A Nonlinear Observer Approach for Concurrent Estimation of Pose, IMU Bias and Camera-to-IMU Rotation

Glauco Garcia Scandaroli1, Pascal Morin1, Geraldo Silveira2

1INRIA Sophia Antipolis-Méditerranée
2004 route des Lucioles,
Sophia Antipolis 06902, France

2CTI Renato-Archer – DRVC Division
Rod. Dom Pedro I, km 143,6, Amarais,
CEP 13069-901, Campinas/SP, Brazil

Glauco.Scandaroli@inria.fr, Pascal.Morin@inria.fr, Geraldo.Silveira@cti.gov.br

Abstract—This paper concerns the problem of pose estimation for an inertial-visual sensor. It is well known that IMU bias, and calibration errors between camera and IMU frames can impair the achievement of high-quality estimates through the fusion of visual and inertial data. The main contribution of this work is the design of new observers to estimate pose, IMU bias and camera-to-IMU rotation. The observers design relies on an extension of the so-called passive complementary filter on SO(3). Stability of the observers is established using Lyapunov functions under adequate observability conditions. Experimental results are presented to assess this approach.

Index Terms—Inertial Estimation, Inertial Vision, Nonlinear Observers, Lyapunov Function.

I. INTRODUCTION

Fusing information obtained from different sensors is ubiquitous in robotics. For some systems and applications, the fusion is necessary in order to obtain an information that cannot be extracted from a single sensor. In other cases, data fusion aims at improving the information quality by exploiting the complementary characteristics of different sensors. The present work is dedicated to data fusion for an inertial-visual system. The objective is to develop observers that can provide high-quality pose estimates via the online estimation of various inertial measurement unit (IMU) biases and self-calibration of camera-to-IMU (C-to-IMU) frames rotation. Inertial-visual fusion has been an active research topic for many years (see, e.g., [1], [2]). 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 an IMU. Pose measurements allow to limit the drift associated with direct integration of the IMU data. On the other side, IMU measurements provide incremental displacements on short time-intervals that can be used to initialize vision algorithms, or to compensate for a momentary loss of vision. Unfortunately, the fusion process can be seriously impaired by measurement errors or 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. Since bias may vary due to several factors (e.g. temperature, battery level), they should be permanently estimated. Another source of difficulties concerns various parameters related to the use of different coordinate frames, e.g. the camera and the IMU frames. Usually these parameters are constant and can be estimated in a preliminary calibration step. A possible solution is to proceed using accelerometers as a measurement of gravity [3]. Nevertheless, this method should assume that IMU bias has been already identified. Some authors have recently proposed the concurrent estimation of pose and IMU bias together with self-calibration [2], [4]. An underlying difficulty is that persistent motion conditions must be satisfied for the system to be completely observable. This property, together with the non-linearities of the system motion equations, make this estimation problem challenging.

This study extends a previous work by the authors on nonlinear filter design for pose and IMU bias estimation. The present contribution is twofold. Firstly, we complement the nonlinear observer proposed in [5] with an estimation of C-to-IMU rotation. This procedure is important because the fusion of visual and inertial data much relies on an accurate estimation of this rotation matrix. Secondly, we validate the proposed approach through experiments with a hand-held inertial-visual sensor.

The present approach is related to recent works on both nonlinear observer design and inertial-visual fusion. Concerning nonlinear observers, the main contribution of this paper is an extension of the passive complementary filter on SO(3) [6] to the case where rotation and angular velocity measurements are obtained in different frames and the rotation between these frames is unknown. This paper is also strongly related to recent works on inertial-visual fusion, e.g. [2], [4], where concurrent pose estimation and self-calibration is also addressed. Comparing to these works, the present paper is less general as only C-to-IMU rotation calibration is addressed while calibration of other parameters is also considered in [2], [4]. However, the present estimation result is stronger since exponential stability of the observer is established under appropriate observability conditions. Extended Kalman filter and unscented Kalman filter approaches used in [4] and [2] respectively do not provide such stability guarantees, especially under the fast motions necessary to ensure good observability properties for the parameters calibration.

Pose measurements are obtained by directly exploiting pixel intensities as proposed in [7], instead of using image
features. That method has been chosen because of its efficiency, and the possibility of using all image information, even from areas where no image features exist. Fusion with inertial data is instrumental in providing a good initialization of the vision algorithm and avoid local minima. We remark that the proposed observer does not depend on any specific vision-based localization method.

II. THEORETICAL BACKGROUND

A. Mathematical Notation and Identities

The unitary vector with 1 in the \( i \)-th position is denoted as \( e_i \in \mathbb{R}^3 \). The special orthogonal group is denoted as \( \mathbb{S} \mathbb{O}(3) \). Its associated Lie algebra is the set of anti-symmetric matrices denoted as \( \mathfrak{s} \mathfrak{o}(3) \). The cross-product can be represented by the product \( S(u)=v \times u \), for all \( u, v \in \mathbb{R}^3 \), where \( S(\cdot) \in \mathfrak{s} \mathfrak{o}(3) \). The inverse of the \( S(\cdot) \) operator is denoted \( \text{vex}^{-1}(\cdot) \), i.e., \( \text{vex}(S(u))=u \). Let \( \mathfrak{a} \in \mathfrak{s} \mathfrak{o}(3) \). With \( \mathfrak{a} \in \mathbb{R}^{3 \times 3} \), the symmetric and anti-symmetric operators are defined as \( P_\mathfrak{a}(\mathfrak{a})=\frac{\mathfrak{a}+\mathfrak{a}^T}{2} \), \( P_\mathfrak{a}(\mathfrak{a})=\frac{\mathfrak{a}-\mathfrak{a}^T}{2} \).

