Visual-inertial structure from motion: observability vs minimum number of sensors

Agostino Martinelli

Visual-inertial structure from motion: observability vs minimum number of sensors

Agostino Martinelli

Abstract— **This paper analyzes the observability properties of the visual inertial structure from motion as the number of inertial sensors is reduced. Specifically, instead of considering the standard formulation where the inertial sensors are 3 orthogonal accelerometers and 3 orthogonal gyroscopes, the sensor system here considered only consists of a monocular camera and 1 or 2 accelerometers. This analysis has never been provided before. The main result achieved in this context is that the observability properties of visual inertial structure from motion do not change by removing all the 3 gyroscopes and 1 accelerometer. By removing a further accelerometer, if the camera is not extrinsically calibrated, the system loses part of its observability properties. On the other hand, if the camera is extrinsically calibrated, the system maintains the same observability properties as in the standard case. This contribution clearly shows that the information provided by a monocular camera, 3 accelerometers and 3 gyroscopes is redundant. Additionally, it provides a new perspective in the framework of neuroscience to the process of vestibular and visual integration for depth perception and self motion perception. Finally, to analyze these systems with a reduced number of inertial sensors, the paper introduces a new method to derive the observability properties of a non linear system when part of its input controls is unknown. This method is a further original paper contribution in control theory.**

I. Introduction

The visual-inertial structure from motion problem (from now on the Vi-SfM problem), has particular interest and has been investigated by many disciplines, both in the framework of computer science [3], [10], [11], [15], [17], [21] and in the framework of neuroscience (e.g., [2], [4], [5]). Vision 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. 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 [9], [10], [11], [12], [13], [15], [18] and [22] the observability properties of the standard Vi-SfM have been investigated in several different scenarios. Very recently, following two independent procedures, the most general result for the standard Vi-SfM problem has been provided in [7] and [16]. 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 this paper we will take a step forward and we will investigate the observability properties when the number of inertial sensors is reduced. We will prove 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, as in the case of three orthogonal accelerometers and gyroscopes. This contribution clearly shows that the information provided by a monocular camera and an IMU is redundant. Additionally, it provides a new perspective in the framework of neuroscience to the process of vestibular and visual integration for depth perception and self motion perception.

To analyze these systems with a reduced number of inertial sensors, the paper introduces a new method to derive the observability properties of a system when part of its input controls is unknown. This method is also an original contribution in the framework of control theory. It is an extension of the theory developed by Herman and Krener [8]. It is based on a suitable state extension. In particular, the extended state includes the unknown inputs together with their time derivatives up to a given order. Note that, this augmented state has already been considered in [1] where a sufficient condition for the state observability has been provided.

The paper is articulated as follows. The system and its basic equations are provided in section II. Section III reminds

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

the reader some basic concepts in non linear observability. In particular, it provides the main results introduced in [8] and [14]. Section IV contains the extension of the theory in [8]. Section V contains the new results about the observability properties when the number of inertial sensors is reduced. This includes the case of a single accelerometer. Finally, conclusions are provided in section VI.

II. The Considered System

We consider a system which consists of a monocular camera and inertial sensors. Specifically, we consider the following three cases:

1) The inertial sensors only consist of a single accelerometer and the camera is extrinsically calibrated (i.e., the transformation between the camera frame and the inertial sensor frame is known);
2) The inertial sensors only consist of two accelerometers and the camera is not extrinsically calibrated;
3) The inertial sensors only consist of one accelerometer and the camera is not extrinsically calibrated;

For the sake of simplicity, in the last two cases we do not consider the extreme case of a single feature. In particular, we assume that the camera is able to provide its position, orientation and speed up to a scale. This is obtained by assuming that the camera is observing at least five point features, simultaneously [19]. This significantly reduces the computational load.

Regarding the first case, we consider the extreme scenario of a single point feature. Without loss of generality, we assume that the camera and the inertial sensor frame have the same orientation and that the accelerometer points towards the $z-$direction of the camera frame. Finally, we denote by $\boldsymbol{R^c}$ the known position of the camera optical center in the inertial sensor frame. The state that characterizes this system is the following $12-$dimensional vector:

$$\boldsymbol{X_1^c} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F},\ \boldsymbol{V},\ q,\ A_z^{bias},\ g\right]^T \tag{1}$$

where we adopt the subscript 1 to denote the case of a single accelerometer and the apex c to denote the case of a calibrated camera. ${}^{\boldsymbol{c}}\boldsymbol{F}$ is the position of the feature in the camera frame, $\boldsymbol{V}$ is the speed of the inertial sensor in its frame, q the unit quaternion which describes the orientation of the camera frame in the global frame, A_z^{bias} the bias of the accelerometer and g the magnitude of the gravity. The dynamics of this state are:

$$\begin{bmatrix} {}^{\boldsymbol{c}}\dot{\boldsymbol{F}} & = M(\boldsymbol{\Omega}){}^{\boldsymbol{c}}\boldsymbol{F} - (\boldsymbol{V} + \boldsymbol{\Omega}\wedge\boldsymbol{R^c}) \\ \dot{\boldsymbol{V}} & = M(\boldsymbol{\Omega})\boldsymbol{V} + \boldsymbol{A} - \boldsymbol{A^{bias}} + \boldsymbol{G} \\ \dot{q} & = \frac{1}{2}q\Omega_q \\ \dot{g} & = \dot{A}_z^{bias} = 0 \end{bmatrix} \tag{2}$$

where $\boldsymbol{\Omega} \equiv [\Omega_x\ \Omega_y\ \Omega_z]$ is the unknown angular speed of the camera, $M(\boldsymbol{\Omega}) \equiv \begin{bmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{bmatrix}$, Ω_q is the quaternion associated with $\boldsymbol{\Omega}$, i.e., $\Omega_q \equiv \Omega_x i + \Omega_y j + \Omega_z k$, $\boldsymbol{A^{bias}} = [0\ 0\ A_z^{bias}]^T$ and $\boldsymbol{A} - \boldsymbol{A^{bias}}$ is the camera acceleration in the local frame, whose first two components are unknown and the third component is known up to the bias thanks to the accelerometer.

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

$$\boldsymbol{h_{cam}}(\boldsymbol{X_1^c}) \equiv [h_u,\ h_v]^T = \left[\frac{F_x}{F_z},\ \frac{F_y}{F_z}\right]^T \tag{3}$$

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

$$\boldsymbol{h_{const}}(\boldsymbol{X_1^c}) \equiv h_q = q^*q \tag{4}$$

Regarding the second and the third case, as previously mentioned, we assume that the camera is observing at least five features. This allows us to consider the camera as a sensor able to provide its orientation, its angular speed and its position and speed up to a scale in a global reference frame attached to these features. Obviously, the gravity in this global frame is not necessarily along the vertical axis. We denote this vector by $\boldsymbol{g} \equiv [g_x,\ g_y,\ g_z]^T$, which is unknown (both in magnitude and direction). Additionally, we denote by $\frac{1}{\mu}$ the unknown absolute scale.

