion to obtain exponential stability.

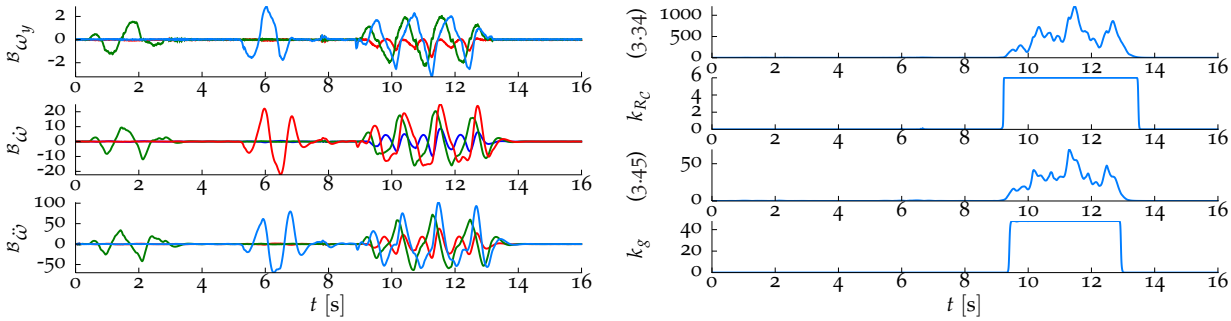
We propose to use variable gains $\overline{k}_{R_{\mathcal{C}}} = k_{R_{\mathcal{C}}} m_{\delta}(t)$, $\overline{k}_{g} = k_{g} m_{\epsilon}(t)$ and $\overline{k}_{p_{\mathcal{C}}} = k_{p_{\mathcal{C}}} m_{\epsilon}(t)$, modulated by some $\delta(t) = |{}^{\mathcal{B}}\dot{\omega}(t) \times {}^{\mathcal{B}}\ddot{\omega}(t)|$ and $\epsilon(t) = |{}^{\mathcal{B}}\omega(t) \times {}^{\mathcal{B}}\dot{\omega}(t)|$. This choice allows the application of the innovation terms for C-to-IMU calibration and local gravitational acceleration only if the observability condition is satisfied. The terms that provide sufficient observability conditions cannot be directly measured, hence we design a secondary filter to identify this condition from gyro measurements. We consider an approximate model of constant angular jerk, *i.e.* ${}^{\mathcal{B}}\dddot{\omega} = 0$ and use a linear Kalman filter with ${}^{\mathcal{B}}\widehat{\omega}$, ${}^{\mathcal{B}}\widehat{\dot{\omega}}$, and ${}^{\mathcal{B}}\widehat{\ddot{\omega}}$ as states, and ${}^{\mathcal{B}}\omega_y$ as measurement. The success of this filter relies on the fact that the bias b_{ω} is constant for relatively long periods of time, therefore this variable does not influence the evaluation of ${}^{\mathcal{B}}\widehat{\dot{\omega}}$ and ${}^{\mathcal{B}}\widehat{\ddot{\omega}}$ instantaneously. Additionally, the goal of such estimation is not to estimate these variables accurately, yet to identify when $|{}^{\mathcal{B}}\dot{\omega} \times {}^{\mathcal{B}}\ddot{\omega}| > \delta$, and $|{}^{\mathcal{B}}\omega \times {}^{\mathcal{B}}\dot{\omega}| > \epsilon$, for $\delta > 0$ and $\epsilon > 0$.

We arbitrarily define

$$m_{\delta}(t) = \left(1 + e^{-\frac{5}{4}(\delta(t) - 100)}\right)^{-1}, \qquad m_{\epsilon}(t) = \left(1 + e^{-\frac{5}{4}(\epsilon(t) - 10)}\right)^{-1}. \tag{4.2}$$

The above functions address the problem caused by movements that do not necessarily yield a fully observable system, we still have to choose the values of $\overline{k}_{R_{\mathcal{C}}}$, $\overline{k}_{g}$ and $\overline{k}_{p_{\mathcal{C}}}$ however. The rationale from Section 3.4 only holds if the system is uniformly observable, and since these parameters are modeled by constant terms, we employ the values of k_{ω} obtained in (3.51) for $\overline{k}_{R_{\mathcal{C}}}$, and k_{a} obtained in (3.55) for $\overline{k}_{g}$ and $\overline{k}_{p_{\mathcal{C}}}$.

To evaluate the proposed classifier for the angular motion, we place the sensor on a tripod and further perform three different angular motions. The first two movements are made



(a) Measurement of $^{\mathcal{B}}\omega$ and its time derivatives

(b) Observability conditions and respective gains

Figure 4.2: Evaluation of the filter for identification of observability condition.

(a) Template 1

(b) Template 2

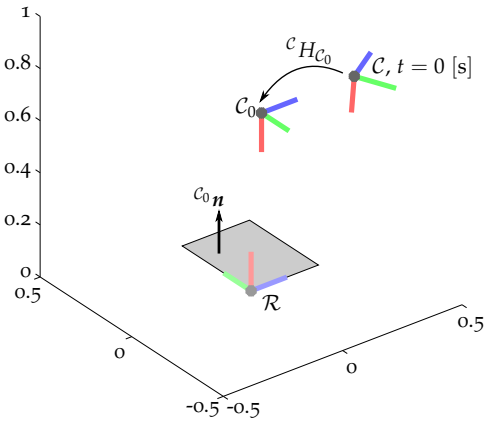
(c) Template 3

Figure 4.3: Template images employed in the experiments.

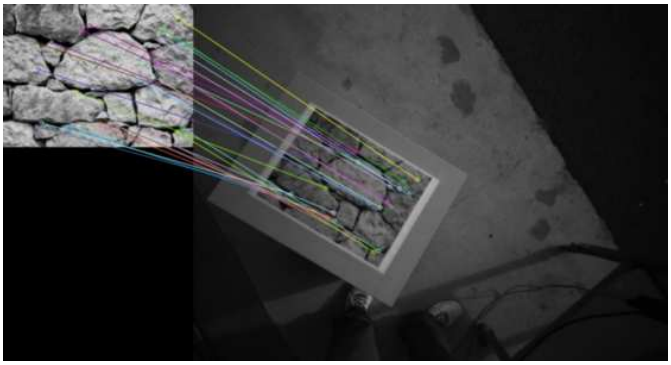
around single axes, from 0.4 to 3 [s], and 5.5 to 8 [s]. A third movement satisfying the observability condition is made from 9 to 13 [s]. The result obtained using the estimator is depicted in Figure 4.2, where Figure 4.2 (a) shows, from top to bottom, the angular velocity $^{\mathcal{B}}\omega_y$ [rad/s], angular acceleration $^{\mathcal{B}}\widehat{\dot{\omega}}$ [rad/s^2], angular jerk $^{\mathcal{B}}\widehat{\ddot{\omega}}$ [rad/s^3], whilst Figure 4.2 (b) displays, from top to bottom, evaluations of the term $|^{\mathcal{B}}\widehat{\dot{\omega}} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$ provided by (3.34) and the resulting $\bar{k}_{R_C}$, the term $|^{\mathcal{B}}\widehat{\omega} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$ provided by (3.45) and the resulting $\bar{k}_g$. It is clear that the estimated evolution of angular acceleration and jerk are mostly parallel for the first two parts. For the third movement, the angular acceleration and jerk are not parallel and $|^{\mathcal{B}}\widehat{\dot{\omega}} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$, $|^{\mathcal{B}}\widehat{\omega} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$ are large. Note that in the end of this movement, around 13 [s], the terms $|^{\mathcal{B}}\widehat{\dot{\omega}} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$ and $|^{\mathcal{B}}\widehat{\omega} \times {}^{\mathcal{B}}\widehat{\ddot{\omega}}|$ decrease slowly, but the gain functions decrease faster. This behavior is substantial to ignore slow motions that would not contribute to the estimation process with respect to the noise value of gyro measurements.

4.4 EXPERIMENTAL SETUP

We performed several hand-held experiments for the estimation of pose and system parameters to evaluate the proposed data fusion algorithm of pose estimation using direct visual tracking methods and the proposed nonlinear observers. The experiments are conducted as follows. First, we printed out versions of the templates depicted in Figures 4.3 (a), (b) and (c) in 37.6 × 28.2 [cm] rectangles, and thus each target can serve as reference image $\boldsymbol{I}_{\mathrm{R}}$ to the direct visual tracking method. For each experiment, a target is placed over a surface parallel to the ground. Hence, we can define a reference frame $\mathcal{R}$ with two axes coincident to two orthogonal sides of the target, and the third axis coincident to the gravitational acceleration. For the sake of simplicity, we define the origin of $\mathcal{R}$ at the corner corresponding to the up-



(a) Important camera frames



(b) Initialization of camera pose via feature matching

Figure 4.4: Initialization setup of the system.

per left pixel of the digital image. Concerning the visual tracking methods, we use a $\boldsymbol{I}_{\mathrm{R}}$ with 320×240 [pixel]. Let us define $\mathcal{C}_0$ as the frame associated with the optical center of the camera measuring $\boldsymbol{I}_{\mathrm{R}}$ hypothetically. Therefore, the configuration of target with the known intrinsic parameters from (4.1) allow us to compute the orientation and position ${}^{\mathcal{R}}(R, p)_{\mathcal{C}_0}$ associated to $\mathcal{C}_0$, and the respective scaled normal vector ${}^{\mathcal{C}_0}\boldsymbol{n}$.

A simplified representation of the setup is depicted in Figure 4.4. Recall that direct visual methods are incapable of computing the pose of the camera on the first frame, since these methods employ local optimization schemes. Therefore, the initialization of the camera pose is made using SIFT features from the template and $\boldsymbol{I}_{\mathrm{C}}$. Using the corresponding features, we can compute the projective homography ${}^{\mathcal{C}}H_{\mathcal{C}_0}$ using the classic technique described in Section 1.3.4. The initial pose ${}^{\mathcal{R}}(R, p)_{\mathcal{C}}$ can be computed straightforwardly using a decomposition of the obtained projective homography ${}^{\mathcal{C}}H_{\mathcal{C}_0}$ and the scaled normal vector ${}^{\mathcal{C}_0}\boldsymbol{n}$.

Each experiment correspond to a sequence of images and IMU measurements that evaluated off-line with using the structure of Algorithm 4. An initial guess for the biases is obtained after leaving the IMU over the same surface for a few seconds, and this bias is subtracted from the raw IMU measurements afterwards. We employ a direct visual method derived from the SSD with Huber weights and ESM optimization. We have employed this method instead of the NCC from Algorithm 1, since the changes in illumination are not substantial. In this application, both techniques provide similar results, however the SSD has slightly less computational effort.

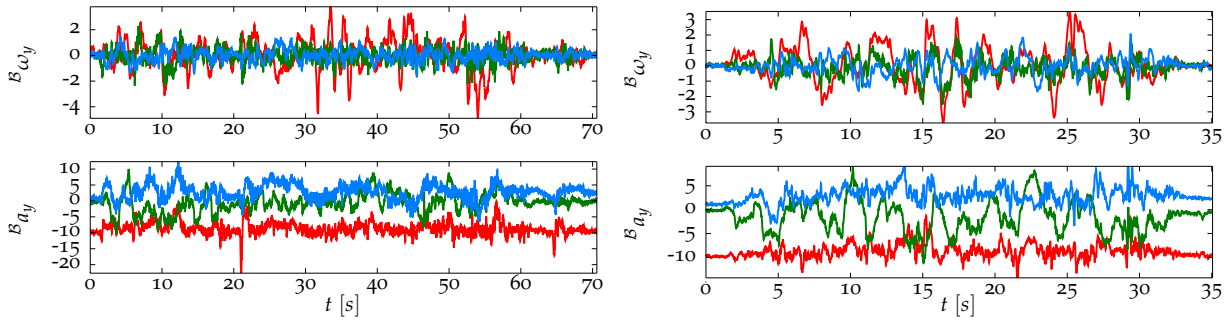
We still need to define the gains of the nonlinear observer. Recall from Section 3.4 the continuous design of the filter allows to define the gains based on arbitrarily small settling times. However, the settling times for the digital implementation are limited by the update-rate of the system. We use the following "rule of the thumb" to define the settling times in function of the update period:

$$\tau_{R_{\mathcal{B}}} = 4\Delta_{\mathcal{C}}, \qquad \tau_{b_\omega} = 15, \qquad \tau_{p_{\mathcal{B}}} = 4\Delta_{\mathcal{C}}, \qquad \tau_v = 8\Delta_{\mathcal{C}}, \qquad \tau_{b_a} = 15.$$

The corresponding gains are computed using (3.51) and (3.55), and, when applicable, we use the reference gains $k_{R_{\mathcal{C}}} = k_{b_\omega}$, and $k_{p_{\mathcal{C}}} = k_g = k_{b_a}$.

4.5 CONCURRENT POSE AND IMU BIAS ESTIMATION

The first couple of experiments consider the pose estimation using the template from Figure 4.3 (a). The movements consist mostly in translational displacement with peaks of high



(a) Sequence 1. (b) Sequence 2.

Figure 4.5: IMU measurements.

angular velocities. Figure 4.5 shows the IMU measurements for these sequences. For these initial experiments, we assume a rough calibration of c-to-IMU calibration, *i.e.*

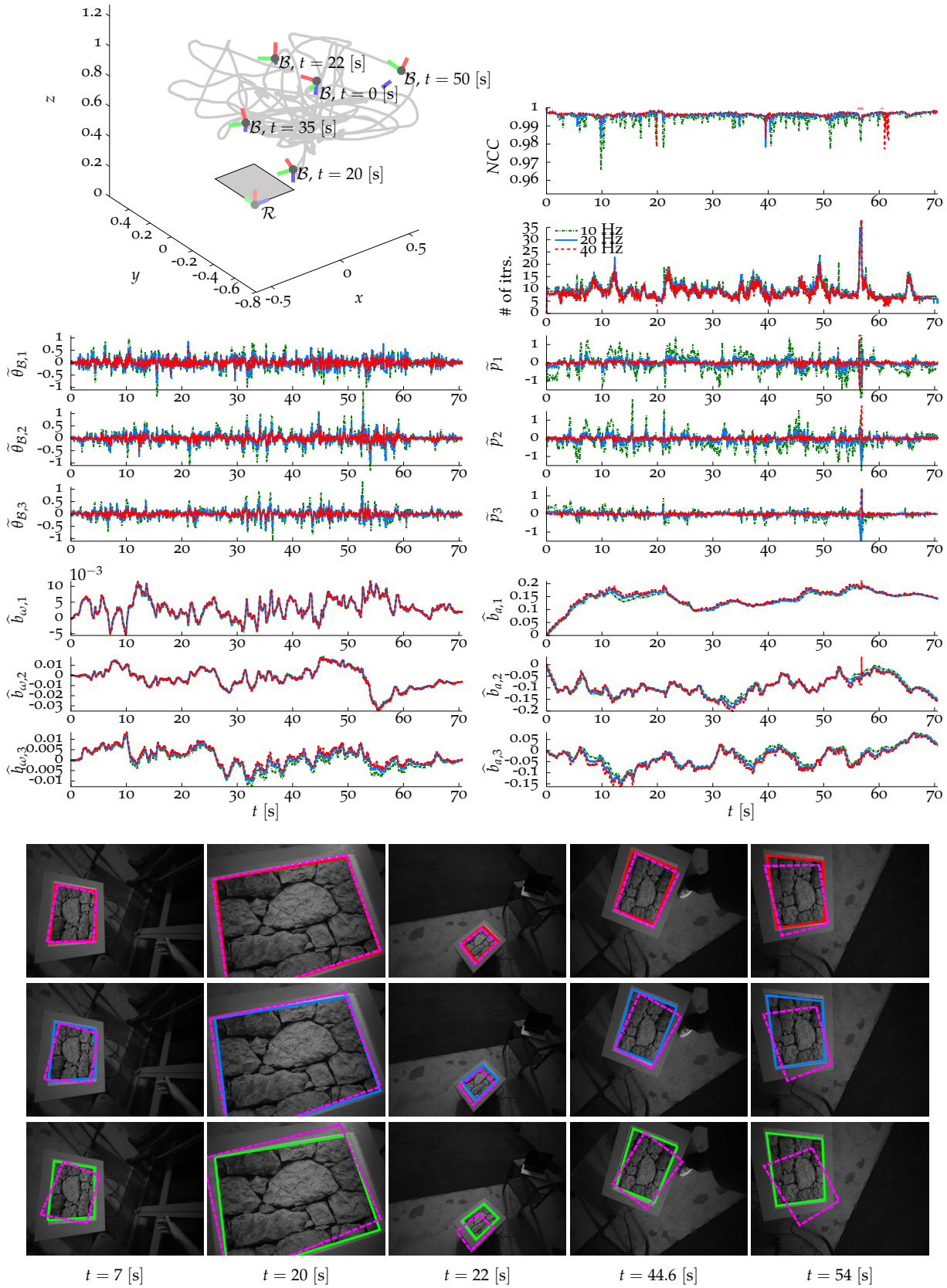
$$^{\mathcal{B}}R_{\mathcal{C}} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \qquad ^{\mathcal{B}}R_{\mathcal{C}}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.3}$$

We thus compute only pose (orientation and position), linear velocity, and IMU biases. Recall that the resulting system is uniformly observable independently of the inputs. The following experiments explore the effects of update-rate reduction in pose estimation, and the ability of the method to recover from complete occlusion of the target. The videos for these sequences are available in `http://goo.gl/68gH3`.

4.5.1 *Comparing multiple update-rates*

We first inspect the effects of different update-rates in the pose estimation. The camera provides, optimally, frames at 40 [Hz]. However, the direct visual tracking method is unable to compute a solution for pose immediately. There are two main implications due to this delay. First, the update-rate of the pose estimation system is given by, at least, the computation time of the visual tracker. Secondly, the obtained solution is delayed by that same period. The accuracy of the pose estimate obtained by the tracking method is unaffected though.

In this experiment, we evaluate using 5 different update rates: 40 [Hz], 20 [Hz], and 10 [Hz]. Figure 4.6 displays the results of this experiment. The upper left plot depicts the trajectory of $\mathcal{B}$ frame as computed by the nonlinear observer updated at 40 [Hz], the upper right plot represents the normalized cross correlation (NCC) coefficient between two frames, and number of iterations computed by the visual tracker. Then, the central right shows the translation estimation error $\widetilde{p}$ [cm], whilst the left plot displays the orientation innovation $\widetilde{\theta}_{\mathcal{B}} = \frac{180}{\pi}\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))$. The latter estimation error is equivalent (locally) to the angle-axis representation of the orientation error in [°]. The bottom figures depict estimated gyro $\widehat{b}_\omega$ [rad/s], and accelerometer $\widehat{b}_a$ [m/s^2] biases at left and right, respectively. There are five curves displayed in each plot, the dashed green curve denotes the response for 10 [Hz], solid blue for 20 [Hz] and dashed red 40 [Hz]. Moreover, fifteen image samples from this sequence are organized in three rows and five columns. The samples collected at a the same instant are displayed column-wise, whilst the results obtained by the nonlinear observer updated at 40, 20 and 10 [Hz] are displayed row-wise. Notice in each sample that two squares are drawn besides the image. Squares in dashed magenta lines represent the projection of the corners of target using the pose measured at the previous frame. For instance, on the first row, the green squares refer to



$t = 7$ [s] $t = 20$ [s] $t = 22$ [s] $t = 44.6$ [s] $t = 54$ [s]

Figure 4.6: Comparing multiple frame-rates.

the projection the delayed by 25 [ms], on the second and third rows by 50 and 100 [ms], respectively. Additionally, red, blue and green squares represent the projection of target corners of using a forecast provided by the nonlinear observer at 40, 20 and 10 [Hz], respectively. These samples provide a rough impression of the current velocity of the body, as we can compare the results of the previous frame to a virtual ground truth provided by the target.

The resulting trajectory shows a rich translational motion throughout the experiment. The positions vary from points of view close and far from the target. We can certify that the high quality of the measurements throughout the experiment via the inter-frame NCC coefficient. Recall that a high NCC coefficient between the frames is related to a good tracking of the target, therefore the high coefficients ensure the high quality of the measurements. Moreover, the integration of inertial data provides a better initialization of the tracking algorithm. In fact, we can assume that each image sample displays the information available at the given time instant. Hence, when the visual tracking method finishes to compute the pose estimate, the camera has already moved. Therefore the provided pose measurement is already "out of date". We obtain indeed better estimates of the current pose using the high frequency inertial data.

Remark from the NCC coefficients that the decreasing the update-rate of the visual tracker does not imply notably worse pose measurements. However, the number of iterations to compute the solution increase considerably with the reduction of the update-rate, the system at 10 [Hz] compute 3 or 4 iterations more than at 40 [Hz], *i.e.* an increase from 10 up to 40%. Besides, the reduction of the frame-rate can be indeed dangerous for the visual tracking method. Notice that innovation errors are related to the initialization errors of the visual tracking methods, since they are computed directly from the difference between the initialization and output of the visual tracking method. Although the errors on orientation are similar for all of the evaluated update-rates, they increase slightly for lower update-rates. The translational errors, on the other hand, degrade significantly in two out of the three axes with a decrease of the update-rate. Therefore, we can verify that the visual tracking method may be already at the limits of the basin of convergence at 10 [Hz], and may not provide the global solution for even lower update rates.

