



Self Calibration of a vision system embedded in a Visual SLAM framework

Cyril Joly, Patrick Rives

► To cite this version:

Cyril Joly, Patrick Rives. Self Calibration of a vision system embedded in a Visual SLAM framework. IEEE International Conference on Robotics and Automation (ICRA), Oct 2011, San Francisco, United States. IEEE, pp.3320-3326, 2011, .

HAL Id: hal-00767402

<https://hal.inria.fr/hal-00767402>

Submitted on 19 Dec 2012

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.

Self Calibration of a vision system embedded in a Visual SLAM framework

Cyril Joly and Patrick Rives
INRIA Sophia Antipolis Méditerranée
2004 route des lucioles BP93
Sophia Antipolis Cedex

Abstract—This paper presents a novel approach to self calibrate the extrinsic parameters of a camera mounted on a mobile robot in the context of fusion with the odometry sensor. Calibrating precisely such system can be difficult if the camera is mounted on a vehicle where the frame is difficult to localize precisely (like on a car for example). However, the knowledge of the camera pose in the robot frame is essential in order to make a consistent fusion of the sensor measurements. Our approach is based on a Simultaneous Localization and Mapping (SLAM) framework: the estimation of the parameters is made when the robot moves in an unknown environment which is only viewed by the camera. First, a study of the observability properties of the system is made in order to characterize conditions that its inputs have to satisfy to make possible the calibration process. Then, we show on 3 real experimentations with an omnidirectional camera the validity of the conditions and the quality of the estimation of the 3D pose of the camera with respect to the odometry frame.

I. INTRODUCTION

In this paper, the problem of fusing *odometry data* and *images from a single camera* is addressed. This association is often considered in mobile robotics since these two sensors are complementary. On one hand, the camera can acquire lots of information about the environment but does not provide metric information; in the other hand, odometry provides an estimation of the robot motion which compensates the absence of metric information of the camera.

Camera and odometry fusion is one of sensor combination used to solve the simultaneous localization and mapping (SLAM) problem when an horizontal robot motion is assumed ([6], [9]+tardos). In the latter papers, the calibration between the odometry frame and the camera frame is assumed to be known. Moreover, latter approaches assume that the camera is aligned with the axis of the robot which can be difficult to verify in practice.

A novel approach to calibrate the camera and odometry frames in 3 dimensions is proposed. It is based on a SLAM framework. The main idea is that precise frame calibration imply to acquire simultaneously data from both sensors. In the case of odometry, this implies to make move the robot. Then, the SLAM framework does not make assumptions about the environment: making the calibration with this framework allows to calibrate the sensor without using a particular pattern which position with respect to the robot trajectory has to be known. Consequently, the SLAM framework is well adapted to perform a calibration with few

hypotheses.

However, this formulation imply to estimate simultaneously many variables of a complex and non linear system. This rises naturally the question of the *observability*: “ is it possible to reconstruct the state given the inputs and measurements? ”. To answer this question, an extended analysis of the observability properties of system is proposed and show that only a specific sequence of inputs can guarantee that the estimation problem is well posed. To our knowledge, such complete analysis does not exist in literature (see related works in section II). Finally, our method and the observability criterions are validated on real experiments.

This paper is organized as follows. Firstly, an overview of related works is presented in section II. Then, the mathematical description of the system is proposed in section III. Section IV presents the observability analysis of the augmented system: a problematic class of input will be fully discussed and used to construct a new class of inputs sequence which provides good observability properties. Section V provides the global implementation, which is based on a SAM (Smoothing And Mapping) algorithm which is more consistent than classical Extended Kalman Filter. The experimental testbed is presented in section VI and results are presented and discussed. We conclude the paper in section VII and give some perspectives.