For the second case, without loss of generality, we assume that the two available accelerometers are along the y and the $z-$axis in the inertial sensor frame. Let us denote with N_f the number of observed features ($N_f \geq 5$). The state is:

$$\boldsymbol{X_2^u} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F^1}, \cdots, {}^{\boldsymbol{c}}\boldsymbol{F^{N_f}}, \boldsymbol{V}, q, A_y^{bias}, A_z^{bias}, \boldsymbol{R^c}, q^c, \boldsymbol{g}, \mu\right]^T \tag{5}$$

where the subscript 2 denotes two accelerometers and the apex u the fact that we are considering the case of a camera extrinsically uncalibrated. ${}^{\boldsymbol{c}}\boldsymbol{F^i}$ ($i = 1, \cdots, N_f$) is the position of the i^{th} feature in the camera frame, $\boldsymbol{V}$ is the speed in the inertial sensor frame and the unit-quaternion q characterizes the attitude of the inertial sensor frame in the global frame. Finally, the unit-quaternion q^c characterizes the attitude of the camera frame in the inertial sensor frame and the vector $\boldsymbol{R^c}$ is included in the state since it is now unknown. We will also denote with R_{q^c} the rotation matrix associated with the unit-quaternion q^c. The dynamics of this state are ($i = 1, \cdots, N_f$):

$$\begin{bmatrix} {}^{\boldsymbol{c}}\dot{\boldsymbol{F}}^{\boldsymbol{i}} & = M({}^{\boldsymbol{c}}\boldsymbol{\Omega}){}^{\boldsymbol{c}}\boldsymbol{F^i} - R_{q^c}[\boldsymbol{V} + \boldsymbol{\Omega}\wedge\boldsymbol{R^c}] \\ \dot{\boldsymbol{V}} & = M(\boldsymbol{\Omega})\boldsymbol{V} + \boldsymbol{A} - \boldsymbol{A^{bias}} + \boldsymbol{G} \\ \dot{q} & = \frac{1}{2}q\Omega_q \quad \dot{\boldsymbol{g}} = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{A}_y^{bias} & = \dot{A}_z^{bias} = \dot{q}^c = \dot{\mu} = 0 \end{bmatrix} \tag{6}$$

where $\boldsymbol{\Omega}$ is the unknown angular speed in the inertial sensor frame, $\boldsymbol{A}$ is the acceleration in the inertial sensor frame, whose first component is unknown and $\boldsymbol{A^{bias}} = [0,\ A_y^{bias},\ A_z^{bias}]^T$. Since at least five point features are available, the angular speed is known in the camera frame, i.e., the vector ${}^{c}\boldsymbol{\Omega}$ $(= R_{q^c}\,\boldsymbol{\Omega})$ can be obtained from the visual measurements [19]. Additionally, the visual measurements provide $3N_f + 9$ scalar functions (system outputs). The first $3N_f$ are the components of the vectors $\mu\ {}^{c}\boldsymbol{F}^{\boldsymbol{i}}$, $i = 1, \cdots, N_f$, which are the positions of the features in the camera frame up to the scale. Hence, we have:

$$h_{F_x^i} \equiv \mu\ {}^{c}F_x^i;\ \ h_{F_y^i} \equiv \mu\ {}^{c}F_y^i;\ \ h_{F_z^i} \equiv \mu\ {}^{c}F_z^i;\ \ i = 1, \cdots, N_f \tag{7}$$

Regarding the remaining nine outputs, three of them are the components of the speed in the camera frame up to a scale, i.e.:

$$h_{V_x} \equiv \mu\ {}^{c}V_x;\quad h_{V_y} \equiv \mu\ {}^{c}V_y;\quad h_{V_z} \equiv \mu\ {}^{c}V_z \tag{8}$$

where ${}^{c}\boldsymbol{V} \equiv [{}^{c}V_x,\ {}^{c}V_y,\ {}^{c}V_z]^T = R_{q^c}[\boldsymbol{V} + \boldsymbol{\Omega} \wedge \boldsymbol{R^c}]$. Since the camera provides its orientation in the global frame, also the component of the quaternion qq^c can be considered system outputs. We have:

$$h_t \equiv (qq^c)_t;\quad h_x \equiv (qq^c)_x;\quad h_y \equiv (qq^c)_y;\quad h_z \equiv (qq^c)_z \tag{9}$$

Finally, both the quaternions q and q^c must be unit quaternions. Hence, we have the two outputs:

$$h_q \equiv q_t^2 + q_x^2 + q_y^2 + q_z^2;\ \ h_{q^c} \equiv (q_t^c)^2 + (q_x^c)^2 + (q_y^c)^2 + (q_z^c)^2; \tag{10}$$

Regarding the third case, the system is described by the following state:

$$\boldsymbol{X_1^u} \equiv \left[{}^{c}\boldsymbol{F^1},\ \cdots,\ {}^{c}\boldsymbol{F^{N_f}},\ \boldsymbol{V},\ q,\ A_z^{bias},\ \boldsymbol{R^c}, q^c,\ \boldsymbol{g},\ \mu\right]^T \tag{11}$$

where we assumed, without loss of generality, that the available accelerometer is along the $z-$axis in the inertial sensor frame.

III. Observability analysis

In the systems defined in the previous section part of the inputs is unknown. For this reason, an observability analysis cannot be performed by using standard methods. In this section we remind the reader some basic concepts in the theory of observability for non linear systems. For the sake of clarity, we will refer to a simple example. This will allow us to better illustrate these concepts and, in section IV-A, to introduce new concepts in order to deal with non linear systems when part (or even all) of the input controls is unknown.

