

Minimalistic sensor design in visual-inertial structure from motion

Agostino Martinelli

► **To cite this version:**

Agostino Martinelli. Minimalistic sensor design in visual-inertial structure from motion. icra 2015, May 2015, Seattle, United States. 2015, <10.1109/ICRA.2015.7139656>. <hal-01248785>

HAL Id: hal-01248785

<https://hal.inria.fr/hal-01248785>

Submitted on 28 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Minimalistic sensor design in visual-inertial structure from motion*

Agostino Martinelli

Abstract—This paper presents a theoretical investigation in the framework of visual-inertial sensor fusion and the results here provided are the extension of our previous contribution in [10], [11]. The general goal of this research is to establish minimalistic visual-inertial sensors settings, which still provide full information even in the most challenging situations, i.e., in the case of unknown camera extrinsic calibration, unknown magnitude of the gravity, unknown inertial sensor bias and when only a single point feature is available. The investigation here provided allows us to conclude that, even in the case of a single point feature, the information provided by a sensor suit composed by a monocular camera and two inertial sensors (along two independent axes and where at least one is an accelerometer) is the same as in the case of a complete inertial measurement unit (i.e., when the inertial sensors consist of three orthogonal accelerometers and three orthogonal gyroscopes). To derive this result, an observability analysis of systems with only one and two inertial sensors is performed. This analysis requires to approach an open problem in control theory, called the Unknown Input Observability (UIO). In this paper we adopt the same method introduced in [10], [11] to solve this UIO problem. The method has been here improved in order to deal with systems more complex than the ones analyzed in [10], [11]. The paper also provides a general discussion on UIO and in particular on the proposed solution.

I. INTRODUCTION

Visual and inertial sensing have received great attention by the mobile robotics community since they require no external infrastructure and this is a key advantage for robots operating in unknown environments where GPS signals are shadowed. Recent works have shown that through these sensors it is possible to perform metric structure from motion (e.g., [1], [5], [6], [9], [14]). In the sequel we will refer to the problem of building the 3D-structure of the environment starting from visual and inertial measurements obtained during navigation as to the visual-inertial structure from motion problem (the Vi-SfM). Inertial sensors usually consist of three orthogonal accelerometers and three orthogonal gyroscopes. All together, they constitute the Inertial Measurement Unit (IMU). We will refer to the fusion of monocular vision with the measurements from an IMU as to the *standard* Vi-SfM problem. In [4], [5], [6], [7], [8], [9], [13] and [15] the observability properties of the standard Vi-SfM have been investigated in several different scenarios. Recently, following two independent procedures, the most general result on the observability of the standard Vi-SfM problem has been provided in [2] and [11]. This

result can be summarized as follows. In the standard Vi-SfM problem all the independent observable modes are: the positions in the local frame of all the observed features, the three components of the speed in the local frame, the biases affecting the inertial measurements, the roll and the pitch angle, the magnitude of the gravity and the transformation between the camera and IMU frames. The fact that the yaw angle is not observable is an obvious consequence of the system invariance under rotation about the gravity vector.

In [10], [11] we investigated the observability properties of the Vi-SfM when the number of inertial sensors is reduced. We considered separately the cases of camera extrinsically calibrated and not calibrated. In the second case, we assumed that at least five point features were available. This allowed us to consider the camera as a sensor able to provide the position and the speed up to a scale. Only in the first case we also considered the case of less than five point features (specifically, in that case we dealt with a single point feature). Under these premises, we proved that the observability properties of Vi-SfM do not change by removing all the three gyroscopes and one of the accelerometers. In other words, exactly the same properties hold when the sensor system only consists of a monocular camera and two accelerometers. By removing a further accelerometer (i.e., by considering the case of a monocular camera and a single accelerometer) the system loses part of its observability properties. In particular, a new symmetry arises. This symmetry corresponds to an internal rotation around the accelerometer axis. This means that some of the internal parameters that define the extrinsic camera calibration, are no longer observable. Although this symmetry does not affect the observability of the absolute scale and the magnitude of the velocity, it reflects in an indistinguishability of all the initial speeds that differ for a rotation around the accelerometer axis. On the other hand, if the camera is extrinsically calibrated (i.e., if the relative transformation between the camera frame and the accelerometer frame is known) this symmetry disappears and the system still maintains full observability.

The goal of this paper is to extend the analysis in [10], [11] by also including the extreme case of a single point feature when the camera is not extrinsically calibrated. The system is described in section II. Even if this extension seems to be simple, the analytic computation must be totally changed. Indeed, by including in the state the camera extrinsic parameters, the computation, as carried out in [10], [11] in the case when the camera is calibrated, becomes prohibitive.