II. RELATED WORK

III. FRAMEWORK

A. Notations

Our method is based on a SLAM framework. So, the classical SLAM notation are used (see Fig.1):

- $\mathbf{x}(t)$ the robot state which corresponds to the current robot pose with respect to a reference frame. It is assumed that the robot moves on the horizontal plane. Thus, the pose is classically parametrized by the center of the rear wheels axis ($x(t)$ and $y(t)$) and the orientation ($\theta(t)$) of the longitudinal axis of the robot. In the following, the 3D position of the robot in the global frame is noted $\mathbf{p}(t) = [x(t) \ y(t) \ 0]^T$, and the 3D rotation matrix is noted $\mathbf{R}_z(\theta(t))$.¹

¹In the following, we use the notation $\mathbf{R}_x(\alpha)$ (resp $\mathbf{R}_y(\alpha)$ or $\mathbf{R}_z(\alpha)$) to designate the 3×3 rotation matrix around the x (resp y or z) axis with an angle α .

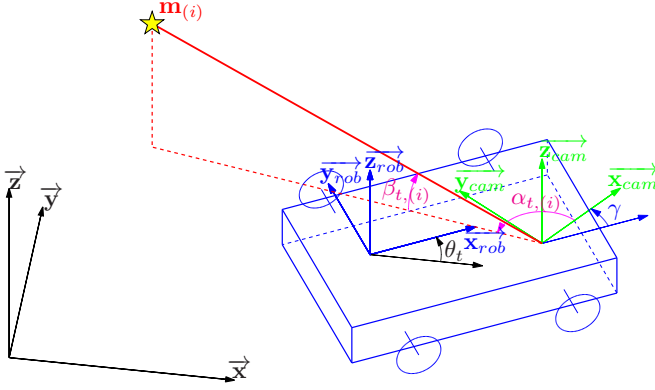


Fig. 1. Notations

- $\mathbf{u}(t)$ (linear and rotational speed: V and ω) the control inputs of the model at time t measured from the wheels encoders.
- $\mathbf{m}_{(i)}$ the i^{th} -landmark state. Each landmark is a 3D-point parametrized by its euclidian coordinates with respect to the reference frame $(x_{(i)}, y_{(i)}$ and $z_{(i)})$. Concatenation of all landmarks state is \mathbf{m}
- $\mathbf{z}_{(i)}(t)$ the measurements of landmark i at time t . If all the landmarks are considered, the notation $\mathbf{z}(t)$ is used. In this paper, we assume that only a bearing information is extracted from the images and the measurement vector is made by two angles (azimuth $\alpha_{(i)}(t)$ and elevation $\beta_{(i)}(t)$).
- \mathbf{x}_c the camera pose expressed in the local robot frame defined by two translations (t_x and t_y) and a 3×3 rotation matrix ($\mathbf{R}_c = \mathbf{R}_z(\gamma_z) \cdot \mathbf{R}_y(\gamma_y) \cdot \mathbf{R}_x(\gamma_x)$). Since the robot moves on the plane, it is impossible to differentiate the z translation coordinate of the camera with the z landmarks coordinates (any value of t_z can be valid if we switch all the z landmarks coordinates). Thus, we assume that the center of the camera is in the plane $z = 0$.

In the following, we omit to precise that \mathbf{x} , \mathbf{p} , \mathbf{u} and \mathbf{z} are time-dependent in order to simplify the notations.

B. Description of the system

The system is classically described by an evolution equation which provides its dynamic in function of the inputs (the odometry readings) and a measurement equation which links the camera measurement to the state.

1) Model equation:

In this work, the evolution equation provides the dynamic of the odometry frame (**and not the camera frame**). We have:

$$\begin{cases} \begin{bmatrix} \dot{x} & \dot{y} & \dot{\theta} \end{bmatrix} = [V \cos \theta & V \sin \theta & \omega] \\ \begin{bmatrix} \dot{t}_x & \dot{t}_y & \dot{\gamma}_x & \dot{\gamma}_y & \dot{\gamma}_z & \dot{\mathbf{m}} \end{bmatrix} = [0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1 \times 3N}] \end{cases} \quad (1)$$

where N stands for the number of landmarks.