A. A simple 2D localization problem

We consider a vehicle moving in a $2D$-environment. The configuration of the vehicle in a global reference frame, can be characterized through the vector $[x_v,\ y_v,\ \theta_v]^T$ where x_v and y_v are the cartesian vehicle coordinates, and θ_v is the vehicle orientation. We assume that the dynamics of this vector satisfy the unicycle differential equations:

$$\begin{cases} \dot{x}_v = v\cos\theta_v \\ \dot{y}_v = v\sin\theta_v \\ \dot{\theta}_v = \omega \end{cases} \tag{12}$$

where v and ω are the linear and the rotational vehicle speed, respectively, and they are the system input controls. We assume that the vehicle is equipped with a GPS able to provide its position, i.e.:

$$\boldsymbol{z} = [x_v,\ y_v]^T \tag{13}$$

Our system is characterized by the previous two equations: the former describes its dynamics, the latter its observations. As the majority of real control systems, the dynamics given in (12) are affine in the controls, i.e. they can be written as follows:

$$\left\{ \dot{\boldsymbol{S}} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{u}) = \boldsymbol{f_0}(\boldsymbol{S}) + \sum_{i=1}^{n_c} \boldsymbol{f_i}(\boldsymbol{S}) u_i \right. \tag{14}$$

with $\boldsymbol{S} = [x_v,\ y_v,\ \theta_v]^T$, $n_c = 2$, $\boldsymbol{u} = [u_1,\ u_2]^T = [v,\ \omega]^T$, $\boldsymbol{f_0}(\boldsymbol{S}) = [0, 0, 0]^T$, $\boldsymbol{f_1}(\boldsymbol{S}) = [\cos\theta_v,\ \sin\theta_v,\ 0]^T$, $\boldsymbol{f_2}(\boldsymbol{S}) = [0,\ 0,\ 1]^T$. Additionally, our system is characterized by two observation functions (system outputs), which are $h_x(\boldsymbol{S}) = x_v$ and $h_y(\boldsymbol{S}) = y_v$.

B. Observability rank criterion

This criterion was introduced in 1977 by Hermann and Krener [8], in order to investigate the observability properties of a non linear system which satisfies (14) and with one or more outputs (observations). It requires to compute the Lie derivatives of all the observation functions with respect to all the vector fields $\boldsymbol{f_0}(\boldsymbol{S}),\ \boldsymbol{f_1}(\boldsymbol{S}),\ \cdots,\ \boldsymbol{f_{n_c}}(\boldsymbol{S})$. The Lie derivatives are defined recursively. The zero order Lie derivative of a given observation function $h(\boldsymbol{S})$ is the function itself, i.e., $\mathcal{L}^0 h \equiv h$. Then, the $(k+1)$-order Lie derivative of the observation function $h(\boldsymbol{S})$ with respect to $\boldsymbol{f_{i_1}}(\boldsymbol{S}),\ \cdots,\ \boldsymbol{f_{i_k}}(\boldsymbol{S}),\ \boldsymbol{f_{i_{k+1}}}(\boldsymbol{S})$ (with $i_1,\ i_2,\ \cdots,\ i_{k+1} = 0, 1, \cdots, n_c$) is $\mathcal{L}^{k+1}_{i_1,\cdots,i_k,i_{k+1}} h \equiv \boldsymbol{\nabla_S} \mathcal{L}^k_{i_1,\cdots,i_k} h \cdot \boldsymbol{f_{i_{k+1}}}(\boldsymbol{S})$. Note that this operation is not commutative with respect to the indexes's order. Hence, for a given observation function, we have $(n_c+1)^k$ $k-$order Lie derivatives.

Let us denote with $\mathcal{V}$ the space of all the Lie derivatives up to the $k-$order and with $\nabla\mathcal{V}$ the vector space spanned by the gradients of these functions. In this notation, the observability rank criterion can be expressed in the following way: *The dimension of the largest observable sub-system at a given $\boldsymbol{S}$ is equal to the dimension of* $\nabla\mathcal{V}$[1]. As a consequence, if for a given $k-$order the dimension of $\nabla\mathcal{V}$ is equal to the dimension of $\boldsymbol{S}$, the state $\boldsymbol{S}$ is observable and it is not necessary to compute higher order Lie derivatives.

Let us apply the observability rank criterion to our example. The system has two outputs: $h_x \equiv x_v$ and $h_y \equiv y_v$. By definition, they coincide with their zero-order Lie derivatives.

[1]Actually, this condition guarantees that the system is *locally weakly* observable. The reader is addressed to [8] for a detailed description of *weak* and *local* observability.

Their gradients with respect to the state $\boldsymbol{S}$ are, respectively: $[1,\ 0,\ 0]^T$ and $[0,\ 1,\ 0]^T$. Hence, the space spanned by the zero-order Lie derivatives has dimension two. In particular, by considering only the zero-order Lie derivatives, we can only conclude that the first 2 state components are observable. We do not know whether the third component, i.e., the vehicle orientation, is observable or not. Let us compute the first order Lie derivatives. We obtain: $\mathcal{L}^1_1 h_x = \cos\theta$, $\mathcal{L}^1_1 h_y = \sin\theta$, $\mathcal{L}^1_2 h_x = \mathcal{L}^1_2 h_y = 0$. Hence, the space spanned by the Lie derivatives up to the first order span the entire configuration space and we conclude that also the vehicle orientation is observable.

IV. Unknown Input Observability

This section introduces a new method to derive the observability properties of a non linear system when all or part of its input controls is unknown. As it is common, we will refer to non linear systems whose dynamics are affine in the controls, i.e., they can be written as in the equation (14). Additionally, we will refer to the case when the observation (system output) is a scalar function of the state, i.e., $z = h(\boldsymbol{S})$. The theory of the observability is based on the assumption that both the system inputs (i.e., u_i in (14), $i = 1, \cdots, n_c$) and the system output (i.e., z) are known during a given time interval. This is a basic assumption. Specifically, the *observability rank criterion* introduced in [8] and used in [14] to define the concept of continuous symmetry, is based on this assumption. In order to extend the observability rank criterion, let us focus on the main steps in the theory introduced in [8]. Let us denote with $[T_{in},\ T_{fin}]$ the interval of time where the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ and $z(t) = h(\boldsymbol{S}(t))$ are known.

The observability rank criterion is obtained by proceeding with the following three steps:

1) The Taylor's theorem is used to obtain the value of any order time derivative for $t = T_{in}$ of the functions u_i $(i = 1, \cdots, n_c)$ and z, starting from the knowledge of the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ and $z(t)$ for $t \in [T_{in},\ T_{fin}]$;
2) the values of all the Lie derivatives of the function $h(\boldsymbol{S})$ in $\boldsymbol{S}(T_{in})$ along all the directions $\boldsymbol{f_i}(\boldsymbol{S})$, $i = 0, 1, \cdots, n_c$ are obtained by inverting a linear system whose coefficients and constant terms are given by the previous time derivatives (see equation (16));
3) the inverse function theorem [20] allows us to identify if the vector $\boldsymbol{S}(T_{in})$ can be obtained starting from the knowledge of the Lie derivatives (which are non linear scalar functions of this vector).