Let us now focus on the results of the translational dynamics. Notice even though the accelerometer data was corrected by a previous calibration step, the estimates converge to a steady state about $\widehat{b}_a = \begin{bmatrix} 0.15 & -0.15 & 0 \end{bmatrix}^T$. That value may correspond to the influence of the gravity within an error of 1 degree. We can point out two origins for this error: bad calibration of c-to-IMU frames, or bad estimate of the local gravitational acceleration. Remark that besides the effects on the bias, we can also verify that translational estimation error increases with higher angular velocities. Hence, we can relate these effects to a bad estimation of c-to-IMU translation. It is important to mention that the proposed algorithm works even with a bad initialization of c-to-IMU frame, however, the estimates can be improved with better sensor-to-sensor parameters.

4.5.2 *Complete occlusion of reference image*

We have verified that the proposed method performs well for different update-rates and, from now on, we display results considering only updates at 20 [Hz]. This frame-rate is easily obtained in current computers. One advantage claimed for the fusion of visual and inertial data is the ability of obtaining a pose estimate even when the target is completely occluded. We verify in the second experiment that the designed system is able to recover from loosing the target out of the field of view. Notice that the direct visual tracking method is unable to

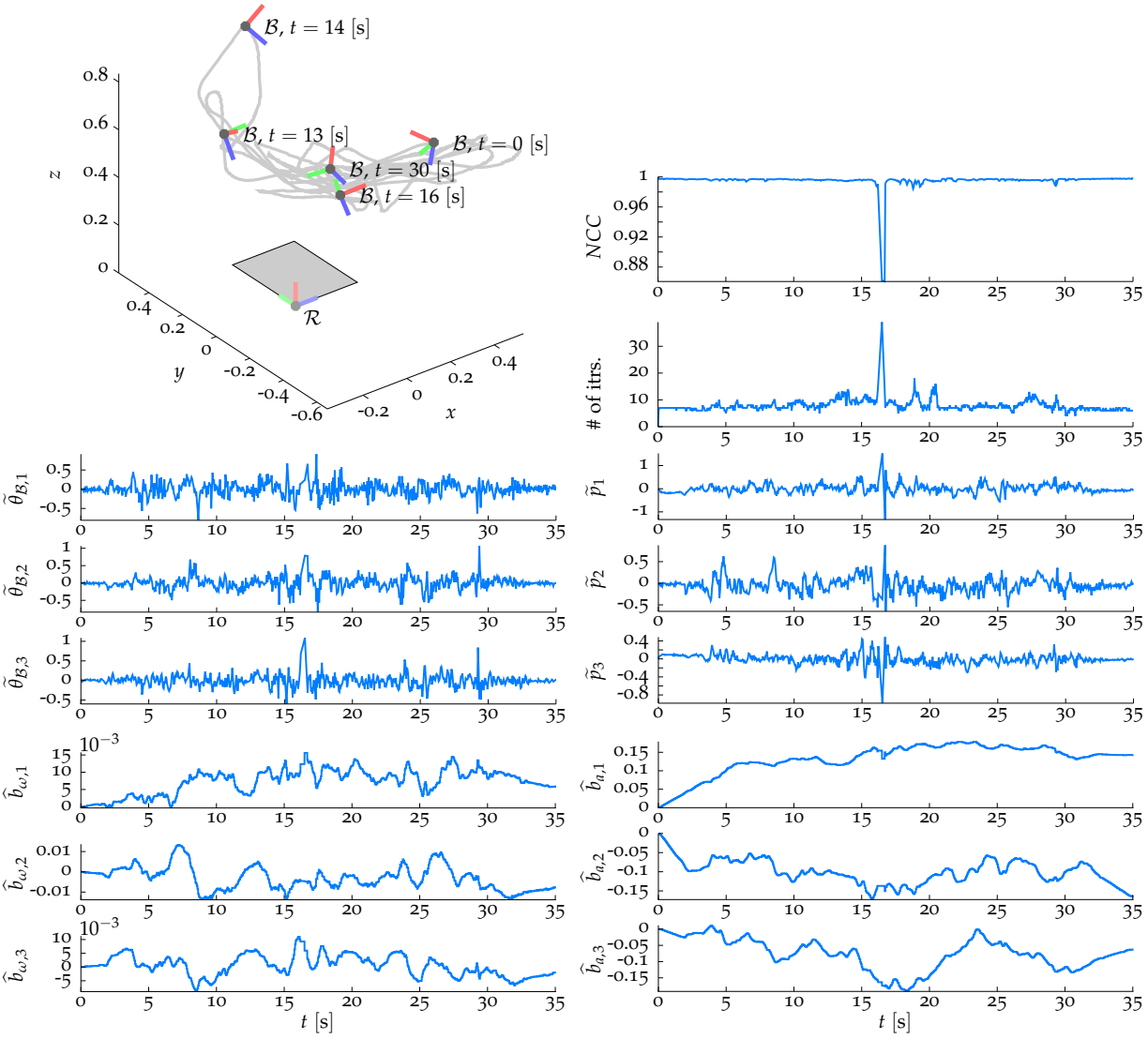
provide pose measurements during this period, therefore the estimates are computed only via the integration of inertial measurements.

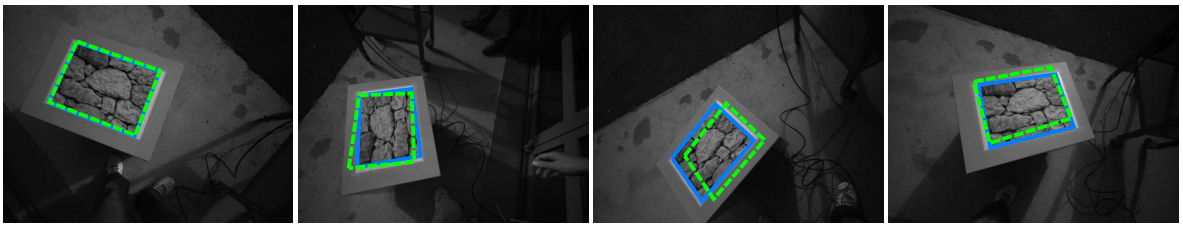
Figure 4.7 displays the results of this experiment similarly to the previous sequence. The upper left plot depicts the trajectory of $\mathcal{B}$ evaluated by the nonlinear observer, and the upper right represents the NCC coefficient between two frames, and the number of iterations computed by the visual tracker. Furthermore, the central right plot shows the translation estimation error $\widetilde{p}$ [cm], whilst the left plot displays the (local) angle-axis representation of the matrix corresponding to the orientation error in [°]. The bottom figures depict the estimated gyro $\widehat{b}_\omega$ [rad/s], and accelerometer $\widehat{b}_a$ [m/s^2] biases at the left and right, respectively. We also present a few image samples of this sequence, where two squares are drawn besides the image itself. The squares in green dashed lines represent the projection of the corners of target using the pose measured at the previous frame, whilst the blue squares represent the projection of target corners using the forecast provided by the nonlinear observer. These samples can indicate the current velocity of the body roughly, since we can compare the results of the previous frame to a virtual ground truth provided by the target.

The estimated trajectory describes the motion of the sensor mostly parallel to the target, however, around 14 [s] the visuo-inertial sensor moves farther from the target. We can verify via the inter-frame NCC that the camera moves such that target indeed exits its field of view around 16 [s], and latter the target is recovered afterwards around 17 [s]. The image samples describe the event with richer details. Moreover, notice that the estimates are accurate even without updates for over 0.5 [s], and the direct visual tracking method is capable of accurately recovering the target once the target is back inside the field of view. Of course, the of orientation and translation estimation errors are larger at the instant that the target returns to the field of view than during the rest of the sequence. However, the system is still able to provide a good initialization for the visual tracking method after a relatively large integration without visual updates. Concluding, let us analyze the results for accelerometer bias estimation. Notice that the results are akin to the previous experiment, as the accelerometer bias converges to a steady state about $\widehat{b}_a = \begin{bmatrix} 0.15 & -0.15 & 0 \end{bmatrix}^T$. Hence we have a stronger evidence of deviation in the rough estimates of c-to-IMU frames.

4.6 CONCURRENT POSE, IMU BIAS AND SENSOR-TO-SENSOR FRAME ESTIMATION

Previous results showed that, even though the method works with rough estimates of the sensor-to-sensor calibration, high angular velocities may impair the calculation of high quality pose estimates. The remaining experiments evaluate the performance of the proposed method for pose estimation with accurate calibration of c-to-IMU frame. We perform evaluate two sequences in Sections 4.6.1 and 4.6.2. The prior experiment computes the c-to-IMU frames using the full Algorithm 3, whilst the latter compares the data fusion algorithm using the obtained calibration frame, and compares the results with a system using the initial rough estimates. Figure 4.8 shows the IMU measurements for these sequences, and the evaluation of $m_\delta(t)$ from Eq. (4.2). The experiment from Section 4.6.1 uses the template from Figure 4.3 (b), and the proposed algorithm to computes c-to-IMU rotation and translational displacement. The experiment from Section 4.6.2 uses the template from Figure 4.3 (c) and we thus verify the benefits of using the calibrated frames over rough c-to-IMU estimates. The videos for these sequences are available in `http://goo.gl/68gH3`.





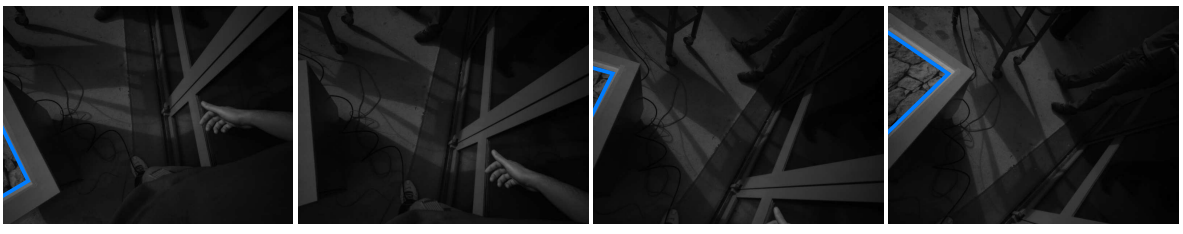
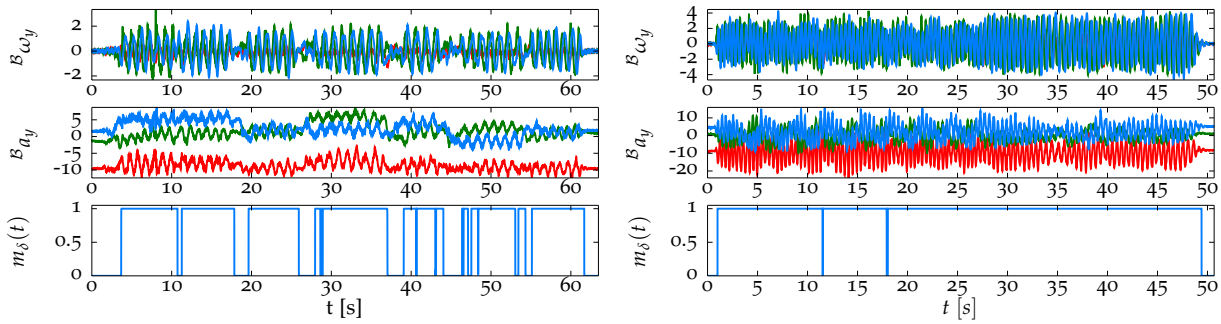


Figure 4.7: Recovering from complete occlusion of reference image.



(a) Sequence 3. (b) Sequence 4.

Figure 4.8: IMU measurements.

4.6.1 *Calibration procedure*

The next sequence corresponds to the c-to-IMU calibration itself. We initialize the c-to-IMU frame parameters as (4.3) and use the full update scheme from Algorithm 3. Furthermore, we analyze the same measurements with a Kalman filter designed for the translational dynamics in order to compare the calibration results obtained. Figure 4.9 displays the results of this experiment. The results are distributed as follows. The upper right plot shows the translation estimation error $\widetilde{p}$ [cm], whilst the upper left plot displays the (local) angle-axis representation of the matrix corresponding to the orientation error in [°]. The central plots depict the estimated gyro $\widehat{b}_\omega$ [rad/s], and accelerometer $\widehat{b}_a$ [m/s^2] biases at the left and right, respectively. Afterwards, the plots refer to results of c-to-IMU orientation and translational displacement. Notice that we represent c-to-IMU rotation $^{\mathcal{B}}R_{\mathcal{C}}$ using Euler angles: roll $\theta_{\mathcal{C}}$, pitch $\phi_{\mathcal{C}}$ and yaw $\psi_{\mathcal{C}}$. The estimation of the local gravitational acceleration is shown in the center. The curves in blue refer to the results obtained by the nonlinear observer and the curves in red the results obtained by the Kalman filter. Finally, at the bottom left we present the trajectory of $\mathcal{B}$ evaluated by the nonlinear observer and, at the bottom right, we present four image samples of this sequence. The squares in green dashed lines represent the projection of the corners of target using the pose measured at the previous frame and the blue squares represent the projection of target corners using the forecast provided by the nonlinear observer.

The resulting trajectory presents regions with richer angular motion, that refer to the self-calibration itself, and transition between these regions. We can identify five regions with richer motion. The evolution of $m_\delta(t)$ in Figure 4.8 (a) shows directly the five periods where the self-calibration is computed. These periods present the larger angular velocity and are also the regions with richer motion depicted in Figure 4.9. Again, we can verify that the initialization provided by the estimator is closer to the current pose than the previous estimate obtained by the visual method. We can verify that c-to-IMU rotation estimates converge towards a steady-state about $\theta_{\mathcal{C}} = -89.8°$, $\phi_{\mathcal{C}} = -2°$ and yaw $\psi_{\mathcal{C}} - 88.5°$ and there are slight gyro bias variations in the order of $2 \cdot 10^{-3}$ [rad/s], *i.e.* $0.12[°/s]$. Moreover, notice that the gyro bias estimates a sinusoidal behavior even with a a calibrated c-to-IMU frame. These effects are probably caused by low frequency errors of the visual tracking method and errors due to scale of the gyroscopes. First, the visual tracking methods have limited precision, and the optimization techniques can yield small low frequency errors. Secondly, we use values of gyro scale factors given by the manufacturers. These values may also vary over time and such changes provide unomdeled effects that can thus influence the final estimates. Concerning the translational dynamics, estimates of c-to-IMU translation converge towards a steady-state of $^{\mathcal{B}}p_{\mathcal{C}} = \begin{bmatrix} 0.0773 & -0.0208 & -0.0292 \end{bmatrix}$.

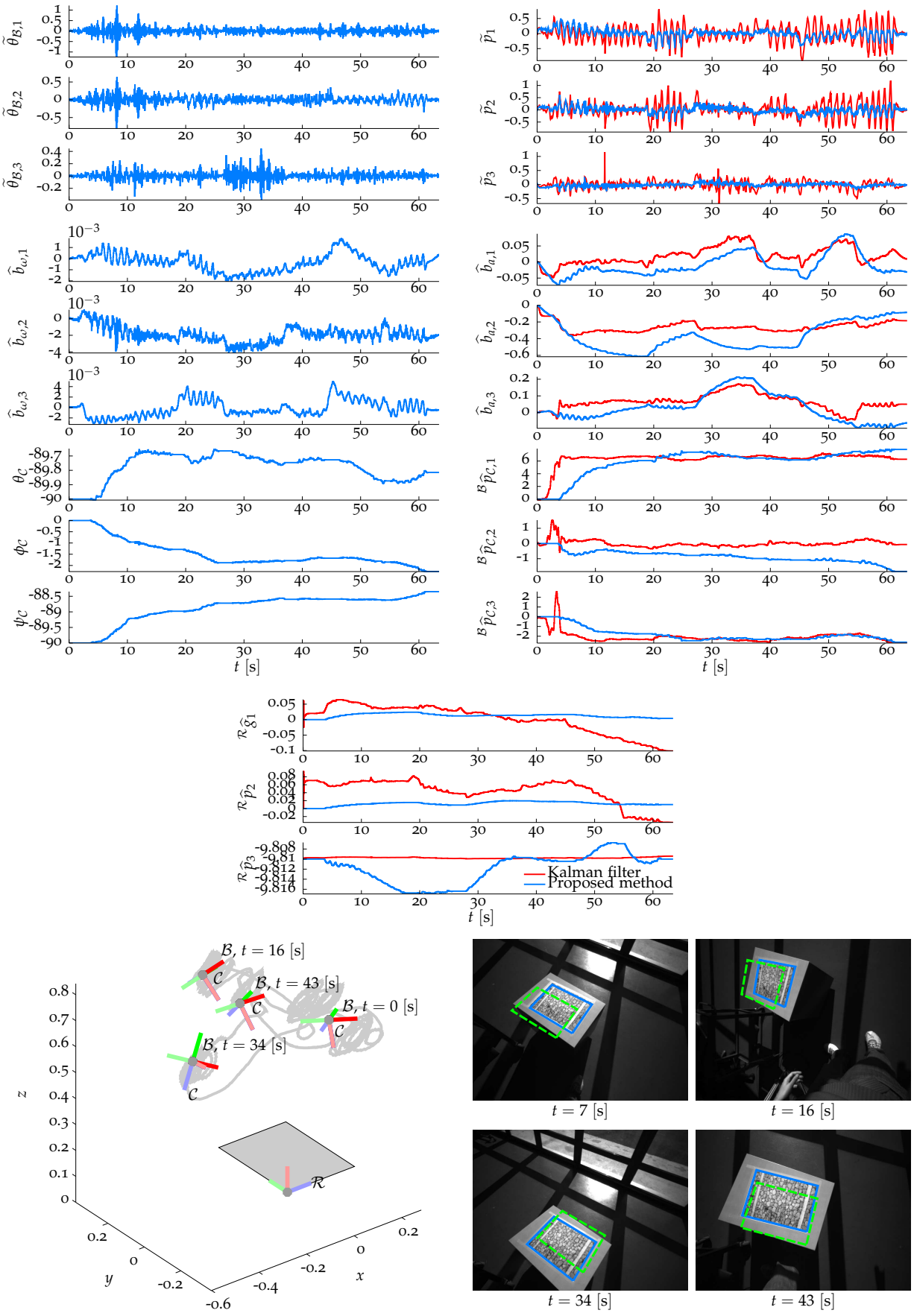


Figure 4.9: Sensor-to-sensor frame calibration.