\[ S(Ru) = RS(u)R^T, \quad \text{vex}(P_a(Ra)) = \text{vex}(P_a(Ra)). \]

Consider any parametrization \( \Theta \) such that \( R \approx I_3 + S(\Theta) \) at first order around the \( 3 \times 3 \) identity matrix \( I_3 \). From the general rotation dynamics \( \frac{\text{d}}{\text{d}t} R = RS(u) \), and around \( R=I_3 \):

\[ \dot{\Theta} \approx u, \quad \text{vex}(P_a(R)) \approx \Theta. \]

The special Euclidean group is denoted as \( \mathbb{S} \mathbb{E}(3) = \{ [ \begin{array}{cc} R & p \\ 0 & 1 \end{array} ] \, | \, R \in \mathbb{S} \mathbb{O}(3), p \in \mathbb{R}^3 \} \), and its associated Lie algebra is the set of twist matrices \( \mathfrak{s} \mathfrak{e}(3) = \{ [ \begin{array}{c} S(u) \\ v \end{array} ] \, | \, u, v \in \mathbb{R}^3 \} \). One can define \( \mathbb{E}(x) \in \mathbb{S} \mathbb{E}(3) \), \( x \in \mathbb{R}^3 \), and one writes \( T(x) = \exp (\mathbb{E}(x)) \), \( \in \mathbb{S} \mathbb{E}(3) \) from the exponential map properties.

B. System description

This work deals with the inertial-visual system depicted in Fig. 1. Denote by \( W \) an inertial world frame as denoted as appropriate. Similarly, let \( B \) and \( C \) denote two different body and camera frames attached to the same rigid body. Measurements relative to angular velocity and linear acceleration are expressed in the former frame, and pose measurements are relative to the later. \( R_i^j \in \mathbb{S} \mathbb{O}(3) \) is the rotation matrix from frame \( i \) to frame \( j \) and \( p_i^j \in \mathbb{R}^3 \) represents the translational displacement of a frame \( i \) w.r.t. frame \( j \). When \( j = W \), the superscripts are omitted. The dynamics for the system comprising attitude and translational displacement of \( B \) with respect to \( W \) is:

\[ \dot{R}_B = R_B S(\omega_B), \quad \dot{p}_B = v, \quad \dot{v} = R_B a_B, \]

where \( v \in \mathbb{R}^3 \) is the body’s translational velocity in world coordinates, \( \omega_B \in \mathbb{R}^3 \) and \( a_B \in \mathbb{R}^3 \) are the rotational velocity and translational acceleration in \( B \) coordinates.

**Assumption 1** There exist five positive constants \( \tau_0, \tau_\omega, \tau_v, \tau_a, \tau \) such that \( \forall t \in [0, \infty) : |\omega_B(t)| \leq \tau_\omega, |\omega_B(t)| \leq \tau_\omega, |\dot{\omega}_B(t)| \leq \tau_\omega, |\ddot{\omega}_B(t)| \leq \tau_\omega, \text{ and } |a_B(t)| \leq \tau_a. \)

This technical assumption is made for the sake of stability analysis. Clearly, it is always satisfied in practice for physical systems. To measure \( \omega_B \) and \( a_B \), an IMU consisting of rate gyroscopes and accelerometers is employed. It is considered that \( B \) and the IMU reference frame coincide, therefore gyroscopes measure the angular velocity \( \omega_B \) and accelerometers the proper acceleration, i.e. body’s acceleration minus gravitational field. Due to MEMS manufacturing, IMU measurements include offset biases that result in:

\[ \omega = \omega_B + \omega_b, \quad a = R_B^T (\dot{v} - g_0) + a_b, \quad \omega_b = 0, \dot{\omega}_b = 0, \]

where \( \omega, a \) are rate gyroscope and accelerometer measurements; \( \omega_b, a_b \) denote the gyroscope and accelerometer bias respectively; and \( g_0 \) is the gravitational field in \( W \) coordinates. It is important to notice that other sensor characteristics are neglected such as limited bandwidth, bias variation w.r.t. temperature, scale errors and additive measurement noise. The experimental results show that highly accurate estimates are obtained by the proposed approach despite these effects have not been considered in the modeling.

As detailed in Section III, a visual algorithm is used to obtain rotation and translation measurements of the camera w.r.t. the world frame. These measurements are \( (\dot{R}_C, \dot{p}_C) \) instead of \( (R_B, \dot{p}_B) \). In order to fuse visual and IMU measurements, the parameters \( (R_C^B, \dot{p}_C^B) \) of the transformation from camera frame to body frame are needed. Poor estimation of these parameters can strongly damage the fusion process. Considering that IMU sensors and the visual system present different sampling frequencies, for several periods of time only estimates obtained from IMU integration are available. If the estimates \( R_B \) are biased, e.g. due to a bad estimation of \( R_C^B \), the translational displacement integrates this bias twice. This misleading double integration leads to a biased position estimation. This can be problematic for the vision algorithm, especially for minimization-based algorithms that require a “good enough” initialization. Obtaining high quality estimates of \( R_C^B \) is one of the objectives of this paper.