2) Measurement equation:

Measurement are constituted by azimuth and elevation angles

with respect to the camera frame.² Let $\mathbf{m}_{t,(i)}^C$ be the euclidian coordinates of a landmark (i) at time t in the camera frame. We have:

$$\mathbf{m}_{t,(i)}^C = \mathbf{R}^C (\mathbf{R}_z(\theta) (\mathbf{m} - \mathbf{p}_t) - \mathbf{t}) \quad (2)$$

where \mathbf{p}_t is made by the 3D coordinate of the robot frame (ie. $\mathbf{p}_t = [x_t \ y_t \ 0]^T$) and \mathbf{t} is made by the 3D coordinates of the camera in the robot frame (ie. $\mathbf{t} = [t_x \ t_y \ 0]^T$).

Finally, the azimuth and elevation angles are given by:

$$\begin{cases} \alpha_{t,(i)} = \text{atan2}(\mathbf{m}_{t,(i)}^C(2), \mathbf{m}_{t,(i)}^C(1)) \\ \beta_{t,(i)} = \text{atan} \left(\frac{\mathbf{m}_{t,(i)}^C(3)}{\sqrt{(\mathbf{m}_{t,(i)}^C(1))^2 + (\mathbf{m}_{t,(i)}^C(2))^2}} \right) \end{cases} \quad (3)$$

IV. OBSERVABILITY ANALYSIS

A. Introduction

Before looking for a solution to the SLAM problem, we want to know if such a solution exists. Thanks to the notion of *observability* we can answer the following question: “*can we find the initial condition of a system given the measurements and its dynamic?*”. Standard observability analysis does not relate to the values of the inputs and just solves the problem of *the existence* of inputs that make the initial condition locally distinguishable (which implies that the system is observable). In the following, a slightly different problem is addressed: “*what are the conditions on the inputs that make the system observable?*”.

B. Classical analysis

At a first step, the problem is stated in terms of the observability of a *continuous system*, with the evolution model defined by:

Moreover, we assume only one landmark in this study.

The system defined by equation (??) is affine with respect to the command $\mathbf{u} = [v \ \omega]^T$. So, we can do a classical analysis observability based on the Lie derivatives ([4]) of \mathbf{z} with respect to the vector fields \mathbf{f}_1 and \mathbf{f}_2 defined by:

$$\begin{cases} \mathbf{f}_1 = [\cos \theta & \sin \theta & 0 & \mathbf{0}_{1 \times 8}]^T \\ \mathbf{f}_2 = [0 & 0 & 1 & \mathbf{0}_{1 \times 8}]^T \end{cases} \quad (4)$$

The system is observable if it exists a full rank observability matrix in which each line is defined by a Lie derivative of \mathbf{z} with respect to \mathbf{f}_1 and/or \mathbf{f}_2 . However, it is now well known that the SLAM problem is not observable: the 3 first columns of the observability matrix (associated to the robot state) is a linear combinaison of the 2 last (landmark coordinates) (see[5]). This can be easily understood since any rigid transformation (2 degrees of freedom in translation and one degree of freedom in rotation) can be applied to the SLAM solution. In practice, the problem is fixed by **setting the initial robot pose to zero without uncertainty**. In the following, we do not consider this “*problem*” as a lack of observability since it has a very simple interpretation.

²It corresponds to high level information which can be deduced from the pixel measurements in the images.