The resulting calibration by the nonlinear observer provides:

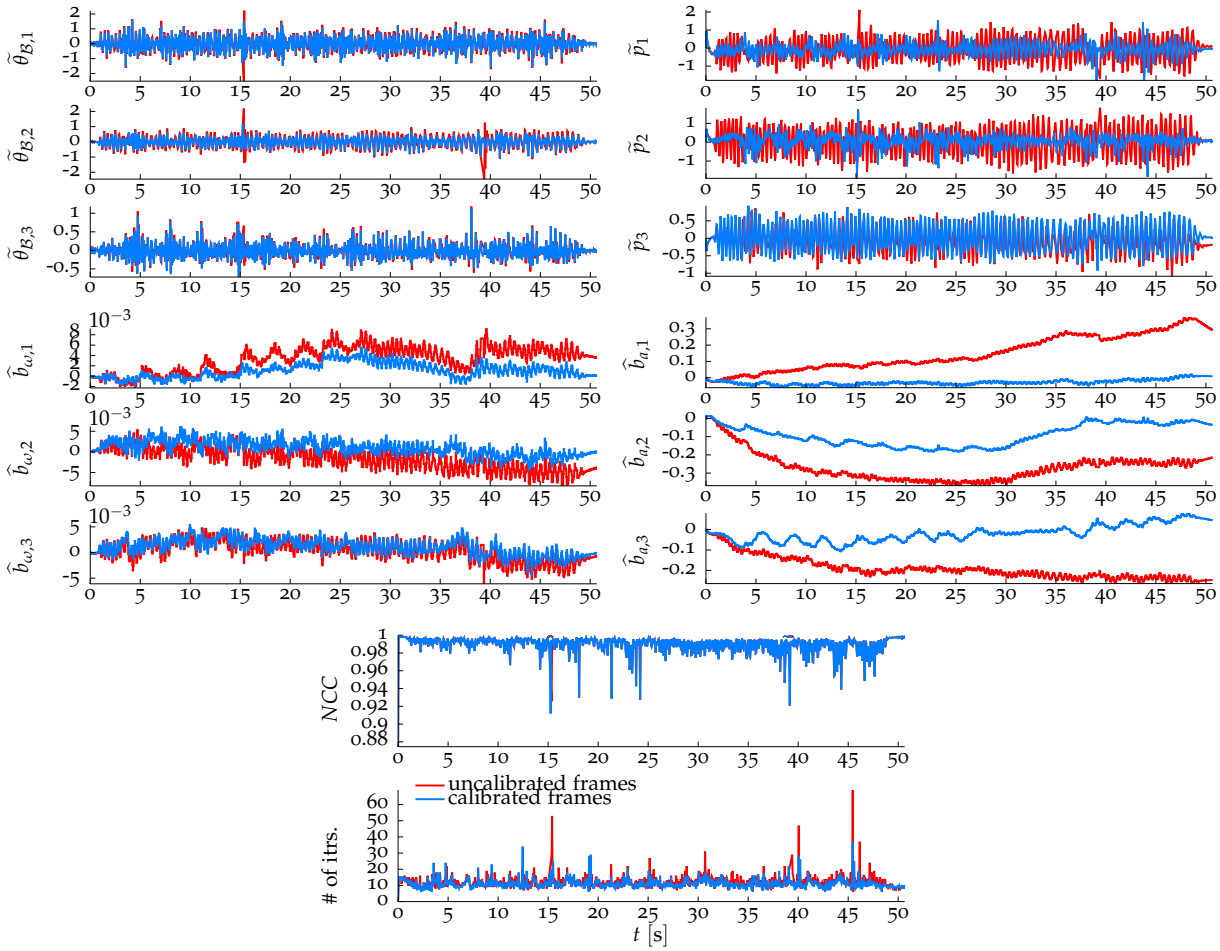
$$ {}^{\mathcal{B}}\widehat{R}_{\mathcal{C}} = \begin{bmatrix} 0.0290 & 0.0043 & 0.9996 \\ -0.9987 & -0.0412 & 0.0291 \\ 0.0415 & -0.9991 & 0.0031 \end{bmatrix}, \qquad {}^{\mathcal{B}}\widehat{p}_{\mathcal{C}} = \begin{bmatrix} 0.0773 \\ -0.0208 \\ -0.0292 \end{bmatrix}. \tag{4.4} $$

Remark that, apart from the second component, the c-to-IMU translation computed by the nonlinear observer is similar to the one computed by the Kalman filter. The results concerning accelerometer bias present the same trend, however, the Kalman filter present more constant results. Differently from the previous experiments, the steady state of the accelerometers converge to values closer to the pre-calibration of the sensors. The estimates on the local gravitational field show largest discrepancy between the nonlinear observer and the KF. The proposed algorithm displays variations mostly on the third component of ${}^{\mathcal{R}}\widehat{g}$, whilst the KF presents a practically constant third component with higher variations on the first and second components. Although there is not enough information to certify which filter provided the "correct" estimate, we can verify through the innovations $\widetilde{p}$ that the proposed filter provides more accurate forecasts than the KF. This difference clearly contradicts our simulation results, where the results had errors with similar order. Let us remark, however, the considered unmodeled effects in the synthetic data consisted mostly of noises, and we knew all of the parameters of simulation. The KF was be finely tuned for synthetic data, whilst this procedure can be considerably more time consuming with real data. Probably, it is possible to tune better the parameters of the KF, however, in this author's opinion, the estimation problem should be overwhelmed by *ad hoc* fine tuning. Even if the KF could be tuned and provide similar or slightly better estimates the our filter, the proposed technique still provides benefits of the simpler tuning and implementation.

4.6.2 *Validation sequence*

The last sequence provides a validation for the c-to-IMU calibration. We compare the results for the proposed algorithm with calibrated c-to-IMU frames, using ${}^{\mathcal{B}}(R, p)_{\mathcal{C}}$ from (4.4). and uncalibrated frames, using ${}^{\mathcal{B}}(R, p)_{\mathcal{C}}$ from (4.3). We use in this experiment a version of the nonlinear observer to estimate pose, linear velocity, IMU bias. This experiments consists mostly of rotational movements, as we can verify from the IMU measurements of Figure 4.8 (b). Figure 4.9 displays the results of this experiment. The red curves correspond to the system with uncalibrated c-to-IMU frames, while the blue curves correspond to a system with calibrated c-to-IMU frames. The distribution of the results is similar to the previous experiment. The upper left and right plots correspond the (local) angle-axis representation the orientation error in [°] and translational estimation error $\widetilde{p}$ [cm]. The central plots depict the estimated gyro $\widehat{b}_{\omega}$ [rad/s], and accelerometer $\widehat{b}_{a}$ [m/s^2] biases at the left and right, respectively. Moreover, there is a plot displaying the normalized cross correlation (NCC) coefficient between two frames, and number of iterations computed by the visual tracker. At last, there are eight image samples of this sequence. The squares in green dashed lines represent the projection of the corners of target using the pose measured at the previous frame, while red and blue squares represent the projection of target corners of using the forecast provided by the nonlinear observer with uncalibrated and calibrated c-to-IMU frames, respectively.

This sequence does not present as much translational motions as the others, however, we can verify from the images samples that angular motion also yields displacements of the reference image inside the field of view. These high angular velocities allows us verify more clearly the effects of calibrated c-to-IMU frame in the estimation. First, notice that errors in position are larger for the uncalibrated case than the calibrated ones. Deficient calibration of c-



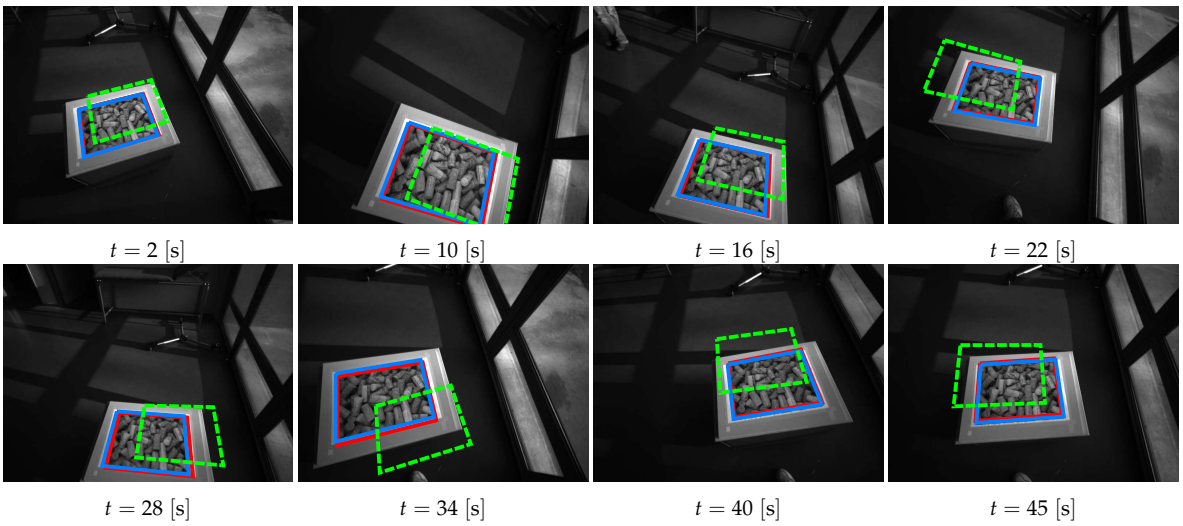


Figure 4.10: Validation of sensor-to-sensor frame calibration.

to-IMU translation associated with high angular velocity yield corrupted estimates ${}^{\mathcal{R}}v$, which in turn imply worse forecasts of ${}^{\mathcal{R}}p_{\mathcal{B}}$. This effect is put into evidence due to the (relatively) large errors of ${}^{\mathcal{B}}p_{\mathcal{C}}$. On the other hand, the slight errors of about 1° are practically unnoticeable in the rotational dynamics. These errors, recalling from the results of Section 4.5, propagate to the estimation of accelerometer bias. We can verify from the results of this sequence that the accelerometer bias estimates with a calibrated c-to-IMU frame are indeed coherent with pre-calibration procedure.

4.7 CONCLUSION

This Chapter discusses several implementation issues and experimental results using real visuo-inertial data. We initially discuss the implementation of pose estimation using direct visual tracking methods. We further present a version of the afore-introduced nonlinear observers for discrete time, and we discuss implementation issues of the gain tuning for systems where the observability is related to the angular motion. We use these techniques for pose estimation in practical situations. We perform four experiments to verify the viability of the our method using real data. We compare the effects due to lower frame rates, and the advantages of using accurate c-to-IMU calibration over roughly calibrated systems.

EPILOGUE

This thesis addressed the problem of pose estimation using visuo-inertial sensors. These systems consist of one camera and inertial sensors that present complementary properties exploited to provide highly accurate estimation with also high frequency.

We have analyzed the details of several direct visual tracking methods, presented a survey of multiple similarity functions, and discussed the robustness properties given by each function. Also, we have discussed optimization techniques for these similarities. We were particularly interested in the normalized-cross correlation (NCC) among the aforementioned similarity functions. This similarity is intrinsically invariant to affine illumination changes and the computation of a gradient-based solution is also simple. We have presented a new method for direct visual tracking based on the NCC, which is built upon three pillars: rejection of bad regions of the image; pixel-wise weighting invariant to affine illumination changes; Newton-like optimization that uses the information from forward and inverse compositional. The proposed method was exhaustively compared to other state of the art methods via an analysis of the basin of convergence, scores obtained using a planar based visual tracking benchmark dataset, and challenging real-world video sequences. We have verified that the choice of the similarity function plays indeed an important role in direct visual tracking, however, the optimization technique is equally important. Moreover, experimental results indicate that our method presents substantial improvements for the tracking of partially occluded objects under severe illumination changes.

Even though direct visual tracking methods provide highly accurate pose information, the pose measurements are computed in lower frequencies than incremental measurements of the IMU. On the one hand, the incremental data can be integrated to provide an accurate initialization for the visual algorithms, the *Achilles' heel* of gradient based direct visual tracking, and also to compensate for momentary loss of sight. IMU measurements are corrupted by additive bias and noise, however, the information from visual pose estimation bounds the drift due to pure integration. A multi-sensory system also has to cope with multiple coordinate systems. For instance, the coordinate frames of the camera and the IMU are not coincident, and, although rough calibration parameters can be usually obtained by a CAD model or inspection, the estimates obtained from a poorly calibrated system are indeed less accurate. Classical algorithms for inertial visual data fusion are typically based on Kalman filters and its extensions. The nonlinear nature of the dynamics may already impair classical solutions, and inherent observability conditions make the estimation problem even more challenging.

We studied the pose estimation and self-calibration problem using a control theory point of view. The main results of this thesis consist in new observers for pose estimation with the concurrent identification of multiple parameters of the system. We have analyzed the rotational dynamics using tools from nonlinear control, provided stable observers on the group of rotation matrices that consist of extensions to the nonlinear complementary filter on $\mathbb{SO}(3)$. First, we develop a filter that whose domain of convergence is independent of the magnitude of the gains. Secondly, we proposed an extension of to that filter so as identify the c-to-IMU rotation. The proposed observers maintain invariance properties of the original systems, and ensure exponential stability of the estimation error under specific (and specified) observability conditions. Moreover, the translational dynamics is studied as a linear time-varying system, and we propose new Luenberger-like observers for several configurations of the system. We were capable of determining motion conditions under which the system is observable. The thorough

observability analyses allow us to prove uniform stability of the proposed observers. We also conjectured an observer for the estimation of the position, linear velocity, accelerometer bias, local gravitational field and c-to-IMU position. That conjecture, however, is not endowed with stability proofs.

The nonlinear observers are tested using synthetic data (simulations) to evaluate properties such as domain of convergence and effects caused by unmodeled noise and parameters. We also evaluated the visuo-inertial data fusion with real data using direct visual tracking methods and the proposed nonlinear observers. We were able to perform experiments from 40 [Hz] (optimally) down to 10 [Hz] (worst case). The estimation of IMU bias is practically unchanged, but we could verify that the pose estimation errors increase substantially. It is possible to use a direct visual tracking even under lower framerates, however, the system is more prone to initialization with large errors. We also verified the quality of pose estimation with self-calibration using the conjectured method, which obtained fair results similar to a Kalman filter. The experimental results support the use of visuo-inertial data fusion over pure vision, and also the improvements due to accurately calibrated sensor-to-sensor frames over roughly calibrated systems.

The development of new data fusion techniques was the main challenge of this thesis, since we had to face the well established theory of Kalman-filtering. The main difficulty to justify novel data fusion methods is that Kalman filters (KFs) work very well in most situations. The proposed methods are simpler than KF because we can use directly pose and IMU measurements with constant gains. A reader familiar with KFs can recall that our methods do not involve solutions of Riccati algebraic nor differential equations, utterly because our filters do not rely on the computation of extra parameters, such as the time varying gain or covariance matrix. The good performance of KFs is also related to the fine tuning of their parameters, specially for the orientation estimation. The effort put into the KF tuning was not discussed here but the work must not be underrated. Let us remark that the KF is severely impaired by a bad choice of its parameters. In some way, the good performance of our filter may be shadowed by the results of a finely tuned KF. On the other hand, the proposed methods are notably easier to tune than KF. We were able to use the proposed filter with simulated and real data with practically the same parameters, whilst the KF had to be tuned for each application. The proposed filters are endowed with proofs of the convergence of the estimates for large initial errors. These properties are specially important for nonlinear estimation of the orientation dynamics, whilst the classical KF fails to guarantee the convergence of the estimates, and fails to keep invariance properties of the group. Concerning the translational dynamics, recall that KFs have convergence proofs for linear time-varying systems. However, the reader should remark that the stability proofs do not hold for the steady-state implementations of the filter, thus the Riccati differential equation must be computed in order to ensure the convergence of the KF estimates. The detailed observability analyses developed in this thesis are contributions not less important than the observers. Remark that stability properties for (part of) the observers rely explicitly on body motion, which is in turn related to observability conditions of the system. Our results provide explicit motion conditions that guarantee observability and allow us to better understand the convergence properties of the proposed observers and also the convergence of Kalman-based filters.

The end of a thesis is also followed by proposal of future outcomes and extensions of the work developed. First, we expect to validate the proposed algorithms (visual tracking and nonlinear observer) on mini-drones in the short future. The visual tracking invariant to illumination changes could be useful in this context. Moreover, although the experimental validation of the data-fusion tried to stimulate fast displacements of the sensors, the IMU measurements obtained in drones are impaired, for instance, by effects due to the vibration of the chassis. This difficulty still needs to be addressed. A direct continuation of this thesis could focus a

gain tuning analysis and investigation of direct relations to optimal criteria given by strict Lyapunov functions (Sepulchre et al., 1997, Ch. 3). One could improve the performance of the proposed techniques, for instance, with simple time varying gains and maybe obtain a similar response to KF with simpler innovation terms. Moreover, we considered that multiplicative parameters of the system were already identified. Surely, bad estimation of these parameters can also impair the pose estimation, however their estimation requires more complex observability conditions and likely more complex observers. Finally, we considered a known structure of the scene, however, one must also estimate the normal vector with respect to the plane in order to employ our method in simultaneous localization and tracking using monocular vision in an unknown environment. It is likely that one cannot write observers with constant gains and stability proofs for these systems without considering either the derivatives of acceleration and angular velocity, or the trajectory of the pose during a period of time.

BIBLIOGRAPHY

Anderson, B. D. O. and Moore, J. B. (1969). New results in linear system stability. *SIAM Journal on Control*, 7(3). (Cited on pages 67 and 149.)

Armesto, L., Chroust, S., Vincze, M., and Tornero, J. (2004). Multi-rate fusion with vision and inertial sensors. In *IEEE International Conference on Robotics and Automation*. (Cited on page 26.)

Arya, K. V., Gupta, P., Kalra, P. K., and Mitra, P. (2007). Image registration using robust M-estimators. *Pattern Recognition Letters*, 28(15). (Cited on pages 42 and 44.)

Azuma, R., Hoff, B., Neely, H., and Sarfaty, R. (1999). A motion-stabilized outdoor augmented reality system. In *IEEE Virtual Reality*. (Cited on page 26.)

Baker, S. and Matthews, I. (2001). Equivalence and Efficiency of Image Alignment Algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*. (Cited on pages 13, 31, 36, 37, 48, 50, 51, and 52.)

Baldwin, G., Mahony, R., and Trumpf, J. (2009). A nonlinear observer for 6 DOF pose estimation from inertial and bearing measurements. In *IEEE International Conference on Robotics and Automation*. (Cited on page 26.)

Baldwin, G., Mahony, R., Trumpf, J., Hamel, T., and Cheviron, T. (2007). Complementary Filter Design on the Special Euclidean Group SE(3). In *European Control Conference*. (Cited on page 23.)

Barczyk, M. and Lynch, A. F. (2012). Invariant Observer Design for a Helicopter UAV Aided Inertial Navigation System. *IEEE Transactions on Control Systems Technology*. (Cited on page 26.)

Bartoli, A. (2008). Groupwise Geometric and Photometric Direct Image Registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12). (Cited on pages 13 and 31.)

Batista, P., Silvestre, C., and Oliveira, P. (2009). Sensor-based complementary globally asymptotically stable filters for attitude estimation. In *IEEE Conference on Decision and Control*. (Cited on page 25.)

Batista, P., Silvestre, C., Oliveira, P., and Cardeira, B. (2011). Accelerometer calibration and dynamic bias and gravity estimation: analysis, design, and experimental evaluation. *IEEE Transactions on Control Systems Technology*, 19(5). (Cited on page 11.)

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *European Conference on Computer Vision*. (Cited on page 13.)

Bell, B. M. and Cathey, F. W. (1993). The iterated Kalman filter update as a Gauss-Newton method. *IEEE Transactions on Automatic Control*, 38. (Cited on page 22.)

Benhimane, S. (2006). *Vers une approche unifiée pour le suivi temps-réel et l'asservissement visuel*. PhD thesis, Université de Nice-Sophia Antipolis. (Cited on page 113.)

Benhimane, S. and Malis, E. (2007). Homography-based 2D Visual Tracking and Servoing. *International Journal of Robotics Research*, 26. (Cited on pages 13, 31, 38, 44, 46, and 48.)

Beravs, T., Podobnik, J., and Munih, M. (2012). Three-axial accelerometer calibration using kalman filter covariance matrix for online estimation of optimal sensor orientation. *IEEE Transactions on Instrumentation and Measurement*, 61(9). (Cited on page 11.)

Besançon, G. (2007). An overview on observer tools for nonlinear systems. In Besançon, G., editor, *Nonlinear observers and applications*. Springer-Verlag. (Cited on pages 23, 62, 63, and 65.)

Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. In *Special Interest Group on Graphics and Interactive Techniques*. (Cited on page 14.)

Bonnabel, S., Martin, P., and Rouchon, P. (2008). Symmetry-Preserving Observers. *IEEE Transactions on Automatic Control*, 53(11). (Cited on pages 23, 72, and 78.)

Bonnabel, S., Martin, P., and Rouchon, P. (2009a). Nonlinear Symmetry-Preserving Observers on Lie Groups. *IEEE Transactions on Automatic Control*, 54(7). (Cited on page 23.)

Bonnabel, S., Martin, P., and Salaün, E. (2009b). Invariant Extended Kalman: Filter Theory and Application to a Velocity-Aided Attitude Estimation Problem. In *IEEE Conference on Decision and Control*. (Cited on pages 22 and 25.)

Bornard, G., Couenne, N., and Celle, F. (1989). Regularly persistent observers for bilinear systems. In Descusse, J., Fliess, M., Isidori, A., and Leborgne, D., editors, *New Trends in Nonlinear Control Theory*. Springer. (Cited on page 68.)

Boyde, S. and VanderbergheL. (2004). *Convex Optimization*. Cambridge University Press. (Cited on page 159.)

Brás, S., Cunha, R., Vasconcelos, J. F., Silvestre, C., and Oliveira, P. (2011). A nonlinear attitude observer based on active vision and inertial measurements. *IEEE Transactions on Robotics*, 27(4). (Cited on page 26.)

Bristeau, P.-J., Petit, N., and Praly, L. (2010). Design of a navigation filter by analysis of local observability. In *IEEE Conference on Decision and Control*. (Cited on page 64.)

Brooks, R. and Arbel, T. (2010). Generalizing Inverse Compositional and ESM Image Alignment. *International Journal of Computer Vision*, 87(3). (Cited on pages 14, 40, and 44.)

Busvelle, E. and Gauthier, J. (2002). High-gain and non-high-gain observers for nonlinear systems. In *Conference on Geometric Control Theory and Applications*. (Cited on page 23.)

Campolo, D., Keller, F., and Guglielmelli, E. (2006). Inertial/Magnetic Sensors Based Orientation Tracking on the Group of Rigid Body Rotations with Application to Wearable Devices. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. (Cited on pages 23 and 25.)

Chen, C.-T. (1984). *Linear System Theory and Design*. CBS College Publishing, 2 edition. (Cited on pages 21, 62, 63, 64, and 67.)

Cheviron, T., Hamel, T., Mahony, R., and Baldwin, G. (2007). Robust Nonlinear Fusion of Inertial and Visual Data for position, velocity and attitude estimation of UAV. In *IEEE International Conference on Robotics and Automation*. (Cited on page 26.)

Cook, R. and Torrance, K. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1). (Cited on page 14.)

Corke, P., Lobo, J., and Dias, J. (2007). An Introduction to Inertial and Visual Sensing. *International Journal of Robotics Research*, 26(6). (Cited on pages 9 and 10.)

Crassidis, J. L. (2006). Sigma-point Kalman filtering for integrated GPS and inertial navigation. *IEEE Transactions on Aerospace and Electronic Systems*, 42(2). (Cited on pages 22 and 25.)

