HAL
open science

Visual-inertial structure from motion: observability and resolvability

Agostino Martinelli

To cite this version:

Agostino Martinelli. Visual-inertial structure from motion: observability and resolvability. IROS 2013, Nov 2013, Japan. hal-00905890

HAL Id: hal-00905890

https://hal.science/hal-00905890v1

Submitted on 18 Nov 2013

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.

Visual-inertial structure from motion: observability and resolvability

Agostino Martinelli

***Abstract*— This paper provides two novel contributions. The former regards the observability of the visual-inertial structure from motion. It is proven that, the information contained in the data provided by a monocular camera which observes a single point-feature and by an Inertial Measurement Unit (IMU) allows estimating the absolute scale, the speed in the local frame, the absolute roll and pitch angles, the biases which affect the accelerometer's and the gyroscope's measurements, the magnitude of the gravitational acceleration and the extrinsic camera-IMU calibration. The latter contribution is the derivation of a new closed form solution to determine some of the previous observable quantities by only using few camera measurements collected during a short time interval and the data provided by the IMU during the same time interval. This closed-solution allows us to investigate the intrinsic properties of the visual-inertial structure from motion and in particular to identify the conditions under which the problem has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions and infinite solutions depending on the trajectory, on the number of point-features and on their layout and on the number of camera images. The proposed closed solution is finally used in conjunction with a filter based approach in order to show its benefit.**

I. Introduction

The structure from motion problem (SfM) consists in determining the three-dimensional structure of the scene by using the measurements provided by one or more sensors over time (e.g. vision sensors, ego-motion sensors, range sensors). In the case of visual measurements only, the SfM problem has been solved up to a scale [3], [4], [10], [16], [22] and a closed form solution has also been derived [10], [16], [22], allowing the determination of the three-dimensional structure of the scene, without the need for any prior knowledge. The case of inertial and visual measurements, i.e., 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 [2], [12], [13], [20], [23] and in the framework of neuroscience (e.g., [1], [5], [8]). 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.

From a theoretical perspective, recent works on Vi-SfM have focused on two separate issues: (i) understanding the observability properties in several contexts and (ii) determining the solution in closed form.

The first issue has been faced in [11], [12], [13], [14], [15], [20], [21] and [24]. In [11], [14] and [15] the authors investigated the estimator inconsistency in the Vi-SfM problem. In [11] the authors found that standard EKF-based estimators lead to spurious information gain along unobservable directions. They also suggested a modification on the basic estimator in order to enforce the unobservable directions and thus to reduce inconsistency. In [12], [13], [21] and [24] the observability properties have been derived by accounting an unknown transformation between the camera and the IMU frames and an unknown magnitude of the gravity. Additionally, in [13] and [24] also the case of biased inertial measurements has been considered. We remark that in [12], [13] and [24] the observability properties have been derived starting from basic results in computer vision. Specifically, in [12] and [13], starting from the results derived in [3], a global frame has been fixed by constraining three directions determined by three points on the image plane. In [24], the camera is considered as a localization sensor up to a scale. This is based on the assumption that the camera is observing a number of features (at least five [22]) which guarantees that its motion can be reconstructed up to a scale. This significantly simplifies the observability analysis since, the expression of the observation provided by the camera consists of three components of the state which defines the system. In [20] the observability properties have been derived without using the previous mentioned results from computer vision and this allowed us to deal with the case when a single point feature is observed by the camera. The analysis was based on the concept of *continuous symmetry* introduced in [19]. Since under these conditions the camera observation has an expression much more complex, the analysis in [20] was limited to the case when the camera extrinsic calibration in the IMU frame is a priori known. The first contribution of this paper (in section III) is precisely the extension of the observability analysis carried out in [20] in order to cope with the case of unknown camera extrinsic calibration. In order to achieve this objective, the theory introduced in [19] has been extended by adding some new techniques which allow us to significantly reduce the load of symbolic computation required to perform the analysis and the derivation of the system symmetries. For the sake of brevity, the details about these new results (as the concept of *quasi-projection*) are available in a separate technical report [17]. It is interesting to note that, thanks to this new analysis, it is shown that two points features are sufficient to fix a global frame where the state which defines the Vi-SfM is observable (included the bias, the magnitude of the gravity and the camera-IMU transformation).

The second theoretical issue previously mentioned, namely the determination of the Vi-SfM in closed form, has been

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

faced in [6], [18] and [20]. The interest of this research comes from the fact that any approach to the Vi-SfM based on a recursive filter or on a smoothing estimator needs to initialize the estimated state. Due to the system non linearities, an erroneous initialization can cause a divergence of the estimation process. A deterministic solution, i.e., a solution which analytically expresses the observable modes in terms of the measurements provided by the sensors during a short time-interval, will avoid this important inconvenient. Closed form solutions to the Vi-SfM have been introduced in [20]. In [6] also the case of an unknown camera-IMU extrinsic calibration has been dealt and a deterministic algorithm able to also determine the parameters characterizing this transformation has been introduced. Here we present a further closed solution to the basic Vi-SfM (section IV). Compared with the solutions proposed in [20], this new solution allows us to investigate the intrinsic properties of the Vi-SfM problem and to identify the conditions under which the problem can be solved in closed form. In particular, these conditions regard the trajectory, the number of point-features and their layout and the number of monocular images where the same point-features are seen. These results are obtained under the assumption of noiseless visual and inertial measurements. Additionally, the measurements provided by the IMU are assumed to be unbiased and the camera-IMU extrinsic calibration is assumed known. On the other hand, Monte Carlo simulations have also been performed by relaxing all these assumptions and the closed form solution is used in conjunction with a filtering approach in order to show its benefit (section VI). For the sake of brevity, also in this case we omit several proofs and several additional properties. They can be found in [18], where also the case of biased measurements from the accelerometers is considered.