The problem of deriving the observability properties of Vi-SfM when the number of inertial sensors is reduced corresponds to solve a problem that in control theory is

*This work has been supported by the French National Research Agency ANR 2014 through the project VIMAD

A. Martinelli is with INRIA Rhone Alpes, Grenoble, France e-mail: agostino.martinelli@ieee.org

known as Unknown Input Observability (UIO). This problem is still unsolved in the nonlinear case. In [10], [11] we introduced a new method able to provide sufficient conditions for the state observability. On the other hand, this method is based on a state augmentation. Specifically, the new extended state includes the original state together with the unknown inputs and their time-derivatives up to a given order. Then, the method introduced in [10], [11] (which is summarized in section IV-A and IV-B) is based on the computation of a codistribution defined in the augmented space. This makes the computation necessary to derive the observability properties dependent on the dimension of the augmented state and consequently prohibitive in our case. Our effort to deal with this fundamental issue, is devoted to separate the information on the original state from the information on its extension. We fully solve this problem only in the case of a single unknown input [12]. For the general case, we partially solve this problem and we suggest a technique able to partially perform this separation (section IV-C). In section V we derive the observability properties for our systems. We obtain the extension of the results derived in [10], [11], i.e., in the case of a single point feature available and when the transformation between the camera frame and the frame attached to the inertial sensors is unknown. Finally, in section VI we provide a discussion of the results and our conclusion.

II. THE CONSIDERED SYSTEM

We consider a system which consists of a monocular camera and a single-axis accelerometer. All the results provided in this paper are obtained starting by the analysis of this system, even if they also regard the case of an additional inertial sensor, that can be both an accelerometer or a gyroscope. We assume that the local frame coincides with the frame attached to the single-axis accelerometer and, without loss of generality, we assume that this local frame has its z-axis coincident with the axis of the accelerometer. The position of the camera optical center in the local frame will be denoted by \mathbf{R}^c and the camera orientation will be characterized through the three Euler angles α , β , γ . Specifically, a vector with orientation $\hat{\tau}$ in the local frame, will have the orientation $R\hat{\tau}$ in the camera frame, where $R = R(\alpha, \beta, \gamma) = R_\alpha^z R_\beta^x R_\gamma^z$ and R_η^z and R_η^x rotates the unit vector $\hat{\tau}$ clockwise through the angle η about the z-axis and the x-axis, respectively. The vector \mathbf{R}^c and the three angles α , β , γ , characterize the extrinsic camera calibration and are assumed to be unknown. We consider the extreme case when a single point feature is available and we denote its position in the camera frame with ${}^c\mathbf{F} \equiv [{}^cF_x \ {}^cF_y \ {}^cF_z]^T$. We also introduce a global frame whose z-axis is along the gravity vector but pointing upward. In other words, in this frame the gravity vector will be $\mathbf{g} = [0, 0, -g]^T$, where g is the magnitude of the gravity which is assumed to be unknown. We characterize our system through the following state:

$$\mathbf{S} \equiv \left[{}^c\mathbf{F}^T, \mathbf{V}^T, q_t, q_x, q_y, q_z, A_z^b, g, \mathbf{R}^{cT}, \alpha, \beta, \gamma \right]^T \quad (1)$$

where \mathbf{V} is the speed in the local frame, q_t, q_x, q_y, q_z are the four components of the unit quaternion $q \equiv q_t + q_x i + q_y j + q_z k$, which characterizes the orientation of the local frame in the global frame and A_z^b is the accelerometer bias. This state is a vector with 18 components, which are not independent since the quaternion must be unit ($qq^* = 1$). The dynamics of this state are the following:

$$\begin{cases} {}^c\dot{\mathbf{F}} = -{}^c\boldsymbol{\Omega} \wedge {}^c\mathbf{F} - R(\mathbf{V} + \boldsymbol{\Omega} \wedge \mathbf{R}^c) \\ \dot{\mathbf{V}} = -\boldsymbol{\Omega} \wedge \mathbf{V} + \mathbf{A} + \mathbf{G} \\ \dot{q} = \frac{1}{2}q\Omega_q \quad \dot{g} = \dot{A}_z^b = \dot{\alpha} = \dot{\beta} = \dot{\gamma} = 0 \end{cases} \quad (2)$$