Crassidis, J. L., Markley, F. L., and Cheng, Y. (2007). A Survey of Nonlinear Attitude Estimation Methods. *Journal of Guidance, Control, and Dynamics*, 30(1). (Cited on pages 24 and 25.)

Cui, J., He, C., Yang, Z., Ding, H., Guo, Z., Hao, Y., and Yan, G. (2012). Virtual Rate-Table Method for Characterization of Microgyroscopes. *IEEE Sensors Journal*, 12(6). (Cited on page 11.)

Dame, A. and Marchand, E. (2010). Accurate real-time tracking using mutual information. In *International Symposium on Mixed and Augmented Reality*. (Cited on pages 14, 35, 40, 48, 50, 53, 54, and 55.)

Daum, F. (2005). Nonlinear filters: beyond the Kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8). (Cited on pages 22 and 23.)

Davenport, P. B. (1968). A vector approach to the algebra of rotations with applications. Technical report, NASA. (Cited on page 24.)

Doucet, A., de Freitas, N., Murphy, K., and Russell, S. (2000a). Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks. In *Conference on Uncertainty in Artificial Intelligence*. (Cited on page 22.)

Doucet, A., Godsill, S. J., and Andrieu, C. (2000b). On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistics and Computing*, 10. (Cited on page 22.)

Draper, C. S. (1981). Origins of Inertial Navigation. *Journal of Guidance and Control*, 4(5). (Cited on page 9.)

Dudek, G. and Jenkin, M. (2008). Inertial Sensors, GPS, and Odometry. In Siciliano, B. and Khatib, O., editors, *Springer Handbook of Robotics*. Springer. (Cited on pages 9 and 10.)

Esfandiari, F. and Khalil, H. K. (1992). Output feedback stabilization of fully linearizable systems. *International Journal of Control*, 56. (Cited on page 23.)

Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10). (Cited on pages 14 and 41.)

Farrel, J., Stuelpnagel, J. C., Wessner, R. H., Velman, J. R., and Brook, J. E. (1966). A least squares estimate of satellite attitude. *SIAM Reviews*, 8(3). (Cited on page 24.)

Faugueras, O. and Lustman, F. (1988). Motion and structure from motion in a piecewise planar environment. Technical report, INRIA. (Cited on page 19.)

Fleps, M., Mair, E., Ruepp, O., Suppa, M., and Burschka, D. (2011). Optimization based IMU camera calibration. In *International Conference on Intelligent Robots and Systems*. (Cited on page 27.)

Foxlin, E. and Naimark, L. (2003). Miniaturization, calibration & accuracy evaluation of a hybrid self-tracker. In *International Symposium on Mixed and Augmented Reality*. IEEE Comput. Soc. (Cited on page 27.)

Gauthier, J. P. and Kupka, I. A. K. (1994). Observability and observers for nonlinear systems. *SIAM Journal on Control and Optimization*, 32(4). (Cited on pages 23 and 63.)

Gay-bellile, V., Bartoli, A., and Sayd, P. (2010). Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 32. (Cited on page 46.)

Hager, G. and Toyama, K. (1998). The XVision System: A General-Purpose Substrate for Portable Real-Time Vision Applications. *Computer Vision and Image Understanding*, 69(1). (Cited on page 13.)

Hager, G. D. and Belhumeur, P. N. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 20(10). (Cited on pages 13, 31, and 37.)

Hammouri, H. and de Leon Morales, J. (1990). Observer synthesis for state-affine systems. In *IEEE Conference on Decision and Control*. (Cited on pages 23 and 68.)

Hammouri, H. and Gauthier, J. P. (1992). Global time-varying linearization up to output. *SIAM Journal on Control and Optimization*, 30(6). (Cited on page 23.)

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey Vision Conference*. (Cited on page 13.)

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press. (Cited on page 14.)

Hermann, R. and Krener, A. (1977). Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22. (Cited on pages 65 and 66.)

Ho, Y. and Lee, R. (1964). A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9(4). (Cited on page 20.)

Hol, J. D., Schon, T. B., and Gustafsson, F. (2010). Modeling and Calibration of Inertial and Vision Sensors. *International Journal of Robotics Research*, 29(2-3). (Cited on page 27.)

Horst, R. and Pardalos, P. M. (1995). *Handbook of Global Optimization*. Springer. (Cited on pages 31 and 36.)

Hua, M.-D. (2009a). Attitude observers for accelerated rigid bodies based on GPS and INS measurements. In *IEEE Conference on Decision and Control*. (Cited on page 25.)

Hua, M.-D. (2009b). *Contributions to the automatic control of aerial vehicles*. PhD thesis, Université de Nice-Sophia Antipolis. (Cited on page 24.)

Hua, M.-d., Ducard, G., Hamel, T., Mahony, R., and Rudin, K. (2013). Implementation of a nonlinear attitude estimator for aerial robotic vehicles. *IEEE Transactions on Control Systems Technology*. (Cited on pages 25 and 74.)

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons. (Cited on pages 42 and 160.)

Ikeda, M., Maeda, H., and Kodama, S. (1972). Stabilization of linear systems. *SIAM Journal on Control*, 4. (Cited on page 67.)

Irani, M. and Anandan, P. (1998). Robust multi-sensor image alignment. In *International Conference on Computer Vision*. (Cited on pages 14, 40, 41, and 43.)

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Dover. (Cited on pages 20, 21, and 22.)

Jones, E. and Soatto, S. (2011). Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *International Journal of Robotics Research*, 30(4). (Cited on page 27.)

Jones, E., Vedaldi, A., and Soatto, S. (2007). Inertial structure from motion with autocalibration. In *Workshop on Dynamical Vision*. (Cited on pages 27, 76, and 81.)

Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3). (Cited on page 22.)

Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. F. (2000). A new method for nonlinear transformations of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3). (Cited on page 22.)

Kailath, T. (1979). *Linear Systems*. Prentice-Hall. (Cited on pages 21 and 62.)

Kalman, R. E. (1960a). A new approach to linear filtering and prediction. *Transactions of the ASME – Journal of Basic Engineering*, 82. (Cited on page 20.)

Kalman, R. E. (1960b). On the general theory of control systems. In *International Congress of Automatic Control*. (Cited on pages 20 and 62.)

Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME – Journal of Basic Engineering*, 83. (Cited on pages 20 and 68.)

Keller, Y. and Averbuch, A. (2004). Fast motion estimation using bidirectional gradient methods. *IEEE Transactions on Image Processing*, 13(8). (Cited on page 44.)

Kelly, J. and Sukhatme, G. S. (2011). Visual-Inertial Sensor Fusion: Localization, Mapping and Sensor-to-Sensor Self-Calibration. *International Journal of Robotics Research*, 30(1). (Cited on pages 27, 76, and 81.)

Khalil, H. K. (2002). *Nonlinear Systems*. Prentice-Hall, 3 edition. (Cited on pages 69, 149, 150, 151, 152, 153, 154, and 156.)

Kim, H.-J., Kim, C., Rho, O.-H., and BLACK, H. D. (1964). A passive system for determining the attitude of a satellite. *AIAA Journal*, 2(7):1350–1351. (Cited on page 24.)

Klein, G. and Murray, D. (2007). Parallel Tracking and Mapping for Small AR Workspaces. In *International Symposium on Mixed and Augmented Reality*. (Cited on page 13.)

Krener, A. and Isidori, A. (1983). Linearization by output injection and nonlinear observers. *System & Control Letters*, 3. (Cited on page 69.)

Kuritsky, M. M. and Goldstein, M. S. (1983). Introduction and Overview of Inertial Navigation. *Proceedings of the IEEE*, 71(10). (Cited on page 8.)

Lageman, C., Trumpf, J., and Mahony, R. (2010). Gradient-Like Observers for Invariant Dynamics on a Lie Group. *IEEE Transactions on Automatic Control*, 55(2). (Cited on page 23.)

Lang, P. and Pinz, A. (2005). Calibration of Hybrid Vision/Inertial Tracking. In *Workshop on Integration of Vision and Inertial Senors*. (Cited on page 27.)

Lefferts, E. J., Markley, F. L., and Shuster, M. D. (1982). Kalman Filtering for Spacecraft Attitude Estimation. *Journal of Guidance, Control, and Dynamics*, 5(5). (Cited on pages 9, 12, 22, 25, 88, and 162.)

Lerman, H. (1983). Terrestrial Stellar-Inertial Navigation Systems. *Proceedings of the IEEE*, 71(10). (Cited on page 9.)

Lieberknecht, S., Benhimane, S., Meier, P., and Navab, N. (2009). A dataset and evaluation methodology for template-based tracking algorithms. In *International Symposium on Mixed and Augmented Reality*. (Cited on pages 32, 49, 50, 51, and 52.)

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2). (Cited on page 41.)

Lobo, J. and Dias, J. (2003). Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12). (Cited on page 26.)

Lobo, J. and Dias, J. (2007). Relative Pose Calibration Between Visual and Inertial Sensors. *International Journal of Robotics Research*, 26(6). (Cited on pages 27 and 61.)

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2). (Cited on page 13.)

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*. (Cited on page 13.)

Luenberger, D. G. (1964). Observing the State of a Linear System. *IEEE Transactions on Military Electronics*, 8(2). (Cited on page 21.)

Luenberger, D. G. (1966). Observers for Multivariable Systems. *IEEE Transactions on Automatic Control*, 11(5). (Cited on pages 21, 67, and 69.)

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. S. (2003). *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer. (Cited on pages 5, 14, 18, and 46.)

Mahony, R., Hamel, T., Morin, P., and Malis, E. (2012). Nonlinear complementary filters on the special linear group. *International Journal of Control*, 85(10). (Cited on page 23.)

Mahony, R., Hamel, T., and Pflimlin, J. M. (2005). Complementary Filter Design on the Special Orthogonal Group SO(3). In *IEEE Conference on Decision and Control*. (Cited on page 23.)

Mahony, R., Hamel, T., and Pflimlin, J. M. (2008). Nonlinear Complementary Filters on the Special Orthogonal Group. *IEEE Transactions on Automatic Control*, 53(5). (Cited on pages 25, 72, 74, and 149.)

Maithripala, D. H. S., Dayawansa, W. P., and Berg, J. M. (2005). Intrinsic observer-based stabilization for simple mechanical systems on lie groups. *SIAM Journal on Control and Optimization*, 44(5). (Cited on page 23.)

Malis, E. (2004). Improving vision-based control using efficient second-order minimization techniques. In *IEEE International Conference on Robotics and Automation*. (Cited on page 38.)

Malis, E. and Vargas, M. (2007). Deeper understanding of the homography decomposition for vision-based control. Technical report, INRIA. (Cited on page 19.)

Marchand, E. (1999). ViSP: A Software Environment for Eye-in-Hand Visual Ser- voing. In *IEEE International Conference on Robotics and Automation*. (Cited on page 13.)

Martin, P., Rouchon, P., and Rudolph, J. (2004). Invariant tracking. *ESAIM: Control, Optimisation and Calculus of Variations*, 10(1). (Cited on pages 72 and 78.)

Martin, P. and Rudolph, J. (1999). Invariant tracking and stabilization: problem formulation and examples. In Aeyels, D., Lamnabhi-Lagarrigue, F., and van der Schaft, A., editors, *Stability and Stabilization of Nonlinear Systems*. Springer. (Cited on page 23.)

Martin, P. and Salaun, E. (2008). An Invariant Observer for Earth-Velocity-Aided Attitude Heading Reference Systems. In *IFAC World Conference*. (Cited on pages 23, 25, and 74.)

Martin, P. and Salaün, E. (2010). Design and implementation of a low-cost observer-based attitude and heading reference system. *Control Engineering Practice*, 18(7). (Cited on page 25.)

Martinelli, A. (2011). State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis : The Case of Calibration. *IEEE Transactions on Robotics*, 27(2). (Cited on page 27.)

Martinelli, A. (2012). Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale, and Bias Determination. *IEEE Transactions on Robotics*, 28(1). (Cited on page 27.)

Mégret, R., Authesserre, J. B., and Berthoumieu, Y. (2008). The bi-directional framework for unifying parametric image alignment approaches. In *European Conference on Computer Vision*. (Cited on page 44.)

Meilland, M., Comport, A. I., and Rives, P. (2010). A spherical robot-centered representation for urban navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. (Cited on page 46.)

Mirzaei, F. M. and Roumeliotis, S. I. (2008). A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation. *IEEE Transactions on Robotics*, 24(5). (Cited on pages 27, 76, and 81.)

Morin, P. and Samson, C. (2008). Motion control of wheeled mobile robots. In *Handbook of Robotics*. Springer-Verlag. (Cited on page 73.)

Moustafa, M. (2001). History of inertial navigation systems in survey applications. *Journal of the American Society for Photogrammetry and Remote Sensing*, 67(11). (Cited on page 9.)

Newcombe, R. A., Lovegrove, S., and Davison, A. J. (2011). DTAM: Dense tracking and mapping in real-time. In *International Conference on Computer Vision*. (Cited on page 13.)

Nijmeijer, H. and Mareels, I. M. Y. (1997). An observer looks at synchronization. *IEEE Transactions on Circuits and Systems*, 44(10). (Cited on page 23.)

Nijmeijer, H. and van der Schaft, A. J. (1990). *Nonlinear Dynamical Control Systems*. Springer. (Cited on pages 23, 62, and 66.)

Nocedal, J. and Wright, S. J. (2000). *Numerical Optimization*. Springer. (Cited on pages 36 and 40.)

Ojeda, L., Chung, H., and Borenstein, J. (2000). Precision-calibration of fiber-optics gyroscopes for mobile robot navigation. In *IEEE International Conference on Robotics and Automation*. (Cited on page 11.)

Olivares, A., Olivares, G., Gorriz, J. M., and Ramirez, J. (2009). High-efficiency low-cost accelerometer-aided gyroscope calibration. In *International Conference on Test and Measurement*. Ieee. (Cited on page 11.)

Ozuysal, M., Fua, P., and Lepetit, V. (2007). Fast Keypoint Recognition in Ten Lines of Code. In *International Conference on Computer Vision*. (Cited on page 13.)

Pickering, M., Muhit, A. A., Scarvell, J. M., and Smith, P. N. (2009). A new multi-modal similarity measure for fast gradient-based 2D-3D image registration. In *IEEE International Conference of the Engineering in Medicine and Biology Society*. (Cited on pages 14, 31, and 33.)

Pitman, G. (1962). *Inertial Guidance*. John Wiley & Sons. (Cited on page 9.)

Rehbinder, H. and Ghosh, B. K. (2003). Pose estimation using line-based dynamic vision and inertial sensors. *IEEE Transactions on Automatic Control*, 48(2). (Cited on pages 23 and 26.)

Reinhard, H. (1989). *Equations différentielles: fondements et applications*. Dunod, 2 edition. (Cited on page 67.)

Richa, R., Sznitman, R., Taylor, R., and Hager, G. (2011). Visual tracking using the sum of conditional variance. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. (Cited on pages 14, 34, 39, 48, 50, 52, and 55.)

Roche, A., Malandain, G., Ayache, N., and Pennec, X. (1998a). Multimodal image registration by maximization of the correlation ratio. Technical report, INRIA. (Cited on page 35.)

Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998b). The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In *International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer. (Cited on pages 14, 31, and 35.)

Roth, P. M. and Winter, M. (2008). Survey of appearance-based methods for object recognition. Technical report, Institute for Computer Graphics and Vision, Graz University of Technology. (Cited on page 13.)

Salcudean, S. (1991). A globally convergent angular velocity observer for rigid body motion. *IEEE Transactions on Automatic Control*, 36(12). (Cited on pages 23 and 25.)

Sanyal, A. K. (2006). Optimal attitude estimation and filtering without using local coordinates, Part I: uncontrolled and deterministic attitude dynamics. In *American Control Conference*. (Cited on page 24.)

Scandaroli, G. G., Meilland, M., and Richa, R. (2012). Improving NCC-based Direct Visual Tracking. In *European Conference on Computer Vision*. (Cited on pages 3, 32, 39, and 41.)

Scandaroli, G. G. and Morin, P. (2011). Nonlinear filter design for pose and IMU bias estimation. In *IEEE International Conference on Robotics and Automation*. (Cited on pages 2, 61, 79, 88, 154, and 156.)

Scandaroli, G. G., Morin, P., and Silveira, G. (2011). A nonlinear observer approach for concurrent estimation of pose, IMU bias and camera-to-IMU rotation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. (Cited on pages 2, 61, 76, 153, and 156.)

Sepulchre, R., Jankovic, M., and Kokotovic, P. (1997). *Constructive Nonlinear Control*. Springer. (Cited on page 131.)

Servant, F., Houlier, P., and Marchand, E. (2010). Improving monocular plane-based SLAM with inertial measures. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Ieee. (Cited on page 26.)

Shi, J. and Tomasi, C. (1994). Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*. (Cited on page 13.)

Shuster, M. D. (1978). Approximate algorithms for fast optimal attitude computation. In *Guidance and Control Conference*. (Cited on page 25.)

Shuster, M. D. (1993). A Survey of Attitude Representations. *Journal of the Astronomical Sciences*, 41. (Cited on pages 7 and 24.)

Silveira, G. and Malis, E. (2010). Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images. *International Journal of Computer Vision*, 89(1). (Cited on pages 13, 31, 46, 50, 51, 52, and 55.)

Silveira, G., Malis, E., and Rives, P. (2008). An efficient direct approach to visual SLAM. *IEEE Transactions on Robotics*, 24(5). (Cited on page 13.)

Silver, M. (1983). RLG Strapdown System Navigation: A System Design Viewpoint. *Proceedings of the IEEE*, 71(10). (Cited on page 9.)

Simon, D. (2006). *Optimal State Estimation: Kalman, H-∞, and Nonlinear Approaches*. John Wiley & Sons. (Cited on pages 21 and 159.)

Song, Y. and Grizzle, J. (1995). The extended Kalman filter as a local asymptotic observer for discrete-time nonlinear systems. *Journal of Mathematical Systems, Estimation and Control*, 5(1). (Cited on page 23.)

Sontag, E. D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer. (Cited on page 72.)

Sorenson, H. W. (1985). *Kalman filtering: theory and application*. IEE Press. (Cited on page 21.)

Sussmann, H. J. (1979). Single-input observability of continuous-time systems. *Mathematical Systems Theory*, 12. (Cited on page 64.)

Szeliski, R. (2012). *Computer Vision Algorithms and Applications*. Springer. (Cited on page 14.)

Tayebi, A., McGilvray, S., Roberts, A., and Moallem, M. (2007). Attitude estimation and stabilization of a rigid body using low-cost sensors. In *IEEE Conference on Decision and Control*. Ieee. (Cited on page 25.)

Thienel, J. and Sanner, R. M. (2003). A coupled nonlinear spacecraft attitude controller and observer with an unknown constant gyro bias and gyro noise. *IEEE Transactions on Automatic Control*, 48. (Cited on pages 23 and 25.)

Tsai, R. Y. (1987). Versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4). (Cited on page 16.)

Tsai, R. Y. and Lenz, R. K. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5. (Cited on page 26.)

Vandyke, M. C., Schwartz, J. L., and Hall, C. D. (2004). Unscented Kalman Filtering for Spacecraft Attitude State and Parameter Estimation. In *AAS/AIAA Space Flight Mechanics Conference*. (Cited on page 25.)

Vasconcelos, J. F., Cunha, R., Silvestre, C., and Oliveira, P. (2007). Landmark based nonlinear observer for rigid body attitude and position estimation. In *IEEE Conference on Decision and Control*. (Cited on page 23.)

Vasconcelos, J. F., Silvestre, C., and Oliveira, P. (2008a). A Nonlinear GPS/IMU based observer for rigid body attitude and position estimation. In *IEEE Conference on Decision and Control*. (Cited on page 26.)

Vasconcelos, J. F., Silvestre, C., and Oliveira, P. (2008b). A Nonlinear Observer for Rigid Body Attitude Estimation using Vector Observations. In *IFAC World Conference*. (Cited on pages 23, 25, 73, and 74.)

Viéville, T. and Faugueras, O. (1989). Computation of inertial information on a robot. In *International Symposium on Robotics Research*. (Cited on pages 8, 9, and 26.)

Vik, B. and Fossen, T. I. (2001). A nonlinear observer for GPS and INS integration. In *IEEE Conference on Decision and Control*. (Cited on pages 23 and 26.)

Viola, P. and Wells, W. M. (1997). Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2). (Cited on pages 14, 31, and 35.)

Wahba, G. (1965). A least squares estimate of satellite attitude. *SIAM Reviews*, 7(3). (Cited on page 24.)

Wang, F. and Vemuri, B. C. (2007). Non-Rigid Multi-Modal Image Registration Using Cross-Cumulative Residual Entropy. *International Journal of Computer Vision*, 74(2). (Cited on pages 14, 31, and 35.)

Warner, F. W. (1987). *Foundations of differential manifolds and Lie groups*. Springer. (Cited on pages 7, 19, 22, 46, and 69.)

Yazdi, N., Ayazi, F., and Najafi, K. (1998). Micromachined Inertial Sensors. *Proceedings of the IEEE*, 86(8). (Cited on page 9.)