II. The Considered System

We consider a system (from now on we call it the *vehicle*) consisting of a monocular camera and an IMU. We introduce a global frame in order to characterize the motion of the vehicle moving in a $3D$ environment. Its z-axis points vertically upwards. We will adopt lower-case letters to denote vectors in this frame (e.g., the gravity is $\boldsymbol{g} = [0,\ 0,\ -g]^T$, where g is the magnitude of the gravitational acceleration). We define the vehicle local frame as the IMU frame. We will adopt upper-case letters to denote vectors in the vehicle frame. The camera frame differs from the local frame. We characterize the transformation between these two frames through $\boldsymbol{R^c}$ and q^c, where $\boldsymbol{R^c} \equiv [X^c,\ Y^c,\ Z^c]^T$ is the position of the camera optical center in the local frame and $q^c \equiv q^c_t + q^c_x i + q^c_y j + q^c_z k$ is the unit quaternion which characterizes the orientation of the camera frame in the local frame. We assume that both $\boldsymbol{R^c}$ and q^c are independent of time and are unknown.

The IMU provides the vehicle angular speed and acceleration. We will denote the measured quantities by $\boldsymbol{\Omega}$ and $\boldsymbol{A}$, respectively. These quantities differ from the true values, $\boldsymbol{\Omega^{true}}$ and $\boldsymbol{A^{true}}$. Regarding the angular speed, the one measured by the gyroscopes includes a bias and a zero-mean error, i.e.: $\boldsymbol{\Omega} = \boldsymbol{\Omega^{true}} + \boldsymbol{\Omega^{bias}} + \boldsymbol{n_\Omega}$. Regarding the acceleration, the one measured by the accelerometers includes the inertial acceleration ($\boldsymbol{A}^{inertial}$), the gravitational acceleration ($\boldsymbol{G}$), a bias and a zero-mean error. In other words: $\boldsymbol{A} = \boldsymbol{A}^{inertial} + \boldsymbol{A^{bias}} - \boldsymbol{G} + \boldsymbol{n_A}$. Note that the gravity comes with a minus since, when the vehicle does not accelerate (i.e. $\boldsymbol{A}^{inertial}$ is zero), the accelerometers perceive an acceleration which is the same of an object accelerated upward in absence of gravity.

Our system is characterized by the state $[\boldsymbol{r},\ \boldsymbol{v},\ q]^T$ where $\boldsymbol{r} = [r_x,\ r_y,\ r_z]^T$ is the $3D$ vehicle position in the global frame, $\boldsymbol{v}$ is its time derivative, i.e. the vehicle speed in the global frame ($\boldsymbol{v} \equiv \frac{d\boldsymbol{r}}{dt}$) and $q \equiv q_t + q_x i + q_y j + q_z k$ is the unit quaternion which characterizes the vehicle orientation in the global frame.

In the following we want to derive the analytical expression of the dynamics and the camera observations. For the sake of simplicity, we consider the case of noiseless measurements. The case with noise can be easily obtained with the substitution $\boldsymbol{A} \to \boldsymbol{A} + \boldsymbol{n_A}$ and $\boldsymbol{\Omega} \to \boldsymbol{\Omega} + \boldsymbol{n_\Omega}$. The dynamics of the previous state can be easily provided by expressing all the $3D$ vectors as imaginary quaternions. In practice, given a $3D$ vector $\boldsymbol{w} = [w_x,\ w_y,\ w_z]^T$ we associate with it the imaginary quaternion $w_q \equiv 0 + w_x i + w_y j + w_z k$. The dynamics of the state $[r_q,\ v_q,\ q]^T$ are:

$$\begin{cases} \dot{r}_q &= v_q \\ \dot{v}_q &= qA_q^{inertial}q^* = qA_q q^* + qA_q^{bias}q^* + g_q \\ \dot{q} &= \dfrac{1}{2}q\Omega_q + \dfrac{1}{2}q\Omega_q^{bias} \end{cases} \tag{1}$$

being q^* the conjugate of q, $q^* = q_t - iq_x - jq_y - kq_z$. By considering the case of unknown biases, unknown magnitude of the gravity and unknown transformation between the IMU and the camera frames, the state which defines our system becomes the following $24-$dimensional vector:

$$\boldsymbol{X} \equiv \left[\boldsymbol{r},\ \boldsymbol{v},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{R^c}, q^c,\ g\right]^T \tag{2}$$

whose dynamics are given in (1) with the following trivial additional equations:

$$\begin{cases} \dot{\boldsymbol{A}}^{\boldsymbol{bias}} = \dot{\boldsymbol{\Omega}}^{\boldsymbol{bias}} = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{g} = \dot{q}^c = 0 \end{cases} \tag{3}$$

Note that this is the state which defines our system when a single point feature is observed by the camera. In this case the origin of the global frame can be chosen as coincident with the observed feature. In the case of multiple features, the state dimension becomes $24 + 3(N_f - 1)$ (N_f being the number of observed features) and the coordinates of the further $N_f - 1$ features are included in the state (see [20] for more details). On the other hand, the state defined in (2) is not a suitable choice to characterize our system. Indeed, the expression of the camera observations in terms of it involves the product of five quaternions: $(q^c)^* q^* r_q q q^c$. This makes impossible to efficiently derive the observability properties. In order to have a simple expression of the camera

observations it is much more convenient to adopt a new state. Let us refer to the case of a single feature. The new state is:

$$\boldsymbol{X_n} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F},\ \boldsymbol{V},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{R^c}, q^c,\ g\right]^T \quad (4)$$

where ${}^{\boldsymbol{c}}\boldsymbol{F} \equiv [{}^cF_x,\ {}^cF_y,\ {}^cF_z]^T$ is the position of the feature in the camera frame and $\boldsymbol{V}$ is the vehicle speed in the local frame (i.e., in the IMU frame). By using the equations in (1) we obtain the following dynamics for the new state:

$$\begin{bmatrix} {}^{\boldsymbol{c}}\dot{\boldsymbol{F}} & = M({}^{\boldsymbol{c}}\boldsymbol{\Omega}){}^{\boldsymbol{c}}\boldsymbol{F} - R_{q^c}[\boldsymbol{V} + (\boldsymbol{\Omega} + \boldsymbol{\Omega^{bias}}) \wedge \boldsymbol{R^c}] \\ \dot{\boldsymbol{V}} & = M(\boldsymbol{\Omega} + \boldsymbol{\Omega^{bias}})\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{A^{bias}} + \boldsymbol{G} \\ \dot{q} & = \frac{1}{2}q\Omega_q + \frac{1}{2}q\Omega_q^{bias} \\ \dot{\boldsymbol{A}}^{\boldsymbol{bias}} & = \dot{\boldsymbol{\Omega}}^{\boldsymbol{bias}} = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{g} & = \dot{q}^c = 0 \end{bmatrix} \quad (5)$$

where:

- $M(\boldsymbol{\Omega}) \equiv \begin{bmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{bmatrix}$;
- ${}^{\boldsymbol{c}}\boldsymbol{\Omega}$ is the angular speed in the camera frame, i.e., ${}^c\Omega_q = (q^c)^*(\Omega_q + \Omega_q^{bias})q^c$;
- R_{q^c} is the rotation matrix associated with the quaternion q^c (i.e., for a $3D$ vector $\boldsymbol{w} = [w_x,\ w_y,\ w_z]^T$, $(R_{q^c}\boldsymbol{w})_q = (q^c)^* w_q q^c$);
- the symbol "$\wedge$" denotes the vectorial product.