C. Planar scenes

Denote \( C_i \) a reference frame attached to the optical center of a camera at an instant \( i \); let \( p_i^C \in \mathbb{R}^3 \) denote the coordinates of a point in space expressed in \( C_i \) frame coordinates. Pinhole cameras [8] are based on the perspective projection...
model, where a point $p^C_i$ is projected onto the image plane $I_i$ with the following transformation
\[
p^C_i = \frac{1}{e^3p^C_i} K p^C_i \in \mathbb{P}^2
\]
where $K \in \mathbb{R}^{3\times 3}$ represents the intrinsic parameters of the camera, i.e., focal length, skew factor and principal point.

Now, consider two images $\mathcal{I}_i, \mathcal{I}_{i+1}$ obtained at instants $i, i+1$, with $C_i, C_{i+1}$ denoting the camera frame at these respective instants. Suppose that a point $p_j$ lies on a plane $\Pi$ of the 3D space, let $p^C_j$ denote the coordinates of $p_j$ in $C_i$ and $n^C_j$ the scaled normal vector to $\Pi$ such that $n^C_j, p^C_j \neq 1$. The coordinate transformation between the two camera frames for any $p_j \in \Pi$ yields:
\[
p^C_{j+1} = (R^C_{i+1} + p^{C_{i+1}} n^C_i, T) p^C_j \triangleq H^C_{i+1} p^C_j,
\]
where $H \in \mathbb{GL}(3)$ is called Euclidean homography transformation. Using the point projection onto an image (5), and the 3D coordinate planar transformation yields
\[
\lambda p^C_j = KH^C_{i+1} K^{-1} p^C_j \triangleq G_{i+1} p^C_j,
\]
where $p^C_j, p^C_{j+1}$ are the projections of the point $p_j$ on the images $\mathcal{I}_i$ and $\mathcal{I}_{i+1}$, respectively, and $G \in \mathbb{GL}(3)$ denotes the projective homography transformation. Note that the relation from the points is made up to a scale factor $\lambda$.

This scalar can be suppressed by defining a group action,
\[
w(G, p) = \left( \begin{array}{c} e^T G p \\ e^T G p + 1 \end{array} \right)^T : \mathbb{GL}(3) \times \mathbb{P}^2 \rightarrow \mathbb{P}^2.
\]

The following properties hold for the group action:
\[
w(I_3, p) = p, \quad w(G_a, w(G_b, p)) = w(G_a G_b, p), \quad w(G, p)^{-1} = w(G^{-1}, p).
\]

An image $\mathcal{I}_i$ consists in a matrix of intensities, and the function $\mathcal{J}_i(p) : \mathbb{P}^2 \rightarrow \mathbb{R}$ maps the point $p$ to its respective intensity in $\mathcal{I}_i$. From (8) it is not difficult to verify that for two images $\mathcal{I}_1, \mathcal{I}_2$:
\[
\mathcal{J}_1(p) = \mathcal{J}_2(w(G_2^2, p)).
\]

### III. VISUAL METHOD

The employed visual method is based on the direct use of the pixel intensity and the image flow. We consider in this work a known planar target so that we can focus only on the localization aspects. For localization in unknown piecewise planar scenes, the reader is referred to [7]. The visual method aims at obtaining the pose $(R_C, p_C)$. This problem can be stated as finding $\hat{x}^C_C = \arg \min_{{\hat{x}^C_C}} |r(x)|^2$, where $r(x)$ is the vector of residuals $r_j(G(x))$ defined from Eq. (9), i.e.,
\[
r_j(G) = \mathcal{J}_j(w(G_0, p_j)) - \mathcal{J}_0(p_j), \quad \forall p_j \in \mathcal{J}_0,
\]
where $\mathcal{J}_0$ is the image obtained at a reference frame $C_0$. Recall from Eqs. (6) and (7) that $G_0^C$ is related to the rotation and translational displacement of two different camera frames. Let $T_C$ denote the current camera pose and $T_{C_0}$ the pose of the reference frame. Then, the parametrization $x^C_C \in \mathbb{R}^6 : T_C = T(x^C_C) = \exp(\Xi(x^C_C))$ is used, and $H(x) : \mathbb{R}^6 \rightarrow \mathbb{GL}(3)$, $G(x) : \mathbb{R}^6 \rightarrow \mathbb{GL}(3)$ can be defined from (6), (7).

It is not a simple task to solve this problem analytically, and the solution is given by nonlinear iterative minimization. The solution is based on the Taylor expansion and a second-order approximation [9]. This second-order method provides that Hessians are never computed explicitly yielding the following approximation of $r(x)$ around $x = 0$
\[
r(x) = r(0) + \frac{1}{2} \left( J(0) + J(x) \right) x + O^3.
\]

where $J(x) = \frac{\partial r}{\partial a} p_{(x)}$ denotes the Jacobian of $r$, and $O^3$ elements of order higher than two. Note that this approximation depends on the unknown $x$, which is the global minimum of the minimization problem. The computation of $J(x) x$ is possible owing to the the known reference image $\mathcal{I}_0$ and the Lie group parametrization. Ultimately, the solution for the minimization is given iteratively by the linear least squares
\[
\frac{1}{2} (J(x_{0}) + J(\hat{x}^C_{C\{n-1\}})) \sigma_{C|n} = -r(\hat{x}^C_{C\{n-1\}}),
\]
where $\sigma_{C|n}$ is the computed increment that updates the pose via $T(\hat{x}^C_{C\{n\}}) = \exp(\Xi(\Delta x^C_C)) T(\hat{x}^C_{C\{n-1\}})$, until the resulting $|\Delta x^C_C|$ is “small enough”.