where $\boldsymbol{\Omega} \equiv [\Omega_x \ \Omega_y \ \Omega_z]^T$ is the unknown angular speed in the local frame, ${}^c\boldsymbol{\Omega} = R\boldsymbol{\Omega}$ is the angular speed in the camera frame, Ω_q is the imaginary quaternion associated with $\boldsymbol{\Omega}$, i.e., $\Omega_q \equiv 0 + \Omega_x i + \Omega_y j + \Omega_z k$, \mathbf{A} is the acceleration that would be perceived by an ideal (i.e., noiseless and unbiased) three-axis accelerometer attached to the local frame and $\mathbf{G} \equiv [G_x, G_y, G_z]^T$ is the gravity vector in the local frame, i.e., by denoting with G_q the corresponding imaginary quaternion ($G_q = 0 + G_x i + G_y j + G_z k$), we have the following quaternion equality $G_q = -gq^*kq$. Note that the first two components of \mathbf{A} are unknown while the third is available up to the bias A_z^b thanks to the single axis accelerometer.

The monocular camera provides the position of the feature in the camera frame (${}^c\mathbf{F}$) up to a scale. Hence, it provides the ratios of the components of ${}^c\mathbf{F}$:

$$h_{cam}(\mathbf{S}) \equiv [h_u, h_v]^T = \left[\frac{{}^cF_x}{{}^cF_z}, \frac{{}^cF_y}{{}^cF_z} \right]^T \quad (3)$$

We have also to consider the constraint $q^*q = 1$. This provides the further observation:

$$h_{const}(\mathbf{S}) \equiv h_q = q^*q \quad (4)$$

III. OBSERVABILITY ANALYSIS

The system described in the previous section is characterized by the state in (1), the dynamics in (2) and the observations in (3) and (4). This can be regarded as an input-output non linear system, where, part of the inputs is unknown. Specifically, the dynamics in (2) have the following structure:

$$\dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x}) + \sum_{i=1}^{m_u} \mathbf{f}_i(\mathbf{x})u_i + \sum_{j=1}^{m_w} \mathbf{g}_j(\mathbf{x})w_j \quad (5)$$

where:

- $m_u = 1$, $m_w = 5$, $u_1 = A_z$, $w_1 = A_x$, $w_2 = A_y$, $w_3 = \Omega_x$, $w_4 = \Omega_y$, $w_5 = \Omega_z$;
- \mathbf{f}_0 is obtained by setting to zero all the components of both \mathbf{A} and $\boldsymbol{\Omega}$ in (2), with the exception of the third component of \mathbf{A} , which is set equal to $-A_z^b$;
- \mathbf{f}_1 is obtained by removing \mathbf{f}_0 from (2) and then by setting $A_x = A_y = \Omega_x = \Omega_y = \Omega_z = 0$ and $A_z = 1$;
- \mathbf{g}_j (for $j = 1, \dots, 5$) are obtained by removing \mathbf{f}_0 from (2) and then by setting to zero all the components

of both A and Ω with the exception of one of them depending on j (e.g., to obtain g_1 we set $A_y = A_z = \Omega_x = \Omega_y = \Omega_z = 0$ and $A_x = 1$).

The observability properties of systems with the dynamics given in (5) cannot be obtained by applying standard methods in control theory (e.g., the observability rank condition in [3]) because part of the inputs is unknown. We apply the same method introduced in [10], [11], which has been better formalized in [12]. On the other hand, the system defined by the dynamics in (2) is more complex than the systems analyzed in [10], [11]. Note that in [10], [11] the case of a single feature was considered only when the camera was extrinsically calibrated, which means that the state had 12 entries instead of 18. It results that the analytical computation becomes prohibitive when we use the Lie derivatives whose order is equal or larger than 3. Note that to deal with the simpler case of camera calibrated, we had to use the Lie derivatives up to the 7th order.

For these reasons, from one side we introduce some general properties, which allow us to reduce the analytic computation demanded to solve any UIO problem (section IV). From another side, we will proceed in three distinct phases, as illustrated in section V.

IV. COMPUTATION OF THE OBSERVABLE CODISTRIBUTION: KEY PROPERTIES

In this section we summarize the method introduced in [10], [11] to solve an UIO problem and we add new properties that allow us to significantly reduce the analytic computational burden.

A. Extended system

According to the observability rank condition [3], the observability properties of a nonlinear system with only known inputs, can be investigated by analyzing the codistribution generated by the gradients of its Lie derivatives. Specifically, if the dimension of this codistribution is equal to the dimension of the state, we conclude that the state is observable¹. The observability rank condition has been introduced and proved that it is a sufficient condition for a system to be observable (theorem 3.1 in [3]). The proof of this theorem is based on the fact that the Lie derivatives of the outputs are constant on the indistinguishable sets.

