Visual-Inertial Structure from Motion Observability

Agostino Martinelli

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Visual-Inertial Structure from Motion Observability

Agostino Martinelli

N° 8272

Mars 2013

Thèmes COG et NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--8272--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche GRENOBLE - RHÔNE-ALPES

Visual-Inertial Structure from Motion Observability

Agostino Martinelli

Thèmes COG et NUM — Systèmes cognitifs et Systèmes numériques
Équipe-Projet Emotion

Rapport de recherche n° 8272 — Mars 2013 — 15 pages

Abstract: This paper investigates the visual inertial structure from motion problem with special focus on its observability properties. The considered system is a sensor assembling consisting of a monocular camera and inertial sensors (i.e., three orthogonal accelerometers and three orthogonal gyroscopes). The considered state contains the vehicle speed and attitude, the biases on the inertial measurements, the position of the features observed by the camera, the parameters characterizing the extrinsic camera-inertial sensors calibration and the magnitude of the gravity. In the case of a single feature, this state consists of 24 components. The observability analysis is carried out by using the method of *continuous symmetries*, which has recently been introduced. Due to the system complexity, a direct derivation of the system continuous symmetries would require an excessive computational cost. For this reason, new theoretical results are derived in order to significantly reduce the load of symbolic computation. The paper contribution is therefore twofold. From one side it provides new theoretical results, which allow significantly reducing the symbolic computational load necessary to investigate the observability properties of any estimation problem by making the use of the concept of continuous symmetry. From the other side, it provides a new result in the framework of the observability of the visual-inertial structure from motion. Specifically, 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.

Key-words: Sensor Fusion, Observability, Inertial Sensors, Vision, Structure from Motion

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Résumé : Cet article étudie le problème *visual inertial structure from motion* avec un accent particulier sur ses propriétés d'observabilité. Le système considéré est un assemblage constitué par une caméra monoculaire et des capteurs inertiels (c'est à dire trois accéléromètres orthogonaux et trois gyroscopes orthogonales). L'état considéré contient la vitesse du véhicule et l'attitude, les bias sur les mesures inertielles, la position des amers observés par la caméra, les paramètres caractérisant la calibration extrinsèque caméra -capteurs inertielles et l'ampleur de la gravité. Dans le cas d'une seule amer, cet état se compose de 24 éléments. L'analyse d'observabilité est effectuée en utilisant la méthode de *symétries continues*, qui a récemment été mis en place. En raison de la complexité du système, une dérivation directe des symétries continues du système nécessiterait un coût excessif de calcul. Pour cette raison, de nouveaux résultats théoriques sont dérivées afin de réduire considérablement la charge de calcul symbolique. La contribution du papier est donc double. D'un côté, il fournit de nouveaux résultats théoriques, qui permettent de réduire considérablement la charge symbolique de calcul nécessaire pour étudier les propriétés d'observabilité de n'importe quel problème d'estimation en faisant l'usage de la notion de symétrie continue. De l'autre côté, il fournit un nouveau résultat dans le cadre de l'observabilité du problème *visual inertial structure from motion*. Plus précisément, il est prouvé que, l'information contenue dans les données fournies par une caméra monoculaire qui observe un seul amer et par les capteurs inertielles permet d'estimer l'échelle absolue, la vitesse dans le repère local, l'angle de roulis et de tangage, les bias qui affectent les mesures des accéléromètres et des gyroscopes, l'ampleur de la gravité et la calibration extrinsèque caméra -capteurs inertielles.

Mots-clés : Fusion Sensoriel, Observabilité, Capteurs inertiels, Vision, Structure from Motion

1 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 [4, 5, 8, 12, 16] and a closed form solution has also been derived [8, 12, 16], 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 [3, 10, 11, 14, 17] and in the framework of neuroscience (e.g., [2, 6, 7]). 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.

In [9, 10, 11, 14] and [18] the observability properties of the Vi-SfM have been investigated. In [9] the authors investigated the estimator inconsistency in the Vi-SfM problem. They 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 [10, 11] and [18] the observability properties have been derived by accounting an unknown transformation between the camera and the IMU[1] frames and an unknown magnitude of the gravity. Additionally, in [11] and [18] also the case of biased inertial measurements has been considered. We remark that in [10, 11] and [18] the observability properties have been derived starting from basic results in computer vision. Specifically, in [10] and [11], starting from the results derived in [4], a global frame has been fixed by constraining three directions determined by three points on the image plane. In [18], 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 [16]) 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 [14] 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 [13]. Since under these conditions the camera observation has an expression much more complex, the analysis in [14] was limited to the case when the camera extrinsic calibration in the IMU frame is a priori known.