Zamani, M., Trumpf, J., and Mahony, R. (2011). Near-Optimal Deterministic Filtering on the Rotation Group. *IEEE Transactions on Automatic Control*, 56(6). (Cited on page 25.)

Zhang, Z. (1997). Parameter estimation techniques: a tutorial with application to conic filtering. *Image and Vision Computing*, 15(1). (Cited on page 160.)

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(11). (Cited on page 16.)

Zhang, Z. and Hanson, A. R. (1995). Scaled Euclidean 3D reconstruction based on externally uncalibrated cameras. In *International Symposium on Computer Vision*. (Cited on page 19.)

A

PROOFS FOR CHAPTER 3

A.1 PROOF OF LEMMA 3.1

In order to show that the linear time varying system (3.3) is observable, we have to verify the existence of constants $\tau, \delta > 0$ such that (3.4), *i.e.*

$$\forall t \geq 0, \quad 0 < \delta I_n \leq W(t, t+\tau) \triangleq \int_t^{t+\tau} \Psi(s,t)^T C^T(s) C(s) \Psi(s,t)\, ds,$$

is satisfied. The above inequality is equivalent to $x^T W(t, t+\tau) x \geq \delta |x|^2$ for any vector $x \in \mathbb{K} = \{x \in \mathbb{R}^n : |x| = 1\}$. Thus, the proof consists in showing the existence of constants $\tau, \delta > 0$ such that

$$\forall t \geq 0, \quad 0 < \delta \leq \min_{x \in \mathbb{K}} \int_t^{t+\tau} |C(s)\Psi(s,t)x|^2\, ds.$$

We proceed by contradiction. Assume that for any $\tau, \delta > 0$, there exists $t(\tau, \delta)$ such that

$$\min_{x \in \mathbb{K}} \int_{t(\tau,\delta)}^{t(\tau,\delta)+\tau} |C(s)\Psi(s, t(\tau,\delta))x|^2\, ds < \delta$$

Take $\tau = \overline{\tau}$ with τ the constant in (3.7), and consider the sequence $(\delta_p = 1/p)$. Thus, for any $p \in \mathbb{N}$, there exists t_p such that

$$\min_{x \in \mathbb{K}} \int_{t_p}^{t_p+\overline{\tau}} |C(s)\Psi(s,t_p)x|^2\, ds < \frac{1}{p}$$

so that there exists $x_p \in K$ such that

$$\int_{t_p}^{t_p+\overline{\tau}} |C(s)\Psi(s,t_p)x_p|^2\, ds < \frac{1}{p}. \tag{A.1}$$

Since $\mathbb{K}$ is compact, a sub-sequence of the sequence (x_p) converges to some $\overline{x} \in \mathbb{K}$. From Assumption 3.1, A is bounded on $[0, +\infty)$. Therefore,

$$\forall x \in \mathbb{R}^n,\ \forall t \leq s, \quad e^{-(s-t)\|A\|_\infty}|x| \leq |\Psi(s,t)x| \leq e^{(s-t)\|A\|_\infty}|x|, \tag{A.2}$$

where $\|A\|_\infty = \sup_{t \geq 0} \|A(t)\|$. Since C is also bounded (from Assumption 3.1) and the interval of integration in (A.1) is of fixed length $\overline{\tau}$, it follows that

$$\lim_{p \to +\infty} \int_{t_p}^{t_p+\overline{\tau}} |C(s)\Psi(s,t_p)\overline{x}|^2\, ds = 0.$$

By a change of integration variable, this equation can be written as

$$\lim_{p \to +\infty} \int_0^{\overline{\tau}} |f_p(s)|^2\, ds = 0, \tag{A.3}$$

where $f_p(t) = C(t+t_p)\Psi(t+t_p,t_p)\overline{x}$. Furthermore, it is easy to verify that

$$f_p^{(k)}(t) = N_k(t+t_p)\Psi(t+t_p,t_p)\overline{x}, \tag{A.4}$$

where $f_p^{(k)}$ the k-th order derivative of f_p and N_k defined by (3.6). The existence of $f_p^{(k)}$, for any $k = 0, \cdots, K+1$, follows by Assumption 1 of Proposition 3.1. The end of the proof relies on the following lemma, proved further.

Lemma A.1. For any $k = 0, \cdots, K$,

$$\lim_{p\to+\infty} \int_0^{\overline{\tau}} |f_p^{(k)}(s)|^2\, ds = 0. \tag{A.5}$$

Since the matrix M in (3.7) is composed of row vectors of $N_0, \cdots, N_K$, it follows from (A.4) that

$$\int_0^{\overline{\tau}} |M(s+t_p)\Psi(s+t_p,t_p)\overline{x}|^2\, ds \leq \sum_{k=0}^{K} \int_0^{\overline{\tau}} |f_p^{(k)}(s)|^2\, ds.$$

Therefore, from Lemma A.1,

$$\lim_{p\to+\infty} \int_{t_p}^{t_p+\overline{\tau}} |M(s)\Psi(s,t_p)\overline{x}|^2\, ds = \lim_{p\to+\infty} \int_0^{\overline{\tau}} |M(s+t_p)\Psi(s+t_p,t_p)\overline{x}|^2\, ds = 0. \tag{A.6}$$

Then, for any $\xi \in \mathbb{R}^n$

$$|M(s)\xi|^2 \geq |\xi|^2 \min_i \lambda_i(M^T(s)M(s)) = |\xi|^2 \lambda_1(M^T(s)M(s)) \tag{A.7}$$

with $\lambda_1(M^T(s)M(s)) \leq \cdots \leq \lambda_n(M^T(s)M(s))$ the eigenvalues of $M^T(s)M(s)$ in increasing order. Furthermore, since M is bounded on $[0,+\infty)$ (as a consequence of Assumption (3.1) and the definition of M), there exists a constant $c > 0$ such that

$$\max_i \lambda_i(M^T(s)M(s)) \leq c, \quad \forall\, s.$$

Thus

$$\lambda_1(M^T(s)M(s)) = \frac{\det(M^T(s)M(s))}{\prod_{j>1} \lambda_j(M^T(s)M(s))} \geq \frac{\det(M^T(s)M(s))}{c^{n-1}}$$

It follows from this inequality, (A.2), (A.7), and the fact that $|\overline{x}| = 1$ that

$$\forall p \in \mathbb{N}, \quad \int_{t_p}^{t_p+\overline{\tau}} |M(s)\Psi(s,t_p)\overline{x}|^2\, ds \geq \overline{c} \int_{t_p}^{t_p+\overline{\tau}} \det(M(s)^T M(s))\, ds \tag{A.8}$$

with $\overline{c} = e^{-2\tau\|A\|}/c^{n-1} > 0$. Thus, it follows from (3.7) and (A.8) that

$$\forall p \in \mathbb{N}, \quad \int_{t_p}^{t_p+\overline{\tau}} |M(s)\Psi(s,t_p)\overline{x}|^2\, ds \geq \overline{c}\delta > 0$$

which contradicts (A.6). To complete the proof, we must prove Lemma A.1.

Proof of Lemma A.1

Let us proceed by induction. From (A.3), (A.5) holds true for $k = 0$. Assuming that it holds true for $j = 0, \cdots, k < K$, we show that it holds true for $k+1$. First, Assumption 1 of Proposition 3.1 implies that for any $j = 1, \cdots K+1$, $f_p^{(j)}$ is well defined and bounded on $[0, \overline{\tau}]$, uniformly w.r.t. p.

We claim that $f_p^{(k)}(0)$ tends to zero as p tends to $+\infty$.

Assume on the contrary that $f_p^{(k)}(0)$ does not tend to zero. Then, there exists $\varepsilon > 0$ and a subsequence $(f_{p_j}^{(k)})$ of $(f_p^{(k)})$ such that $|f_{p_j}^{(k)}(0)| > \varepsilon$ for all $j \in \mathbb{N}$. Since $|f_{p_j}^{(k+1)}(0)|$ is bounded uniformly w.r.t. j (because $f_p^{(k+1)}$ is bounded on $[0, \overline{\tau}]$ uniformly w.r.t. p), there exists a constant $t' > 0$ such that

$$\forall j \in \mathbb{N},\ \forall t \in [0, t'], \qquad |f_{p_j}^{(k)}(t)| > \varepsilon/2$$

This contradicts the induction hypothesis (A.5) for k. Therefore, $f_p^{(k)}(0)$ tends to zero as p tends to $+\infty$. By a similar arguments, one can show that $f_p^{(k)}(\overline{\tau})$ tends to zero as p tends to $+\infty$. Now,

$$\begin{aligned}
\int_0^{\overline{\tau}} |f_p^{(k+1)}(s)|^2 \, ds &= \sum_{i=1}^n \int_0^{\overline{\tau}} \left(f_{p,i}^{(k+1)}(s)\right)^2 ds \\
&= -\sum_{i=1}^n \int_0^{\overline{\tau}} f_{p,i}^{(k)}(s) f_{p,i}^{(k+2)}(s)\, ds + \sum_{i=1}^n \left[f_{p,i}^{(k)}(s) f_{p,i}^{(k+1)}(s)\right]_0^{\overline{\tau}}, \\
&\leq \sum_{i=1}^n \left(\int_0^{\overline{\tau}} \left(f_{p,i}^{(k)}(s)\right)^2 ds\right)^{1/2} \left(\int_0^{\overline{\tau}} \left(f_{p,i}^{(k+2)}(s)\right)^2 ds\right)^{1/2} + \sum_{i=1}^n \left[f_{p,i}^{(k)}(s) f_{p,i}^{(k+1)}(s)\right]_0^{\overline{\tau}}.
\end{aligned}$$

Concluding, we have that each term

$$\left(\int_0^{\overline{\tau}} \left(f_{p,i}^{(k)}(s)\right)^2 ds\right)^{1/2} \left(\int_0^{\overline{\tau}} \left(f_{p,i}^{(k+2)}(s)\right)^2 ds\right)^{1/2}$$

in the first sum tends to zero as p tends to infinity due to (A.5) for k and the fact that $f_p^{(k+2)}$ is bounded uniformly w.r.t. p. Boundary terms in the second sum also tend to zero as p tends to infinity since $f_p^{(k)}(0)$ and $f_p^{(k)}(\overline{\tau})$ tend to zero, and $f_p^{(k+1)}$ is bounded.

A.2 COMPUTING IMPORTANT DETERMINANTS

Several proofs on this chapter rely on the widely known decomposition using Schur's complement. Let $A \in \mathbb{M}(n)$, $B \in \mathbb{M}(n,m)$, $C \in \mathbb{M}(m,n)$, $D \in \mathbb{M}(m,n)$, with nonsingular A. We can write

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_n & 0_{n\times m} \\ CA^{-1} & I_n \end{bmatrix} \begin{bmatrix} A & 0_{n\times m} \\ 0_{m\times n} & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I_n & A^{-1}B \\ 0_{m\times n} & I_n \end{bmatrix},$$

such that

$$\det(M) = \det(A)\det(D - CA^{-1}B), \tag{A.9}$$

and we can also verify for nonsingular D that

$$\det(M) = \det(D)\det(A - BD^{-1}C). \tag{A.10}$$

so that for the specific case when $D = -I_m$, we immediately obtain

$$\det(A + BC) = \det(A)\det(I_m + CA^{-1}B). \tag{A.11}$$

A.2.1 *On the determinant of* $\mathrm{S}(a)^2 + \mathrm{S}(b)^2$

We claim that

$$\forall\, a, b \in \mathbb{R}^3, \quad \det(\mathrm{S}(a)^2 + \mathrm{S}(b)^2) = -(|a|^2 + |b|^2)|a \times b|^2. \tag{A.12}$$

The relation is clearly satisfied when $a = 0_{3\times 1}$, thus let us focus on the case when $a \neq 0_{3\times 1}$. For any vector a, we have that $\mathrm{S}(a)^2 = -|a|^2 I_3 + aa^T$, thus

$$\det\left(\mathrm{S}(a)^2 + \mathrm{S}(b)^2\right) = \det\left(-(|a|^2 + |b|^2)I_3 + \begin{bmatrix} a & b \end{bmatrix}\begin{bmatrix} a & b \end{bmatrix}^T\right). \tag{A.13}$$

Thus, using (A.11), one obtains after (A.13) that

$$\begin{aligned}\det\left(\mathrm{S}(a)^2 + \mathrm{S}(b)^2\right) &= -(|a|^2 + |b|^2)\det\left(\begin{bmatrix} |a|^2 + |b|^2 & 0 \\ 0 & |a|^2 + |b|^2 \end{bmatrix} - \begin{bmatrix} |a|^2 & a^Tb \\ b^Ta & |b|^2 \end{bmatrix}\right), \\ &= -(|a|^2 + |b|^2)(|a|^2|b|^2 - (a^Tb)^2). \end{aligned} \tag{A.14}$$

Finally, let θ denote the angle between the vectors a and b, then $(a^Tb)^2 = |a|^2|b|^2\cos(\theta)^2$ and $|a \times b|^2 = |a|^2|b|^2\sin(\theta)^2 = |a|^2|b|^2(1 - \cos(\theta)^2)$. Hence, the relation (A.12) follows directly from (A.14).

A.2.2 *On the determinant of* $\mathrm{S}(a)^2 + \mathrm{S}(b)$

We claim that

$$\forall\, a, b \in \mathbb{R}^3, \quad \det\left(\mathrm{S}(a)^2 - \mathrm{S}(b)\right) = -|a \times b|^2 \tag{A.15}$$

The relation is clearly satisfied when $a = 0_{3\times 1}$. Let us assume that $a \neq 0_{3\times 1}$, then for any rotation matrix $R \in \mathbb{SO}(3)$,

$$R\left(\mathrm{S}(a)^2 + \mathrm{S}(b)\right)R^T = \mathrm{S}(Ra)^2 + \mathrm{S}(Rb) = \mathrm{S}(\bar{a})^2 + \mathrm{S}(\bar{b}) \tag{A.16}$$

with $\bar{a} = Ra$ and $\bar{b} = Rb$. Thus, let us consider the case with R such that $Ra = \bar{a} = (\bar{a}_1, 0, 0)^T$. Then,

$$\mathrm{S}(\bar{a})^2 + \mathrm{S}(\bar{b}) = \begin{bmatrix} 0 & \bar{b}_3 & -\bar{b}_2 \\ -\bar{b}_3 & -\bar{a}_1{}^2 & \bar{b}_1 \\ \bar{b}_2 & -\bar{b}_1 & -\bar{a}_1{}^2 \end{bmatrix}$$

We obtain after some straightforward calculation that

$$\det\left(\mathrm{S}(\bar{a})^2 + \mathrm{S}(\bar{b})\right) = -\bar{a}_1{}^2(\bar{b}_2{}^2 + \bar{b}_3{}^2) = -|\bar{a} \times \bar{b}|^2 = -|a \times b|^2.$$

Since $\mathrm{S}(\bar{a})^2 + \mathrm{S}(\bar{b}) = R(\mathrm{S}(a)^2 + \mathrm{S}(b))R^T$ and $\det(AB) = \det(A)\det(B)$, we have that Eq. (A.15) follows directly from (A.16) and the above equality.

A.3 OBSERVABILITY OF VISUO-INERTIAL SYSTEMS

A.3.1 *Proposition 3.2*

We claim in Proposition 3.2 that the system comprising body orientation ${}^{\mathcal{R}}R_{\mathcal{B}}$, angular rate gyro bias b_ω and c-to-IMU rotation ${}^{\mathcal{B}}R_{\mathcal{C}}$ with angular rate gyro measurements ${}^{\mathcal{B}}\omega_y$ in

the body frame and orientation ${}^{\mathcal{R}}R_{\mathcal{C}}$ in camera frame and is instantaneously observable if Assumption 3.2 and Eq. (3.30) hold, *i.e.*

$$\forall\, t \geq 0, \quad |{}^{\mathcal{B}}\dot{\omega}(t) \times {}^{\mathcal{B}}\ddot{\omega}(t)|^2 > 0.$$

Let us first recall the dynamics given in (3.29)

$$\begin{cases} {}^{\mathcal{R}}\dot{R}_{\mathcal{B}} = {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega_y - b_\omega), \\ \dot{b}_\omega = 0_{3\times 1}, \\ {}^{\mathcal{B}}\dot{R}_{\mathcal{C}} = 0_{3\times 3}, \end{cases}$$

with measurement $R_y = {}^{\mathcal{R}}R_{\mathcal{C}} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}R_{\mathcal{C}}$. We can rewrite this system in

$$X \triangleq ({}^{\mathcal{R}}R_{\mathcal{C}}, {}^{\mathcal{C}}R_{\mathcal{B}}({}^{\mathcal{B}}\omega_y - b_\omega), {}^{\mathcal{C}}R_{\mathcal{B}}) = (R, w, Q)$$

coordinates as

$$\begin{cases} \dot{R} = R\mathrm{S}(w), \\ \dot{w} = Q^{\mathcal{B}}\dot{\omega}, \\ \dot{Q} = 0_{3\times 3}, \end{cases} \tag{A.17}$$

with input $u = {}^{\mathcal{B}}\dot{\omega}$, and measurement $Y \triangleq R_y = R$. We can show that $\nexists$ two states X_1, X_2: $X_1 \neq X_2$, that generate the same output map if the inputs satisfy (3.30). First, let us compute the expression of $Y(t)$ and its derivatives

$$Y = R, \tag{A.18}$$
$$\dot{Y} = R\mathrm{S}(w), \tag{A.19}$$
$$\ddot{Y} = R\mathrm{S}(w)^2 + R\mathrm{S}(Q^{\mathcal{B}}\dot{\omega}), \tag{A.20}$$
$$\dddot{Y} = R\mathrm{S}(w)\ddot{Y} + R\mathrm{S}(Q^{\mathcal{B}}\ddot{\omega}). \tag{A.21}$$

The output map $R = Y$ in (A.18), $\forall\, t \geq 0$, is unique independently of the inputs. Furthermore, we obtain from (A.19) that $\mathrm{S}(\omega) = Y^T\dot{Y} \Longrightarrow \omega = \mathrm{vex}(Y^T\dot{Y})$, $\forall\, t \geq 0$, is also unique independently of the inputs. The same property is verified for Q using (A.20) and (A.21), we obtain that

$$\mathrm{S}(Q^{\mathcal{B}}\dot{\omega}) = Y^T\ddot{Y} - \mathrm{S}(w)^2 = \mathrm{P_a}(Y^T\ddot{Y} - \mathrm{S}(w)^2) = \mathrm{P_a}(Y^T\ddot{Y}),$$
$$\mathrm{S}(Q^{\mathcal{B}}\ddot{\omega}) = Y^T\dddot{Y} - Y^T\ddot{Y} = \mathrm{P_a}(Y^T\dddot{Y} - \ddot{Y})$$

and

$$\begin{bmatrix} \mathrm{vex}\big(\mathrm{P_a}(Y^T\ddot{Y})\big) \\ \mathrm{vex}\big(\mathrm{P_a}(Y^T\dddot{Y} - \ddot{Y})\big) \end{bmatrix} = \begin{bmatrix} Q\dot{\omega} \\ Q\ddot{\omega} \end{bmatrix}. \tag{A.22}$$

The above equation provides a linear map of that associates Q, ${}^{\mathcal{B}}\dot{\omega}$, ${}^{\mathcal{B}}\ddot{\omega}$ to a function of the outputs $\mathrm{vex}\big(\mathrm{P_a}(Y^T\ddot{Y})\big)$, $\mathrm{vex}\big(\mathrm{P_a}(Y^T\dddot{Y} - \ddot{Y})\big)$. Moreover, this output map is unique if the condition given by Eq. (3.30) holds. This concludes the observability analysis of (A.17), and consequently the analysis for the original system.