In [10], [11] and [12], we proposed an extension of this method in order to deal with systems driven by unknown inputs. Specifically, our approach can be used to investigate the observability properties of systems that satisfy the dynamics in (5), namely where part of the inputs is unknown. We defined a new system, denoted by $\Sigma^{(k)}$. It is simply obtained by extending the original state by including the unknown inputs together with their time derivatives up to the $(k-1)$ -order. The extended state is:

$$\mathbf{x}^{(k)} \equiv [\mathbf{x}^T, \mathbf{w}^T, \mathbf{w}^{(1)T}, \dots, \mathbf{w}^{(k-1)T}]^T \quad (6)$$

¹Actually, it is weakly locally observable (see [3] for the definition of weak local observability).

where $\mathbf{w}^{(i)} \equiv \left. \frac{d^i \mathbf{w}}{dt^i} \right|_{t=0}$, $i = 0, 1, \dots, k-1$. The dimension of the extended state is $n + km_w$. Its basic equations can be obtained starting from (5) and are available in [12] together with the main properties satisfied by $\Sigma^{(k)}$. One fundamental property is that the Lie derivatives up to the k -order are constant on the indistinguishable sets (see property 4 in [12]). In other words, the Lie derivatives up to the k -order satisfy exactly the same property satisfied by all the Lie derivatives in the standard case of known inputs. As mentioned above, this property allows us to prove the theorem 3.1 in [3]. As a result, we can adopt the observability rank condition to investigate the observability properties of a system driven by also unknown inputs, provided that we suitably augment the state.

B. Extension of the Observability Rank Condition

On the basis of the previous discussion, the observability properties of a general nonlinear system, where part of the inputs is unknown, can be obtained by extending the state and by computing the codistribution that is the span of the gradients of all the Lie derivatives, up to the k -order, in the extended system $\Sigma^{(k)}$. In the sequel we denote this codistribution by $\bar{\Omega}_k$. Additionally, let us denote by $\mathbf{f}_0^{(k)}, \mathbf{f}_1^{(k)}, \dots, \mathbf{f}_{m_u}^{(k)}$ the vector fields that characterize the dynamics of $\Sigma^{(k)}$ (their expression is provided in [12] and it is easily obtained starting from (5)). The first step to do is to compute the codistribution $\bar{\Omega}_k$. This is carried out recursively, by using the following algorithm:

Algorithm 1 (Computation of $\bar{\Omega}_k$)

Set $\bar{\Omega}_0$ the span of the gradient of the output

Set $m = 0$

while $m < k$ **do**

Set $m = m + 1$

Compute all the m -order Lie derivatives of the output along the directions $\mathbf{f}_0^{(k)}, \mathbf{f}_1^{(k)}, \dots, \mathbf{f}_{m_u}^{(k)}$

Set $\bar{\Omega}_m$ the span of the gradients of the previous Lie derivatives and the generators of $\bar{\Omega}_{m-1}$

end while

We are interested in deriving the observability properties of the original state \mathbf{x} and not in the observability of the entire extended state. We can check the observability of the original state component by component, i.e., by proceeding as follows. Let us consider the i^{th} component of the original state, i.e., x_i ($i = 1, \dots, n$). We check if the gradient of x_i with respect to the extended state belongs to $\bar{\Omega}_k$. In other words, we check if the row vector of dimension $n + km_w$, with all the entries equal to zero with the exception of the i^{th} entry equal to 1, is in the span of the gradients of the Lie derivatives up to the k -order. If there exists a k such that this is true, we conclude that x_i is observable. If this holds for $i = 1, \dots, n$, we conclude that the original state is observable.

C. Properties of $\bar{\Omega}_k$

The method described in the previous subsection requires to compute the codistribution $\bar{\Omega}_k$. Then, the computation

demanded to check if a given covector belongs to $\bar{\Omega}_k$ can be very complex because by increasing k we also increase the dimension of the extended state. In the sequel, we will focus our attention on this fundamental issue. Specifically, we want to solve the following problem: separate the information on the original state from the information on its extension. We fully solve this problem only in the special case of a single unknown input and when the vector $f_0(x)$ in (5) vanishes. For this case we found the following fundamental result:

Theorem 1 (Information Separation when $m_u = m_w = 1$)
For every integer m , the codistribution $\bar{\Omega}_m$ is the span of two sets of covectors. The first set consists of gradients of scalar functions that only depend on the original state. The second set consists of gradients of scalar functions that depend on the entire augmented state but that do not contain additional information on the original state.

