Observability, Identifiability and Sensitivity of Vision-Aided Inertial Navigation

Joshua Hernandez and **Konstantine Tsotsos** and **Stefano Soatto**
University of California, Los Angeles
{jheez,ktsotsos,soatto}@ucla.edu

Abstract

We analyze the observability of 3-D position and orientation from the fusion of visual and inertial sensors. The model contains unknown parameters, such as sensor biases, and so the problem is usually cast as a mixed filtering/identification problem, with the resulting observability analysis providing necessary conditions for convergence to a unique point estimate. Most models treat sensor bias rates as "noise," independent of other states, including biases themselves, an assumption that is violated in practice. We show that, when this assumption is lifted, the resulting model is not observable, and therefore existing analyses cannot be used to conclude that the set of states that are indistinguishable from the measurements is a singleton. We re-cast the analysis as one of sensitivity: Rather than attempting to prove that the set of indistinguishable trajectories is a singleton, we derive bounds on its volume, as a function of characteristics of the sensor and other sufficient excitation conditions. This provides an explicit characterization of the indistinguishable set that can be used for analysis and validation purposes.

1 Introduction

We present a novel approach to the analysis of observability/identifiability of three-dimensional (3-D) position and orientation (termed *pose*) in visually-assisted navigation, whereby inertial sensors (accelerometers and gyrometers, jointly referred to as an Inertial Measurement Unit, or IMU) are used in conjunction with optical sensors (vision) to yield an estimate of the 3-D pose of the sensor platform. It is customary to frame this as a filtering problem, where the time-series of positions and orientations of the sensor platform is modeled as the state trajectory of a dynamical system, that produces sensor measurements as outputs, up to some uncertainty. Observability/identifiability analysis refers to the characterization of the set of possible state trajectories that produce the same measurements, and therefore are indistinguishable given the outputs [Soatto, 1994; Kelly and Sukhatme, 2009; Mourikis and Roumeliotis, 2007; Jones *et al.*, 2007; Martinelli and others, 2014].

The parameters in the model are either treated as unknown constants (e.g., calibration parameters) or as random processes (e.g., accelerometer and gyro biases) and included in the state of the model, which is then driven by some kind of *uninformative* ("noise") input. Because noise does not affect the observability of a model, for the purpose of analysis it is usually set to zero. However, the input to the model of accelerometer and gyro bias is typically *small* but *not independent* of the state. Thus, it should be treated as an *unknown input*, which is known to be "small" in some sense, rather than "noise."

Our first contribution is to show that while (a prototypical model of) assisted navigation is *observable* in the absence of unknown inputs, it is *not* observable when unknown inputs are taken into account. Our second contribution is to reframe observability as a *sensitivity* analysis, and to show that while the set of indistinguishable trajectories is *not* a singleton (as it would be if the model was observable), it is nevertheless bounded. *We explicitly characterize this set and bound its volume* as a function of the characteristics of the inputs, which include sensor characteristics (bias rates) and the motion undergone by the platform (sufficient excitation).

Related work

In addition to the above-referenced work on visual-inertial observability, our work relates to general unknown-input observability of linear time-invariant systems addressed in [Basile and Marro, 1969; Hamano and Basile, 1983], for affine systems [Hammouri and Tmar, 2010], and non-linear systems in [Dimassi *et al.*, 2010; Tanwani, 2011; Bezzaoucha *et al.*, 2011]. The literature on robust filtering and robust identification is relevant, if the unknown input is treated as a disturbance. However, the form of the models involved in vision-aided navigation does not fit in the classes treated in the literature above, which motivates our analysis. The model we employ includes alignment parameters for the (unknown) pose of the inertial sensor relative to the camera.

1.1 Notation

We adopt the notation of [Murray *et al.*, 1994], where a reference frame is represented by an orthogonal 3×3 positive-determinant (rotation) matrix $R \in \mathrm{SO}(3) \doteq \{R \in \mathbb{R}^{3\times 3} \mid R^T R = RR^T = I,\ \det(R) = +1\}$ (the *special orthogonal group*) and a translation vector $T \in \mathbb{R}^3$. They are collectively indicated by $g = (R, T) \in \mathrm{SE}(3)$ (the *special Euclidean group*). When g represents the change of coordinates from a reference frame "a" to another ("b"), it is indicated by g_{ba}. Then the columns of R_{ba} are the coordinate axes of a relative to the reference frame b, and T_{ba} is the origin of a in the reference frame b. If p_a is a point relative to the reference frame a, then its representation relative to b is $p_b = g_{ba} p_a$. If X_a are the coordinates of p_a, then $X_b = R_{ba} X_a + T_{ba}$ are the coordinates of p_b. A time-varying transformation (or pose) is indicated with $g(t) = (R(t), T(t))$