Figure 1 displayes the three reference frames together with some of the previous vectors.

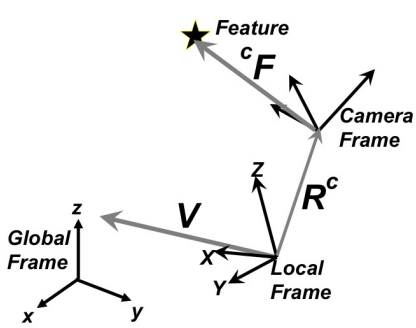

Fig. 1. Global frame, local (IMU) frame and camera frame with the feature position (${}^{\boldsymbol{c}}\boldsymbol{F}$) in the camera frame and the vehicle speed ($\boldsymbol{V}$) and the camera position ($\boldsymbol{R^c}$) in the local frame.

The expression of the camera observations in terms of the new state is trivial. Indeed, the camera provides the direction of the observed feature in its own frame. Hence, it provides the vector ${}^{\boldsymbol{c}}\boldsymbol{F}$ up to a scale, or, equivalently, the two following ratios:

$$\boldsymbol{h_{cam}}(\boldsymbol{X_n}) \equiv [h_u,\ h_v]^T = \left[\frac{{}^cF_x}{{}^cF_z},\ \frac{{}^cF_y}{{}^cF_z}\right]^T \quad (6)$$

We have also to consider the two constraints $q^*q = 1$ and $(q^c)^*q^c = 1$. These can be dealt as further observations:

$$\boldsymbol{h_{const}}(\boldsymbol{X_n}) \equiv [h_q,\ h_{q^c}]^T = [q^*q,\ (q^c)^*q^c]^T \quad (7)$$

Finally, the case of multiple features can be characterized by including in the state the position of each feature in the camera frame, i.e., ${}^{\boldsymbol{c}}\boldsymbol{F} \rightarrow {}^{\boldsymbol{c}}\boldsymbol{F^i}$, $i = 1, 2, \cdots, N_f$. The resulting state has dimension $24 + 3(N_f - 1)$.

III. Observability Properties

In [20] we investigated the observability properties of the Vi-SfM in several contexts which include the case of a single and multiple features, the case of known and unknown magnitude of the gravity, the case of biased and unbiased inertial measurements. In all the considered cases, the extrinsic camera-IMU transformation was assumed known. In the case of a single feature, the state adopted to characterize the case of biased inertial measurements and unknown magnitude of the gravity was $\left[\boldsymbol{r},\ \boldsymbol{v},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ g\right]^T$, whose dimension is 17. The results of the observability analysis carried out in [20] were obtained by using the method of *continuous symmetries* developed in [19] and can be summarized as follows:

Theorem 1 (Known extrinsic calibration) *Let us consider the Vi-SfM with biased inertial measurements, unknown magnitude of the gravity and known camera-IMU transformation. All the independent observable modes are: the positions in the local frame of all the observed features, the* 3 *components of the speed of the vehicle in the local frame, the two biases affecting the accelerometer and gyroscope measurements, the roll and the pitch angle and the magnitude of the gravity.*

The derivation of this result required to analytically compute the Lie derivatives up to the third order. In general, the complexity of the computation of the Lie derivatives and the determination of their dependence or independence dramatically depends on the state dimension.

Let us consider now the case when the extrinsic camera-IMU transformation is unknown. In order to solve the structure from motion we also need to estimate the parameters which characterize this transformation. In other words, the state which defines our system, is the one given in (2) or in (4), for the case of a single feature. In this case, even by using the state in (4), we found prohibitive to analytically deal with second-order Lie derivatives. Specifically, by using the symbolic computation tool of Matlab running on a 2.7GHz dual-core Intel Core i7 processor with 4MB shared L3 cache, the time demanded to compute the rank of the matrix whose lines are the gradients of all the Lie derivatives up to the second order, is equal to $101734s$ and the analytical determination of its null space required $127683s$. In order to apply the theory developed in [19] we derived some new theoretical results (see properties 2 and 3 and the operation of quasi-projection in [17]). By using these new results we were able to deal with third-order Lie derivatives and the total time of computation requested to determine the number of independent Lie derivatives up to the third order does

not exceed $200s$ (see [17] for all the details). The result of this analysis allows us to extend the result in theorem 1 to the case of unknown camera-IMU calibration. We have the following new result[1]:

Theorem 2 (Unknown extrinsic calibration) *Let us consider the Vi-SfM with biased inertial measurements, unknown magnitude of the gravity and unknown camera-IMU transformation. All the independent observable modes are: the positions in the local frame of all the observed features, the* 3 *components of the speed of the vehicle in the local frame, the two biases affecting the accelerometer and gyroscope measurements, the roll and the pitch angle, the magnitude of the gravity and the transformation between the camera and IMU frames.*

IV. The closed form solution

In the following we assume that the camera and the IMU frames coincide and that the IMU biases and the magnitude of the gravity are known. These assumptions will be relaxed in section VI. Since the local frame is time dependent, we will adopt the following notation: $\boldsymbol{W}_t(\tau)$ will be the vector with global coordinates $\boldsymbol{w}(\tau)$ in the local frame at time t. Additionally, we will denote with $C_{t_1}^{t_2}$ the matrix which characterizes the rotation occurred during the time interval $(t_1,\ t_2)$ and with $C_{t_2}^{t_1}$ its inverse (i.e., $(C_{t_1}^{t_2})^{-1} = C_{t_2}^{t_1}$). Let us refer to vectors which are independent of the origin of the reference frame (e.g., speed, acceleration, etc.). For these vectors we have: $\boldsymbol{W}_{t_1}(\tau) = C_{t_1}^{t_2}\boldsymbol{W}_{t_2}(\tau)$. Finally, C^t will denote the rotation matrix between the global frame and the local frame at time t, i.e., $\boldsymbol{w}(\tau) = C^t\boldsymbol{W}_t(\tau)$. We assume that the camera is observing one or more point-features during the time interval $[T_{in},\ T_{fin}]$. Our goal is to express in closed-form the observable modes at a given time T_{in} only in terms of the visual and inertial measurements obtained in the time interval $[T_{in},\ T_{fin}]$.