The goal of this section is to extend the observability rank criterion in order to deal with the case when all or part of the n_c input controls of our system are unknown. In this case, we do not know some of the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ $t \in [T_{in},\ T_{fin}]$. Let us denote with n_k and n_u the number of known and unknown input controls, respectively. We have $n_k + n_u = n_c$. Additionally, we order the inputs such that the known input are the first n_k. In other words, $u_i(t)$ $(i = 1, \cdots, n_k)$ are known for any $t \in [T_{in},\ T_{fin}]$, while $u_i(t)$ $(i = n_k + 1, \cdots, n_c)$ are unknown. Hence, the time derivatives of $u_i(t)$ $(i = n_k+1, \cdots, n_c)$ are not available and the second step mentioned above cannot be used to obtain the Lie derivatives. Our basic idea consists in modifying the original state in order to be able to select some Lie derivatives, which can be obtained even without knowing all the time derivatives of $u_i(t)$ $(i = n_k + 1, \cdots, n_c)$ at $t = T_{in}$. This allows obtaining sufficient conditions for the state observability. We will the new criterion the *extended observability rank criterion*. It will be introduced in IV-A. Then, to better illustrate the proposed method, we consider again the localization problem discussed in III-A.

A. Extended observability rank criterion

Let us refer to the non linear system described by equation (14) and a given observation function $z = h(\boldsymbol{S})$. It is possible to analytically derive the expression of the n^{th} time derivative of the observation function in $t = T_{in}$ in terms of all the Lie derivatives of the function h along all the directions $\boldsymbol{f_i}(\boldsymbol{S})$, $i = 0, 1, \cdots, n_c$ computed in $\boldsymbol{S_0} \equiv \boldsymbol{S}(T_{in})$ up to the n^{th} order and all the time derivatives of the functions u_i $(i = 1, \cdots, n_c)$ computed in $t = T_{in}$. We have:

$$\left.\frac{d^n h(\boldsymbol{S}(t))}{dt^n}\right|_{t=T_{in}} = \sum_{p=1}^{n} \sum_{i_1 i_2 \cdots i_p = 0}^{n_c} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) \tag{15}$$
$$\sum_{k_1, k_2, \cdots, k_p = 0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1, k_2, \cdots, k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)}$$

where:

- $u_i^{(k)} \equiv \frac{d^k u_i}{dt^k}$, $k = 0, 1, \cdots, n$; $i = 1, \cdots, n_c$;
- $u_0 \equiv 1$ and $u_0^{(k)} = 0$, $k > 1$;
- $C^{n,\ p}_{k_1, k_2, \cdots, k_p}$, are real numbers satisfying a recursive equation which can be obtained by directly differentiating the expression in (15) with respect to time.

The expression in (15) allows us to perform the second step mentioned above, i.e., it allows us to obtain the Lie derivatives of h starting from the knowledge of the time derivatives of the system inputs and output by inverting a linear system. When n_u inputs are unknown, this step cannot be performed starting directly from (15). We split the expression in (15) as follows:

$$\left.\frac{d^n h(\boldsymbol{S}(t))}{dt^n}\right|_{t=T_{in}} = \sum_{p=1}^{n} \left\{ \sum_{i_1 i_2 \cdots i_p = 0}^{n_k} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) \right. \tag{16}$$
$$\sum_{k_1, k_2, \cdots, k_p = 0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1, k_2, \cdots, k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)} +$$
$$+ \sum_{i_1 i_2 \cdots i_p = remaining} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0})$$
$$\left. \sum_{k_1, k_2, \cdots, k_p = 0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1, k_2, \cdots, k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)} \right\}$$

The first sum only contains the know controls (i.e., $i_1, i_2, \cdots, i_p = 0, 1, \cdots n_k$) while, for each addend in the second sum, i.e., the sum where the indexes $i_1, i_2, \cdots, i_p$ take the *remaining* values, at least one control is unknown. In the special case when all the Lie derivatives $\mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0})$ vanish when at least one index $i_1, i_2, \cdots, i_p$ is larger than n_k, the second sum, which also contains unknown controls, vanishes as well. Hence, the expression in (16) can still be used to obtain all the Lie derivatives and the observability rank criterion can still be adopted. Obviously, this is a very special case. Our idea is to extend the original state in order to artificially reproduce such a situation. In particular, we include the unknown inputs in the state together with their time derivatives. By including the time derivatives up to the $(n-1)^{th}$ order, we will obtain $\mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) = 0$ when at least one index $i_1, i_2, \cdots, i_p$ is larger than n_k and for all $p = 0, 1, \cdots, n$. Let us illustrate this by referring to the case when $n_k = n_c - 1$ (and, consequently, $n_u = 1$). Obviously, the zero order Lie derivative can be obtained without the necessity to know the inputs (it is trivially the output at $t = T_{in}$, $\mathcal{L}^0 h(\boldsymbol{S_0}) = h(\boldsymbol{S}(T_{in}))$). Let us consider the first order time derivative ($n = 1$). The expression in (16) becomes:

$$\left.\frac{dh(\boldsymbol{S}(t))}{dt}\right|_{t=T_{in}} = \sum_{i=0}^{n_c-1} \mathcal{L}^1_i h(\boldsymbol{S_0}) u_i + \mathcal{L}^1_{n_c} h(\boldsymbol{S_0}) u_{n_c} \quad (17)$$

Let us include the unknown u_{n_c} in the state, i.e., $\boldsymbol{S} \rightarrow \boldsymbol{S^e} \equiv [\boldsymbol{S},\ u_{n_c}]^T$. We have:

$$\left\{ \dot{\boldsymbol{S^e}} = \boldsymbol{f^e_0}(\boldsymbol{S^e}) + \sum_{i=1}^{n_c} \boldsymbol{f^e_i}(\boldsymbol{S^e}) u^e_i \right. \quad (18)$$

where

- $\boldsymbol{f^e_0}(\boldsymbol{S^e}) \equiv [\boldsymbol{f_0}(\boldsymbol{S})^T + \boldsymbol{f_{n_c}}(\boldsymbol{S})^T u_{n_c},\ 0]^T$;
- $\boldsymbol{f^e_i}(\boldsymbol{S^e}) \equiv [\boldsymbol{f_i}(\boldsymbol{S})^T,\ 0]^T$, $i = 1, \cdots, n_c - 1$;
- $\boldsymbol{f^e_{n_c}}(\boldsymbol{S^e}) \equiv [\boldsymbol{0},\ 1]^T$, with $\boldsymbol{0}$ the line vector whose entries are all zero and whose dimension is equal to the one of $\boldsymbol{S}^T$;
- $u^e_i \equiv u_i$, $i = 1, \cdots, n_c - 1$;
- $u^e_{n_c} \equiv u^{(1)}_{n_c} = \dot{u}_{n_c}$.