A.3.2 *Proposition 3.5*

We claim in Proposition 3.5 that the concurrent estimation body-position in the reference frame ${}^{\mathcal{R}}p_{\mathcal{B}}$, accelerometer bias b_a and the acceleration due to the gravitational field ${}^{\mathcal{R}}g$ in the reference frame is uniformly observable if Assumption 3.2, (3.42), *i.e.*

$$\forall\, t \geq 0, \quad \exists\, \tau, \delta > 0 : \int_t^{t+\tau} |{}^{\mathcal{B}}\omega(s) \times {}^{\mathcal{B}}\dot{\omega}(s)|^2 ds > \delta\,,$$

hold. We verify this statement using Proposition 3.1. Let us recall the dynamics given in (3.41) for the variables relating to the translational motion:

$$\begin{cases} {}^{\mathcal{R}}\dot{p}_{\mathcal{B}} = {}^{\mathcal{R}}v\,, \\ {}^{\mathcal{R}}\dot{v} = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a_y - z + {}^{\mathcal{R}}g, \\ \dot{z} = \mathrm{S}({}^{\mathcal{R}}\omega)z\,, \\ {}^{\mathcal{R}}\dot{g} = 0_{3\times 1}, \end{cases}$$

with measurements $(p_{\mathrm{y}}, R_{\mathrm{y}}, {}^{\mathcal{B}}\omega_{\mathrm{y}}) = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}R_{\mathcal{B}}, {}^{\mathcal{B}}\omega)$. We can define a state-affine system with states $x = ({}^{\mathcal{R}}p_{\mathcal{B}}, {}^{\mathcal{R}}v, z, {}^{\mathcal{R}}g)$, inputs $u{=}({}^{\mathcal{B}}R_{\mathcal{R}}, {}^{\mathcal{B}}\omega, {}^{\mathcal{B}}a_y)$ and outputs $y = {}^{\mathcal{R}}p_{\mathcal{B}}$ which yields a state-affine system

$$\begin{cases} \dot{x} = A(u)x + b(u)\,, \\ y = Cx + d(u) \end{cases}$$

with

$$A(u) = \begin{bmatrix} 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \\ 0_{3\times3} & 0_{3\times3} & \mathrm{S}({}^{\mathcal{R}}\omega) & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \quad b(u) = \begin{bmatrix} 0_{3\times1} \\ {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}a_y \\ 0_{3\times1} \\ 0_{3\times1} \end{bmatrix},$$

$$C(u) = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \quad d(u) = 0_{3\times1}.$$

We conclude the analysis using Proposition 3.1. Let us compute the elements of observable space directly via (3.6), *i.e.*

$$\begin{aligned} N_0 = C(u) &= \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \\ N_1 = N_0A + \dot{N}_0 &= \begin{bmatrix} 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \\ N_2 = N_1A + \dot{N}_1 &= \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \end{bmatrix}, \\ N_3 = N_2A + \dot{N}_2 &= \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -\mathrm{S}({}^{\mathcal{R}}\omega) & 0_{3\times3} \end{bmatrix}, \\ N_4 = N_3A + \dot{N}_3 &= \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -\big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{R}}\dot{\omega})\big) & 0_{3\times3} \end{bmatrix}, \end{aligned}$$

Moreover, we can define $M(t)$ stacking the N_0, N_1, N_2, and N_4 *i.e.*

$$M(t) = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \\ 0_{3\times3} & 0_{3\times3} & -\big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{R}}\dot{\omega})\big) & 0_{3\times3} \end{bmatrix}, \tag{A.23}$$

so that M is a square matrix, and directly via (A.11) we obtain $\det(M) = \det\big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{B}}\dot{\omega})\big)$. Moreover, let $u, v \in \mathbb{R}^3$ then $R \in \mathbb{SO}(3)$, $|(Ra) \times (Rb)| = |a \times b|$ and using (A.15), we simplify $\det(M) = -|{}^{\mathcal{B}}\omega \times {}^{\mathcal{B}}\dot{\omega}|^2$.

We conclude this observability analysis verifying the requirements from Proposition 3.1. Let ${}^{\mathcal{B}}\omega(t)$ such that ${}^{\mathcal{B}}\omega(t)$ and ${}^{\mathcal{B}}\dot{\omega}(t)$ are continuous and bounded, as Assumption 3.2 states, then $\det(M^TM) = \det(M)^2$, and we have that the system is uniformly observable if the original hypothesis (3.42) holds, which concludes this proof.

A.4 NONLINEAR OBSERVERS

The following result (Anderson and Moore, 1969, Th. 5) concerns the link between uniform observability and uniform exponential stability.

Lemma A.2. Consider an autonomous linear system

$$\dot{x} = A(t)x \tag{A.24}$$

with $A(t)$ continuous and bounded on $[0, +\infty)$. Assume that there exists a smooth matrix-valued function P satisfying the following Lyapunov inequalities for some constants $c_1, c_2 > 0$:

$$\begin{aligned} 0 < c_1 I \leq P \leq c_2 I \\ \dot{P} + PA + A^TP = -C^TC \end{aligned} \tag{A.25}$$

with C a bounded and continuous matrix-valued function. Then, System (A.24) is uniformly exponentially stable if the pair (A, C) is uniformly observable.

A.4.1 *Proof of Proposition 3.1*

This proof refers to the stability of the equilibrium point of the error dynamics (3.24) using the innovation terms from Proposition 3.1. This proof is similar to the proof of Theorem 3.3 in (Mahony et al., 2008). It is given for completeness. Let us recall the dynamics obtained by replacing the proposed innovation terms (3.26) in the (3.24)

$$\begin{cases} \dot{\widetilde{R}} = -\widetilde{R}\,\mathrm{S}\left({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\widetilde{b}_\omega + k_{R_{\mathcal{B}}} \dfrac{\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big)}{\big(1+\mathrm{tr}(\widetilde{R})\big)^2} \right), \\ \dot{\widetilde{b}}_\omega = k_\omega {}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} \mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big). \end{cases} \tag{A.26}$$

and defining $\mathbb{E}_\mathrm{s} = (I_3, 0_{3\times 1})$, and $\mathbb{E}_\mathrm{u} = \{(\widetilde{R}, \widetilde{b}_\omega) \in \mathbb{SO}(3) \times \mathbb{R}^3 \,|\, \mathrm{tr}(\widetilde{R}) = -1\}$, the proof thus consists of three statements:

1. Local exponential stability of the equilibrium set $\mathbb{E}_\mathrm{s}$;
2. Every solution stating in $\big(\widetilde{R}(0), \widetilde{b}_\omega(0)\big) \notin \mathbb{E}_\mathrm{u}$ converges to $\mathbb{E}_\mathrm{s}$.

We can prove the first statement using Lyapunov's indirect method, *c.f.* (Khalil, 2002, p. 161) together with Lemma A.2. We use a variable transformation $\widetilde{z} = -{}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_w$, ${}^{\mathcal{R}}\omega = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}\omega$ and some parametrization $\widetilde{R} \approx I_3 + \mathrm{S}(\widetilde{\theta})$ around the identity, *e.g.* an element $\widetilde{\theta} \in \mathbb{R}^3$ of $\mathfrak{so}(3)$ writes $\widetilde{R} = \exp(\mathrm{S}(\theta)) \approx I_3 + \mathrm{S}(\widetilde{\theta})$ around I_3. We obtain after linearizing (A.26)

$$\dot{\begin{bmatrix} \widetilde{\theta} \\ \widetilde{z} \end{bmatrix}} = \begin{bmatrix} -(k_{R_{\mathcal{B}}}/16) I_3 & I_3 \\ -k_\omega I_3 & \mathrm{S}({}^{\mathcal{R}}\omega) \end{bmatrix} \begin{bmatrix} \widetilde{\theta} \\ \widetilde{z} \end{bmatrix}, \tag{A.27}$$

now, for $k_{R_{\mathcal{B}}}, k_\omega > 0$, we consider the following Lyapunov candidate function

$$\mathcal{V} = \frac{16}{k_{R_{\mathcal{B}}}} \left(|\widetilde{\theta}|^2 + \frac{1}{k_\omega} |\widetilde{z}|^2 \right), \tag{A.28}$$

which along the solutions of (A.27)

$$\dot{\mathcal{V}} = -|\widetilde{\theta}|^2 \leq 0. \tag{A.29}$$

It is not very difficult to show using Barbalat's Lemma (Khalil, 2002, p. 323) that $(\widetilde{\theta}, \widetilde{z}) = 0$ is an asymptotically stable equilibrium of the linearized system, however, this is not even sufficient to prove local asymptotic stability of the original system. Lemma A.2 relates the uniform observability of a system to the uniform exponentially stability. Thus, let us define the states $x = (\widetilde{\theta}, \widetilde{z})$ and input $u = {}^{\mathcal{R}}\omega$, then $\dot{x} = \big(A(u) + LC\big)x$ with

$$A(u) = \begin{bmatrix} 0_{3\times3} & I_3 \\ 0_{3\times3} & \mathrm{S}({}^{\mathcal{R}}\omega) \end{bmatrix}, \quad L = \begin{bmatrix} -(k_{R_\mathcal{B}}/16)I_3 \\ -k_\omega I_3 \end{bmatrix}, \quad C = \begin{bmatrix} I_3 & 0_{3\times3} \end{bmatrix}.$$

Furthermore, let $P \in \mathbb{M}(6)$ a diagonal matrix where $V = x^T P x$ as Eq. (A.28), P and Eq. (A.29) satisfy indeed (A.25) and we can show via Lemma A.2 that System (A.27) is uniformly exponentially stable if the pair (A, C) is uniformly observable. Considering that Assumption 3.2 holds, such that that $A(u)$ and its derivatives are bounded, we can trivially verify that the pair (A, C) is uniformly observable using Proposition 3.1 independently of the system inputs. Therefore, the origin $(\widetilde{\theta}, \widetilde{z}) = (0_{3\times1}, 0_{3\times1})$ is an exponentially stable equilibrium of the linearized system, and using Lyapunov's indirect method (Khalil, 2002, p. 161), we verify that $(\widetilde{R}, \widetilde{b}_\omega) = (I_3, 0_{3\times1})$ is a locally exponentially stable equilibrium of (A.26), which concludes the proof of the first statement.

We continue the proof analyzing the Lyapunov candidate function

$$\mathcal{V} = \mathrm{tr}(I_3 - \widetilde{R}) + \frac{1}{k_\omega}|\widetilde{z}|^2, \tag{A.30}$$

and, along the solutions of (A.26), we obtain

$$\begin{aligned} \dot{\mathcal{V}} &= -\mathrm{tr}\big(\dot{\widetilde{R}}\big) + \frac{2}{k_\omega}\dot{\widetilde{z}}^T\widetilde{z} \\ &= \mathrm{tr}\big(\widetilde{R}\mathrm{S}({}^{\mathcal{R}}\widehat{R}_\mathcal{B}\widetilde{b}_\omega)\big) + k_{R_\mathcal{B}}\frac{\mathrm{tr}\big(\widetilde{R}\mathrm{P_a}(\widetilde{R})\big)}{(1+\mathrm{tr}(\widetilde{R}))^2} + 2\,\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))^T\,{}^{\mathcal{R}}\widehat{R}_\mathcal{B}\widetilde{b}_\omega \end{aligned} \tag{A.31}$$

The above derivative can be simplified recalling for $u, v \in \mathbb{R}^3$ and $R \in \mathbb{SO}(3)$ the properties

$$\begin{aligned} &\mathrm{tr}(R\mathrm{S}(u)) = \mathrm{tr}((\mathrm{P_s}(R) + \mathrm{P_a}(R))\mathrm{S}(u)) = \mathrm{tr}(\mathrm{P_a}(R)\mathrm{S}(u)), \\ &\mathrm{tr}(\mathrm{S}(u)\mathrm{S}(v)) = -2u^Tv, \\ &R\,\mathrm{vex}(\mathrm{P_a}(R)) = \mathrm{vex}(\mathrm{P_a}(R)), \end{aligned}$$

thus we can write

$$\begin{aligned} \dot{\mathcal{V}} =& -2\,\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))^T{}^{\mathcal{R}}\widehat{R}_\mathcal{B}\widetilde{b}_\omega - 2k_{R_\mathcal{B}}\frac{|\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))|^2}{(1+\mathrm{tr}(\widetilde{R}))^2} + 2\,\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))^T\,{}^{\mathcal{R}}\widehat{R}_\mathcal{B}\widetilde{b}_\omega \\ =& -2k_{R_\mathcal{B}}\frac{|\mathrm{vex}(\mathrm{P_a}(\widetilde{R}))|^2}{(1+\mathrm{tr}(\widetilde{R}))^2} \leq 0. \end{aligned} \tag{A.32}$$

Now, let us consider the angle axis parametrization for $\mathbb{SO}(3)$, *i.e.* $\widetilde{\theta} = \theta\widetilde{r}$ with $\theta \in (-\pi, \pi]$ and $\widetilde{r} \in \mathbb{R}^3$: $|r|^2 = 1$ such that $\widetilde{R} = \exp(\theta\widetilde{R})$. Using the Taylor expansion for the exponential matrix, it is easy to verify that $\mathrm{P_a}(\widetilde{R}) = \sin(\theta)\mathrm{S}(\widetilde{r})$ and $\mathrm{tr}(\widetilde{R}) = 1 + 2\cos(\theta)$. Moreover, the derivative (A.32) can be written as

$$\dot{\mathcal{V}} = -\frac{k_{R_\mathcal{B}}}{2}\frac{\sin(\theta)^2}{(1+\cos(\theta))^2} = -\frac{k_{R_\mathcal{B}}}{2}\tan(\theta/2)^2 \leq 0. \tag{A.33}$$

We can conclude so far that the errors are bounded since $\dot{\mathcal{V}}$ decreases, although not strictly. In order to continue and use Barbalat's lemma (Khalil, 2002) to show the asymptotic stability, we must first verify that $\dot{\mathcal{V}}$ is uniformly continuous along the solutions of (A.26). Let us analyze the angular part of the Lyapunov candidate function:

$$\mathcal{V}_\theta = \mathrm{tr}(I_3 - \widetilde{R}) = 1 - \cos(\theta),$$

then, along the solutions of (A.26)

$$\dot{\mathcal{V}}_\theta = -\frac{k_{R_\mathcal{B}}}{2}\tan(\theta/2)^2 - 2\sin(\theta)(\widetilde{r} \times \widetilde{b}_\omega) \leq -\frac{k_{R_\mathcal{B}}}{2}\tan(\theta/2)^2 + 2|\widetilde{b}_\omega|. \tag{A.34}$$

We have shown that the states are bounded in (A.33), therefore the rightmost term of (A.34) has an upper bound given by $|\widetilde{b}_\omega| \leq \overline{\widetilde{b}}_\omega$. Moreover, no matter how large but finite is $\overline{\widetilde{b}}_\omega$, there always exist some $\varepsilon > 0$: $|\theta| = \pi - \varepsilon \Rightarrow \tan(\theta/2)^2 > \overline{\widetilde{b}}_\omega$. Analogously, there will always exist some $|\theta| < \pi$ such that $\dot{\mathcal{V}}_\theta$ and, consequently, θ will decrease. We thus verified that $\mathcal{V}$ is uniformly continuous starting from any $\theta \in (-\pi, \pi)$ and never reach $\theta = \pm\pi$. This concludes the proof of the second statement.

We can now use Barbalat's lemma to continue the analysis of the asymptotic stability. Since the function $\mathcal{V}$ is bounded and $\dot{\mathcal{V}}$ is uniformly continuous, we obtain that $\dot{\mathcal{V}} \to 0$ as $t \to \infty$. Hence, the term $\tan(\theta/2) \to 0$, which in turn implies that the orientation error $\widetilde{R} \to I_3$ as $t \to \infty$. Furthermore, we can verify that every higher order derivative of $\dot{\mathcal{V}}$ is uniformly continuous and $\ddot{\mathcal{V}} \to 0$ as $t \to \infty$. Analogously, $\widetilde{b}_\omega \to 0_{3\times 1}$ as $t \to \infty$, which concludes the third statement and the proof of Proposition 3.1.

A.4.2 *Proof of Corollary 3.3*

In order to show that the error dynamics (3.24) using (3.28) is equivalent to the dynamics (3.24) using (3.26), let us recall the proposed innovation terms

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} \dfrac{{}^\mathcal{B}\widehat{R}_\mathcal{R} \sum_{i=1}^N k_i \big((V^{-1})^\mathcal{R}\beta_i\big) \times \big({}^\mathcal{R}\widehat{R}_\mathcal{B}{}^\mathcal{B}\beta_i\big)}{\big(1 + \sum_{i=1}^N k_{\beta_i}\big((V^{-1})^\mathcal{R}\beta_i\big)^T\big({}^\mathcal{R}\widehat{R}_\mathcal{B}{}^\mathcal{B}\beta_i\big)\big)^2}, \\ \alpha_\omega = -k_\omega {}^\mathcal{B}\widehat{R}_\mathcal{R} \sum_{i=1}^N k_i \big((V^{-1})^\mathcal{R}\beta_i\big) \times \big({}^\mathcal{R}\widehat{R}_\mathcal{B}{}^\mathcal{B}\beta_i\big). \end{cases}$$

where $V = \sum_{i=1}^N k_{\beta_i}{}^\mathcal{R}\beta_i{}^\mathcal{R}\beta_i{}^T$. Remark that ${}^\mathcal{R}\widehat{R}_\mathcal{B}{}^\mathcal{B}\beta_i = \widetilde{R}^T{}^\mathcal{R}R_\mathcal{B}{}^\mathcal{B}\beta = \widetilde{R}^T{}^\mathcal{R}\beta$, and considering $u, v \in \mathbb{R}^3$, recall that $\mathrm{S}(u \times v) = \mathrm{P_a}(uv^T)$, $\mathrm{tr}(uv^T) = v^T u$. We can thus rewrite the innovation terms (3.28) as

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} \dfrac{{}^\mathcal{B}\widehat{R}_\mathcal{R}\mathrm{vex}\big(\mathrm{P_a}\big(\widetilde{R}\big(\sum_{i=1}^N k_i{}^\mathcal{R}\beta_i{}^\mathcal{R}\beta_i{}^T\big)V^{-1}\big)\big)}{\Big(1 + \mathrm{tr}\big(\widetilde{R}\big(\sum_{i=1}^N k_{\beta_i}{}^\mathcal{R}\beta_i{}^\mathcal{R}\beta_i)^T\big)V^{-1}\big)\Big)^2}, \\ \alpha_\omega = -k_\omega {}^\mathcal{B}\widehat{R}_\mathcal{R}\mathrm{vex}\big(\mathrm{P_a}\big(\widetilde{R}\big(\sum_{i=1}^N k_i{}^\mathcal{R}\beta_i{}^\mathcal{R}\beta_i{}^T\big)V^{-1}\big)\big) \end{cases}$$

and from the definition of V we obtain

$$\begin{cases} \alpha_{R_\mathcal{B}} = k_{R_\mathcal{B}} \dfrac{{}^\mathcal{B}\widehat{R}_\mathcal{R}\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big)}{(1 + \mathrm{tr}(\widetilde{R}))^2}, \\ \alpha_\omega = -k_\omega {}^\mathcal{B}\widehat{R}_\mathcal{R}\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R})\big). \end{cases}$$

We immediately obtain that the error dynamics (3.24) using (3.28) is identical to the dynamics (3.24) using (3.26).

A.4.3 *Proof of Proposition 3.3*

This proof refers to the local exponential stability of equilibrium point of dynamics (3.32) using the innovation terms from Proposition 3.3. Let us recall the dynamics obtained by replacing the innovation terms (3.33) in (3.32)

$$\begin{cases} \dot{\widetilde{R}} &= -\widetilde{R}\mathrm{S}\Big({}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\big(\widetilde{b}_\omega + k_{R_{\mathcal{B}}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\big) - k_{R_{\mathcal{C}}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)\big)\Big), \\ \dot{\widetilde{b}}_\omega &= k_\omega{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\mathrm{vex}\big(\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})\big), \\ \dot{\widetilde{Q}} &= -\widetilde{Q}\mathrm{S}(k_{R_{\mathcal{C}}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega)). \end{cases} \tag{A.35}$$

We claim in (3.30) if

$$\exists\, \tau, \delta > 0 :\ \forall\, t \geq 0, \quad \int_t^{t+\tau} |{}^{\mathcal{B}}\dot{\omega}(s) \times {}^{\mathcal{B}}\ddot{\omega}(s)|^2 ds > \delta,$$

then $(\widetilde{R},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ is a locally exponentially stable equilibrium of (A.35). We verify this statement via Lyapunov's indirect method, *c.f.* (Khalil, 2002, p. 161), showing the uniform stability of a linearized system. In this case, we prove that the linearized system is uniformly exponentially stable using Lemma A.2 and Proposition 3.1.