### IV. ESTIMATOR DESIGN

From Eqs. (3) and (4), the dynamics writes:
\[
\dot{R}_B = R_B S(\omega - \omega_b), \quad \dot{\omega}_b = 0, \quad \dot{R}_C^B = 0,
\]
\[
\dot{p}_B = v, \quad \dot{v} = R_B (a - \bar{a}_b) + g_0, \quad \dot{\bar{a}}_b = 0.
\]

Assuming that $R_B^B$ can be neglected, the visual system and IMU together provide the following measurements
\[
y = (p_C, a, R_C, \omega) = (p_B, a, R_B R_C^B, \omega).
\]

The problem addressed in this section is the state estimation of System (13)-(14). The observers
\[
\dot{\hat{R}}_B = \hat{R}_B S(\omega - \omega_b + \alpha_R), \quad \dot{\hat{\omega}}_b = \alpha_\omega, \quad \dot{\hat{R}}_B^B = \hat{R}_B^B S(\alpha_C),
\]
\[
\dot{\hat{p}}_B = \dot{v} + \alpha_p, \quad \dot{\hat{v}} = \hat{R}_B(a - \bar{a}_b) + g_0 + \alpha_v, \quad \dot{\hat{\bar{a}}}_b = \alpha_a
\]
are defined for the attitude and position estimation, respectively, where $\alpha_R, \alpha_\omega, \alpha_C, \alpha_p, \alpha_v, \alpha_a$ are innovation terms defined further on. Denote the estimation errors as
\[
\tilde{R}_B = \hat{R}_B R_B^T, \quad \tilde{\omega}_b = \omega_b - \hat{\omega}_b, \quad \tilde{R}_C^B = \hat{R}_B^B S(\alpha_C),
\]
\[
\tilde{p}_B = \tilde{v} - \alpha_p, \quad \tilde{v} = R_B(a - \bar{a}_b) + g_0 - \alpha_v, \quad \tilde{\bar{a}}_b = a - \hat{\bar{a}}_b.
\]

The estimation error dynamics is thus given by
\[
\begin{cases}
\dot{\tilde{R}}_B = \tilde{R}_B S(-\tilde{R}_B \tilde{\omega}_b - \tilde{R}_B \alpha_R), \quad \dot{\tilde{\omega}}_b = -\alpha_\omega, \\
\dot{\tilde{R}}_C^B = \tilde{R}_C^B S(-\tilde{R}_C^B \alpha_C),
\end{cases}
\]
\[
\begin{cases}
\dot{\tilde{p}}_B = \tilde{v} - \alpha_p, \quad \dot{\tilde{v}} = R_B(\bar{a}_b - (I - \tilde{R}_B) \hat{R}_B(a - \bar{a}_b) - \alpha_v), \\
\dot{\tilde{\bar{a}}}_b = -\alpha_a.
\end{cases}
\]

The objective is to define innovation terms so that $(\tilde{R}_B, \tilde{\omega}_b, \tilde{R}_C^B, \tilde{p}_B, v, \bar{a}_b) = (I_3, 0, I_3, 0, 0, 0)$ is an asymptotically stable equilibrium for the error dynamics. Since the rotation dynamics is independent of the translation dynamics, the estimation of the former is addressed first.
A. Rate gyro bias and C-to-IMU rotation estimation

Let \( \hat{R}_C = \hat{R}_B \hat{R}_B^T C \) denote the estimate of \( R_C \), as deduced from the estimates \( \hat{R}_B \) and \( \hat{R}_B^T \). With this notation, the main theoretical result of this paper is given next.

**Theorem 1** Let

\[
\begin{align*}
\alpha_R &= k_1 \hat{R}_B^T \text{vec}(P_a(\hat{R}_C)) - \hat{R}_B^T \omega_C, \\
\omega_x &= -k_2 \hat{R}_B^T \text{vec}(P_a(\hat{R}_C)), \\
\alpha_C &= k_3 (\hat{R}_B^T)^T \hat{R}_B^T P_a(\hat{R}_C) \hat{R}_B (\omega - \omega_b),
\end{align*}
\]

with \( k_1, k_2, k_3 > 0 \) and \( \hat{R}_C = \hat{R}_C \hat{R}_C^T \). Assume that the following condition is satisfied:

\[
\exists \delta > 0 : |\omega_b(t) \times \omega_b(t)| > \delta, \ \forall \ t. \tag{22}
\]

Then, \( (\hat{R}_B, \omega_b, \bar{R}_C) = (I_3, 0, I_3) \) is a locally exponentially stable equilibrium point of the error dynamics (19).

**Proof.** See Appendix.

This result calls for several remarks.