It is immediate to realize that the first order Lie derivative along the direction $\boldsymbol{f^e_{n_c}}(\boldsymbol{S^e})$ is identically zero. This allows us to obtain all the other first order Lie derivatives. By including in the state also the first time derivative of u_{n_c} (namely, $u^{(1)}_{n_c}$) we can obtain all the Lie derivatives, up to the second order, along the first $n_c - 1$ directions. By including higher order time derivatives of the unknown input control u_{n_c} we can obtain higher order Lie derivatives along the first $n_c - 1$ directions. At this point, the third step in the Herman & Krener theory previously mentioned can be performed by using the Lie derivatives which are available. By analyzing these Lie derivatives it is possible to detect potential symmetries according to the theory developed in [14] and for a given mode to be observable. In particular, we will use the following property, which is a sufficient condition for a scalar function to be an observable mode:

Property 1 *The function $m(\boldsymbol{S})$ is observable if its gradient is spanned by the gradients of a set of Lie derivatives*

B. Observability of the system in III-A with unknown inputs

We illustrate the method introduced in the previous section by deriving the observability properties of the simple system introduced in III-A when part, or all, of the input controls is unknown. We already know that the state $\boldsymbol{S} = [x_v,\ y_v,\ \theta_v]^T$ is observable when all the input controls are known (i.e., when the functions $v(t)$ and $\omega(t)$ are known for any $t \in [T_{in},\ T_{fin}]$). Intuitively, we know that the knowledge of both the inputs is unnecessary in order to have the full observability of the entire state. Indeed, the first two state components can be directly obtained from the GPS. By knowing these two components in a given time interval, we also know their time derivatives. In particular, we known $\dot{x}_v(T_{in})$ and $\dot{y}_v(T_{in})$. From (12) we easily obtain: $\theta_v(T_{in}) = atan\left(\frac{\dot{y}_v(T_{in})}{\dot{x}_v(T_{in})}\right)$. Hence, also the initial orientation is observable, by only using the GPS measurements. By applying the method introduced in the previous section we obtain exactly the same result. We start by including in the original state the unknown v, i.e., $\boldsymbol{S^e} = [x_v,\ y_v,\ \theta_v,\ v]^T$. We obtain:

$$\begin{bmatrix} \dot{x}_v = v\cos\theta_v \\ \dot{y}_v = v\sin\theta_v \\ \dot{\theta}_v = \omega \\ \dot{v} = v^{(1)} \end{bmatrix} \quad (19)$$

We have: $n_c = 2$, $\boldsymbol{u} = [u^e_1,\ u^e_2]^T = [\omega,\ v^{(1)}]^T$, $\boldsymbol{f^e_0}(\boldsymbol{S^e}) = [v\cos\theta_v,\ \ v\sin\theta_v,\ \ 0,\ \ 0]^T$, $\boldsymbol{f^e_1}(\boldsymbol{S^e}) = [0,\ 0,\ 1,\ 0]^T$, $\boldsymbol{f^e_2}(\boldsymbol{S^e}) = [0,\ 0,\ 0,\ 1]^T$. The first order Lie derivatives are: $\mathcal{L}^1_0 h_x(\boldsymbol{S^e}) = v\cos\theta_v$, $\mathcal{L}^1_1 h_x(\boldsymbol{S^e}) = \mathcal{L}^1_2 h_x(\boldsymbol{S^e}) = 0$. By chance, also $\mathcal{L}^1_1 h_x(\boldsymbol{S^e}) = 0$ and we do not need to include also ω in the state. By using (16) up to the first order (i.e., $n \leq 1$), we can determine $\mathcal{L}^1_0 h_x(\boldsymbol{S^e})$. In other words, we can determine $v\cos\theta_v$. By considering the second observation function (i.e., h_y) we find that we can also determine $\mathcal{L}^1_0 h_y(\boldsymbol{S^e}) = v\sin\theta_v$. The gradients of the functions $\mathcal{L}^0 h_x(\boldsymbol{S^e})$, $\mathcal{L}^0 h_y(\boldsymbol{S^e})$, $\mathcal{L}^1_0 h_x(\boldsymbol{S^e})$ and $\mathcal{L}^1_0 h_y(\boldsymbol{S^e})$ span the entire configuration space of the state $\boldsymbol{S^e}$ meaning that this extended state is observable.

V. Observability of Vi-SfM with unknown inputs

We use the method described in section IV to analyze the Vi-SfM problem when the number of accelerometers and gyroscopes is reduced. In other words, we analyze the three systems defined in section II.

A. Single Accelerometer and Camera extrinsically calibrated

The dynamics in (2) provide seven independent directions along with the Lie derivatives can be computed. On the other hand, only two directions are available. They are the vector $\boldsymbol{f_0}(\boldsymbol{X^c_1})$, which is obtained by setting $\boldsymbol{\Omega} = \boldsymbol{A} = [0\ 0\ 0]^T$ in (2) and the vector $\boldsymbol{f_3}(\boldsymbol{X^c_1})$, which is obtained by setting $\boldsymbol{\Omega} = [0\ 0\ 0]^T$ and $\boldsymbol{A} = [0\ 0\ 1]^T$ in the dynamics in (2), once $\boldsymbol{f_0}(\boldsymbol{X^c_1})$ has been removed. Since the Lie derivatives along the other five directions are not null, we have to proceed as in section IV-B. We must proceed in several subsequent steps. In each step we check, first of all, which highest order of Lie derivatives of the observations can be used. This is obtained by checking that, for a given order, all the Lie derivatives up to this order, computed along at least one of the directions which are not available (i.e., $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, $\boldsymbol{f_4}$, $\boldsymbol{f_5}$ and $\boldsymbol{f_6}$) are identically zero. Once this highest order is identified, we find the largest number of independent Lie derivatives up to this order. Then, we compute the set of all of vectors which are orthogonal to the gradients of these Lie derivatives. Finally, we apply the property 1 in order to detect which components of the vector in (1) are observable. Specifically, we compute the gradient of each state component and we check if it is orthogonal to all the previous vectors (in which case it means that the gradient of this component is spanned by the gradients of the considered Lie derivatives). We include new time derivatives of the unknown inputs (i.e., $A_x,\ A_y,\ \Omega_x,\ \Omega_y,\ \Omega_z$) in order to make usable higher order Lie derivatives, as explained in section IV.

1) First step: We start with the 12−dimensional state given by the vector in (1). Since the first order Lie derivatives along $\boldsymbol{f_4}$, $\boldsymbol{f_5}$ and $\boldsymbol{f_6}$ are different from zero both for h_u and h_v, we can only use zero-order Lie derivatives. On the other hand, the first order Lie derivatives along $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ (and also $\boldsymbol{f_3}$) vanish. Hence, it suffices to include $\Omega_x,\ \Omega_y,\ \Omega_z$ in the state in order to use the Lie derivatives up to the first order. The zero-order Lie derivatives are the three functions in (3, 4). They are independent. Hence the system has three observable modes. The set of vectors orthogonal to the gradients of these three functions can be determined. Property 1 does not allow us to prove the observability for any component of the state in (1).