The proof of this theorem is complex and is based on several tricky steps (see theorem 1 in [12] for a complete proof). On the basis of the previous theorem, we can investigate the observability properties of the original state by only considering the first set of functions. This makes the computation independent of the dimension of the extended state².

Unfortunately, so far, we have not derived a similar result for the case of multiple inputs. On the other hand, we have proved the following weaker property, which still significantly simplifies the computation and which has been fundamental in deriving the observability properties of our system in (2).

Property 1 *Let us suppose that, for a given integer $m \leq k$, the gradient of a scalar function of the original state (i.e., $m(x)$) belongs to $\bar{\Omega}_{m-1}$. Then, the gradients of all the Lie derivatives of $m(x)$ along the directions $f_0^{(k)}, f_1^{(k)}, \dots, f_{m_u}^{(k)}$ belong to $\bar{\Omega}_m$.*

Proof: Let us denote with ∇ the gradient with respect to the extended state and by $\mathcal{L}_{f_i^{(k)}}$ the Lie derivative operator along the direction $f_i^{(k)}$, $i = 0, \dots, m_u$. Additionally, we assume that the dimension of $\bar{\Omega}_{m-1}$ is d and we denote with γ_i ($i = 1, \dots, d$) d independent Lie derivatives (whose order is less or equal to m). In other words, we have: $\bar{\Omega}_{m-1} = \text{span}\{\nabla\gamma_1, \dots, \nabla\gamma_d\}$. By assumption we have $\nabla m \in \bar{\Omega}_{m-1}$, namely: $\nabla m = \sum_{j=1}^d c_j \nabla\gamma_j$, where the coefficients c_j depend on the entire extended state. We have, for every $i = 0, 1, \dots, m_u$: $\nabla \mathcal{L}_{f_i^{(k)}} m = \mathcal{L}_{f_i^{(k)}} \nabla m = \sum_{j=1}^d \mathcal{L}_{f_i^{(k)}} (c_j \nabla\gamma_j) = \sum_{j=1}^d (\mathcal{L}_{f_i^{(k)}} c_j) \nabla\gamma_j + \sum_{j=1}^d c_j (\mathcal{L}_{f_i^{(k)}} \nabla\gamma_j)$ and the statement follows by remarking that $\mathcal{L}_{f_i^{(k)}} \nabla\gamma_j \in \bar{\Omega}_m$ ■

²Note that in [12] we provide a simple automatic method to directly generate these functions, starting from the dynamics in (5). Note also that in [12] we proved a further fundamental result that allows us to detect the largest integer m to be considered in the algorithm 1 to obtain all the observability properties (see theorem 2 in [12]).

The advantage of this property is that, if we are able to detect such a scalar function $m(x)$, we are allowed to use in the algorithm 1 ∇m as one of the generators of the codistribution $\bar{\Omega}_{m-1}$ and its Lie derivatives along $f_0^{(k)}, f_1^{(k)}, \dots, f_{m_u}^{(k)}$ as a part of the generators of $\bar{\Omega}_m$ (provided that they are independent). This partially separates the information on the original state from its extension.

V. OBSERVABILITY PROPERTIES

We extend the results stated by theorems 2 and 3 in [10] to the case of a single feature. Because of the huge computational burden, we proceed into three separate steps. In V-A we prove that the parameters α and β are identifiable. We know, from the analysis carried out in [10], that the parameter γ is unobservable when 5 or more features are available. Therefore, also in the case of a single feature it remains unobservable. In section V-B we consider the case when γ is known (since α and β are identifiable, the entire rotation matrix R is known). For the resulting system we prove that all the remaining states are observable, with the exception of the absolute yaw (which is unobservable even in the standard case, i.e., when a complete IMU is available and for every number of features). Finally, in V-C we perform a system decomposition and we provide the final results.

A. Identifiability of the parameters α and β

We apply the algorithm 1 to compute the codistribution $\bar{\Omega}_m$ for $m = 0, 1, 2, 3$ for the system characterized by the state in (1), the dynamics in (2) and the observations in (3) and (4). At each step of the algorithm, we check the observability of α and β , respectively. By a direct computation, it is possible to verify that the gradients of both α and β belong to $\bar{\Omega}_3$. More precisely, we do not need to fully compute $\bar{\Omega}_3$ since the gradients of both α and β belong to the span of all the Lie derivatives up to the third order with the exception of the ones along the direction $f_0^{(3)}$ all the three times (note that the expression of these Lie derivatives is the most complex). We proved the following property:

Property 2 *In the Vi-SfM problem with a single accelerometer, no gyroscope, unknown transformation between the camera and the accelerometer frames, the relative roll and pitch of the camera frame in the accelerometer frame are identifiable. This holds even in the case of one point feature.*

B. Observability properties when the matrix R is known

We investigate the observability properties of the system characterized by the state in (1) where we removed the last three entries, the dynamics given in (2) with R equal to the identity matrix and the observations in (3) and (4). We apply the algorithm 1 and, at each step, we check the observability of all the state entries. We obtain that the gradients of the three components of ${}^c F$ belong to $\bar{\Omega}_3$. Hence, we apply three times the property 1 with the scalar function equal to these three components. This significantly reduce the computation and allows us to manage with higher order Lie derivatives (specifically, up to the six order). By a direct

computation it is possible to verify that the gradients of six scalar functions belong to Ω_6 . These functions are the three components of the angular speed Ω and the three components of the vector $\mathbf{V} + \Omega \wedge \mathbf{R}^c$. We apply again property 1. We compute the codistribution that is the span of the following scalar functions:

- the three components of ${}^c\mathbf{F}$;
- the constraint in (4);
- the three components of the angular speed Ω , together with their Lie derivatives up to the 7th order;
- the three components of the vector $\mathbf{V} + \Omega \wedge \mathbf{R}^c$, together with their Lie derivatives up to the 6th order.

We obtain that the gradients of the following scalar functions belong to the previous codistribution: the first six components and the last five components of the state; the three functions of the quaternion components that correspond to the roll and the pitch angles and its norm (i.e., $Q_r \equiv \frac{q_t q_x + q_y q_z}{1 - 2(q_x^2 + q_y^2)}$, $Q_p \equiv q_t q_y - q_z q_x$ and $q_t^2 + q_x^2 + q_y^2 + q_z^2$; actually, the roll and pitch are $\arctan(2Q_r)$ and $\arcsin(2Q_p)$, respectively). Hence, we proved the following property:

Property 3 (Known camera-accelerometer rotation)

In the Vi-SfM problem with a single accelerometer, no gyroscope, known rotation between the camera and the accelerometer frames and unknown shift between these frames, all the independent observable modes are the same as in the standard Vi-SfM problem. This holds even in the extreme case of a single point feature.

C. Results in the general case

We are now ready to investigate the observability properties of the system characterized by the state in (1), the dynamics given in (2) and the observations in (3) and (4). From the analysis described in section V-A we know that the two parameters α and β are observable. Hence, we are allowed to simplify the system by eliminating from the state these two parameters. Additionally, we rotate the camera frame, i.e., we rotate all the images provided by the camera by pre-multiplying ${}^c\mathbf{F}$ by the rotation matrix $(R_\alpha^z R_\beta^x)^{-1}$. By introducing the vector ${}^n\mathbf{F} \equiv (R_\alpha^z R_\beta^x)^{-1} {}^c\mathbf{F}$, we obtain the new state: $\mathbf{S} \equiv [{}^n\mathbf{F}^T, \mathbf{V}^T, q_t, q_x, q_y, q_z, A_z^b, g, \mathbf{R}^c{}^T, \gamma]^T$. From (2), we obtain its dynamics as follows:

$$\begin{cases} {}^n\dot{\mathbf{F}} = -{}^n\Omega \wedge {}^n\mathbf{F} - R_\gamma^z (\mathbf{V} + \Omega \wedge \mathbf{R}^c) \\ \dot{\mathbf{V}} = -\Omega \wedge \mathbf{V} + \mathbf{A} + \mathbf{G} \\ \dot{q} = \frac{1}{2} q \Omega_q \quad \dot{g} = \dot{A}_z^b = \dot{\gamma} = 0 \end{cases} \quad (7)$$

where ${}^n\Omega \equiv (R_\alpha^z R_\beta^x)^{-1} {}^c\Omega$ is the angular speed in the new camera frame. Now, we perform the following coordinates change: $\mathbf{S} \rightarrow \mathbf{S}' \equiv [{}^n\mathbf{F}^T, \mathbf{V}'^T, q'_t, q'_x, q'_y, q'_z, A_z^b, g, \mathbf{R}'^c{}^T, \gamma]^T$, with $\mathbf{V}' = R_\gamma^z \mathbf{V}$, $\mathbf{R}'^c = R_\gamma^z \mathbf{R}^c$, $q' \equiv q'_t + q'_x i + q'_y j + q'_z k = qp$, and p is the unit quaternion associated to the rotation matrix R_γ^z ($p = \cos \frac{\gamma}{2} + \sin \frac{\gamma}{2} k$). Starting from (7), we obtain the following dynamics in the new coordinates:

$$\begin{cases} {}^n\dot{\mathbf{F}} = -{}^n\Omega \wedge {}^n\mathbf{F} - (\mathbf{V}' + {}^n\Omega \wedge \mathbf{R}'^c) \\ \dot{\mathbf{V}}' = -{}^n\Omega \wedge \mathbf{V}' + \mathbf{A}' + \mathbf{G}' \\ \dot{q}' = \frac{1}{2} q'^n \Omega_q \quad \dot{g} = \dot{A}_z^b = \dot{\gamma} = 0 \end{cases} \quad (8)$$

where \mathbf{G}' is obtained from the quaternion product $-gq'^* kq'$ and $\mathbf{A}' = R_\gamma^z \mathbf{A}$ (note that the rotation implemented by the matrix R_γ^z leaves unvaried the acceleration along the z -axis. We remark the following two properties:

- 1) In the new coordinates the dynamics of the first 15 state entries are independent of γ and also the observations are independent of γ ;
- 2) The system characterized by the first 15 entries of \mathbf{S}' is precisely the system analyzed in section V-B.

The first remark allows us to decompose the state into two parts. Specifically, the state that defines the former consists of the first 15 state entries while the latter only consists of γ . The first consequence of this decomposition is that the parameter γ is unobservable. This is not a new result since we already know from theorem 3 in [10] that, in the case of at least 5 features, the system is characterized by a rotation symmetry that corresponds to a rotation around the accelerometer axis. The second remark allows us to conclude that the remaining states are observable up to the yaw (property 3). By adding to the system an additional inertial sensor along an axis that does not coincide with the axis of the accelerometer, the symmetry that corresponds to a rotation around the z -axis is broken and also the parameter γ becomes observable. On the other hand, if the new inertial sensor shares the same axis of the first one, the symmetry is not broken and the parameter γ remains unobservable. We summarize this new result with the following theorem:

Theorem 2 *In the Vi-SfM problem with only two inertial sensors, the observability properties are the same as in the standard case provided that the two inertial sensors are along two distinct axes and with at least one of them that is an accelerometer. This holds even in the most challenging scenario, i.e., in the case of unknown camera-inertial sensor transformation, unknown magnitude of the gravity, unknown biases and single point feature available. In the case when the inertial sensors only consist of a single accelerometer, a new internal symmetry arises. As a consequence, the initial speed and orientation and the camera-inertial sensor transformation are not fully observable: all these quantities cannot be distinguished from the same quantities rotated around the accelerometer axis. All the remaining states are observable as in the standard Vi-SfM problem.*

VI. DISCUSSION AND CONCLUSION

This paper provided new theoretical results on the Vi-SfM problem. Specifically, the investigation aimed to discover how the observability properties change when the number of inertial sensors is reduced and the new results here provided extended our previous results presented in [10], [11]. With respect to the contributions in [10], [11], here we

considered simultaneously the case when the transformation between the camera frame and the frame attached to the inertial sensors is unknown and only a single point feature is available. To deal with this case we had to derive new general results on a theoretical problem that, in the framework of control theory, is known as Unknown Input Observability (UIO). Thanks to these results, we could derive the observability properties of the Vi-SfM in the case of a single feature available and unknown camera extrinsic calibration. Our results are summarized by theorem 2, which basically states that by using two inertial sensors along two independent axes and where at least one of them is an accelerometer, the observability properties do not change with respect to the standard Vi-SfM. In the case of a single accelerometer, exactly as in the case considered in [10], [11] (i.e., with at least 5 point features), it is not possible to distinguish all the physical quantities rotated around the accelerometer axis, independently of the accomplished trajectory. This means that, in this setting, it is not possible to fully perceive self-motion. Note that the case of a single point features is much more complex than the case when the number of features is at least five. In this latter case, the monocular vision alone allows us to build the structure of the environment up to a scale. This means that the angular speed is provided by the visual measurements and consequently, the fact that the gyroscopes can be removed without altering the observability properties is an obvious result. In the case of a single point feature, the monocular vision does not allow us to observe the angular speed and the fact that the gyroscopes can be removed without altering the observability properties is not obvious. We remark that all the results obtained so far, clearly show that the observability properties are independent of the number of features, provided that the inertial sensors include at least one accelerometer (see table I). From our derivations, it is also possible to conclude that they are independent of the knowledge of the bias and the magnitude of the gravity. In order to complete our investigation, we need to consider the case without accelerometers. Obviously, in this case we lose the absolute scale and the roll and pitch angles. On the other hand, it must be investigated how the number of features impacts the observability properties, depending on the gyroscopes that are available. We are currently investigating these cases.