Notice that one form of the estimation error for ${}^{\mathcal{R}}R_{\mathcal{C}}$ can be expressed by

$$\widetilde{R}_{\mathcal{C}} = {}^{\mathcal{R}}R_{\mathcal{C}}{}^{\mathcal{C}}\widehat{R}_{\mathcal{B}}{}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} = {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{Q}^{\mathcal{B}}\widehat{R}_{\mathcal{R}} = \widetilde{R}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\widetilde{Q}^{\mathcal{B}}\widehat{R}_{\mathcal{R}},$$

moreover, if $(\widetilde{R}_{\mathcal{C}},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ denotes a stable equilibrium point then $(\widetilde{R},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ is also a stable equilibrium point of dynamics (3.32). We continue the analysis considering the linearized system given by a parametrization of $\mathbb{SO}(3)$, *i.e.*

$$\widetilde{\xi} \in \mathbb{R}^3 : \widetilde{R}_{\mathcal{C}} \approx I_3 + \mathrm{S}(\widetilde{\xi}), \qquad \widetilde{\phi} \in \mathbb{R}^3 : \widetilde{R} \approx I_3 + \mathrm{S}(\widetilde{\phi}), \qquad \widetilde{\psi} \in \mathbb{R}^3 : \widetilde{Q} \approx I_3 + \mathrm{S}(\widetilde{\psi}).$$

Around the equilibrium, we can assume that

$${}^{\mathcal{B}}\omega \approx {}^{\mathcal{B}}\omega_y - \widehat{b}_\omega, \qquad {}^{\mathcal{R}}\widehat{R}_{\mathcal{B}} \approx {}^{\mathcal{R}}R_{\mathcal{B}},$$

then using these approximations and the expression, $\widetilde{R}_{\mathcal{C}} = \widetilde{R}^{\mathcal{R}}\widehat{R}_{\mathcal{B}}\widetilde{Q}^{\mathcal{B}}\widehat{R}_{\mathcal{R}}$ we obtain that, around the equilibrium, $\widetilde{R}_{\mathcal{C}} \approx I_3 + \mathrm{S}(\widetilde{\xi}) \approx (I_3 + \mathrm{S}(\widetilde{\phi}))(I_3 + \mathrm{S}({}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{\psi}))$. Thus,

$$\mathrm{S}(\dot{\widetilde{\xi}}) \approx \mathrm{S}(\dot{\widetilde{\phi}}) + \mathrm{S}({}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\widetilde{\psi} + {}^{\mathcal{R}}R_{\mathcal{B}}\dot{\widetilde{\psi}})$$

and using Eq. (A.35) and neglecting the higher order terms, we obtain

$$\begin{aligned} \dot{\widetilde{\xi}} \approx\ & -k_{R_{\mathcal{B}}}\widetilde{\xi} - {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_\omega + k_{R_{\mathcal{C}}}\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega) \\ & + {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}({}^{\mathcal{B}}\omega)\widetilde{\psi} - k_{R_{\mathcal{C}}}\mathrm{P_a}(\widetilde{R}_{\mathcal{C}})^{\mathcal{R}}\widehat{R}_{\mathcal{B}}({}^{\mathcal{B}}\omega_y - \widehat{b}_\omega) \end{aligned}$$

and the linearized system for $(\widetilde{R}_{\mathcal{C}},\, \widetilde{Q},\, \widetilde{b}_\omega)$ writes

$$\begin{cases} \dot{\widetilde{\xi}} &= -k_{R_{\mathcal{B}}}\widetilde{\xi} - {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_\omega + {}^{\mathcal{R}}R_{\mathcal{B}}\mathrm{S}(\widetilde{\psi})^{\mathcal{B}}\omega, \\ \dot{\widetilde{b}}_\omega &= k_\omega{}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{\xi}, \\ \dot{\widetilde{\psi}} &= k_{R_{\mathcal{C}}}\mathrm{S}({}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{\xi})^{\mathcal{B}}\omega. \end{cases} \tag{A.36}$$

Now, consider the following variable change $\widetilde{\theta} = {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{\xi}$, then in coordinates $(\widetilde{\theta}, \widetilde{b}_\omega,\, \widetilde{\psi})$, System (A.36) is thus given by

$$\begin{cases} \dot{\widetilde{\theta}} &= -\big(k_{R_{\mathcal{B}}}I_3 + \mathrm{S}({}^{\mathcal{B}}\omega)\big)\widetilde{\theta} - \widetilde{b}_\omega - \mathrm{S}({}^{\mathcal{B}}\omega)\widetilde{\psi}, \\ \dot{\widetilde{b}}_\omega &= k_\omega\widetilde{\theta}, \\ \dot{\widetilde{\psi}} &= -k_{R_{\mathcal{C}}}\mathrm{S}({}^{\mathcal{B}}\omega)\widetilde{\theta}. \end{cases} \tag{A.37}$$

The stability of the linearized system is verified via the Lyapunov candidate function

$$\mathcal{V} = \frac{1}{2k_{R_\mathcal{B}}}|\widetilde{\theta}|^2 + \frac{1}{2k_{R_\mathcal{B}}k_\omega}|\widetilde{b}_\omega|^2 + \frac{1}{2k_{R_\mathcal{B}}k_{R_\mathcal{C}}}|\widetilde{\psi}|^2. \tag{A.38}$$

Then, we obtain along the solutions of (A.37)

$$\begin{aligned}\dot{\mathcal{V}} =& -|\widetilde{\theta}|^2 - \frac{1}{k_{R_\mathcal{B}}}\widetilde{\theta}^T\widetilde{b}_\omega - \frac{1}{k_{R_\mathcal{B}}}\widetilde{\theta}^T\mathrm{S}({}^\mathcal{B}\omega)\widetilde{\psi} + \frac{1}{k_{R_\mathcal{B}}}\widetilde{b}_\omega{}^T\widetilde{\theta} - \frac{1}{k_{R_\mathcal{B}}}\widetilde{\psi}^T\mathrm{S}({}^\mathcal{B}\omega)\widetilde{\theta},\\ =& -|\widetilde{\theta}|^2 \leq 0.\end{aligned} \tag{A.39}$$

From this point, it is not very difficult to show using Barbalat's Lemma (Khalil, 2002, p. 323) and the observability condition (3.30) that $(\widetilde{\theta}, \widetilde{b}_\omega, \widetilde{\psi}) = 0$ is an asymptotically stable equilibrium of the linearized system. However, this is not even sufficient to prove local asymptotic stability of the original system.

We show in (Scandaroli et al., 2011) that it is possible to obtain a strictly decreasing Lyapunov function, however, the original result states the stability when

$$\forall\, t > 0, \qquad |\dot{\omega}(t) \times \ddot{\omega}(t)| > 0\,.$$

We can verify uniform stability of the solution via Lemma A.2, that relates the uniform observability to uniform exponential stability of an observer, and the observability condition given by Proposition 3.1.

Notice that System (A.37) defines a state-affine system $\dot{x} = \big(A(u) + L(u)C\big)x$ with states $x = (\widetilde{\theta}, \widetilde{b}_\omega, \widetilde{\psi})$ and inputs $u = {}^\mathcal{B}\omega$ where

$$A(u) = \begin{bmatrix} 0_{3\times3} & -I_3 & -\mathrm{S}({}^\mathcal{B}\omega) \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix},\ L(u) = \begin{bmatrix} -k_{R_\mathcal{B}}I_3 - \mathrm{S}({}^\mathcal{B}\omega) \\ k_\omega I_3 \\ k_{R_\mathcal{C}}\mathrm{S}({}^\mathcal{B}\omega) \end{bmatrix},\ C = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}.$$

Let $P \in \mathbb{M}(9)$ the diagonal matrix associated with the Lyapunov candidate function in (A.38), *i.e.* $\mathcal{V} = x^T P x$. Therefore P and Eq. (A.39) indeed satisfy (A.25), and we can show via Lemma A.2 that System (A.37) is uniformly exponentially stable if the pair (A, C) is uniformly observable. Considering that Assumption 3.2 holds, we obtain that $A(u,t)$ and its derivatives are bounded, we can verify a sufficient condition for the observability using Proposition 3.1. We can compute the components of the observable space as

$$\begin{aligned} N_0 = C &= \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \\ N_1 = N_0A + \dot{N}_0 &= \begin{bmatrix} 0_{3\times3} & -I_3 & -\mathrm{S}({}^\mathcal{B}\omega) \end{bmatrix}, \\ N_2 = N_1A + \dot{N}_1 &= \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -\mathrm{S}({}^\mathcal{B}\dot{\omega}) \end{bmatrix}, \\ N_3 = N_2A + \dot{N}_2 &= \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -\mathrm{S}({}^\mathcal{B}\ddot{\omega}) \end{bmatrix}, \end{aligned}$$

and define $M(t)$ stacking the N_0, N_1, N_2, and N_3, *i.e.*

$$M = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & -I_3 & -\mathrm{S}({}^\mathcal{B}\omega) \\ 0_{3\times3} & 0_{3\times3} & -\mathrm{S}({}^\mathcal{B}\dot{\omega}) \\ 0_{3\times3} & 0_{3\times3} & -\mathrm{S}({}^\mathcal{B}\ddot{\omega}) \end{bmatrix},\quad M^TM = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_3 & \mathrm{S}({}^\mathcal{B}\omega) \\ 0_{3\times3} & -\mathrm{S}({}^\mathcal{B}\omega) & -\mathrm{S}({}^\mathcal{B}\omega)^2 - \mathrm{S}({}^\mathcal{B}\dot{\omega})^2 - \mathrm{S}({}^\mathcal{B}\ddot{\omega})^2 \end{bmatrix}.$$

We thus obtain from (3.7) that the pair (A, C) is observable if

$$\exists\, \tau, \delta > 0 : \ \forall\, t \geq 0, \quad \delta \leq \int_t^{t+\tau} |\det(M(s)^T M(s))|\, ds.$$

Thus, using (A.9), we directly compute that $\det(M^T M) = -\det\big(\mathrm{S}({}^{\mathcal{R}}\dot{\omega})^2 + \mathrm{S}({}^{\mathcal{B}}\ddot{\omega})^2\big)$, and, via (A.12), $\det(M^T M) = (|{}^{\mathcal{B}}\dot{\omega}|^2 + |{}^{\mathcal{B}}\ddot{\omega}^2|)|{}^{\mathcal{B}}\dot{\omega} \times {}^{\mathcal{B}}\ddot{\omega}|^2$. We thus obtain using Proposition 3.1 that the pair (A, C) is observable if condition (3.30) holds. This concludes the proof of uniform exponential stability of the origin $(\widetilde{\theta},\, \widetilde{b}_\omega,\, \widetilde{\psi})$=0 for System (A.37), and consequently for the linearized System (A.36). As the origin of (A.36) is a uniformly exponentially stable, then, using Lyapunov's indirect method, $(\widetilde{R}_{\mathcal{C}},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ is a locally exponentially stable equilibrium point of the nonlinear system and so is $(\widetilde{R},\, \widetilde{b}_\omega,\, \widetilde{Q}) = (I_3,\, 0_{3\times 1},\, I_3)$ for the dynamics (A.35).

A.4.4 *Proof of Proposition 3.4*

This proof concerns the exponential stability of the equilibrium point of (3.37) using the innovation terms from Proposition 3.4. In this specific case, we consider $\widetilde{R} \approx I_3$ and $\widetilde{b}_\omega \approx 0_{3\times 1}$. Hence, we obtain the following dynamics from (3.37) and the innovation terms given by (3.38):

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - k_{p_{\mathcal{B}}} \widetilde{p}\,, \\ \dot{\widetilde{v}} = \ -{}^{\mathcal{R}}R_{\mathcal{B}} \widetilde{b}_a - k_v \widetilde{p}\,, \\ \dot{\widetilde{b}}_a = k_a \big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}} \mathrm{S}({}^{\mathcal{B}}\omega)\big)\, {}^{\mathcal{B}}R_{\mathcal{R}}\, \widetilde{p}\,. \end{cases} \tag{A.40}$$

Let us consider the following variable change

$$\widetilde{z} = {}^{\mathcal{R}}R_{\mathcal{B}} \widetilde{b}_a + \tfrac{k_a}{k_{p_{\mathcal{B}}}} \widetilde{p}, \tag{A.41}$$

and $k'_v = k_v - \frac{k_a}{k_{p_{\mathcal{B}}}}$, ${}^{\mathcal{R}}\omega = {}^{\mathcal{R}}R_{\mathcal{B}}\, {}^{\mathcal{B}}\omega$. Hence, we can write system (A.40) in $(\widetilde{p},\, \widetilde{v},\, \widetilde{z})$ coordinates as

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - k_{p_{\mathcal{B}}} \widetilde{p}\,, \\ \dot{\widetilde{v}} = \ -\widetilde{z} - k'_v \widetilde{p}\,, \\ \dot{\widetilde{z}} = \mathrm{S}({}^{\mathcal{R}}\omega)\widetilde{z} + \frac{k_a}{k_{p_{\mathcal{B}}}} \widetilde{v}\,. \end{cases} \tag{A.42}$$

Now, let us define the following Lyapunov candidate function:

$$\mathcal{V} = \frac{1}{2k_{p_{\mathcal{B}}}} |\widetilde{p}|^2 + \frac{1}{2k_{p_{\mathcal{B}}} k'_v} |\widetilde{v}|^2 + \frac{1}{2k'_v k_a} |\widetilde{z}|^2, \tag{A.43}$$

and remark that $\mathcal{V}$ is indeed a definite positive function due to the constraint $k_{p_{\mathcal{B}}},\, k_v,\, k_a > 0$ with $k_a < k_{p_{\mathcal{B}}} k_v$ on the gains. We obtain along the solutions of System (A.42)

$$\begin{aligned} \dot{\mathcal{V}} = \frac{1}{k_{p_{\mathcal{B}}}} \widetilde{v}^T \widetilde{p} - |\widetilde{p}|^2 - \frac{1}{k_{p_{\mathcal{B}}} k'_v} \widetilde{z}^T \widetilde{v} - \frac{1}{k_{p_{\mathcal{B}}}} \widetilde{p}^T \widetilde{v} + \frac{1}{k_{p_{\mathcal{B}}} k'_v} \widetilde{v}^T \widetilde{z} \\ = -|\widetilde{p}|^2 \leq 0. \end{aligned} \tag{A.44}$$

At this point, we could use Barbalat's Lemma (Khalil, 2002, p. 323) and Assumption 3.2 to prove that $(\widetilde{p},\, \widetilde{v},\, \widetilde{z}) = (0_{3\times 1},\, 0_{3\times 1},\, 0_{3\times 1})$ is a globally *asymptotically* stable equilibrium point of the system. However, we want to establish *exponential* stability. We show in (Scandaroli and Morin, 2011) how to obtain a strictly decreasing Lyapunov function for System (A.42),

but in this thesis we proceed using Lemma A.2, since the proof is straightforward using Proposition 3.1.

Remark that System (A.42) defines a state-affine system $\dot{x} = \big(A(u) + LC\big)x$ with states $x = (\widetilde{p}, \widetilde{v}, \widetilde{z})$ and inputs $u = {}^{\mathcal{R}}\omega$ where

$$A(u) = \begin{bmatrix} 0_{3\times3} & I_3 & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -I_3 \\ 0_{3\times3} & -\frac{k_a}{k_{p_\mathcal{B}}} I_3 & \mathrm{S}({}^{\mathcal{R}}\omega) \end{bmatrix}, \; L(u) = \begin{bmatrix} -k_{p_\mathcal{B}} I_3 \\ -k_v' I_3 \\ 0_{3\times3} \end{bmatrix}, \; C = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}.$$

Let $P \in \mathbb{M}(9)$ denote the diagonal matrix associated with the Lyapunov candidate function in Eq. (A.43), *i.e.* $\mathcal{V} = x^T P x$. Therefore, P and Eq. (A.44) satisfy indeed (A.25) and we can show via Lemma A.2 that System (A.42) is uniformly exponentially stable if the pair (A, C) is uniformly observable. Considering that Assumption 3.2 holds, we have that $A(u)$ and its derivatives are bounded, and it is straightforward to verify using Proposition 3.1 that the pair (A, C) is uniformly observable independently of the system inputs. Therefore, the origin $(\widetilde{p}, \widetilde{v}, \widetilde{z}) = (0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is a uniformly exponentially stable equilibrium of the system (A.42), and $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a) = (0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is also a uniformly exponentially stable equilibrium of the original system. This concludes the proof of Proposition 3.4.

A.4.5 *Proof of Corollary 3.4*

In the proof of Proposition 3.4, we assumed that $\widetilde{R} \approx I_3$ and $\widetilde{b}_\omega \approx 0_{3\times1}$, and the equilibrium of the resulting error dynamics is globally exponentially stable. That result can be initially extended to the case where $\widetilde{b}_\omega \to 0_{3\times1}$. We obtain the following dynamics from (3.37) and the innovation terms given by (3.39):

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - k_{p_\mathcal{B}} \widetilde{p}\,, \\ \dot{\widetilde{v}} = \; - {}^{\mathcal{R}}R_\mathcal{B} \widetilde{b}_a - k_v \widetilde{p}\,, \\ \dot{\widetilde{b}}_a = k_a \big(I_3 + \frac{1}{k_{p_\mathcal{B}}} \mathrm{S}({}^{\mathcal{B}}\omega_y - \widetilde{b}_\omega)\big) {}^{\mathcal{B}}R_\mathcal{R} \widetilde{p}\,. \end{cases}$$

Moreover, defining the states the states $X = (\widetilde{p}, \widetilde{v}, \widetilde{b}_a)$, the above dynamics can be written in the form

$$\dot{X} = A(t)X + f_0(X, \widetilde{b}_\omega, t), \tag{A.45}$$

where $A(t)$ is given by the right-hand side of System (A.42), and $f_0(X, \widetilde{b}_\omega, t)$ is a "perturbation term" that satisfies

$$|f_0(X, \widetilde{b}_\omega, t)| \leq c|X||\widetilde{b}_\omega| \tag{A.46}$$

for some constant $c > 0$. The above relation implies that the solutions of the system are well defined for all time. Recall from the proof of nonlinear observer 3.4, that since the system is uniformly exponentially stable, there exists a positive definite Lyapunov function $\mathcal{V}$ such that, along the solution of $\dot{X} = A(t)X$,

$$\dot{\mathcal{V}} \leq -\eta \mathcal{V}\,, \quad \eta > 0. \tag{A.47}$$

Furthermore, if $\widetilde{b}_\omega(t)$ converges asymptotically to zero regardless of the initial conditions, then we deduced from (A.46) and (A.47) that, along any solution of the System (A.45), there exists $T \geq 0$ such that for $t \geq T$, $\dot{\mathcal{V}} \leq -\frac{1}{2}\eta \mathcal{V}$. Convergence to zero of X readily follows from this inequality.

A.4.6 *Proof of Proposition 3.6*

This proof refers to the local exponential stability of equilibrium point of dynamics (3.43) using the innovation terms from Proposition 3.6. Let us recall the dynamics obtained by replacing the innovation terms (3.44) in (3.43)

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - k_{p_{\mathcal{B}}}\widetilde{p}\,, \\ \dot{\widetilde{v}} = -{}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_a - k_v\widetilde{p} + \widetilde{g}\,, \\ \dot{\widetilde{b}}_a = k_a\big(I_3 + \frac{1}{k_{p_{\mathcal{B}}}}\mathrm{S}({}^{\mathcal{B}}\omega)\big){}^{\mathcal{B}}R_{\mathcal{R}}\widetilde{p}\,, \\ \dot{\widetilde{g}} = -k_g\widetilde{p}\,. \end{cases} \tag{A.48}$$

Moreover, we claim in (3.45) if

$$\exists\, \tau, \delta > 0 :\ \forall\, t \geq 0, \quad \int_t^{t+\tau} |\omega_{\mathcal{B}}(s) \times \dot{\omega}_{\mathcal{B}}(s)| ds > \delta.$$

then $(\widetilde{p},\, \widetilde{v},\, \widetilde{b}_a,\, \widetilde{g}) = (0_{3\times1},\, 0_{3\times1},\, 0_{3\times1},\, 0_{3\times1})$ is a globally exponentially stable equilibrium of the dynamics (A.48). We verify that this system is uniformly exponentially stable using Lemma A.2 and Proposition 3.1. Let us consider the following variable change

$$\widetilde{z} = {}^{\mathcal{R}}R_{\mathcal{B}}\widetilde{b}_a + \frac{k_a}{k_{p_{\mathcal{B}}}}\widetilde{p}, \qquad \widetilde{w} = \widetilde{g} - \frac{k_g}{k_{p_{\mathcal{B}}}}\widetilde{p}, \tag{A.49}$$

and $k'_v = k_v - \frac{k_a + k_g}{k_{p_{\mathcal{B}}}}$, ${}^{\mathcal{R}}\omega = {}^{\mathcal{R}}R_{\mathcal{B}}{}^{\mathcal{B}}\omega$. Hence, we can write system (A.48) in $(\widetilde{p},\, \widetilde{v},\, \widetilde{z},\, \widetilde{w})$ coordinates as

$$\begin{cases} \dot{\widetilde{p}} = \widetilde{v} - k_{p_{\mathcal{B}}}\widetilde{p}\,, \\ \dot{\widetilde{v}} = -\widetilde{z} + \widetilde{w} - k'_v\widetilde{p}\,, \\ \dot{\widetilde{z}} = \mathrm{S}({}^{\mathcal{R}}\omega)\widetilde{z} + \frac{k_a}{k_{p_{\mathcal{B}}}}\widetilde{v} \\ \dot{\widetilde{w}} = -\frac{k_g}{k_{p_{\mathcal{B}}}}\widetilde{v}. \end{cases} \tag{A.50}$$

Now, let us define the following Lyapunov candidate function:

$$\mathcal{V} = \frac{1}{2k_{p_{\mathcal{B}}}}|\widetilde{p}|^2 + \frac{1}{2k_{p_{\mathcal{B}}}k'_v}|\widetilde{v}|^2 + \frac{1}{2k'_v k_a}|\widetilde{z}|^2 + \frac{1}{2k'_v k_g}|\widetilde{w}|^2, \tag{A.51}$$

and remark that $\mathcal{V}$ is indeed a definite positive function due to the constraint $k_{p_{\mathcal{B}}},\, k_v,\, k_a\, k_g > 0$ on the gains with $k_a + k_g < k_{p_{\mathcal{B}}}k_v$. We obtain along the solutions of System (A.50)

$$\begin{aligned} \dot{\mathcal{V}} &= \frac{1}{k_{p_{\mathcal{B}}}}\widetilde{v}^T\widetilde{p} - |\widetilde{p}|^2 - \frac{1}{k_{p_{\mathcal{B}}}k'_v}\widetilde{z}^T\widetilde{v} + \frac{1}{k_{p_{\mathcal{B}}}k'_v}\widetilde{w}^T\widetilde{v} - \frac{1}{k_{p_{\mathcal{B}}}}\widetilde{p}^T\widetilde{v} + \frac{1}{k_{p_{\mathcal{B}}}k'_v}\widetilde{v}^T\widetilde{z} - \frac{1}{k_{p_{\mathcal{B}}}k'_v}\widetilde{v}^T\widetilde{w} \\ &= -|\widetilde{p}|^2 \leq 0. \end{aligned} \tag{A.52}$$

Similarly to the proofs of Proposition 3.3 and 3.4, from this point, it is not very difficult to show using Barbalat's Lemma (Khalil, 2002, p. 323) and the observability condition (3.45) that $(\widetilde{p},\, \widetilde{v},\, \widetilde{z},\, \widetilde{w}) = 0$ is an asymptotically stable equilibrium of system (A.50). However, we want to establish the exponential stability of the filter. We could use similar strategies to (Scandaroli and Morin, 2011) and (Scandaroli et al., 2011) and obtain a strictly decreasing Lyapunov function, but instead we consider Lemma A.2 so that we conclude the proof using Proposition 3.1 and obtain a uniform condition.