2) Second step: We include $\Omega_x,\ \Omega_y,\ \Omega_z$ in the state. The dimension of the new state is 15. We can use all the Lie derivatives up to the first order. We detect the additional independent functions from them: $\mathcal{L}^1_0 h_u$ and $\mathcal{L}^1_0 h_v$. Hence the system has five independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 10 vectors. Again, property 1 does not allow us to prove the observability for any component of the state in (1).

3) Third step: In order to use the second order Lie derivatives we need to include $A_x,\ A_y,\ \Omega_x^{(1)} \equiv \dot{\Omega}_x,\ \Omega_y^{(1)} \equiv \dot{\Omega}_y,\ \Omega_z^{(1)} \equiv \dot{\Omega}_z$ in the state. The dimension of the new state is 20. We can use all the Lie derivatives up to the second order. We detect the following additional independent functions: $\mathcal{L}^2_{00} h_u$, $\mathcal{L}^2_{03} h_u$ and $\mathcal{L}^2_{00} h_v$. Hence the system has eight independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 12 vectors. This time, property 1 allows us to conclude that the first 3 components of the vector in (1) (i.e., ${}^{\boldsymbol{c}}\boldsymbol{F}$) are observable.

4) Forth step: In order to use the third order Lie derivatives we need to include also $A_x^{(1)} \equiv \dot{A}_x,\ A_y^{(1)} \equiv \dot{A}_y,\ \Omega_x^{(2)} \equiv \dot{\Omega}_x^{(1)},\ \Omega_y^{(2)} \equiv \dot{\Omega}_y^{(1)},\ \Omega_z^{(2)} \equiv \dot{\Omega}_z^{(1)}$ in the state. The dimension of the new state is 25. We detect the following additional independent functions: $\mathcal{L}^3_{000} h_u$, $\mathcal{L}^3_{003} h_u$, $\mathcal{L}^3_{030} h_u$, $\mathcal{L}^3_{000} h_v$ and $\mathcal{L}^3_{003} h_v$. Hence the system has 13 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 12 vectors. Property 1 allows us to conclude that also the sixth component of the vector in (1) (i.e., V_z) is observable.

5) Fifth step: In order to use the fourth order Lie derivatives we need to include also $A_x^{(2)} \equiv \dot{A}_x^{(1)},\ A_y^{(2)} \equiv \dot{A}_y^{(1)},\ \Omega_x^{(3)} \equiv \dot{\Omega}_x^{(2)},\ \Omega_y^{(3)} \equiv \dot{\Omega}_y^{(2)},\ \Omega_z^{(3)} \equiv \dot{\Omega}_z^{(2)}$ in the state. The dimension of the new state is 30. We detect the following additional independent functions: $\mathcal{L}^4_{0000} h_u$, $\mathcal{L}^4_{0003} h_u$, $\mathcal{L}^4_{0030} h_u$, $\mathcal{L}^4_{0300} h_u$, $\mathcal{L}^4_{0000} h_v$ and $\mathcal{L}^4_{0003} h_v$. Hence the system has 19 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 11 vectors. Property 1 allows us to conclude that the first six components of the vector in (1) (i.e., both the vector ${}^{\boldsymbol{c}}\boldsymbol{F}$ and $\boldsymbol{V}$) are observable.

6) Sixth step: In order to use the fifth order Lie derivatives we need to include also $A_x^{(3)} \equiv \dot{A}_x^{(2)},\ A_y^{(3)} \equiv \dot{A}_y^{(2)},\ \Omega_x^{(4)} \equiv \dot{\Omega}_x^{(3)},\ \Omega_y^{(4)} \equiv \dot{\Omega}_y^{(3)},\ \Omega_z^{(4)} \equiv \dot{\Omega}_z^{(3)}$ in the state. The dimension of the new state is 35. We detect the following additional independent functions: $\mathcal{L}^5_{00000} h_u$, $\mathcal{L}^5_{00003} h_u$, $\mathcal{L}^5_{00030} h_u$, $\mathcal{L}^5_{03000} h_u$, $\mathcal{L}^5_{00000} h_v$ and $\mathcal{L}^5_{00003} h_v$. Hence the system has 25 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 10 vectors. By using property 1 we find the same properties obtained in the previous step.

7) Seventh step: In order to use the sixth order Lie derivatives we need to include also $A_x^{(4)} \equiv \dot{A}_x^{(3)},\ A_y^{(4)} \equiv \dot{A}_y^{(3)},\ \Omega_x^{(5)} \equiv \dot{\Omega}_x^{(4)},\ \Omega_y^{(5)} \equiv \dot{\Omega}_y^{(4)},\ \Omega_z^{(5)} \equiv \dot{\Omega}_z^{(4)}$ in the state. The dimension of the new state is 40. We detect the following additional independent functions: $\mathcal{L}^6_{000000} h_u$, $\mathcal{L}^6_{000003} h_u$, $\mathcal{L}^6_{000030} h_u$, $\mathcal{L}^6_{030000} h_u$, $\mathcal{L}^6_{000000} h_v$ and $\mathcal{L}^6_{000003} h_v$. Hence the system has 31 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 9 vectors. Again, by using property 1, we find the same properties obtained in the previous step.

8) Eighth step: In order to use the seventh order Lie derivatives we need to include also $A_x^{(5)} \equiv \dot{A}_x^{(4)},\ A_y^{(5)} \equiv \dot{A}_y^{(4)},\ \Omega_x^{(6)} \equiv \dot{\Omega}_x^{(5)},\ \Omega_y^{(6)} \equiv \dot{\Omega}_y^{(5)},\ \Omega_z^{(6)} \equiv \dot{\Omega}_z^{(5)}$ in the state. The dimension of the new state is 45. We detect the following additional independent functions: $\mathcal{L}^7_{0000000} h_u$, $\mathcal{L}^7_{0000003} h_u$, $\mathcal{L}^7_{0000030} h_u$, $\mathcal{L}^7_{0300000} h_u$, $\mathcal{L}^7_{0000000} h_v$ and $\mathcal{L}^7_{0000003} h_v$. Hence the system has 37 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 8 vectors. By using property 1, we find that the first six components of the vector in (1) (i.e., both the vector ${}^{\boldsymbol{c}}\boldsymbol{F}$ and $\boldsymbol{V}$) and the last two components of this vector (i.e., the accelerometer bias A_z^{bias} and the magnitude of the gravity g) are observable. Additionally, also the roll and pitch are observable. The unique unobservable mode is the yaw. Since this unobservable mode is a consequence of the system invariance with respect to rotations about the vertical axis, it is useless to include higher order Lie derivatives: the observability properties of the state in (1) would not change. We summarize this section with the following theorem:

Theorem 1 (Single accelerometer, calibrated camera) *In the Vi-SfM problem with a single accelerometer, no* *gyroscope and known camera-inertial sensor transformation, all the independent observable modes arethe same as in the standard Vi-SfM problem. This holds even in the extreme case of a single point feature.*

B. Two Accelerometers and Uncalibrated Camera

We start by investigating the observability properties of a simplified system, which is obtained by referring to the state in (5) with $N_f = 1$ and with the $3 + 9 = 12$ outputs given by (7) for a single feature and (8-10). Note that we are using the four observations in (9): this implicitly assumes that we are actually exploiting the camera observations related to at least five features, simultaneously.