We indicate with $\widehat{\omega}$ a skew-symmetric matrix $\widehat{\omega} \in \mathfrak{so}(3) \doteq \{S \in \mathbb{R}^{3\times 3} \mid S^T = -S\}$ corresponding to the cross product with the vector $\omega \in \mathbb{R}^3$, so that $\widehat{\omega} v = \omega \times v$ for any vector $v \in \mathbb{R}^3$. In homogeneous coordinates, we write $\bar{X}_b = g_{ba} \bar{X}_a$ where $\bar{X}^T = [X^T \ 1]$ and $\bar{X}_b^T = [(R_{ba} X_a + T_{ba})^T 1]$.

1.2 Motion Model

There are several reference frames to be considered in a navigation scenario. The *spatial frame* s is typically attached to the Earth and oriented so that gravity γ takes the form $\gamma^T = [0 \ 0 \ 1]^T \|\gamma\|$ where $\|\gamma\|$ can be read from tabulates based on location and is typically around $9.8 m/s^2$. The *body frame* b is attached to the IMU. The *camera frame* c, relative to which image measurements are captured, is also unknown, although we will assume that *intrinsic calibration* has been performed, so that measurements on the image plane are provided in metric units [Ma *et al.*, 2003]. The motion of a sensor platform is represented as the time-varying pose g_{sb} of the body relative to the spatial frame.

The equations of motion (known as mechanization equations) are usually described in terms of the body frame at time t relative to the spatial frame $g_{sb}(t)$. Since the spatial frame is arbitrary (other than for being aligned to gravity), it is often chosen to be co-located with the body frame at time $t = 0$. To simplify the notation, we indicate this time-varying frame $g_{sb}(t)$ simply as g, and so for $R_{sb}, T_{sb}, \omega_{sb}, v_{sb}$, thus effectively omitting the subscript sb wherever it appears. This yields $\dot{T} = v$, $\dot{R} = R\widehat{\omega}$, $\dot{v} = \alpha, \dot{\omega} = w$, $\dot{\alpha} = \xi$ where $w \in \mathbb{R}^3$ is the rotational acceleration, and $\xi \in \mathbb{R}^3$ the translational jerk (derivative of acceleration). For further details, see [Jones and Soatto, 2011].

1.3 Sensor Model

Although the acceleration α defined above corresponds to neither body nor spatial acceleration, it is conveniently related to accelerometer measurements α_{imu}:

$$\alpha_{\mathrm{imu}}(t) = R^T(t)(\alpha(t) - \gamma) + \underbrace{\alpha_b(t) + n_\alpha(t)}_{} \tag{1}$$

where the measurement error in bracket includes a slowly-varying mean ("bias") $\alpha_b(t)$ and a residual term n_α that is commonly modeled as a zero-mean (its mean is captured by the bias), white, homoscedastic and Gaussian noise process. In other words, it is assumed that n_α is independent of α, hence uninformative. Measurements from a gyro, ω_{imu}, can be similarly modeled as

$$\omega_{\mathrm{imu}}(t) = \omega(t) + \underbrace{\omega_b(t) + n_\omega(t)}_{} \tag{2}$$

where the measurement error in bracket includes a slowly-varying bias $\omega_b(t)$ and a residual "noise" n_ω also assumed zero-mean, white, homoscedastic and Gaussian, independent of ω.

Other than the fact that the biases α_b, ω_b change *slowly*, they can change arbitrarily. One can therefore consider them an *unknown input* to the model, or a *state* in the model, in which case one has to hypothesize a dynamical model for them. For instance,

$$\dot{\omega}_b(t) = w_b(t), \quad \dot{\alpha}_b(t) = \xi_b(t) \tag{3}$$

for some unknown inputs w_b, ξ_b that can be safely assumed to be *small*, but not (white, zero-mean and, most importantly) independent of the biases. Nevertheless, it is common to consider them to be realizations of a Brownian motion that is *independent* of ω_b, α_b. This is done for convenience as one can then consider all unknown inputs as "noise." Unfortunately, however, this has implications on the analysis of the observability and identifiability of the resulting model.