Relation (22) is a “persistent excitation” condition related to the system’s observability properties. Indeed, System (13) with \( R_C \) and \( \omega \) as measurements is not observable for every input \( \omega_b \). Trivially, it can be verified that the state is not observable when \( \omega_b = 0 \). One can show that (22) is a sufficient condition for the system’s uniform observability. The system is still uniformly observable under slightly weaker conditions, but it is not observable when \( \omega_b(t) \times \omega_b(t) = 0 \).

The proposed observer can be viewed as an extension of the passive complementary filter on \( \text{SO}(3) \) proposed in [6]. More precisely, setting \( k_3 = 0 \) in (21) and assuming that \( \hat{R}_C^T = \hat{R}_B^T = I_3 \), the observer reduces to an attitude and rate gyro bias estimator. In this special case, it has been shown in [6] that this estimator is semi-globally exponentially stable, independently of \( \omega_b \). Despite the fact that semi-global exponential stability seems more difficult to prove for the present observer, simulation results suggest that its domain of attraction is also quite large.

In practice, Condition (22) will not always be satisfied and some care is necessary in 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 of \( \hat{R}_C^T \), and then setting \( k_3 = 0 \) in order to use, as explained above, the observer as an attitude and rate gyro bias estimator. A second possibility consists in using the stability condition (22) so as to tune the gain \( k_3 \) 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 \( |\omega_b(t) \times \omega_b(t)| \) is significantly larger than zero, so as to avoid possible drift of \( \hat{R}_C^T \) in case of weak motion excitation and measurement noise. Although this quantity is not directly measured, it can be estimated from the measurement of \( \omega \). This is detailed in Section V-A.

B. Accelerometer bias estimation

The authors proposed in a previous work [5] a position and accelerometer bias observer based on the dynamics (20):

**Lemma 1 (Position and accelerometer bias observer) [5]** Assume that \( \hat{R}_B = R_B \) and \( \omega_b = \omega_b \). Let

\[
\alpha_p = k_4 \bar{p}_B, \ \alpha_v = k_5 \bar{p}_B, \ \alpha_a = -k_6 (I_3 + \frac{1}{3} S(\omega_b)) \hat{R}_B^T \bar{p}_B, \tag{23}
\]

with \( k_4, k_5, k_6 > 0 \), such that \( k_6 < k_4 k_5 \). Then, \( (\bar{p}_B, \bar{v}, \bar{a}_b) = (I_3, 0, I_3) \) is a globally exponentially stable equilibrium point of the estimation error dynamics (20).

This Lemma provides a global exponentially stable observer for position and accelerometer bias. However, remark that attitude \( R_B \) and \( \omega_B \) must be directly available.

C. Full state estimation

Full state estimation is easily obtained by coupling the two observers presented above.

**Corollary 1** Let

\[
\begin{align*}
\hat{R}_B &= \hat{R}_B S(\omega - \omega_b + \alpha_R), \\
\hat{\omega}_b &= \alpha_x, \\
\hat{R}_B^C &= \hat{R}_B^C S(\alpha_C), \\
\hat{p}_B &= \bar{v} + \alpha_p, \\
\hat{v} &= \hat{R}_B (a - \bar{a}_b) + g_0 + \alpha_v, \\
\hat{\bar{a}}_b &= \alpha_a,
\end{align*}
\]

where,

\[
\alpha_p = k_4 \bar{p}_B, \ \alpha_v = k_5 \bar{p}_B, \ \alpha_a = -k_6 (I_3 + \frac{1}{3} S(\omega - \omega_b)) \hat{R}_B^T \bar{p}_B \tag{26}
\]

and the innovation terms \( \alpha_R, \alpha_x, \alpha_C \) are given by (21), with \( k_1, \cdots, k_6 > 0 \), and \( k_6 < k_4 k_5 \). Assume that condition (22) is satisfied. Then, \( (\hat{R}_B, \hat{\omega}_B, \hat{R}_B^C, \bar{p}_B, \bar{v}, \bar{a}_b) = (I_3, 0, I_3, 0, 0, 0) \) is a locally exponentially stable equilibrium point of the estimation error dynamics (19)–(20).

The proof is similar to the proof of [5, Corollary 2].

When \( \alpha_p \) is not negligible and an estimate \( \hat{p}_C \) of this term is available, the term \( \hat{p}_C \) in (26) should be replaced either by \( p_C - R_C(\hat{R}_B^T)^T \hat{p}_B - \bar{p}_B \) or by \( p_C - \hat{R}_B \hat{p}_B - \bar{p}_B \).

V. EXPERIMENTAL RESULTS

A. System setup

The proposed method is evaluated for the estimation of pose, IMU bias and C-to-IMU rotation for an inertial-visual sensor. We make use of the sensor depicted in Fig. 2(a) consisting of a xSens MTI–G IMU working at a frequency of 200 [Hz], and an AVT Stingray 125B camera that provides 20 images of \( 800 \times 600 \) [pixel] resolution per second. The camera uses Theia SY125M wide-angle lenses, and a previous calibration of intrinsic parameters resulted in focal length (448.85, 450.26) [pixel] and principal point (394.30, 292.82) [pixel]. A version of Fig. 2(b) is printed out on a 376 × 282 [mm] sheet of paper to serve as a reference for the visual system. The target is placed over a surface parallel to the ground, and this configuration allows...
a direct calculation of the scaled normal vector $n_C$, and 
the reference frame $C_0$ considering the reference image with 
320 x 240 [pixel] and the camera’s intrinsic parameters.