The position of the vehicle $\boldsymbol{r}$ at any time $t \in [T_{in},\ T_{fin}]$ satisfies the equation $\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \int_{T_{in}}^{t}\int_{T_{in}}^{\tau}\boldsymbol{a}(\xi)d\xi d\tau$. The last term contains a double integral over time, which can be simplified in a single integral by integrating by parts. We obtain:

$$\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \int_{T_{in}}^{t}(t-\tau)\boldsymbol{a}(\tau)d\tau \tag{8}$$

where $\boldsymbol{a} \equiv \frac{d\boldsymbol{v}}{dt}$ and $\Delta t \equiv t - T_{in}$. We write equation (8) by highlighting the vector $\boldsymbol{A}_\tau(\tau)$ provided by the accelerometer:

$$\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \boldsymbol{g}\frac{\Delta t^2}{2} + C^{T_{in}}\boldsymbol{S}_{T_{in}}(t) \tag{9}$$

where:

$$\boldsymbol{S}_{T_{in}}(t) \equiv \int_{T_{in}}^{t}(t-\tau)C_{T_{in}}^{\tau}\boldsymbol{A}_\tau(\tau)d\tau$$

[1]We realized that this result has independently been proven very recently in [9], by following a completely different methodology.

The matrix $C_{T_{in}}^{\tau}$ can be obtained from the angular speed during the interval $[T_{in},\ \tau]$ provided by the gyroscopes [7]. Hence, the vector $\boldsymbol{S}_{T_{in}}(t)$ can be obtained by integrating the data provided by the gyroscopes and the accelerometers delivered during the interval $[T_{in},\ t]$.

Let us suppose that N_f point-features are observed, simultaneously. Their position in the local frame are $\boldsymbol{F}_t^i(t)$ and, in the global frame $\boldsymbol{f}^i$.

$$\boldsymbol{f}^i = \boldsymbol{r}(t) + C^{T_{in}}C_{T_{in}}^{t}\boldsymbol{F}_t^i(t) \tag{10}$$

We write this equation at time $t = T_{in}$ obtaining:

$$\boldsymbol{f}^i - \boldsymbol{r}(T_{in}) = C^{T_{in}}\boldsymbol{F}_{T_{in}}^i(T_{in}) \tag{11}$$

By inserting the expression of $\boldsymbol{r}(t)$ provided in (9) into equation (10), by using (11) and by pre multiplying by the rotation matrix $(C^{T_{in}})^{-1}$ (we remind the reader that, according to our notation, $\boldsymbol{v}(T_{in}) = C^{T_{in}}\boldsymbol{V}_{T_{in}}(T_{in})$ and $\boldsymbol{g} = C^{T_{in}}\boldsymbol{G}_{T_{in}}$) we finally obtain the following equation:

$$\begin{aligned} C_{T_{in}}^{t}\boldsymbol{F}_t^i(t) = \boldsymbol{F}_{T_{in}}^i(T_{in}) - \boldsymbol{V}_{T_{in}}(T_{in})\Delta t - \boldsymbol{G}_{T_{in}}\frac{\Delta t^2}{2} + \\ - \boldsymbol{S}_{T_{in}}(t); \qquad i = 1,\ 2,\ ...,\ N_f \end{aligned} \tag{12}$$

A single image provides the bearing angles of all the point-features in the local frame. In other words, an image taken at time t provides all the vectors $\boldsymbol{F}_t^i(t)$ up to a scale. Since the data provided by the gyroscopes during the interval $(T_{in},\ T_{fin})$ allow us to build the matrix $C_{T_{in}}^{t}$, having the vectors $\boldsymbol{F}_t^i(t)$ up to a scale, allows us to also know the vectors $C_{T_{in}}^{t}\boldsymbol{F}_t^i(t)$ up to a scale.

We assume that the camera provides n_i images of the same N_f point-features at the consecutive times: $t_1 = T_{in} < t_2 < ... < t_{n_i} = T_{fin}$. From now on, for the sake of simplicity, we adopt the following notation:

- $\boldsymbol{F}_j^i \equiv C_{T_{in}}^{t_j}\boldsymbol{F}_{t_j}^i(t_j)$, $i = 1,\ 2,\ ...,\ N_f$; $j = 1,\ 2,\ ...,\ n_i$
- $\boldsymbol{F}^i \equiv \boldsymbol{F}_{T_{in}}^i(T_{in})$, $i = 1,\ 2,\ ...,\ N_f$
- $\boldsymbol{V} \equiv \boldsymbol{V}_{T_{in}}(T_{in})$
- $\boldsymbol{G} \equiv \boldsymbol{G}_{T_{in}}$
- $\boldsymbol{S}_j \equiv \boldsymbol{S}_{T_{in}}(t_j)$, $j = 1,\ 2,\ ...,\ n_i$

We remark that the difference $\boldsymbol{F}_j^i - \boldsymbol{F}_1^i,\ i = 1, 2, ..., N,\ j = 2, ..., n_i$, is independent of i (see equation (12), where, by definition, $C_{T_{in}}^{t_j}\boldsymbol{F}_{t_j}^i(t_j) = \boldsymbol{F}_j^i$). Hence, we will set $\boldsymbol{\chi}_j \equiv \boldsymbol{F}_j^i - \boldsymbol{F}_1^i$. This quantity characterizes the motion of the vehicle. We will denote with $\boldsymbol{\mu}_j^i$ the unit vector with the same direction of $\boldsymbol{F}_j^i$ and we introduce the unknowns λ_j^i such that $\boldsymbol{F}_j^i = \lambda_j^i\boldsymbol{\mu}_j^i$. Finally, without loss of generality, we can set $T_{in} = 0$, i.e., $\Delta t = t$. Our sensors provide $\boldsymbol{\mu}_j^i$ and $\boldsymbol{S}_j$ for $i = 1,\ 2,\ ...,\ N_f$; $j = 1,\ 2,\ ...,\ n_i$. Equation (12) can be written as follows:

$$\boldsymbol{F}^i - \boldsymbol{V}t_j - \boldsymbol{G}\frac{t_j^2}{2} - \lambda_j^i\boldsymbol{\mu}_j^i = \boldsymbol{S}_j \tag{13}$$