1.4 Model Reduction

The equations above define a dynamical model having as output the IMU measurements. In this standard model, data from the IMU are considered as (output) *measurements*. However, it is customary to treat them as (known) *input* to the system, by writing ω in terms of ω_{imu} and α in terms of α_{imu}. Including the initial conditions and biases, the resulting mechanization model is

$$\begin{cases} \dot{T} = v & T(0) = 0 \\ \dot{R} = R(\widehat{\omega}_{\mathrm{imu}} - \widehat{\omega}_b) + n_R & R(0) = R_0 \\ \dot{v} = R(\alpha_{\mathrm{imu}} - \alpha_b) + \gamma + n_v & \\ \dot{\omega}_b = w_b & \\ \dot{\alpha}_b = \xi_b & \end{cases} \tag{4}$$

with $n_R = -n_\omega$ and $n_v = -R n_\alpha$, both typically considered independant of the state.

1.5 Imaging Model and Alignment

Initially we assume there is a collection of points X^i, $i = 1, \dots, N$, visible from time $t = 0$ to the current time t. If $\pi : \mathbb{R}^3 \to \mathbb{R}^2; X \mapsto [X_1/X_3, \ X_2/X_3]$ is a canonical central (perspective) projection, assuming that the camera is *calibrated*,[1] *aligned*,[2] and that the spatial frame coincides with the body frame at time 0, we have

$$y^i(t) = \frac{R_{1:2}^T(t)(X^i - T_{1:2}(t))}{R_3^T(t)(X^i - T_3(t))} \doteq \pi(g^{-1}(t) X^i) + n^i(t) \tag{5}$$

In practice, the measurements $y(t)$ are known only up to a transformation g_{cb} mapping the body frame to the camera, often referred to as "alignment":

$$y^i(t) = \pi\left(g_{cb} g^{-1}(t) X_s^i\right) + n^i(t) \in \mathbb{R}^2 \tag{6}$$

Which we can add, along with the points X_s^i, to the state with trivial dynamics $\dot{g}_{cb} = 0$:

$$\begin{cases} \dot{X}_s^i = 0, \quad i = 1, \dots, N \\ \dot{g}_{cb} = 0 \end{cases} \tag{7}$$

[1] Intrinsic calibration parameters are known.

[2] The pose of the camera relative to the IMU is known.

2 Analysis of the Model

The goal here is to exploit imaging and inertial measurements to infer the sensor platform trajectory. For this problem to be well-posed, a (sufficiently exciting) realization of $\omega_{\rm imu}, \alpha_{\rm imu}$ and y should constrain the set of trajectories that satisfy (4)-(7) to be unique. If there are different trajectories that satisfy (4) with the same outputs and inputs, they are *indistinguishable*. If the set of indistinguishable trajectories is a singleton (contains only one element, presumably the "true" trajectory), the model (4) is *observable*, and one may be able to retrieve a unique point-estimate of the state using a filter, or observer.

While it is commonly accepted that the model (specifically, its equivalent reduced realization) (4), is observable, this is the case only when *biases are exactly constant.* But if biases are allowed to change, however slowly, the observability analysis conducted thus far cannot be used to conclude that the indistinguishable set is a singleton. Indeed, we show that this is the not the case, by computing the indistinguishable set explicitly. The following claim is proven in [Hernandez *et al.*, 2013].

Claim 1 (Indistinguishable Trajectories) *Let* $g(t) = (R(t), T(t)) \in \mathrm{SE}(3)$ *satisfy* (4)-(7) *for some known constant* γ *and functions* $\alpha_{\rm imu}(t)$, $\omega_{\rm imu}(t)$ *and for some unknown functions* $\alpha_b(t), \omega_b(t)$ *that are constrained to have* $\|\dot\alpha_b(t)\| \leq \epsilon$, $\|\dot\omega_b(t)\| \leq \epsilon$, *and* $\|\ddot\omega_b(t)\| \leq \epsilon$ *at all t, for some* $\epsilon < 1$.