Concerning the observability condition (22) obtained for 
the observer, we use a variable gain $k_3 = f(\delta(t))$, with 
$\delta(t) = |\omega_B(t)\times\omega_B(t)|$. This choice allows the application of 
the C-to-IMU innovation only if the observability condition 
is satisfied. However, the term $|\omega_B(t)\times\omega_B(t)|$ cannot be 
directly measured. Hence, a secondary filter is designed in 
order to identify this condition from gyro measurements. We 
consider an approximate model of constant angular jerk, i.e. 
$\omega_B=0$, and a linear Kalman filter is used defining $\hat{\omega}_B$, 
$\hat{\omega}_B$, and $\hat{\omega}_B$ as states, and $\omega$ as measurement. The success of 
this filter relies on the fact that the bias $\omega_b$ is constant for 
short periods of time, therefore it should not influence the 
evaluation of $\hat{\omega}_B$ and $\hat{\omega}_B$. Also the goal is not to precisely 
estimate these variables, yet to identify when $|\hat{\omega}_B\times\hat{\omega}_B| > \delta$, 
for $\delta > 0$. We use $k_3(\delta) = k_3(1 + e^{-\frac{20}{3}(\delta-10)})^{-1}$.

B. Results and discussion

To evaluate the motion classification, the sensor is placed on 
a tripod and three different angular motions are performed. 
The two firsts are made around single axes, from 
0.4 to 3 [s], and 5.5 to 8 [s]. A third motion satisfying the 
observability condition is made from 9 to 13 [s]. The result 
obtained using the estimator is depicted in Fig. 3, where 
angular acceleration $\hat{\omega}_B$ [rad/s^2], angular jerk $\hat{\omega}_B$ [rad/s^3], 
$\hat{\omega}_B\times\hat{\omega}_B$, and the resulting $k_3(\delta)$ are displayed from top 
to bottom. It is visually clear that the estimated evolution 
of angular acceleration and jerk are mostly parallel for 
the first two motions. For the third motion, the angular 
acceleration and jerk are not parallel and $|\hat{\omega}_B\times\hat{\omega}_B|$ is large. 
Note that in the end of this motion, around 13 [s], $|\hat{\omega}_B\times\hat{\omega}_B|$ 
slowly decreases, but the gain function decreases faster. This 
behavior is substantial to ignore slow motions that would not 
contribute to the estimation process w.r.t. the noise value of 
gyro measurements.

After being able to detect when the observability condition 
is satisfied, we performed a hand-held experiment for the 
estimation of pose, IMU bias and C-to-IMU rotation. An initial 
guess for the biases is obtained after leaving the IMU over 
the same surface for a few seconds. Measuring the C-to-IMU 
distance, we obtain an approximate $B = (8,0.5,3)$ [cm], and 
the initialization of the frame rotation $R_B^C$ in Euler angles 
is manually performed resulting $\psi = (-90,0,-90)$ [°]. The 
observed of Corollary 1 is employed using $k_1=3.33$, $k_2=1$, 
k_4=7.85, $k_5=15.14$, $k_6=4.11$, and the proposed variable 
gain function $k_3(\delta)$ with $k_3=1$. This gain tuning follows the 
procedure described in [5] for settling times of 1, 9, 0.6, 1.2, 
and 8.5 [s] for $\hat{R}_B$, $\hat{\omega}_b$, $\hat{p}_B$, $\hat{v}$, $\hat{a}_b$, respectively.

Results obtained for this experiment are depicted in 
Figs. 4, 5, and 6. The supplemental multimedia material 
presents the image sequence together with the evolution of 
trajectory. This supplemental material also shows a sample of 
the same sequence without C-to-IMU self-calibration leading 
to the divergence of the visual algorithm. Fig. 4 presents 
the estimated position trajectory during the experiment. The 
resulting trajectory presents regions with richer motion, that 
refer to the self-calibration itself, and transition between 
these regions. One can identify three regions with richer 
motion. Fig. 5 depicts four sample images obtained from 
the camera at 1.5, 9, 20, and 25 [s], where the red lines represent 
the last target measured by the visual system, yellow is the 
initialization provided by the nonlinear observer and cyan the 
current measurement. From these images, one can see the 
 improvement yielded by the data fusion and self-calibration.
Firstly, at 1.5 [s], the previous measurement presents a medium-sized distance w.r.t. the current, and the estimates obtained improve the visual method’s initialization, however some distance can still be noticed mainly due to initial IMU bias and C-to-IMU rotation errors. The other samples are obtained after time has passed. Despite faster motion of the camera, the initialization provided by the estimator is closer to the solution achieved by the visual method. Hence, this corroborates the improvement obtained with the estimation of IMU bias and C-to-IMU rotation. Fig. 6 shows the values obtained from top to bottom for body orientation, and C-to-IMU rotation evolution from the initial state, both in Euler angle [°], gyro bias [rad/s], accelerometer bias [m/s²], gyro measurements [rad/s] and gain k₃(δ). The components 1, 2, and 3 of ˜Φ, ˜ψ, ˜ω_b, ˜RB_C, and ˜ω are presented in blue, green and red.