Consequently, we search for an observability matrix of rank 8 and not of rank 11 (since it is impossible). Now, let us consider the matrix defined by:

$$\mathbf{O} = \begin{bmatrix} dL^0 \mathbf{z} \\ dL_{f_1}^1 \mathbf{z} \\ dL_{f_2}^1 \mathbf{z} \\ dL_{f_1 f_2}^2 \mathbf{z} \\ dL_{f_1 f_1}^2 \mathbf{z} \\ dL_{f_2 f_2}^2 \mathbf{z} \end{bmatrix} \quad (5)$$

It can be shown that the rank of \mathbf{O} is equal to 8, so the system is observable. Thus, there exists a sequence of inputs that make possible the computation of the \mathbf{x} , \mathbf{x}_C and \mathbf{m} (up to a global translation and rotation for \mathbf{x} and \mathbf{m}). If the map has more than one landmark, the system remains observable since the computation of the rest of the map can be viewed as a single mapping problem. However, this result is not informative about the kind of trajectory which really makes the system observable. In the following, a problematic sequence of input is studied in order to construct a sequence of inputs which makes the system observable.

C. Example of a problematic set of inputs

1) Observability analysis:

Let us consider the classical case in which the inputs are constant (V_0 and ω_0). In this case, the robot moves along a circle with a radius equal to V_0/ω_0 . Moreover, we assume for this example that the angles γ_x and γ_y of the camera are known and equal to zero (so, the axis of the camera is aligned with the z axis of the robot). Let us perform an observability analysis based on Lie bracket of this particular case. There is only one vector field for the derivation:³

$$\mathbf{f}_3 = \begin{bmatrix} V_0 \cos \theta & V_0 \sin \theta & \omega_0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \quad (6)$$

In this particular case, two vectors can be identified in the observability matrix, whatever the degree of derivation:

$$\mathbf{n}_1 = \begin{bmatrix} \mathbf{0}_{3,1} \\ 0 \\ 1 \\ \omega^2 t_x \\ \frac{(V_0 - \omega_0 t_y)^2 + \omega_0^2 t_x^2}{-V_0^2 \sin \theta - \omega_0 V_0 \widehat{x}_{(i)} - t_x \omega_0^2 (y_{(1)} - y) + t_y \omega_0^2 (x_{(1)} - x)} \\ \frac{(V_0 - \omega_0 t_y)^2 + \omega_0^2 t_x^2}{V_0^2 \cos \theta - \omega_0 V_0 \widehat{y}_{(i)} + t_x \omega_0^2 (y_{(1)} - y) + t_y \omega_0^2 (x_{(1)} - x)} \\ 0 \end{bmatrix} \quad (7)$$

and

$$\mathbf{n}_2 = \begin{bmatrix} \mathbf{0}_{3,1} \\ 1 \\ 0 \\ -\omega_0^2 t_t + V_0 \omega_0 \\ \frac{(V_0 - \omega_0 t_y)^2 + \omega_0^2 t_x^2}{V_0^2 \cos \theta - \omega_0 V_0 \widehat{y}_{(i)} + t_x \omega_0^2 (x_{(1)} - x) + t_y \omega_0^2 (y_{(1)} - y)} \\ \frac{(V_0 - \omega_0 t_y)^2 + \omega_0^2 t_x^2}{V_0^2 \sin \theta + \omega_0 V_0 \widehat{x}_{(i)} + t_x \omega_0^2 (y_{(1)} - y) - t_y \omega_0^2 (x_{(1)} - x)} \\ 0 \end{bmatrix} \quad (8)$$

³The values associated to γ_x and γ_y are removed.

where $\widehat{x}_{(i)} = x_{(i)} - x - t_x \cos \theta + t_y \sin \theta$ and $\widehat{y}_{(i)} = y_{(i)} - y - t_y \cos \theta - t_x \sin \theta$. If more landmarks are considered, two similar vectors in the kernel of the observability matrix will be present. So, in any case, a loss of rank of 2 in the observability matrix occurs when the inputs are constant and the system is not observable. A direct consequence is that the full system (when γ_x and γ_y are considered) is not observable with this set of inputs.

2) Global proof of the non-observability:

In this paragraph, a global proof of the non-observability property in the case of constant inputs is provided. It will be shown that several values for the states can lead to exactly the same outputs. In the following, the full system (with γ_x and γ_y) is considered.