Suppose $\tilde g(t) \doteq \sigma(g_B g(t) g_A)$ *for some constant* $g_A = (R_A, T_A)$, $g_B = (R_B, T_B)$, $\sigma > 0$, *with bounds on the configuration space such that*[3] $\|T_A\| \leq M_A$ *and* $0 < m_\sigma \leq |\sigma| \leq M_\sigma$. *Then, under sufficient excitation conditions,* $\tilde g(t)$ *satisfies (4)-(7) if and only if*

$$\|I - R_A\| \leq \frac{2\epsilon}{m(\dot\omega_{\rm imu}\!:\!\mathbb{R}^+)} \tag{8}$$

$$|\sigma - 1| \leq \frac{k_1\epsilon + M_\sigma\|I - R_A\|}{m(\dot\alpha_{\rm imu}\!:\!\mathcal{I}_1)} \tag{9}$$

$$\|T_A\| \leq \frac{\epsilon(k_2 + (2M_\sigma + 1)M_A)}{m_\sigma\, m(\ddot\omega_{\rm imu}\!:\!\mathcal{I}_2)} \tag{10}$$

$$\|(1 - R_B^T)\gamma\| \leq \frac{\epsilon(k_3 + M_\sigma M_A)}{m_\sigma\, m(\omega_{\rm imu} - \omega_b\!:\!\mathcal{I}_3)} + \\ + \frac{(|\sigma - 1| + \epsilon)M(\omega_{\rm imu} - \omega_b\!:\!\mathcal{I}_3)\|\gamma\|}{m_\sigma\, m(\omega_{\rm imu} - \omega_b\!:\!\mathcal{I}_3)} \tag{11}$$

for $\mathcal{I}_i$ *and* k_i *determined by the sufficient excitation conditions.*

Here, *sufficient excitation* (m and M) refers to the nature of the motion, with more dynamic motion leading to larger values, as described in [Hernandez *et al.*, 2013]. From these bounds, we find that the set of indistinguishable trajectories in the limit where $\epsilon \to 0$ is parametrized by an arbitrary $T_B \in \mathbb{R}^3$ and rotation $\theta \in \mathbb{R}$ about gravity, termed the Gauge ambiguity, which can be explicitly fixed [Hernandez *et al.*, 2013]. This immediately implies the following

[3]Here $\sigma(g)$ is a scaled rigid motion: if $g = (R, T)$, then $\sigma(g) = (R, \sigma T)$.

Claim 2 (unknown-input observability) *The model (4)*-(7) *is* not *observable, even after fixing the Gauge ambiguity, as the indistinguishable set is not a singleton, unless biases are constant (*$\epsilon = 0$*) or their derivative is known exactly.*

We refer the reader to [Hernandez *et al.*, 2013] for additional details and proofs, which are articulated into several steps. In practice, once the Gauge transformations are fixed, a properly designed filter can be designed to converge to a point estimate, but there is no guarantee that such an estimate coincides with the true trajectory. Instead, the estimate can deviate from the true trajectory depending on the biases. The analysis above quantifies how far from the true trajectory the estimated one can be, provided that the estimation algorithm uses bounds on the bias drift rates and the characteristics of the motion. Often these bounds are not strictly enforced but rather modeled through the driving noise covariance.

3 Empirical Validation

To validate the analysis, we run repeated trials to estimate the state of the platform under different motion but identical alignment (the camera is rigidly connected to the IMU) using our experimental platform [Tsotsos *et al.*, 2015]. If alignment parameters (T_{cb} translational and Ω_{cb} parametrizing R_{cb}) were identifiable (or the augmented state observable), we would expect convergence to the same parameters across all trials. Instead, Fig. 1 shows that the estimates of the parameters stabilize, but to different values at each run. Nevertheless, the parameter values are in a set, whose volume can be bounded based on the analysis above and the characteristics of the sensor. In particular, less stable biases, and less exciting motions, result in a larger indistinguishable set: Fig. 2 shows the same experiments with more gentle (hence less exciting) motions. Fig. 3 shows the same where

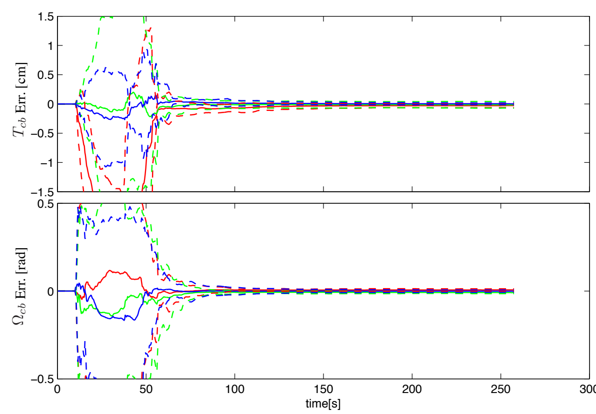

Figure 1: Convergence of alignment parameters to a set, rather than a unique point estimate, due to the lack of unknown-input observability in the presence of (realistic) non-constant biases. The mean (solid) and twice the std. dev. (dashed) of the change in estimated parameters relative to their initial nominal values across multiple trials on real data, show that different trials converge to different parameter values, but to within a bounded set.