In this paper we extend the observability analysis carried out in [14] in order to cope with the case of unknown camera extrinsic calibration. In order to achieve this objective, the theory introduced in [13] 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

[1]In this paper IMU is the Inertial Measurement Unit, which consists of three orthogonal accelerometers and three orthogonal gyroscopes.

the system symmetries. Specifically, we introduce the *quasi-projection* operation which allows us to reduce the space dimensionality when deriving the observability properties of any estimation problem. The considered system together with its basic equations are provided in section 2. In section 3 the observability properties of this system are derived by firstly introducing the quasi-projection operation (3.1) and then by applying this operation to our system (3.2). Finally, conclusions are provided in section 4.

2 The Considered System

We consider a system (from now on we call it the *vehicle*) consisting of a monocular camera and an Inertial Measurement Unit (*IMU*). The IMU consists of three orthogonal accelerometers and three orthogonal gyroscopes. 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:

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

being q^* the conjugate of q, $q^* = q_t - i q_x - j q_y - k q_z$. 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 [14] 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:

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

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:

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

where:

- $M(\boldsymbol{\Omega}) \equiv \begin{bmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{bmatrix}$;

- ${}^{c}\boldsymbol{\Omega}$ 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.

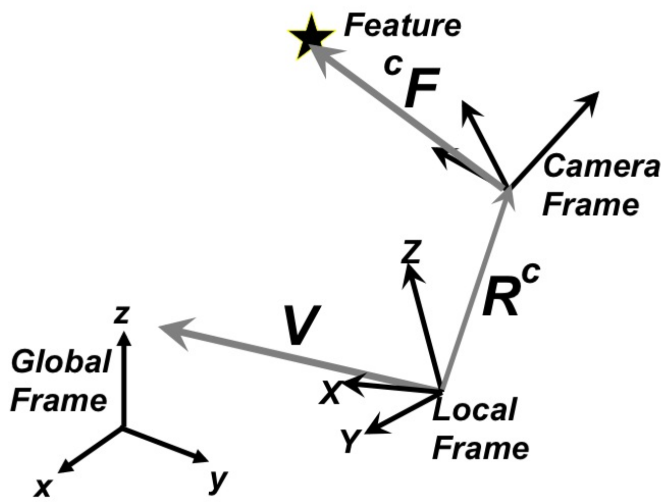

Figure 1: Global frame, local (IMU) frame and camera frame with the feature position (${}^{c}\boldsymbol{F}$) in the camera frame and the vehicle speed ($\boldsymbol{V}$) and the camera position ($\boldsymbol{R^c}$) in the local frame.

As the majority of real control systems, the dynamics given in (5) are affine in the controls, i.e. they can be written as follows:

$$\begin{cases} \dot{\boldsymbol{S}} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{u}) = \boldsymbol{f_0}(\boldsymbol{S}) + \sum_{i=1}^{n_c} \boldsymbol{f_i}(\boldsymbol{S}) u_i \end{cases} \tag{6}$$

with $\boldsymbol{S} = \boldsymbol{X_n}$, $n_c = 6$ and $\boldsymbol{u} = [u_1,\ u_2,\ ...,\ u_{n_c}]^T = [A_x,\ A_y,\ A_z,\ \Omega_x,\ \Omega_y,\ \Omega_z]^T$. The expression of the vector fields $f_i,\ i = 0, 1, ..., n_c$, which is necessary for the computation of the Lie derivatives, can be obtained by a direct computation[2].

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 ${}^{c}\boldsymbol{F}$ up to a scale, or, equivalently, the two following ratios:

$$\boldsymbol{h_{cam}}(\boldsymbol{X_n}) \equiv [h_u,\ h_v]^T = \left[\frac{{}^{c}F_x}{{}^{c}F_z},\ \frac{{}^{c}F_y}{{}^{c}F_z}\right]^T \tag{7}$$

[2]The reader is addressed to [13] for the definition and details about the computation of the Lie derivatives.

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

$$\boldsymbol{h_{const}}(\boldsymbol{X_n}) \equiv [h_q,\ h_{q^c}]^T = [q^*q,\ (q^c)^*q^c]^T \tag{8}$$

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., $\boldsymbol{{}^cF} \to \boldsymbol{{}^cF^i}$, $i = 1, 2, \cdots, N_f$. The resulting state has dimension $24 + 3(N_f - 1)$.