The dynamics in (6) provide seven independent directions along which the Lie derivatives can be computed. On the other hand, one of these directions is not available. This is the vector $\boldsymbol{f_1}(\boldsymbol{X_2^u})$, which is obtained by setting ${}^c\boldsymbol{\Omega} = [0\ 0\ 0]^T$ and $\boldsymbol{A} = [1\ 0\ 0]^T$ in (6), once $\boldsymbol{f_0}(\boldsymbol{X_2^u})$ has been removed[2]. Since the Lie derivatives along this direction are not null, we have to proceed as in section V-A. We must proceed in several subsequent steps. In each step we check, first of all, which highest order of Lie derivatives of the observations can be used. This is obtained by checking that, for a given order, all the Lie derivatives up to this order, computed at least once along $\boldsymbol{f_1}$, are identically zero. Once this highest order is identified, we find the largest number of independent Lie derivatives up to this order. We include new time derivatives of the unknown inputs (i.e., A_x) in order to make usable higher order Lie derivatives, as explained in section IV. We will show that, by including A_x and its first time derivative, we can prove the observability of the entire state.

1) First step: We start with the $23-$dimensional state given by the vector in (5), with a single feature. By chance, the first order Lie derivatives of the functions h_{F_x}, h_{F_y} and h_{F_z} along $\boldsymbol{f_1}$ are null. Regarding the other nine outputs, the first order Lie derivatives along this direction are different from zero. Among the usable Lie derivatives, we detect 14 independent functions, which are: $\mathcal{L}^0 h_{F_x}$, $\mathcal{L}^0 h_{F_y}$, $\mathcal{L}^0 h_{F_z}$, $\mathcal{L}^0 h_{V_x}$, $\mathcal{L}^0 h_{V_y}$, $\mathcal{L}^0 h_{V_z}$, $\mathcal{L}^0 h_t$, $\mathcal{L}^0 h_x$, $\mathcal{L}^0 h_y$, $\mathcal{L}^0 h_q$, $\mathcal{L}^0 h_{q^c}$, $\mathcal{L}^1_5 h_{F_x}$, $\mathcal{L}^1_6 h_{F_x}$ and $\mathcal{L}^1_6 h_{F_y}$.

2) Second step: We include A_x in the state. The new state has dimension 24. Now we can use the first order Lie derivatives of all the outputs and the second order Lie derivatives of the first three outputs. We detect seven additional independent Lie derivatives which are: $\mathcal{L}^2_{00} h_{F_x}$, $\mathcal{L}^2_{02} h_{F_x}$, $\mathcal{L}^2_{03} h_{F_x}$, $\mathcal{L}^2_{00} h_{F_y}$, $\mathcal{L}^2_{02} h_{F_y}$, $\mathcal{L}^2_{03} h_{F_y}$ and $\mathcal{L}^2_{00} h_{F_z}$.

3) Third step: We include $A_x^{(1)} \equiv \dot{A}_x$. The new state has dimension 25. Now we can use the second order Lie derivatives of all the outputs. We detect four additional independent Lie derivatives which are: $\mathcal{L}^2_{00} h_{V_x}$, $\mathcal{L}^2_{05} h_{V_x}$, $\mathcal{L}^2_{06} h_{V_x}$ and $\mathcal{L}^2_{06} h_{V_y}$. Hence, the total number of independent Lie derivatives which are usable is 25, which coincides with the dimension of the state. We proved the theorem:

[2]Note that in this case $\boldsymbol{u} = [u_1,\ u_2,\ \cdots,\ u_{nc}]^T = [A_x,\ A_y,\ A_z,\ {}^c\Omega_x,\ {}^c\Omega_y,\ {}^c\Omega_z]^T$, i.e., the last three inputs are the components of the angular speed in the camera frame, which is known.

Theorem 2 (Two accelerometers, uncalibrated camera) *In the Vi-SfM problem with 2 accelerometers, no gyroscope, unknown camera-inertial sensor transformation and at least five features available, all the independent observable modes are the same as in the standard Vi-SfM problem.*

C. Single Accelerometer and Uncalibrated Camera

Before computing the Lie derivatives in order to apply the extended observability rank criterion introduced in section IV, we derive a continuous symmetry by using an intuitive procedure. Let us suppose to collect the data from the camera and the accelerometer during a given time interval for a generic vehicle motion, starting from a given initial state. We remark that, independently of the motion, by rotating the initial state around the accelerometer axis (i.e., around the $z-$axis of the inertial sensor frame) we obtain exactly the same measurements. Let us derive how this rotation changes the initial state by referring to an infinitesimal rotation of an angle ϵ. We rotate all the features, the camera frame (namely its position and orientation in the inertial sensor frame) the initial vehicle speed and orientation, simultaneously, around the $z-$axis of the inertial sensor frame, by the angle ϵ. The camera configuration in the inertial sensor frame changes as follows [6]: $\boldsymbol{R^c} \rightarrow \boldsymbol{R'^c} = \boldsymbol{R^c} + \epsilon[Y^c, -X^c, 0]^T$ and $q^c \rightarrow q'^c = q^c + \epsilon/2(q^c_z + q^c_y i - q^c_x j - q^c_t k)$. The initial speed in the inertial sensor frame ($\boldsymbol{V} \equiv [V_x, V_y, V_z]^T$) changes as follows: $\boldsymbol{V} \rightarrow \boldsymbol{V'} = \boldsymbol{V} + \epsilon[V_y, -V_x, 0]^T$. Let us derive how the initial orientation changes. The state in (11) contains the quaternion q, which describes the orientation of the inertial sensor frame in the global frame and not the orientation of the global frame in the inertial sensor frame. This last orientation is described by the quaternion $p \equiv p_t + ip_x + jp_y + kp_z \equiv q^* = qt - iq_x - jq_y - kq_z$. The quaternion p changes as q^c, namely: $p \rightarrow p' = p + \epsilon/2(p_z + p_y i - p_x j - p_t k)$. Hence, we have: $q^*_t \rightarrow q^*_t + \frac{\epsilon}{2} q^*_z$, $q^*_x \rightarrow q^*_x + \frac{\epsilon}{2} q^*_y$, $q^*_y \rightarrow q^*_y - \frac{\epsilon}{2} q^*_x$ and $q^*_z \rightarrow q^*_z - \frac{\epsilon}{2} q^*_t$. By using $q^*_t = q_t$, $q^*_x = -q_x$, $q^*_y = -q_y$ and $q^*_z = -q_z$ we obtain: $q \rightarrow q' = q + \epsilon/2(-q_z + q_y i - q_x j + q_t k)$. The rotation does not affect all the remaining quantities in the state in (11). Indeed, μ and A_z^{bias} are scalar quantities. The vectors ${}^c\boldsymbol{F^1}, \cdots, {}^c\boldsymbol{F^{N_f}}$ are the relative positions of the features in the camera frame. Since, by definition, we are both rotating the features and the camera frame, these relative positions are unvaried. Finally, the vector $\boldsymbol{g}$ remains unvaried since we are rotating the global frame and the gravity, simultaneously. The rotation described above, is characterized by the symmetry:

$$\boldsymbol{w^{int}} \equiv \Big[\boldsymbol{0}_{3N_f},\ V_y, -V_x, 0, -\frac{q_z}{2}, \frac{q_y}{2}, -\frac{q_x}{2}, \frac{q_t}{2}, \tag{20}$$

$$0,\ Y^c, -X^c, 0,\ \frac{q^c_z}{2}, \frac{q^c_y}{2}, -\frac{q^c_x}{2}, -\frac{q^c_t}{2},\ 0,0,0,\ 0\Big]^T$$

where $\boldsymbol{0}_{3N_f}$ is the zero $1 \times 3N_f$ vector. Namely, the transformation: $\boldsymbol{X_1^u} \rightarrow \boldsymbol{X_1^u} + \epsilon \boldsymbol{w^{int}}$ on the initial state, cannot be detected by analyzing the measurements delivered by the camera and the accelerometer independently of the trajectory.

By proceeding as in the previous section in several subsequent steps, it is possible to show that $\boldsymbol{w}^{int}$ is the only system symmetry. This is obtained by augmenting the original state in order to include A_x, A_y, $A_x^{(1)} \equiv \dot{A}_x$ and $A_y^{(1)} \equiv \dot{A}_y$. We proved the theorem:

Theorem 3 (Single accelerometer, uncalibrated camera) *In the Vi-SfM problem with a single accelerometer, no gyroscope, unknown camera-inertial sensor transformation and at least five features available, there is a continuous internal symmetry. 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 as the number of inertial sensors is reduced. The case of a single accelerometer and no gyroscope was firstly investigated. Theorem 1 basically states that, if the camera is extrinsically calibrated, the observability properties remain the same as in the case of 3 accelerometers and 3 gyroscopes. If the camera is not extrinsically calibrated, an internal symmetry arises (see theorem 3). As a result, 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. If an additional accelerometer is introduced, the system gains again full observability (theorem 2). These results show that, the information provided by an IMU together with a monocular camera, is redundant. Additionally, these results are consistent with our knowledge about the vestibular system, which provides balance in most mammals. Indeed, the otoliths, which indicate linear accelerations, consist of two organs (the *utricle* and the *saccule*) able to sense the acceleration only along two independent axes (see fig. 1).

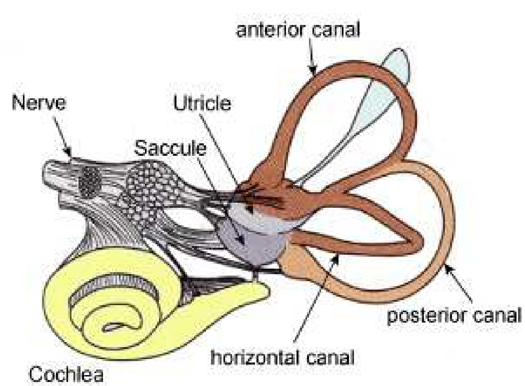

Fig. 1. The otoliths perceive acceleration only along two independent axes

Finally, to analyze these systems with a reduced number of sensors, the paper introduced a new method that allows us to derive the observability properties of a non linear system when part of its input controls is unknown.

References

[1] F. A. W. Belo, P. Salaris, and A. Bicchi, 3 Known Landmarks are Enough for Solving Planar Bearing SLAM and Fully Reconstruct Unknown Inputs, The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems October 18-22, 2010, Taipei, Taiwan

[2] A. Berthoz, B. Pavard and L.R. Young, Perception of Linear Horizontal Self-Motion Induced by Peripheral Vision Basic Characteristics and Visual-Vestibular Interactions, Exp. Brain Res. 23, 471–489 (1975).

[3] M. Bryson and S. Sukkarieh, Observability Analysis and Active Control for Airbone SLAM, IEEE Transaction on Aerospace and Electronic Systems, vol. 44, no. 1, 261–280, 2008

[4] Dokka K., MacNeilage P. R., De Angelis G. C. and Angelaki D. E., Estimating distance during self-motion: a role for visual-vestibular interactions, Journal of Vision (2011) 11(13):2, 1-16

[5] C. R. Fetsch, G. C. DeAngelis and D. E. Angelaki, Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory, Eur J Neurosci. 2010 May ; 31(10): 1721-1729

[6] Goldstein, H. Classical Mechanics, MA: Addison-Wesley, 1980

[7] 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

[8] Hermann R. and Krener A.J., 1977, Nonlinear Controllability and Observability, Transaction On Automatic Control, AC-22(5): 728-740

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] M. Li and A. I. Mourikis, Improving the accuracy of EKF-based visualinertial odometry, ICRA 2012, St. Paul, MN.

[14] A. Martinelli, State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, IEEE Transactions on Robotics, vol. 27, no. 2, pp. 239–255, 2011

[15] 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.

[16] A. Martinelli, Visual-inertial structure from motion: observability and resolvability, iros 2013, Tokyo, Japan

[17] A. Martinelli, Closed-form solution of visual-inertial structure from motion, Int. J. of Computer Vision, Vol. 106, N. 2, 138–152, 2014.

[18] Mirzaei F.M. and Roumeliotis S.I., A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation, Transactions on Robotics, 2008, Vol. 24, No. 5, 1143–1156

[19] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 26(6):756-770, June 2004

[20] Michael Spivak, Calculus on Manifolds

[21] D. Strelow and S. Singh, Motion estimation from image and inertial measurements, International Journal of Robotics Research, 23(12), 2004

[22] Weiss., S., Vision Based Navigation for Micro Helicopters, PhD thesis, Diss. ETH No. 20305