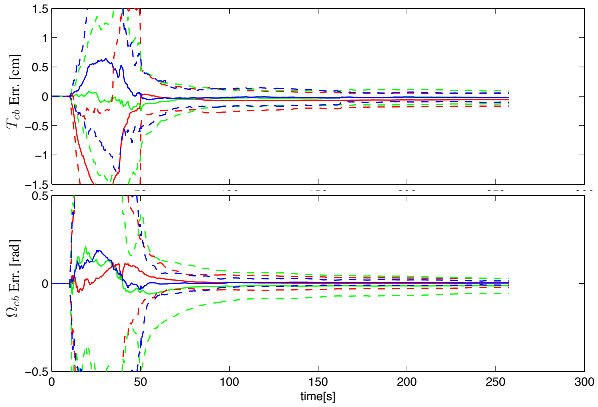

Figure 2: The indistinguishable set is bounded depending on the characteristic of the motion. Gentler motion produces multiple trials that converge to a set of larger volume compared to Fig. 1.

the accel and gyro biases have been artificially inflated by adding a slowly time-varying offset. Additionally, we con-

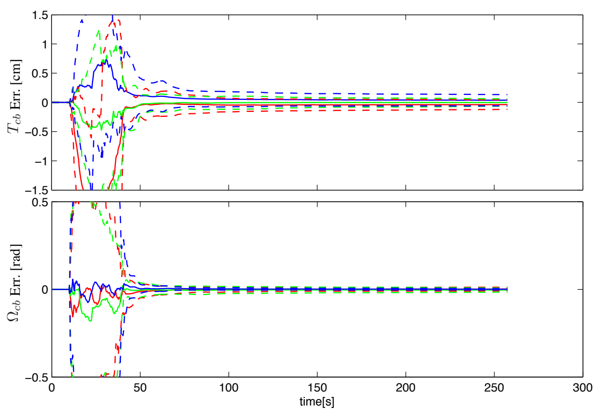

Figure 3: The indistinguishable set also depends on the characteristics of the sensor, and its volume is directly proportional to the sensor bias rate. Here artificial bias drift is added, resulting in a larger indistinguishable set compared to Fig. 1.

ducted Monte-Carlo experiments on the model in simulation using stationary and time-varying biases while undergoing sufficiently exciting motion. Figures 4 and 5 show the resulting estimation errors of alignment states for 50 trials each using a constant and white-noise driven bias respectively.

4 Discussion

This paper presents a overview of the analysis presented in [Hernandez *et al.*, 2015]. We have shown that when inertial sensor biases are included as model parameters in the state of a filter used for navigation estimates, with bias rates treated as unknown inputs, the resulting model is *not observable*, that is, the set of indistinguishable states is not a singleton.

Consequently, we have re-formulated the problem of analyzing the convergence characteristics of (any) filters for

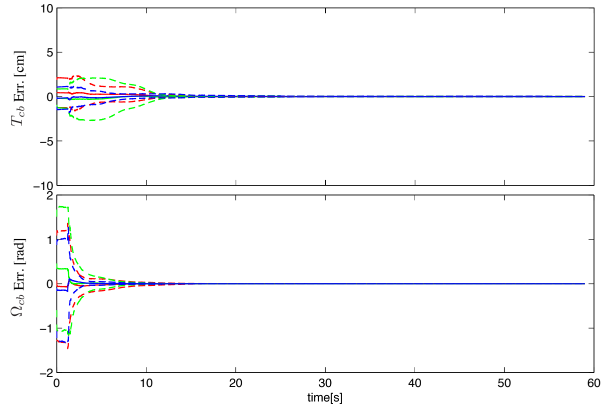

Figure 4: Mean (solid line) and twice the standard deviation (dashed lines) of estimation errors of alignment parameters aggregated over 50 Monte-Carlo trials with a constant bias.