VI. CONCLUSION AND FUTURE WORK

This article proposes a new method for estimating pose, IMU bias and C-to-IMU rotation. An exponentially stable nonlinear observer is developed together with its proof obtained under observability conditions. The satisfaction of this observability condition can be identified using gyroscope measurements through a method also described. Experimental results using an inertial-visual sensor support the effectiveness of the proposed method. Future work will consist in extending the approach to the estimation of other parameters such as the C-to-IMU translational displacement, and scaled normal vector of the visual target.

REFERENCES


APPENDIX: PROOF OF THEOREM 1

The error for C frame orientation is defined as ˜RC = R_C( ˜RB_B)ᵀ ˜RB_B = ˜RB_B ˜RB_B ˜RB_B ˜RB_B ˜RB_B ˜RB_B ˜RB_B. We remark that if ( ˜RC, ˜ω_b, ˜RB_C) = (I₃, 0, I₃) is a stable equilibrium point, then ( ˜RB_B, ˜ω_b, ˜RB_C) = (I₃, 0, I₃) is also a stable equilibrium point of dynamics (17). Let ˜Θ ∈ ℝ³ : ˜RC ≈ I₃ + S( ˜Θ),
\( \tilde{\Phi} \in \mathbb{R}^3 : \tilde{R}_B \approx I_3 + S(\tilde{\Phi}), \tilde{\psi} \in \mathbb{R}^3 : \tilde{R}_C^B \approx I_3 + S(\tilde{\psi}), \omega_b \approx \omega - \tilde{\omega}_b. \) Using these approximations and the expression \( R_C = \tilde{R}_B R_B R_C^B R_B^T \) yields, around the equilibrium, \( R_C \approx I_3 + S(\tilde{\Theta}) \approx (I_3 + S(\tilde{\Phi}))(I_3 + S(\tilde{R}_B \tilde{\psi})). \) Thus,

\[
S(\tilde{\Theta}) \approx S(\tilde{\Phi}) + S(R_B S(\omega_B) \tilde{\psi} + R_B \tilde{\psi})
\]

Using Eqs. (2), (17), (21) and neglecting the higher order terms, one can write

\[
\begin{align*}
\tilde{\Theta} & \approx \tilde{\Phi} + k_1 \tilde{\Theta} - R_B \tilde{\omega}_b + R_B R_C^B \omega_c + R_B S(\omega_B) \psi - R_B R_C^B \omega_c,
\end{align*}
\]

and the linearized system for \((\tilde{R}_C, \tilde{R}_C^B, \tilde{\omega}_b)\) writes

\[
\begin{align*}
\dot{\tilde{\Theta}} &= -k_1 \tilde{\Theta} - R_B \tilde{\omega}_b + R_B R_C^B \omega_c + R_B S(\omega_B) \psi - R_B R_C^B \omega_c, \\
\dot{\tilde{\omega}} &= k_2 R_B^T \tilde{\Theta}, \\
\frac{1}{\tilde{\psi}} &= k_3 S(\tilde{R}_B^T \tilde{\Theta}) \omega_B.
\end{align*}
\]

Consider the following variable change \( \tilde{\Theta} = R_B^T \tilde{\Theta} \), then in coordinates \((\tilde{\Theta}, \tilde{\omega}_b, \tilde{\psi})\), System (27) is thus given by

\[
\begin{align*}
\dot{\tilde{\Theta}} &= - (k_1 I_3 + S(\omega_B)) \tilde{\Theta} - \tilde{\omega}_b - S(\omega_B) \tilde{\psi},
\dot{\tilde{\omega}} &= k_2 \tilde{\Theta}, \\
\frac{1}{\tilde{\psi}} &= k_3 S(\omega_B) \tilde{\Theta}.
\end{align*}
\]

Define the Lyapunov candidate function

\[
L_d = \frac{1}{2} |\tilde{\Theta}|^2 + \frac{1}{2k_1} |\tilde{\omega}_b|^2 + \frac{1}{2k_3} |\tilde{\psi}|^2.
\]

Then along the solutions of (28)

\[
\dot{L}_d = -k_1 |\tilde{\Theta}|^2 - \tilde{\Theta}^T \tilde{\omega}_b - \tilde{\Theta}^T S(\omega_B) \tilde{\psi} + \tilde{\omega}_b^T \tilde{\Theta} + \tilde{\psi}^T S(\omega_B) \tilde{\Theta},
\]

\[
= -k_1 |\tilde{\Theta}|^2 \leq 0.
\]

From this point, using Barbalat’s Lemma [10, p. 323] and the observability condition (22) it is not very difficult to show that \((\tilde{\Theta}, \tilde{\omega}_b, \tilde{\psi}) = 0\) 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 prove below, via the design of a strict Lyapunov function, that the linearized system is uniformly exponentially stable. Then, local exponential stability of the original nonlinear system follows from [10, p. 161]. We proceed by defining the auxiliary variable \( \gamma \) and its derivative

\[
\gamma = \tilde{\omega}_b + S(\omega_B) \tilde{\psi}, \quad \dot{\gamma} = (k_2 I_3 - k_3 S(\omega_B)) \tilde{\Theta} + S(\omega_B) \tilde{\psi},
\]