3 Observability Properties

In [14] we investigated the observability properties of the Vi-SfM problem 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 [14] can be summarized as follows:

Theorem 1 (Known extrinsic calibration) *Let us consider the Vi-SfM problem 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 $127683s$. One of the goal of this paper is to show how the theory developed in [13] can be used to deal with such complex cases, namely, how the observability properties can be derived by avoiding the computation of high-order Lie derivatives and how the space dimensionality where the rank must be computed can be reduced.

3.1 Quasi-Projection Operation and Matrix

In this section we introduce new theoretical results which allow significantly reducing the load of symbolic computation requested in order to derive the observability properties of a given system starting from the analysis of its continuous symmetries. Throughout this section, we denote by n the dimension of the state. We start by reminding the reader the definition of *continuous symmetry* and the basic result derived in [13], which characterizes the observable modes.

Definition 1 (Continuous Symmetry) *The vector field $\boldsymbol{w_s}(\boldsymbol{S})$ is a continuous symmetry in $\boldsymbol{S}$ for the system defined in (6) if and only if it is a non null vector belonging to the null space of the matrix whose lines are the gradients of all the Lie derivatives computed in $\boldsymbol{S}$.*

Property 1 *$g(\boldsymbol{S})$ is an observable mode if and only if its gradient is orthogonal to all the symmetries.*

According to this result, the derivation of all the observable modes requires, first of all, to derive all the symmetries. The following remark provides an upper bound on the number of Lie derivatives which must be computed in order to detect the system symmetries:

Remark 1 *In order to detect the symmetries of the system in (6), only the first $(n-1)-$order Lie derivatives must be computed.*

The reader is addressed to [1] to verify the validity of this remark. On the other hand, this remark is not useful to deal with our case since it is prohibitive even to deal with the second-order Lie derivatives.

Let us denote with $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$ $(1 \leq n' < n)$ a set of independent column vectors of dimension n which are orthogonal to the gradients of $n - n'$ independent Lie derivatives. We have the following result:

Property 2 *A sufficient condition for $g(\boldsymbol{S})$ to be an observable mode is that its gradient is orthogonal to all the vectors $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$.*

Proof: According to definition 1, if $n - n'$ is the number of all the independent Lie derivatives, $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$ are all the symmetries of the system. In this case, as stated in property 1, the previous condition is also necessary. When $n - n'$ is smaller than the number of all the independent Lie derivatives, we only know that the vector space generated by the symmetries is a subspace of the space generated by $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$. Hence, if the gradient of $g(\boldsymbol{S})$ is orthogonal to this latter vector space it is also orthogonal to any of its subspace. From property 1 we conclude that $g(\boldsymbol{S})$ is an observable mode ■

This result is useful since it can be used by computing any number of independent Lie derivatives and not necessarily all of them. Additionally, this simple result allows us to reduce the system dimensionality in the special case when a component of the state verifies the sufficient condition stated in the previous property. In this case, since we know that this component is observable, in order to detect the other observable modes we can consider the reduced system, which is obtained by removing this component from the state.

The second new result allows us to determine the dependence or independence of a given set of Lie derivatives by computing the rank of matrices whose dimension is smaller than the dimension of the original state. Let us suppose that we detected $n - n'$ independent Lie derivatives ($\mathcal{L}^1$, ..., $\mathcal{L}^{n-n'}$) and, as before, let us denote with $\boldsymbol{\phi}^{\mathbf{1}}$, ..., $\boldsymbol{\phi}^{\boldsymbol{n'}}$ ($1 \leq n' < n$) a set of independent column vectors orthogonal to their gradients. Let us suppose that we compute m additional Lie derivatives ($\mathcal{P}^1$, ..., $\mathcal{P}^m$). We need to determine if they are independent from the previous ones and also if they are independent one each other. In other words, we want to know the number of independent Lie derivatives in the set $\mathcal{L}^1$, ..., $\mathcal{L}^{n-n'}$, $\mathcal{P}^1$, ..., $\mathcal{P}^m$. The easiest way to determine this is to compute their gradients and then the rank of the matrix which contains these gradients. The dimension of this matrix is $(n - n' + m) \times n$. We know that this rank is larger or equal to $n - n'$ since we assumed that $\mathcal{L}^1$, ..., $\mathcal{L}^{n-n'}$ are independent. Let us denote this rank by $n-n'+p$. We remark that determining p is very easy if the vectors $\boldsymbol{\phi}^{\mathbf{1}}$, ..., $\boldsymbol{\phi}^{\boldsymbol{n'}}$ are orthonormal. In this special case, it is immediate to project the gradients of the Lie derivatives ($\mathcal{P}^1$, ..., $\mathcal{P}^m$) onto the vectors $\boldsymbol{\phi}^{\mathbf{1}}$, ..., $\boldsymbol{\phi}^{\boldsymbol{n'}}$. This allows us to reduce the space dimensionality: instead of working in the original $n-$dimensional space we can work in a $n'-$dimensional space. In other words, p is the rank of the $m \times n'$ matrix, whose lines are the projections of the gradients of the Lie derivatives ($\mathcal{P}^1$, ..., $\mathcal{P}^m$) onto the vectors $\boldsymbol{\phi}^{\mathbf{1}}$, ..., $\boldsymbol{\phi}^{\boldsymbol{n'}}$. On the other hand, even in the case when the vectors $\boldsymbol{\phi}^{\mathbf{1}}$, ..., $\boldsymbol{\phi}^{\boldsymbol{n'}}$ are not orthonormal, p can be computed by computing a rank of a $m \times n'$ matrix. We have the following result:

Property 3 *p is the rank of the $m \times n'$ matrix, whose i, j entry is the scalar product $\boldsymbol{\nabla_S}\mathcal{P}^i \cdot \boldsymbol{\phi}^{\boldsymbol{j}}$.*

Proof: In this proof we will use the following two results from linear algebra about the matrix rank [15]:

1. $rank(AB) = rank(A)$ when $rank(B) = k$, where k is the number of lines of B;
2. If A is a matrix over the real numbers, $rank(AA^T) = rank(A^TA) = rank(A) = rank(A^T)$.

Let us introduce the following notation:

- $\mathcal{D}$ is the $(n - n') \times n$ matrix whose lines are the gradients of the functions $\mathcal{L}^1$, ..., $\mathcal{L}^{n-n'}$, i.e.:
$$\mathcal{D} \equiv \begin{bmatrix} \boldsymbol{\nabla_S}\mathcal{L}^1 \\ \cdots \\ \boldsymbol{\nabla_S}\mathcal{L}^{n-n'} \end{bmatrix}$$
- $\mathcal{E}$ is the $m \times n$ matrix whose lines are the gradients of the functions $\mathcal{P}^1$, ..., $\mathcal{P}^m$, i.e.:
$$\mathcal{E} \equiv \begin{bmatrix} \boldsymbol{\nabla_S}\mathcal{P}^1 \\ \cdots \\ \boldsymbol{\nabla_S}\mathcal{P}^m \end{bmatrix}$$

- $\mathcal{M}$ is the $(n-n'+m)\times n$ matrix whose lines are all the previous gradients, i.e.: $\mathcal{M} \equiv \begin{bmatrix} \mathcal{D} \\ \mathcal{E} \end{bmatrix}$
- $\mathcal{F}$ is the $n\times n'$ matrix whose columns are the vectors $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$;
- $\mathcal{N}$ is the following $n\times n$: $\mathcal{N} \equiv \begin{bmatrix} \mathcal{D}^T & \mathcal{F} \end{bmatrix}$

We have: $n-n'+p = rank(\mathcal{M})$. Since $rank(\mathcal{N}) = n$, by using the first property mentioned at the beginning of this proof, we also have $rank(\mathcal{M}) = rank(\mathcal{M}\mathcal{N})$. We have:

$$\mathcal{M}\mathcal{N} = \begin{bmatrix} \mathcal{D}\mathcal{D}^T & 0_{(n-n')\times n'} \\ \mathcal{E}\mathcal{D}^T & \mathcal{E}\mathcal{F} \end{bmatrix} \tag{9}$$

where $0_{(n-n')\times n'}$ is the zero $(n-n')\times n'$ matrix. Because of the second property mentioned at the beginning, $rank(\mathcal{D}\mathcal{D}^T) = rank(\mathcal{D})$, which is $n-n'$, because of the independence of $\mathcal{L}^1$, ..., $\mathcal{L}^{n-n'}$. In order to compute the rank of the matrix in (9), we detect the largest number of independent columns. Let us consider any linear combination of its columns which includes all the first $n-n'$ columns. A necessary condition to be the null vector is that the first $n-n'$ coefficients are equal to zero. Hence, a necessary and sufficient condition such that this linear combination vanishes if and only if all the coefficients are equal to zero, is that the columns selected among the last m columns of the matrix $\mathcal{M}\mathcal{N}$ are independent among them. This proves that $rank(\mathcal{M}\mathcal{N}) = n-n'+rank(\mathcal{E}\mathcal{F})$. But the i, j entry of $\mathcal{E}\mathcal{F}$ is precisely the scalar product $\boldsymbol{\nabla_S}\mathcal{P}^i \cdot \boldsymbol{\phi^j}$ ■