Remark that System (A.50) defines a state-affine system $\dot{x} = \big(A(u) + LC\big)x$ with states $x = (\widetilde{p}, \widetilde{v}, \widetilde{z}, \widetilde{w})$ and inputs $u = {}^{\mathcal{R}}\omega$ where

$$A(u) = \begin{bmatrix} 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \\ 0_{3\times3} & \frac{k_a}{k_{p_\mathcal{B}}} I_3 & \mathrm{S}({}^{\mathcal{R}}\omega) & 0_{3\times3} \\ 0_{3\times3} & -\frac{k_g}{k_{p_\mathcal{B}}} I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \quad L = \begin{bmatrix} -k_{p_\mathcal{B}} I_3 \\ -k'_v I_3 \\ 0_{3\times3} \\ 0_{3\times3} \end{bmatrix},$$
$$C = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}.$$

Moreover, let $P \in \mathbb{M}(9)$ denote the diagonal matrix associated with the Lyapunov candidate function in Eq. (A.51), *i.e.* $\mathcal{V} = x^T P x$. Thus P and Eq. (A.44) satisfy indeed (A.25) and we can show via Lemma A.2 that System (A.42) is uniformly exponentially stable if the pair (A, C) is uniformly observable. Considering that Assumption 3.2 holds, we have that $A(u)$ and its derivatives are bounded and continue, thus the first requisite of Proposition 3.1 is satisfied.

Next, let us compute the components of the observable space

$$\begin{aligned} N_0 &= C = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \\ N_1 &= N_0 A + \dot{N}_0 = \begin{bmatrix} 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \\ N_2 &= N_1 A + \dot{N}_1 = \begin{bmatrix} 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \end{bmatrix}, \\ N_3 &= N_2 A + \dot{N}_2 = \begin{bmatrix} 0_{3\times3} & -\frac{k_a+k_g}{k_{p_\mathcal{B}}} I_3 & -\mathrm{S}({}^{\mathcal{R}}\omega) & 0_{3\times3} \end{bmatrix}, \\ N_4 &= N_3 A + \dot{N}_3 = \begin{bmatrix} 0_{3\times3} & \frac{k_a}{k_{p_\mathcal{B}}}\mathrm{S}({}^{\mathcal{R}}\omega) & \frac{k_a+k_g}{k_{p_\mathcal{B}}} I_3 - \big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{R}}\omega)\big) & -\frac{k_a+k_g}{k_{p_\mathcal{B}}} I_3 \end{bmatrix}. \end{aligned}$$

This observable space is computed similarly to (A.23) in Section A.3.2, *i.e.* , we can define $M(t)$ stacking the N_0, N_1, N_2, and N_4:

$$M(t) = \begin{bmatrix} I_3 & 0_{3\times3} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_3 & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & -I_3 & I_3 \\ 0_{3\times3} & \frac{k_a}{k_{p_\mathcal{B}}}\mathrm{S}({}^{\mathcal{R}}\omega) & \frac{k_a+k_g}{k_{p_\mathcal{B}}} I_3 - \big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{R}}\omega)\big) & -\frac{k_a+k_g}{k_{p_\mathcal{B}}} I_3 \end{bmatrix},$$

so that M is a square matrix. Since $\frac{k_a+k_g}{k_{p_\mathcal{B}}} > 0$, we obtain directly via (A.11)

$$\det(M) = \det\big(\mathrm{S}({}^{\mathcal{R}}\omega)^2 + \mathrm{S}({}^{\mathcal{B}}\dot{\omega})\big).$$

Moreover, let $u, v \in \mathbb{R}^3$ then $R \in \mathbb{SO}(3)$, $|(Ra) \times (Rb)| = |a \times b|$ and using (A.15), we simplify $\det(M) = -|{}^{\mathcal{B}}\omega \times {}^{\mathcal{B}}\dot{\omega}|^2$. We thus obtain using Proposition 3.1 that the pair (A, C) is observable if condition (3.45) holds. This concludes the proof of uniform exponential stability of the origin $(\widetilde{p}, \widetilde{v}, \widetilde{z}, \widetilde{w})$=0 for System (A.50), and $(\widetilde{p}, \widetilde{v}, \widetilde{b}_a, \widetilde{g}) = (0_{3\times1}, 0_{3\times1}, 0_{3\times1}, 0_{3\times1})$ is also a uniformly exponentially stable equilibrium of the original system. This concludes the proof of Proposition 3.6.

B

PARAMETER ESTIMATION ROBUST TO OUTLIERS

Many computer vision tasks relate to the problem of parameter estimation, where we must separate data that actually belong to the model, *i.e.* inliers, from ones that do not, *i.e.* outliers. The optimal parameter estimation problem is one of computing

$$x^* = \underset{x \in \mathbb{R}^n}{\arg\min}\, V(x).$$

The above problem is a convex (resp. quasi-convex) minimization if $V(x) \in \mathbb{R}^+$ is a convex (resp. quasi-convex) function (Boyde and VanderbergheL., 2004). Furthermore, the parameter x_i^* is the global minimum of a convex (resp. quasi-convex) $V(x)$ iif:

$$\text{a) } \partial_x V(x) = 0\,; \qquad \text{b) } \partial_x{}^2 V(x) > 0\,. \tag{B.1}$$

More specifically, we are interested in the optimization of convex functions of the linear residuals

$$r_i(x) \triangleq y_i - {j_i}^T x, \tag{B.2}$$

where $y_i \in \mathbb{R}$ denote the i-th measurement, $j_i \in \mathbb{R}^n$ refers to the measurement model, and $x \in \mathbb{R}^n$ are candidate parameters.

B.1 WEIGHTED LEAST-SQUARES

The weighted least-squares is a problem with objective given by the weighted sum of squared residuals, *i.e.*

$$V(x) = \frac{1}{2}\sum_i w_i {r_i}^2 = \frac{1}{2}\sum_i w_i (y_i - {j_i}^T x)^2\,,$$

where $w_i \in \mathbb{R}^+$. This problem derives from the classic least-squares, *i.e.* with $w_i = 1$, $\forall i$, however, in this case, some measurements are more reliable than others and w_i can be tuned to represent the level of confidence. The solution is given by computing (B.1, a),

$$\big(\sum w_i j_i {j_i}^T\big) x - \sum w_i j_i y_i = 0\,. \tag{B.3}$$

Furthermore, if condition (B.1, b) is verified, *i.e.* $\partial_x^2 V(x) = \sum w_i j_i {j_i}^T > 0$, the optimal solution is then given by

$$x^* = \big(\sum w_i j_i {j_i}^T\big)^{-1} \big(\sum w_i j_i y_i\big)\,. \tag{B.4}$$

A possible choice for the weights w_i is based on the variance of each measurement y_i (Simon, 2006), this procedure is not robust to outliers however.

Table B.1: Examples of robust loss functions

type	$\rho(r)$	$\psi(r)$	$w(r)$
L_1	$\lvert r\rvert$	$\operatorname{sign}(r)$	$1/\lvert r\rvert$
L_2	$r^2/2$	r	1
Huber $\begin{cases}\text{if } \lvert r\rvert < k\,;\\ \text{otherwise.}\end{cases}$	$\begin{cases} r^2/2\,;\\ k(\lvert r\rvert - k/2)\,. \end{cases}$	$\begin{cases} r\,;\\ k\operatorname{sign}(r)\,. \end{cases}$	$\begin{cases} 1\,;\\ k/\lvert r\rvert\,. \end{cases}$
Pseudo-Huber	$k^2(\sqrt{1+(r/k)^2}-1)$	$r/\sqrt{1+(r/k)^2}$	$1/\sqrt{1+(r/k)^2}$
Tukey $\begin{cases}\text{if } \lvert r\rvert < c\,;\\ \text{otherwise.}\end{cases}$	$\begin{cases} \frac{c^2}{6}\left(1-(1-(r/c)^2)^3\right);\\ \frac{c^2}{6}\,. \end{cases}$	$\begin{cases} r(1-(r/c)^2)^2\,;\\ 0\,. \end{cases}$	$\begin{cases} (1-(r/c)^2)^2\,;\\ 0,. \end{cases}$

B.2 M-ESTIMATORS

M-estimators are a class of robust estimators whose objective is given by

$$V(x) = \frac{1}{2}\sum_i \rho(r_i),$$

where $\rho(\cdot) \in \mathbb{R}^+$ is convex symmetric function with a unique minimum at zero. M-estimators make use of nonlinear $\rho(\cdot)$ to reduce the vulnerability of the cost function to outliers. Instead of solving the nonlinear problem, we can solve an iterated weighted-least squares (Zhang, 1997), *i.e.* we obtain:

$$\sum \psi(r_i(x))j_i = 0$$

after computing condition (B.1,a), where $\psi(r) = \partial_r \rho(r)$ denotes the M-estimator's *influence function*, for it represents the influence of each residue r_i in the estimation. We can further rewrite the above equation as

$$\left(\sum w(r_i(x^{\mathrm{p}}))j_i j_i^T\right)x - \sum w(r_i(x^{\mathrm{p}}))j_i y_i = 0\,. \tag{B.5}$$

where the *weight function* $w(r) = \psi(r)/r$, and x^{p} denotes an *a priori* estimate of x. Remark that (B.5) is equivalent to the weighted least-squares (B.3), such that the minimum is given by (B.4) with $w_i = w(r_i(x^{\mathrm{p}}))$. Table B.1 shows five examples of loss functions, further details and other robust functions can be found in, *e.g.* , (Huber, 1981) or (Zhang, 1997). We can relate the tuning parameters to the standard deviation σ of the measurement's distribution.

- L_1 estimators are not stable as the weight function is unbounded and may yield an indeterminate solution.
- L_2 estimators are not robust since their influence function is unbounded.
- Huber's function behaves like the L_2 estimator close to $r = 0$ and similarly to the L_1 after $|r| > k$. This estimator has been recommended for almost all situations with constant $k = 1.345\sigma$.
- The pseudo-Huber function is an approximation with continuous derivatives of the original Huber's function.
- Tukey's function cancels the outliers $|r| > c$ instead of reducing their influence, and the constant $c = 2.9846\sigma$ is recommended.

C

INTRODUCTION TO THE MULTIPLICATIVE EKF

C.1 QUATERNION REPRESENTATION FOR $\mathbb{SO}(3)$

An element of the unitary quaternion group $\mathbb{Q}$ is given by the complex number $q = s + ir_1 + jr_2 + kr_3$, with $s^2 + {r_1}^2 + {r_2}^2 + {r_3}^2 = 1$. A unitary quaternion can be also expressed in the vector form $q = \begin{bmatrix} s & r^T \end{bmatrix}^T \in \mathbb{R}^4$ with $s \in \mathbb{R}$ and $r \in \mathbb{R}^3$ such that $s^2 + |r|^2 = 1$. The quaternion $q_e = \begin{bmatrix} 1 & 0_{1\times 3} \end{bmatrix}^T$ denotes the identity element of $\mathbb{Q}$, and $q^{-1} = \begin{bmatrix} s & -r^T \end{bmatrix}^T$ is the inverse element of q. The group operation of two quaternions q_a and q_b yielding q_c is defined by:

$$q_c = q_a \cdot q_b = \begin{bmatrix} s_a s_b - {r_a}^T r_b \\ s_a r_b + s_b r_a + \mathrm{S}(r_a) r_b \end{bmatrix} = \mathcal{S}(q_a) q_b = \mathcal{D}(q_b) q_a,$$

with

$$\mathcal{S}(q) = \begin{bmatrix} s & -r^T \\ r & sI_3 + \mathrm{S}(r) \end{bmatrix}, \qquad \mathcal{D}(q) = \begin{bmatrix} s & -r^T \\ r & sI_3 - \mathrm{S}(r) \end{bmatrix}. \tag{C.1}$$

A quaternion q is uniquely related to a rotation matrix $R \in \mathbb{SO}(3)$ as

$$R = I_3 + 2s\mathrm{S}(r) + 2\mathrm{S}(r)^2,$$

and we can further write the dynamics for quaternion representing rigid body motion defining the quaternion ${}^{\mathcal{B}}\boldsymbol{\omega} = \begin{bmatrix} 0 & \omega^T \end{bmatrix}^T$ such that

$${}^{\mathcal{R}}\dot{q}_{\mathcal{B}} = \frac{1}{2} {}^{\mathcal{R}}q_{\mathcal{B}} \cdot {}^{\mathcal{B}}\boldsymbol{\omega} = \frac{1}{2} \mathcal{S}({}^{\mathcal{R}}q_{\mathcal{B}}) {}^{\mathcal{B}}\boldsymbol{\omega}, = \frac{1}{2} \mathcal{D}({}^{\mathcal{B}}\boldsymbol{\omega}) {}^{\mathcal{R}}q_{\mathcal{B}}.$$

For simplicity of notation, we can define

$$T_{\mathcal{S}}(q) = \begin{bmatrix} -r^T \\ sI_3 + \mathrm{S}(r) \end{bmatrix}, \qquad T_{\mathcal{D}}(\omega) = \begin{bmatrix} 0 & -\omega^T \\ \omega & -\mathrm{S}(\omega) \end{bmatrix}, \tag{C.2}$$

so that the dynamics also writes

$${}^{\mathcal{R}}\dot{q}_{\mathcal{B}} = \frac{1}{2} T_S({}^{\mathcal{R}}q_{\mathcal{B}}) {}^{\mathcal{B}}\omega = \frac{1}{2} T_D({}^{\mathcal{B}}\omega) {}^{\mathcal{R}}q_{\mathcal{B}}. \tag{C.3}$$

C.2 ORIENTATION AND GYRO BIAS ESTIMATION VIA EKF

In order to estimate orientation and gyro bias, we can define $x = ({}^{\mathcal{R}}q_{\mathcal{B}}, b_\omega)$, $u = {}^{\mathcal{B}}\omega_y$, $f(x,u) = \left(\frac{1}{2} T_{\mathcal{D}}({}^{\mathcal{B}}\omega_y - b_\omega) {}^{\mathcal{R}}q_{\mathcal{B}}, 0\right)$ and $y = h(x) = {}^{\mathcal{R}}q_{\mathcal{B}}$. The EKF is the straightforward solution for this system, *c.f.* Section 3.1.5, *i.e.* the estimates are given by

$$\widehat{x} = f(\widehat{x}, u) - K(t)(y - h(\widehat{x}))$$

with $K(t) = P(t)C^T(t)W^{-1}(t)$ and P given by the Riccati differential equation

$$\dot{P} = A(\widehat{x}, u)P + PA^T(\widehat{x}, u) - PC^T(\widehat{x})W^{-1}(t)C(\widehat{x})P + Q(t), \tag{C.4}$$

matrices $Q \geq 0$, $W > 0$ with proper dimension and A, C computed by

$$A(x,u) = \begin{bmatrix} \frac{1}{2}T_{\mathcal{D}}({}^{\mathcal{B}}\omega_y - b_\omega) & -\frac{1}{2}T_{\mathcal{S}}({}^{\mathcal{R}}q_{\mathcal{B}}) \\ 0_{3\times 4} & 0_{3\times 3} \end{bmatrix}, \qquad C(x) = \begin{bmatrix} I_4 & 0_{4\times 3} \end{bmatrix}.$$

There are two problems with the above formulation. First, ${}^{\mathcal{R}}\widehat{q}_{\mathcal{B}}$ is not represented by a unitary quaternion since the state structure assumes ${}^{\mathcal{R}}q_{\mathcal{B}} \in \mathbb{R}^4$. Therefore, it is usual to incorporate the unitary norm constraint to the measurement. That measurement augmentation can lead, in turn, to the second problem, since, if proper care is not taken, the matrix $P(t)$ (originally positive definite) can become singular or even develop a negative eigenvalue (Lefferts et al., 1982, Sec. 8).

C.3 ORIENTATION AND GYRO BIAS ESTIMATION VIA MEKF

The multiplicative EKF can be implemented in three different versions: using a reduced representation of the covariance matrix, using a truncated covariance representation or using the representation of the body-fixed covariance (Lefferts et al., 1982). We use in this thesis the reduced representation of the covariance matrix (Lefferts et al., 1982, Sec. 9). In this method, the reduced covariance matrix P^- is computed using (C.4) with

$$A(x,u) = M(x)^T \begin{bmatrix} \frac{1}{2}T_{\mathcal{D}}({}^{\mathcal{B}}\omega_y - b_\omega) & -\frac{1}{2}T_{\mathcal{S}}({}^{\mathcal{R}}q_{\mathcal{B}}) \\ 0_{3\times 4} & 0_{3\times 3} \end{bmatrix} M(x), \qquad M(x) = \begin{bmatrix} T_{\mathcal{S}}({}^{\mathcal{R}}q_{\mathcal{B}}) & 0_{4\times 3} \\ 0_{3\times 3} & I_3 \end{bmatrix},$$

and the complete covariance P retrieved computing $P = M(\widehat{x})P^-M(\widehat{x})^T$. Moreover, the estimation error is employed using the quaternion representation, that is the measurement $y = q_e$ and the measurement function $h(\widehat{x}) = {}^{\mathcal{B}}q_{\mathcal{R}}{}^{-1}.{}^{\mathcal{R}}\widehat{q}_{\mathcal{B}}$, so that $C(\widehat{x}) = \begin{bmatrix} \mathcal{S}({}^{\mathcal{B}}q_{\mathcal{R}}{}^{-1}) & 0 \end{bmatrix}$.

ABSTRACT

Systems with multiple sensors can provide information unavailable from a single source, and complementary sensory characteristics can improve accuracy and robustness to many vulnerabilities as well. Explicit pose measurements are often performed either with high frequency or precision, however visuo-inertial sensors present both features. Vision algorithms accurately measure pose at low frequencies, but limit the drift due to integration of inertial data. Inertial measurement units yield incremental displacements at high frequencies that initialize vision algorithms and compensate for momentary loss of sight.

This thesis analyzes two aspects of that problem. First, we survey direct visual tracking methods for pose estimation, and propose a new technique based on the normalized cross-correlation, region and pixel-wise weighting together with a Newton-like optimization. This method can accurately estimate pose under severe illumination changes. Secondly, we investigate the data fusion problem from a control point of view. Main results consist in novel observers for concurrent estimation of pose, IMU bias and self-calibration. We analyze the rotational dynamics using tools from nonlinear control, and provide stable observers on the group of rotation matrices. Additionally, we analyze the translational dynamics using tools from linear time-varying systems, and propose sufficient conditions for uniform observability. The observability analyses allow us to prove uniform stability of the observers proposed. The proposed visual method and nonlinear observers are tested and compared to classical methods using several simulations and experiments with real visuo-inertial data.

RÉSUMÉ

Les systèmes multi-capteurs exploitent les complémentarités des différentes sources sensorielles. Par example, le capteur visuo-inertiel permet d'estimer la pose à haute fréquence et avec une grande précision. Les méthodes de vision mesurent la pose à basse fréquence mais limitent la dérive causée par l'intégration des données inertielles. Les centrales inertielles mesurent des incréments du déplacement à haute fréquence, ce que permet d'initialiser la vision et de compenser la perte momentanée de celle-ci.

Cette thèse analyse deux aspects du problème. Premièrement, nous étudions les méthodes visuelles directes pour l'estimation de pose, et proposons une nouvelle technique basée sur la corrélation entre des images et la pondération des régions et des pixels, avec une optimisation inspirée de la méthode de Newton. Notre technique estime la pose même en présence des changements d'illumination extrêmes. Deuxièmement, nous étudions la fusion des données a partir de la théorie de la commande. Nos résultats principaux concernent le développement d'observateurs pour l'estimation de pose, biais IMU et l'autocalibrage. Nous analysons la dynamique de rotation d'un point de vue nonlinéaire, et fournissons des observateurs stables dans le groupe des matrices de rotation. Par ailleurs, nous analysons la dynamique de translation en tant que système linéaire variant dans le temps, et proposons des conditions d'observabilité uniforme. Les analyses d'observabilité nous permettent de démontrer la stabilité uniforme des observateurs proposés. La méthode visuelle et les observateurs sont testés et comparés aux méthodes classiques avec des simulations et de vraies données visuo-inertielles.