Let us assume that the robot inputs are constant (V_0 and ω_0). Only one landmark $\mathbf{m}_{(i)}$ is considered in a first step. Moreover, as we assume that the initial condition of the robot state is known, the whole trajectory is assumed to be known. So, if two solutions provide the same outputs, they will share the same values of $\mathbf{p} = [x \ y \ 0]^T$ and θ . Let us consider the true solution for the landmark parameters ($\mathbf{m}_{(i)}$), the camera translation (\mathbf{t}) and the camera rotation matrix (\mathbf{R}^C). An alternative solution which provides the same outputs will be noted with a hat.

If two solutions produce the same outputs, the projection of the landmarks coordinates on the unit sphere centered on the camera frame have to be equal in the 2 cases at every time. Thus, the landmarks coordinates in the camera frame for the first solution are proportional to the landmarks coordinates in the camera frame for the second solution. Moreover, the coefficient of proportionality has to be positive. So we have:

$$\forall t \neq 0, \mathbf{m}_{(i)}^C = k(t) \widehat{\mathbf{m}}_{(i)}^C \text{ and } k(t) > 0 \quad (9)$$

where $\mathbf{m}_{(i)}^C = \mathbf{R}^C \cdot (\mathbf{R}_z(-\theta) \cdot (\mathbf{m}_{(i)} - \mathbf{p} - \mathbf{t}))$ and $\widehat{\mathbf{m}}_{(i)}^C = \widehat{\mathbf{R}}^C \cdot (\mathbf{R}_z(-\theta) \cdot (\widehat{\mathbf{m}}_{(i)} - \mathbf{p} - \widehat{\mathbf{t}}))$. At a first step, we search if (9) can be satisfied with a constant value k_0 for $k(t)$:

$$\dot{k}(t) = 0 \text{ and } k = k_0 \quad (10)$$

Differentiating 2 times $\mathbf{m}_{(i)}^C$ with respect to the time yields:

$$\ddot{\mathbf{m}}_{(i)}^C = -\mathbf{R}^C \cdot \left(\omega^2 \mathbf{R}_z(-\theta) \cdot (\mathbf{m}_{(i)} - \mathbf{p}) - 2\omega \mathbf{R}_z(-\theta + \pi/2) \dot{\mathbf{p}} + \mathbf{R}_z(-\theta) \ddot{\mathbf{p}} \right) \quad (11)$$

Substituting $\omega^2 \mathbf{R}_z(-\theta) \cdot (\mathbf{m}_{(i)} - \mathbf{p}) = \mathbf{R}^{C^T} \cdot \mathbf{m}_{(i)}^C + \mathbf{t}$ in (11) leads to:

$$\ddot{\mathbf{m}}_{(i)}^C = -\omega^2 \mathbf{m}_{(i)}^C - \omega^2 \mathbf{R}^C \cdot \mathbf{t} + \mathbf{R}^C \cdot \left(2\omega \mathbf{R}_z(-\theta + \pi/2) \dot{\mathbf{p}} - \mathbf{R}_z(-\theta) \ddot{\mathbf{p}} \right) \quad (12)$$

In consequence, differentiating (9) 2 times with respect to the time and taking into account that $\mathbf{m}_{(i)}^C = k_0 \widehat{\mathbf{m}}_{(i)}^C$ yields:

$$\frac{1}{k_0} \widehat{\mathbf{R}}^{C^T} \mathbf{R}^C \cdot \left(2\omega \mathbf{R}_z(-\theta + \pi/2) \dot{\mathbf{p}} - \mathbf{R}_z(-\theta) \ddot{\mathbf{p}} - \omega^2 \mathbf{t} \right) = 2\omega \mathbf{R}_z(-\theta + \pi/2) \dot{\mathbf{p}} - \mathbf{R}_z(-\theta) \ddot{\mathbf{p}} - \omega^2 \widehat{\mathbf{t}} \quad (13)$$