and the secondary function

\[
L_s = \varepsilon \tilde{\Theta}^T \gamma - \varepsilon^2 \tilde{\Theta}^T S(\omega_B) \gamma - \varepsilon^3 \tilde{\psi}^T S(\omega_B) S(\omega_B) \tilde{\psi},
\]

where \( \varepsilon \in (0, 1) \). The derivative for each term results

\[
\begin{align*}
\frac{d}{dt} (\tilde{\Theta}^T \gamma) &= -|\gamma|^2 + \tilde{\Theta}^T (W_1 \tilde{\Theta} - W_2 \gamma + S(\omega_B) \tilde{\psi)), \\
\frac{d}{dt} (\tilde{\psi}^T S(\omega_B) \gamma) &= k_3 \tilde{\Theta}^T S(\omega_B) S(\omega_B) \gamma + \tilde{\psi}^T S(\omega_B) W_1 \tilde{\Theta} + \tilde{\psi}^T S(\omega_B) ^2 \tilde{\psi}, \\
\frac{d}{dt} (\tilde{\psi}^T S(\omega_B) S(\omega_B) \tilde{\psi}) &= \tilde{\psi}^T (S(\omega_B)^2 + S(\omega_B) S(\omega_B)) \tilde{\psi} - k_3 \tilde{\psi}^T S(\omega_B) S(\omega_B) S(\omega_B) \tilde{\psi},
\end{align*}
\]

where \( W_1 = k_2 I_3 - k_3 S(\omega_B)^2, W_2 = k_1 I_3 + S(\omega_B) \). From Assumption 1, \(|W_1 \tilde{\Theta}| \leq k_0 \tilde{\Theta}, |W_2 \tilde{\psi}| \leq k'' \tilde{\psi} \), with \( k_0 = k_2 + k_3 \tilde{c}_2, k'' = k_1 + \tilde{c}_3 \).

In order to continue the analysis, note the inequalities

\[
|\delta \tilde{\Theta}^T | \leq \frac{1}{2}(\sqrt{3}|u|^2 + \delta^2 |v|^2), \quad (34a)
\]

\[
k_u |a_1 \xi + u_2 | + k_b |a_2 \xi + u_2 | \leq \frac{k_b k_1}{k_1 + k_b} |a_1 | |a_2 |, \quad (34b)
\]

where \( \delta \in (0, 1), u, v, \epsilon \in \mathbb{R}^3, a_1, a_2 \in \mathbb{R}^3 \) with \(|a_1 | = |a_2| = 1, \) and \( k_u, k_b > 0 \). Eq. (34a) follows from the triangular inequality, and the proof for (34b) can be found in [11]. Denoting by \( O(\varepsilon^k) \) any term locally bounded by \(|\varepsilon|^k \) in the neighborhood of \( \varepsilon = 0 \), one can write from (33) and (34a)

\[
\varepsilon \frac{d}{dt} (\tilde{\Theta}^T \gamma) \leq \frac{1}{2} (1 + O(\varepsilon)) |\tilde{\Theta}|^2 - \varepsilon (1 + O(\varepsilon)) |\tilde{\Theta}|^2 + \frac{1}{2} \varepsilon^2 |S(\omega_B) \tilde{\psi}|^2 \leq 0,
\]

and the proof for (34b) one obtains

\[
\varepsilon \frac{d}{dt} (\tilde{\psi}^T S(\omega_B) \tilde{\psi}) \leq \frac{1}{2} (1 + O(\varepsilon)) |\tilde{\psi}|^2 - \varepsilon (1 + O(\varepsilon)) |\tilde{\psi}|^2 + \frac{1}{2} \varepsilon^2 |S(\omega_B) \tilde{\psi}|^2 \leq 0.
\]

Using (34a) one obtains

\[
\begin{align*}
\varepsilon \frac{d}{dt} (\tilde{\psi}^T S(\omega_B) \tilde{\psi}) &\leq \frac{1}{2} (1 + O(\varepsilon)) |\tilde{\psi}|^2 - \varepsilon (1 + O(\varepsilon)) |\tilde{\psi}|^2 + \frac{1}{2} \varepsilon^2 |S(\omega_B) \tilde{\psi}|^2 \leq 0,
\end{align*}
\]

with \( \delta'' > 0 \). Finally, defining \( \mathcal{L} = L_d + k_1 L_s \), one verifies from (29), (30), (32), and (38) that there exists \( \varepsilon_0 > 0 \) such that for any \( \varepsilon \in (0, \varepsilon_0) \),

\[
\mathcal{L} \leq \varepsilon (\eta \mathcal{L} - \varepsilon \mathcal{L}), \quad \eta \mathcal{L} > 0,
\]

since \( \mathcal{L} \) is a definite positive function for \( \varepsilon > 0 \) small enough. This concludes the proof of uniform exponential stability of the origin \((\tilde{\Theta}, \tilde{\omega}_b, \tilde{\psi}) = 0\) for System (28), and consequently for the linearized System (27). As the origin of (27) is an uniformly exponentially stable, then \((\tilde{R}_C, \tilde{\omega}_b, \tilde{R}_C^B) = (I_3, 0, I_3)\) is a locally exponentially stable equilibrium point of the nonlinear system and so is \((\tilde{R}_B, \tilde{\omega}_b, \tilde{R}_B^B) = (I_3, 0, I_3)\) for the dynamics (19).