The Vi-SfM problem is the determination of the vectors: $\boldsymbol{F}^i$, $(i = 1,\ 2,\ ...,\ N_f)$, $\boldsymbol{V}$, $\boldsymbol{G}$. We can use the equations in (13)

to determine these vectors. On the other hand, the use of (13) requires to also determine the quantities λ_j^i. By considering $j = 1$ in (13), i.e. $t_j = t_1 = T_{in} = 0$, we easily obtain: $\boldsymbol{F}^i = \lambda_1^i \boldsymbol{\mu}_1^i$. Then, we write the linear system in (13) as follows:

$$\left[\begin{array}{ll} -\boldsymbol{G}\frac{t_j^2}{2} - \boldsymbol{V}t_j + \lambda_1^1\boldsymbol{\mu}_1^1 - \lambda_j^1\boldsymbol{\mu}_j^1 & = \boldsymbol{S}_j \\ \lambda_1^1\boldsymbol{\mu}_1^1 - \lambda_j^1\boldsymbol{\mu}_j^1 - \lambda_1^i\boldsymbol{\mu}_1^i + \lambda_j^i\boldsymbol{\mu}_j^i & = 0_3 \end{array} \right. \tag{14}$$

where $j = 2, ..., n_i$, $i = 2, ..., N_f$ and 0_3 is the 3×1 zero vector. This linear system consists of $3(n_i - 1)N_f$ equations in $N_f n_i + 6$ unknowns. Let us define the two column vectors $\boldsymbol{X}$ and $\boldsymbol{S}$:

$$\boldsymbol{X} \equiv [\boldsymbol{G}^T,\ \boldsymbol{V}^T,\ \lambda_1^1,\ ...,\ \lambda_1^{N_f},\ ...,\ \lambda_{n_i}^1,\ ...,\ \lambda_{n_i}^{N_f}]^T$$

and

$$\boldsymbol{S} \equiv [\boldsymbol{S}_2^T,\ 0_3,\ ...,\ 0_3,\ \boldsymbol{S}_3^T,\ 0_3,\ ...,\ 0_3,\ ...,\ \boldsymbol{S}_{n_i}^T,\ 0_3,\ ...,\ 0_3]^T$$

and the matrix Ξ (see equation (15) at the next page), where $T_j \equiv -\frac{t_j^2}{2} I_3$, $S_j \equiv -t_j I_3$ and I_3 is the identity 3×3 matrix; 0_{33} is the 3×3 zero matrix (note that the third set of columns disappear in absence of bias). The linear system in (14) can be written in the following compact format:

$$\Xi \boldsymbol{X} = \boldsymbol{S} \tag{16}$$

The sensor information is completely contained in the above linear system. Additionally, we assume that the magnitude of the gravitational acceleration is a priori known. This extra information is obtained by adding to our linear system the following quadratic equation: $|\boldsymbol{G}| = g$. By introducing the following $3 \times (N_f n_i + 6)$ matrix, $\Pi \equiv [I_3,\ 0_3\ ...\ 0_3]$, this quadratic constraint can be written in terms of $\boldsymbol{X}$ as follows:

$$|\Pi \boldsymbol{X}|^2 = g^2 \tag{17}$$

The Vi-SfM problem can be solved by finding the vector $\boldsymbol{X}$, which satisfies (16) and (17).

V. Existence and number of distinct solutions

We are interested in understanding how the existence and the number of solutions of the Vi-SfM problem depend on the motion, on the number of observed point-features, on the point-features layout and on the number of camera images. The resolvability of the Vi-SfM problem can be investigated by computing the null space of the matrix Ξ in (15). Let us denote with $\mathcal{N}(\Xi)$ this space. In [18] we prove the following theorem, which allows us to obtain all the properties of the Vi-SfM problem by investigating the null space of Ξ:

Theorem 3 (Number of Solutions) *The Vi-SfM problem has a unique solution if and only if $\mathcal{N}(\Xi)$ is empty. It has two solutions, if and only if $\mathcal{N}(\Xi)$ has dimension 1 and, for any $\boldsymbol{n} \in \mathcal{N}(\Xi)$, $|\Pi \boldsymbol{n}| \neq 0$. It has infinite solutions in all the other cases.*

In [17] we also prove the following important property:

Property 1 *When $n_i \leq 2$ the dimension of $\mathcal{N}(\Xi)$ is at least 3. When $n_i = 3$ the dimension of $\mathcal{N}(\Xi)$ is at least 1. Finally, when $n_i \geq 4$ and the vehicle moves with constant acceleration the dimension of $\mathcal{N}(\Xi)$ is at least 1.*

Regarding the cases $n_i = 3$ and $n_i \geq 4$ with constant acceleration, the proof is obtained by showing that there is at least one $3D-$vector $\boldsymbol{\alpha}_0$ and one $3D-$vector $\boldsymbol{\nu}_0$ such that the following vector always belongs to $\mathcal{N}(\Xi)$:

$$\boldsymbol{n}_0 = [\boldsymbol{\alpha}_0,\ \boldsymbol{\nu}_0,\ \bar{n}_1^1,\ ..\bar{n}_1^i..,\ ..\bar{n}_j^1..,\ ..\bar{n}_j^i..]^T \tag{18}$$

where $\bar{n}_1^1 = -1$, $\bar{n}_j^1 = 1$, $\bar{n}_1^i = 1$, $\bar{n}_j^i = -1$ ($j = 2, 3$; $i = 2, ..., N_f$). In the case of constant acceleration, $\boldsymbol{\alpha}_0$ is precisely its value. In the following, we discuss the number of solutions of the Vi-SfM problem depending on the number of camera images (n_i).

A. $n_i \leq 2$

From property 1 we know that the dimension of $\mathcal{N}(\Xi)$ is at least 3 and, consequently, by using theorem 3, we conclude that the Vi-SfM has always infinite solutions.