This result is very important since allows us to verify the independence of a set of Lie derivatives by computing the rank of matrices with low dimension. Specifically, the gradients of the Lie derivatives are $n-$dimensional vectors. Once we have detected $n-n'$ independent Lie derivatives and a basis for its null space (i.e., $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$), in order to detect further independent Lie derivatives, we can work with $n'-$dimensional vectors. Indeed, for a given Lie derivative $\mathcal{P}$, instead of working with the $n-$dimensional vector $\boldsymbol{\nabla_S}\mathcal{P}$, we can work with the $n'-$dimensional vector, which is defined as the vector whose j^{th} entry is $\boldsymbol{\nabla_S}\mathcal{P} \cdot \boldsymbol{\phi^j}$ $(j = 1,\cdots,n')$. This operation would be a projection if the basis $\boldsymbol{\phi^1}$, ..., $\boldsymbol{\phi^{n'}}$ were orthonormal. For this reason we will call this operation a *quasi-projection*. This result is very powerful since it can be applied more consecutive times. Specifically, let us suppose that, by quasi-projecting the gradients of additional Lie derivatives in the $n'-$dimensional space we detect additional $n'-n''$ independent Lie derivatives $(1 \le n'' < n')$. Let us denote with $\boldsymbol{\phi'^1}$, ..., $\boldsymbol{\phi'^{n''}}$ a basis of the null space of the matrix made by these gradients quasi-projected in the $n'-$dimensional space. Now, in order to work in the new $n''-$dimensional space, we must quasi-project the gradient of a given Lie derivative, first on the $n'-$dimensional space and then on the $n''-$dimensional space. By computing the matrix whose i, k entry is $\sum_j \boldsymbol{\phi_i^j}\boldsymbol{\phi_j'^k}$ we can simply multiply the gradient of the Lie derivative by this matrix. We we will call this matrix the *quasi-projection* matrix. Therefore, each time we apply the result in property 3, we need to compute the new quasi-projection matrix.

3.2 Vi-SfM Observability

We illustrate the power of the previous two properties by referring to our case, where the original state has dimension $n = 24$. We remind the reader that 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 $101734s$. We start by considering the matrix whose lines are the gradients of the following Lie derivatives: $L^0 h_u$, $L^0 h_v$, $L^0 h_q$, $L^0 h_p$, $L^1_{\boldsymbol{f_5}} h_u$, $L^1_{\boldsymbol{f_6}} h_u$, $L^1_{\boldsymbol{f_6}} h_v$. The computation of its rank and its null space demands $0.37s$ and $0.24s$, respectively. In particular, the rank is 7 proving that these functions are independent. In a second step we quasi-project the gradient of the following Lie derivatives: $L^1_{\boldsymbol{f_0}} h_u$ and $L^1_{\boldsymbol{f_4}} h_u$. We build with them a 2×17 matrix. The computation of its rank and null space requires $0.027s$ and $3.11s$, respectively. The rank is 2. The computation of the quasi-projection matrix requires $6.48s$. In a third step we quasi-project the gradient of the following Lie derivatives: $L^1_{\boldsymbol{f_4}} h_v$ and $L^1_{\boldsymbol{f_5}} h_v$. We build with them a 2×15 matrix. The computation of its rank and null space requires $64s$ and $81s$, respectively. The rank is 2. The computation of the quasi-projection matrix requires $0.34s$. Finally, we quasi-project the gradient of the Lie derivative $L^1_{\boldsymbol{f_0}} h_v$. We build with it a 1×13 matrix. The computation of its rank and null space requires $0.041s$ and $0.089s$, respectively. The computation of the quasi-projection matrix requires $0.44s$. We can proceed further in order to detect all the independent Lie derivatives. However, it is more convenient to use property 2. Specifically, we can check if some of the components of the state are observable (property 2 is a sufficient condition). We compute the gradient of the components of the original state and we quasi-project them on the final $12-$dimensional space. Regarding the four components of the quaternion q^c we obtain the null vector, meaning that these components are four observable modes. Hence, we remove these coponents from the original state and we start again the observability analysis of a new system whose state has now dimension equal to 20. As previously mentioned, the complexity of the computation dramatically depends on the dimension of the state. Specifically, with the new state, it is possible to work with third-order Lie derivatives. In particular, we found the following 19 independent Lie derivatives: $L^0 h_u$, $L^0 h_v$, $L^0 h_q$, $L^1_{\boldsymbol{f_0}} h_u$, $L^1_{\boldsymbol{f_5}} h_u$, $L^1_{\boldsymbol{f_6}} h_u$, $L^1_{\boldsymbol{f_0}} h_v$, $L^1_{\boldsymbol{f_6}} h_v$, $L^2_{\boldsymbol{f_0 f_0}} h_u$, $L^2_{\boldsymbol{f_0 f_1}} h_u$, $L^2_{\boldsymbol{f_0 f_4}} h_u$, $L^2_{\boldsymbol{f_0 f_5}} h_u$, $L^2_{\boldsymbol{f_0 f_6}} h_u$, $L^2_{\boldsymbol{f_0 f_0}} h_v$, $L^2_{\boldsymbol{f_0 f_4}} h_v$, $L^3_{\boldsymbol{f_4 f_0 f_0}} h_u$, $L^3_{\boldsymbol{f_0 f_0 f_4}} h_u$, $L^3_{\boldsymbol{f_0 f_0 f_5}} h_u$, $L^3_{\boldsymbol{f_0 f_0 f_6}} h_v$. The computation of the null space of the matrix whose lines are their gradients requires $0.46s$ and provides the vector:

$$\boldsymbol{w_s^{Rot_z}} = [0_{1\times 6}, -q_z, -q_y, q_x, q_t, 0_{1\times 10}]^T$$

where $0_{i\times j}$ denotes the $i\times j$ zero matrix. This vector expresses the system invariance under rotations around the vertical axis. Indeed, an infinitesimal rotation of magnitude ϵ about the vertical axis leaves the vectors ${}^{\boldsymbol{c}}\boldsymbol{F}$, $\boldsymbol{V}$, $\boldsymbol{A^{bias}}$, $\boldsymbol{\Omega^{bias}}$ and $\boldsymbol{R^c}$ unchanged and the quaternion q changes as follows [?]:

$$\begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} \rightarrow \begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} + \frac{\epsilon}{2} \begin{bmatrix} -q_z \\ -q_y \\ q_x \\ q_t \end{bmatrix}$$

i.e., the reduced 20−dimensional state $\boldsymbol{X_r} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F},\ \boldsymbol{V},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{R^c},\ g\right]^T$ changes as follows: $\boldsymbol{X_r} \to \boldsymbol{X_r} + \frac{\epsilon}{2}\boldsymbol{w_s^{Rot_z}}$. Hence, $\boldsymbol{w_s^{Rot_z}}$ is a continuous symmetry for our system and the only non-observable mode is the yaw angle. In other words, we extended the result in theorem 1 to the case of unknown camera-IMU calibration. We have the following new result:

Theorem 2 (Unknown extrinsic calibration) *Let us consider the Vi-SfM problem 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 three 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.*

Note that this result holds even in the case when the camera observes a single point-feature.

4 Conclusion

In this paper we extended the results of the state of the art about the Vi-SfM observability. Specifically, it has been proven that, even in the case of a single point feature, monocular vision and inertial sensors 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, we introduced the operation of *quasi projection* and new results which allow us to perform an observability analysis with a reduced load of symbolic computation. It is remarkable to note that, by using these new results, 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.

References

[1] Anguelova M., Non linear Observability and Identifiability: General Theory and a Case Study of a Kinetic Model, PhD thesis, Chapter 4, Goteborg, April 2004

[2] 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).

[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] 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

[5] Andrew J. Davison, Ian D. Reid, Nicholas D.Molton and Olivier Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), pp 1052–1067, 2007

[6] 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

[7] 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

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

[9] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, Towards consistent vision-aided inertial navigation, in Proc. of the Int. Workshop on the Algorithmic Foundations of Robotics, Cambridge, MA, Jun. 1315, 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] H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature 293: 133–135.

[13] A. Martinelli, State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, Transactions on Robotics, Vol. 27, No. 2, pp 239–255, April 2011

[14] A. Martinelli, Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, Transaction on Robotics, Volume 28 (2012), Issue 1 (February), pp 44–60.

[15] Leon Mirsky: An Introduction to Linear Algebra, 1990, ISBN 0-486-66434-1

[16] 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

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

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

Contents

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399