Finally, we wish to emphasize that, this research on visual-inertial structure from motion with missing inertial inputs, pushed us to investigate the UIO problem. Currently, our results on UIO (available in [12]) provide an exhaustive answer to the case of a single unknown input. In this case, we introduced a simple algorithm to directly obtain the entire observable codistribution. As in the standard case of only known inputs (i.e., without unknown inputs), the observable codistribution is obtained by recursively computing the Lie derivatives along the vector fields that characterize the dynamics. However, in correspondence of the unknown input, the corresponding vector field must be suitably rescaled. Additionally, the Lie derivatives must be computed also along a new set of vector fields that are obtained by recursively

Inertial Sensors	Nb of points	Extrinsic Calibration	Observability Properties
Single accelerometer	≥ 1	yes	Full Observability
Single accelerometer	≥ 1	no	internal symmetry: rotation about the accelerometer-axis
Single accelerometer + 1 sensor along a different axis	≥ 1	no	Full Observability

TABLE I
OBSERVABILITY PROPERTIES IN SEVERAL SETTINGS

performing suitable Lie bracketing of the vector fields that define the dynamics. In practice, the entire observable codistribution is obtained by a very simple recursive algorithm. However, the analytic derivations required to prove that this codistribution fully characterizes the weak local observability of the state are very complex (theorem 1 in [12]). Finally, in [12] we showed that the recursive algorithm converges in a finite number of steps and we also provided the criterion to establish that the convergence has been reached. Also this proof is based on several tricky analytical steps (theorem 2 in [12]). We are currently extending these results to the case of multiple unknown inputs.

REFERENCES

- [1] M. Bryson and S. Sukkarieh, Observability Analysis and Active Control for Airborne SLAM, IEEE Transaction on Aerospace and Electronic Systems, vol. 44, no. 1, 261–280, 2008
- [2] Chao X. Guo and Stergios I. Roumeliotis, IMU-RGBD Camera 3D Pose Estimation and Extrinsic Calibration: Observability Analysis and Consistency Improvement, ICRA 2013, Karlsruhe, Germany.
- [3] Hermann R. and Krener A.J., 1977, Nonlinear Controllability and Observability, Transaction On Automatic Control, AC-22(5): 728-740
- [4] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, Towards consistent vision-aided inertial navigation, Int. Workshop on the Algorithmic Foundations of Robotics, Cambridge, MA, Jun, 2012.
- [5] E. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach", The International Journal of Robotics Research, vol. 30, no. 4, pp. 407–430, Apr. 2011.
- [6] J. Kelly and G. Sukhatme, Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration, Int. Journal of Robotics Research, vol. 30, no. 1, pp. 56–79, 2011.
- [7] D. G. Kottas, J. A. Hesch, S. L. Bowman, and S. I. Roumeliotis, On the consistency of vision-aided inertial navigation, in Proc. of the Int. Symposium on Experimental Robotics, Canada, Jun 2012.
- [8] M. Li and A. I. Mourikis, Improving the accuracy of EKF-based visualinertial odometry, ICRA , St. Paul, MN, May 2012.
- [9] A. Martinelli, Vision and IMU data fusion: closed-form solutions for attitude, speed, absolute scale and bias determination, IEEE Transactions on Robotics, Volume 28 (2012), Issue 1 (February), pp 44–60.
- [10] A. Martinelli, Visual-inertial structure from motion: observability vs minimum number of sensors, ICRA Hong Kong, June 2014
- [11] A. Martinelli, Observability Properties and Deterministic Algorithms in Visual-Inertial Structure from Motion, Found. and Trends in Robotics, Vol 3, 139–209, <http://dx.doi.org/10.1561/23000000030>, 2013
- [12] A. Martinelli, Nonlinear Unknown Input Observability: Extension of the Observability Rank Condition and the case of a single unknown input, INRIA Technical report, <http://hal.archives-ouvertes.fr/>
- [13] Mirzaei F.M. and Roumeliotis S.I., A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation, IEEE Transactions on Robotics, 2008, Vol. 24, October 2008
- [14] D. Strelow and S. Singh, Motion estimation from image and inertial measurements, International Journal of Robotics Research, 23(12), 2004
- [15] Weiss., S., Vision Based Navigation for Micro Helicopters, PhD thesis, Diss. ETH No. 20305