B. $n_i = 3$

From property 1 we know that the dimension of $\mathcal{N}(\Xi)$ is at least 1, independently of the number of point-features. When $N_f = 1$, Ξ is a 6×9 matrix. Hence, the dimension of $\mathcal{N}(\Xi)$ is at least 3. Let us consider the case when $N_f = 2$. In this case Ξ is a 12×12 matrix. In [18] we prove the following properties:

Property 2 ($\boldsymbol{n_i = 3,\ N_f = 2}$) *The dimension of $\mathcal{N}(\Xi)$ is 1 if and only if the following two conditions are met:*

1) *for a given j (e.g., for $j = 2$), the three vectors $\boldsymbol{F}_1^1$, $\boldsymbol{F}_1^2$ and $\boldsymbol{\chi}_j$ span the entire $3D-$space;*
2) *for the other value of j (e.g., for $j = 3$) $\boldsymbol{F}_j^i$ is not proportional to $\boldsymbol{F}_j^k$, $\forall i, k = 1, 2, ..., N_f$.*

From now on, we will say that a condition is satisfied *in general* when the probability that it is not satisfied is zero. We remark that both conditions (i) and (ii) are met in general. For $N_f \geq 2$ we have:

Property 3 ($\boldsymbol{n_i = 3,\ N_f \geq 2}$) *When $n_i = 3$ and $N_f \geq 2$ the Vi-SfM problem has in general two distinct solutions. In some special cases it has infinite solutions.*

C. $n_i \geq 4$

When $n_i \geq 4$ the number of equations is larger than the number of unknowns, except when $n_i = 4$ and $N_f = 1$. In this case the matrix Ξ is 9×10 and the dimension of its null space is at least 1. We have the following property (see [18] for its proof):

Property 4 ($\boldsymbol{n_i = 4,\ N_f = 1}$) *The dimension of $\mathcal{N}(\Xi)$ is 1 if and only if the four vectors $\boldsymbol{F}_1^1$, $\boldsymbol{\chi}_2$, $\boldsymbol{\chi}_3$ and $\boldsymbol{\chi}_4$ span the entire $3D-$space.*

$$\Xi \equiv \left[\begin{array}{c|c|c|c|c|c|c|c|c|c|c}
T_2 & S_2 & \boldsymbol{\mu}_1^1 & 0_3 & 0_3 & -\boldsymbol{\mu}_2^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\
0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & -\boldsymbol{\mu}_1^2 & 0_3 & -\boldsymbol{\mu}_2^1 & \boldsymbol{\mu}_2^2 & 0_3 & 0_3 & 0_3 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & 0_3 & -\boldsymbol{\mu}_1^{N_f} & -\boldsymbol{\mu}_2^1 & 0_3 & \boldsymbol{\mu}_2^{N_f} & 0_3 & 0_3 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
T_{n_i} & S_{n_i} & \boldsymbol{\mu}_1^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & 0_3 & 0_3 \\
0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & -\boldsymbol{\mu}_1^2 & 0_3 & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & \boldsymbol{\mu}_{n_i}^2 & 0_3 \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & 0_3 & -\boldsymbol{\mu}_1^{N_f} & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & 0_3 & \boldsymbol{\mu}_{n_i}^{N_f}
\end{array}\right] \tag{15}$$

We do not derive necessary and sufficient conditions for any value of n_i and N_f. The following property holds (see [18] for its proof):

Property 5 ($\boldsymbol{n_i} \geq 4$) *When $n_i = 4$ and $N_f = 1$ the Vi-SfM problem has in general two distinct solutions. If $n_i = 4,\ N_f \geq 2$ or if $n_i \geq 5,\ \forall N_f$ it has in general a unique solution.*

D. Constant acceleration and constant speed

From property 1 we know that when the vehicle moves with constant acceleration, the dimension of $\mathcal{N}(\Xi)$ is at least 1. In [18] we also provide a sufficient condition which guarantees that the dimension of $\mathcal{N}(\Xi)$ is exactly 1. Hence, the Vi-SfM has two distinct solutions. A special case of constant acceleration occurs when the vector $\boldsymbol{\alpha}_0$ vanishes, i.e., when the vehicle moves with constant speed. Since $|\Pi \boldsymbol{n}_0| = |\boldsymbol{\alpha}_0| = 0$, according to theorem 3, the Vi-SfM has infinite solutions. In [18] we also prove that in this case the roll and pitch angles can be uniquely determined. Hence, we have:

Property 6 (Constant speed) *Let us suppose that the vehicle moves with constant speed. The Vi-SfM has infinite solutions. Additionally, the orientation of the vehicle with respect to the horizontal plane can be uniquely determined.*

Table I summarizes the results of this section by providing the number of solutions case by case. Note that this table does not account the point-features layout. Specifically, the motion and the point-features are not supposed to be either coplanar or collinear. Regarding these cases, necessary conditions are provided in [18] where the previous analysis is also extended in order to cope the case of biased accelerometer's measurements.

Cases	Number of Solutions
Varying Acceleration $n_i = 4,\ N_f \geq 2$; $n_i \geq 5,\ \forall N_f$	Unique Solution
Varying Acceleration $n_i = 3,\ N_f \geq 2$; $n_i = 4,\ N_f = 1$	Two Solutions
Constant and non null Acceleration $n_i = 3,\ N_f \geq 2$; $n_i \geq 4,\ \forall N_f$	Two Solutions
Constant Speed $\forall n_i,\ \forall N_f$	Infinite Solutions
Any Motion $n_i \leq 2,\ \forall N_f$; $n_i = 3,\ N_f = 1$	Infinite Solutions

TABLE I

NUMBER OF DISTINCT SOLUTIONS FOR THE VI-SFM PROBLEM.