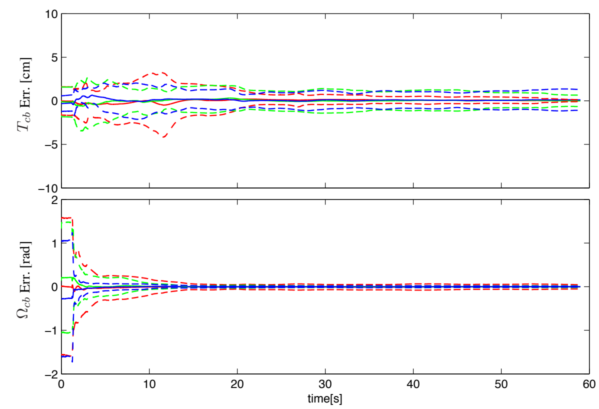

Figure 5: Mean (solid line) and twice the standard deviation (dashed lines) of estimation errors of alignment parameters aggregated over 50 Monte-Carlo trials with a time-varying bias.

vision-aided inertial navigation *not* as one of observability or identifiability, but one of *sensitivity*, by bounding the set of indistinguishable trajectories to a set whose volume depends on motion characteristics.

The advantage of this approach, compared to the standard observability analysis based on rank conditions, is that we characterize the indistinguishable set explicitly. We quantify the "degree of unobservability" as the sensitivity of the solution set to the input; provided that sufficient-excitation conditions are satisfied, the unobservable set can be bounded and effectively be treated as a singleton. More generally, however, the analysis provides an estimate of the uncertainty surrounding the solution set, as well as a guideline on how to limit it by enforcing certain gauge transformations.

Acknowledgements

This work was supported by the Air Force Office of Scientific Research (grant no. AFOSR FA9550-12-1-0364) and the Office of Naval Research (grant no. ONR N00014-13-1-034).

References

[Basile and Marro, 1969] G. Basile and G. Marro. On the observability of linear, time-invariant systems with unknown inputs. *Journal of Optimization theory and applications*, 3(6):410–415, 1969.

[Bezzaoucha *et al.*, 2011] S. Bezzaoucha, B. Marx, D. Maquin, J. Ragot, et al. On the unknown input observer design: a decoupling class approach with application to sensor fault diagnosis. In *1st International Conference on Automation and Mechatronics, CIAM'2011*, 2011.

[Dimassi *et al.*, 2010] H. Dimassi, A. Loría, and S. Belghith. A robust adaptive observer for nonlinear systems with unknown inputs and disturbances. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2602–2607. IEEE, 2010.

[Hamano and Basile, 1983] F Hamano and G Basile. Unknown-input present-state observability of discrete-time linear systems. *Journal of Optimization Theory and Applications*, 40(2):293–307, 1983.

[Hammouri and Tmar, 2010] H. Hammouri and Z. Tmar. Unknown input observer for state affine systems: A necessary and sufficient condition. *Automatica*, 46(2):271–278, 2010.

[Hernandez *et al.*, 2013] J. Hernandez, K. Tsotsos, and S. Soatto. Observability, identifiability, sensitivity, and model reduction for vision-assisted inertial navigation. *UCLA CSD TR130022, http://arxiv.org/abs/1311.7434*, 2013.

[Hernandez *et al.*, 2015] Joshua Hernandez, Konstantine Tsotsos, and Stefano Soatto. Observability, identifiability and sensitivity of vision-aided inertial navigation. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 2319–2325. IEEE, 2015.

[Jones and Soatto, 2011] Eagle S Jones and Stefano Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4):407–430, 2011.

[Jones *et al.*, 2007] E. S. Jones, A. Vedaldi, and S. Soatto. Inertial structure from motion and autocalibration. In *Workshop on Dynamical Vision*, October 2007.

[Kelly and Sukhatme, 2009] J. Kelly and G. Sukhatme. Fast Relative Pose Calibration for Visual and Inertial Sensors. In *Experimental Robotics*, pages 515–524, 2009.

[Ma *et al.*, 2003] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3D vision, from images to geometric models*. Springer Verlag, 2003.

[Martinelli and others, 2014] Agostino Martinelli et al. Visual-inertial structure from motion: observability and resolvability. *Foundations and Trends® in Computer Graphics and Vision*, 1(1), 2014.

[Mourikis and Roumeliotis, 2007] A. Mourikis and S. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3565–3572. IEEE, 2007.

[Murray *et al.*, 1994] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[Soatto, 1994] Stefano Soatto. Observability/identifiability of rigid motion under perspective projection. In *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, volume 4, pages 3235–3240. IEEE, 1994.

[Tanwani, 2011] Aneel Tanwani. *Invertibility and Observability of Switched Systems with Inputs and Outputs*. PhD thesis, University of Illinois at Urbana-Champaign, 2011.

[Tsotsos *et al.*, 2015] Konstantine Tsotsos, Alessandro Chiuso, and Stefano Soatto. Robust inference for visual-inertial sensor fusion. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 5203–5210. IEEE, 2015.