VI. CLOSE SOLUTION IN CONJUCTION WITH AN EKF

In this section we show the benefit of using the closed solution for initializing a filter based approach to solve the Vi-SfM problem. Specifically, we generate noisy visual and inertial measurements through Monte Carlo simulations. Additionally, we corrupt the measurements provided by the accelerometers and the gyroscopes with a time dependent bias and we consider the case when the transformation between the visual and inertial sensors is not perfectly known. In section IV we formulated the Vi-SfM problem as the problem of determining the vectors: $\boldsymbol{F}^i$, $(i = 1,\ 2,\ ...,\ N_f)$ and $\boldsymbol{V}$, $\boldsymbol{G}$. For the sake of clarity, in this section we choose to display the results in a global frame. For this reason, we need to consider at least two point-features. Indeed, two is the minimum number of point-features to uniquely define a global frame, provided that they do not lie on the same vertical axis (defined by the gravity). We define the global frame as follows: first, we define one of the point-feature as the origin of the frame. The z-axis coincides with the gravity axis but with opposite direction. Finally, the x-axis is defined by requiring that the second point-feature belongs to the xz-plane. In other words, the second point-feature has zero $y-$coordinate. In these settings, the Vi-SfM can be defined as the estimation of the vehicle configuration and the estimation of the x and the z coordinate of the second point-feature (from now on, p_x and p_z, respectively). By adding more point-features, the state to be estimated also includes all the three coordinates of each point-feature. We adopt an Extended Kalman Filter (EKF) to perform this estimation. The state to be estimated is:

$$\boldsymbol{x^e} \equiv [\boldsymbol{r},\ \boldsymbol{v},\ q,\ p_x,\ p_z,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{f}^3,\ ...,\ \ \boldsymbol{f}^{N_f}]^T$$

By collecting the sensor measurements during the time-interval $[T_{in},\ T_{fin}]$, the closed solution discussed in the previous sections allows us to determine the vectors $\boldsymbol{F}^i$, $(i = 1,\ 2,\ ...,\ N_f)$, $\boldsymbol{V}$ and $\boldsymbol{G}$ at the time T_{in}. Note that, when $N_f \geq 2$, having the vectors $\boldsymbol{F}^i$, $\boldsymbol{V}$ and $\boldsymbol{G}$ at the time T_{in}, allows us to build the state $\boldsymbol{x^e}$ at time T_{in} (with

the exception of $\boldsymbol{A^{bias}}$ and $\boldsymbol{\Omega^{bias}}$). Since our simulated measurements are corrupted by noise and also include a bias on the IMU and an error on the extrinsic camera-IMU calibration, the values obtained with the closed solution will differ from the true values.

In this section, we investigate how the performance of the EKF depends on its initialization and how this performance can be improved by using the closed solution to initialize the state. Since the closed solution does not provide the initial $\boldsymbol{A^{bias}}$ and $\boldsymbol{\Omega^{bias}}$, their initial values will be set to zero.

1) Simulated Trajectories: All the trajectories are randomly generated starting from the following initial true state:

$$\boldsymbol{r}(T_{in}) = [0.5,\ 0.5,\ 0.5]m;\ \boldsymbol{v}(T_{in}) = [0.1,\ 0.1,\ 0.1]ms^{-1};$$

$q(T_{in}) = 1$, which corresponds to the vehicle attitude $roll = pitch = yaw = 0\ deg$; $\boldsymbol{A^{bias}}(T_{in}) = 0.05\ \boldsymbol{\hat{\mu}}\ m\ s^{-2}$, $\boldsymbol{\Omega^{bias}}(T_{in}) = 0.5\ \boldsymbol{\hat{\mu}}\ deg\ s^{-1}$ where $\boldsymbol{\hat{\mu}}$ is the unit vector pointing in the direction $[1,\ 1,\ 1]$; $p_x = 2m$ and $p_z = 1m$. Both the biases are time-dependent. Specifically, they are modelled as independent random walks (for all the three components of both), whose mean values are the initial ones and their variances increase linearly with time. For the gyroscopes, the three variances are set equal to $(50\ deg/h)^2$ at $100\ s$ and for the accelerometers are set equal to $(1\ m/h^2)^2$ at $100\ s$ (see [25]). We assume that the camera and the IMU frame coincide (i.e., they have the same origin and the same orientation). We characterize an error in the extrinsic calibration by setting the actual position of the origin of the camera frame in the IMU frame to $[0.002,\ -0.003,\ 0.004]m$ and the actual orientation $q_{cam} = 1 - 2.3\ 10^{-5} + (3.5i - 5.2j + 2.6k)\ 10^{-3}$, which corresponds to the attitude $roll = 0.4\ deg$, $pitch = -0.6\ deg$ and $yaw = 0.3\ deg$.

We also considered the case of more than two point-features ($N_f \geq 3$), obtaining similar results in terms of performance and, for the sake of brevity, in the following we only refer to the case of $N_f = 2$.

The trajectories are generated by randomly generating the linear and angular acceleration of the vehicle at $100\ Hz$. In particular, at each time step, the three components of the linear acceleration and the angular speed are generated as zero-mean Gaussian independent variables whose covariance matrices are equal to $(1ms^{-2})^2 I_3$ and $(10\ deg\ s^{-1})^2 I_3$, respectively.

2) Simulated Sensors: Starting from the accomplished trajectory, the true angular speed and the linear acceleration are computed at each time step of $0.01s$ (respectively, at the jth time step, we denote them with $\boldsymbol{\Omega_j^{true}}$ and $\boldsymbol{A_j^{true}}$). Starting from them, the IMU sensors are simulated by randomly generating the angular speed and the linear acceleration at each step according to the following:

$$\boldsymbol{\Omega_j} = N\left(\boldsymbol{\Omega_j^{true}} - \boldsymbol{\Omega^{bias}}(t_j),\ P_\Omega\right)$$

$$\boldsymbol{A_j} = N\left(\boldsymbol{A_j^{true}} - \boldsymbol{G}(t_j) - \boldsymbol{A^{bias}}(t_j),\ P_A\right)$$

where:

- $N(.,\ .)$ indicates the Normal distribution whose first entry is the mean value and the second its covariance matrix;
- P_Ω and P_A are the covariance matrices characterizing the accuracy of the IMU.

In all the simulations we set both the matrices P_Ω and P_A diagonal and in particular: $P_\Omega = (1\ deg\ s^{-1})^2\ I_3$ and $P_A = (1\ cm\ s^{-2})^2\ I_3$.

Regarding the camera, the provided readings are generated in the following way. By knowing the true trajectory and the true camera-IMU transformation, the true bearing angles of the two point-features in the camera frame are computed. They are computed each $0.1s$. Then, the camera readings are generated by adding to the true values zero-mean Gaussian errors whose variance is equal to $(1\ deg)^2$ for all the readings.

3) Simulation Results: We first investigate the convergence of the EKF vs the initialization of the state. In all the considered initializations we set the initial accelerometer and gyroscope biases to zero. In general, the EKF diverges when: (a) the initial scale factor error exceeds 20%; (b) the initial attitude error exceeds $4\ deg$. These conclusions on the EKF convergence have been obtained by running many simulations with the settings specified in VI-.1 and VI-.2. As an illustration, we display here the results obtained with a particular trial. Figures $2a - d$ display the trajectories estimated by the EKF when the initial state differs from the true state because of an error on the absolute scale and on the attitude (as said, the initial state is also affected by an error on the inertial sensors' biases since they are always initialized to zero). Figure $2a$ displays the true trajectory (blue) together with the one estimated by only using inertial measurements (black) and the one estimated by the EKF with an initial absolute scale set to 1.1 times the true value and an error of $1\ deg$ on the roll, pitch and yaw angles. Figure $2b$ displays the trajectories estimated by the EKF with an initial state affected by an error on the attitude (same error on the roll, pitch and yaw) and correct absolute scale. Figures $2c$ and d display the trajectories estimated by the EKF with an initial state affected by an error on the absolute scale and correct attitude.

By using the first 6 camera observations (i.e. by considering the time interval $[T_{in} = 0,\ T_{fin} = 0.6]s$) we obtain the initial position $[0.4961,\ 0.4975,\ 0.5017]m$, the initial speed $[0.1024,\ 0.1028,\ 0.1222]m\ s^{-1}$ and the initial attitude $q = 1 - 4.3\ 10^{-6} + (1.0i - 2.3j + 1.6k)\ 10^{-3}$, which corresponds to the attitude $roll = 0.11\ deg$, $pitch = -0.26\ deg$ and $yaw = 0.18\ deg$. By running many simulations, we found that the initial state determined through the closed solution is never affected by an error larger than 8% regarding the absolute scale and than $0.7\ deg$ regarding the attitude.

VII. Conclusion

In this paper we introduced two novel contributions in the framework of the Vi-SfM . The first contribution extends the results of the state of the art about the Vi-SfM observability.

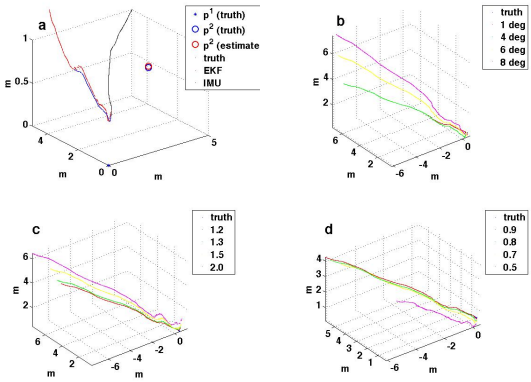

Fig. 2. Fig a displays the true trajectory (blue), the trajectory estimated by the EKF (red) with initial errors (10% on the scale and $1\ deg$ on the attitude) and the trajectory estimated by only using inertial measurements (black). Figs b, c and d display the trajectories estimated by the EKF when the initial state is affected by an error on the attitude (b) and on the absolute scale (larger and smaller than the true one, in c and d, respectively).

Specifically, it has been proven that, even in the case of a single point feature, the visual and inertial sensor provide the necessary information to determine the scale, the vehicle speed, the absolute roll and pitch, the inertial biases, the magnitude of the gravity and the camera extrinsic calibration in the IMU frame. To achieve this result, new techniques able to perform an observability analysis have been introduced. It is remarkable to note that, by using these techniques, it is possible to check the independence of the Lie derivatives up to the third order in less than 200 seconds while, on the same processor and for the same problem, the computational time required to check the independence of the Lie derivatives up to the second order is larger than 10^5 seconds without using these techniques. The second contribution is the derivation of a simple and intuitive closed solution to the Vi-SfM problem. We used this derivation to investigate the intrinsic properties of the Vi-SfM problem and to identify the conditions under which the problem can be solved in closed form. In particular, we showed that the problem can have a unique solution or two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and their layout and on the number of monocular images where the same point-features are seen. The most useful applications of the closed-form solution here derived will be in all the applicative domains which need to solve the structure from motion problem with low-cost sensors and which do not demand any infrastructure (e.g., in GPS denied environment). Additionally, our results could also play an important role in the framework of neuroscience by providing a new insight on the process of vestibular and visual integration for depth perception and self-motion perception.

REFERENCES

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

[2] 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

[3] Alessandro Chiuso, Paolo Favaro, Hailin Jin and Stefano Soatto, "Structure from Motion Causally Integrated Over Time", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), pp 523–535, 2002

[4] A. J. Davison, I. D. Reid, N. D. Molton and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), pp 1052–1067, 2007

[5] 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

[6] T.C Dong-Si, A.I. Mourikis, Initialization in Vision-aided Inertial Navigation with Unknown Camera-IMU Calibration," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal, October 7-12 2012, pp. 1064-1071.

[7] J. A. Farrell, Aided Navigation: GPS and High Rate Sensors. McGraw-Hill, 2008.

[8] 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

[9] Chao X. Guo and Stergios I. Roumeliotis, IMU-RGBD Camera 3D Pose Estimation and Extrinsic Calibration: Observability Analysis and Consistency Improvement, 2013 IEEE International Conference on Robotics and Automation (ICRA) Karlsruhe, Germany, May 6-10, 2013

[10] Richard I. Hartley (June 1997). "In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence 19 (6): 580–593.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] M. Li and A. I. Mourikis, Improving the accuracy of EKF-based visualinertial odometry, in Proc. of the IEEE Int. Conf. on Robotics and Automation, St. Paul, MN, May 2012.

[16] H. C. Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature 293: 133–135.

[17] A. Martinelli, Visual-inertial structure from motion observability, INRIA technical Report (2013)

[18] A. Martinelli, Resolvability of Visual-Inertial Structure from Motion in Closed-form, INRIA technical Report (2012)

[19] 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

[20] 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.

[21] 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, No. 5, October 2008, pages 1143–1156

[22] 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

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

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

[25] Woodman, Oliver J., An introduction to inertial navigation, Technical Report, University of Cambridge, 2007, UCAM-CL-TR-696