Substituting $\dot{\mathbf{p}}$ and $\ddot{\mathbf{p}}$ by their particular values in (13) yields:

$$\frac{1}{k_0} \widehat{\mathbf{R}}^C T \mathbf{R}^C \begin{bmatrix} -\omega_0^2 t_x \\ -\omega_0^2 t_y + V_0 \omega_0 \\ 0 \end{bmatrix} = \begin{bmatrix} -\omega_0^2 \widehat{t}_x \\ -\omega_0^2 \widehat{t}_y + V_0 \omega_0 \\ 0 \end{bmatrix} \quad (14)$$

Eq. (14) shows that $\widehat{\mathbf{R}}^C T \mathbf{R}^C$ is a constant rotation matrix which transforms a vector of the horizontal plane to another vector of the horizontal plane. In consequence, we have:

$$\widehat{\mathbf{R}}^C T \mathbf{R}^C = \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{with } a^2 + b^2 = 1 \quad (15)$$

Substituting (15) in (14) leads to a 2×2 system in $k_0 a$ and $k_0 b$ which solution is:

$$\begin{cases} k_0 a &= \frac{V_0^2 - \omega_0 V_0 (t_y + \widehat{t}_y) \omega_0^2 (t_x \widehat{t}_x + t_y \widehat{t}_y)}{\omega_0^2 t_x^2 + (V_0 - \omega_0 t_y)^2} \\ k_0 b &= \frac{-\omega_0 (V_0 (t_x - \widehat{t}_x) + \omega_0 (t_y \widehat{t}_x - t_x \widehat{t}_y))}{\omega_0^2 t_x^2 + (V_0 - \omega_0 t_y)^2} \end{cases} \quad (16)$$

So, the value of k_0 is given by using the relation $a^2 + b^2 = 1$ and the constraint $k_0 > 0$:

$$k_0 = \frac{\sqrt{\widehat{t}_x^2 + \left(\widehat{t}_y - \frac{V_0}{\omega_0}\right)^2}}{\sqrt{t_x^2 + \left(t_y - \frac{V_0}{\omega_0}\right)^2}} \quad (17)$$

Then, \mathbf{R}^C is completely defined by (15-17) and depends only on t_x , t_y , \widehat{t}_x , \widehat{t}_y , V_0 and ω_0 . Finally, a unique solution for $\widehat{\mathbf{m}}_{(i)}^C$ can be deduced from (9), the value of \mathbf{R}^C and assuming that the initial robot state is equal to zero:

$$\widehat{\mathbf{m}}_{(i)}^C = \frac{1}{k_0} \left(\widehat{\mathbf{R}}^C T \mathbf{R}^C \cdot (\mathbf{m}_{(i)} - \mathbf{t}) + \widehat{\mathbf{t}} \right) \quad (18)$$

The previous constraints on the alternative solution for the calibration and the map was obtained with the assumption that $k(t)$ is constant with respect to the time. Let us show that no other solution exists for $k(t)$. Let us define $k(t) = k_0 + k'(t)$. Applying (9) leads to:

$$\mathbf{m}_{(i)}^C = k_0 \widehat{\mathbf{m}}_{(i)}^C + k'(t) \widehat{\mathbf{m}}_{(i)}^C \quad (\text{with } k_0 \widehat{\mathbf{m}}_{(i)}^C = \mathbf{m}_{(i)}^C) \quad (19)$$

Thus, we have:

$$k'(t) \widehat{\mathbf{m}}_{(i)}^C = 0 \quad (20)$$

Eq (20) implies that $k'(t) = 0$ or $\widehat{\mathbf{m}}_{(i)}^C = 0$. However, $\widehat{\mathbf{m}}_{(i)}^C = 0$ implies that we made a measure for which the landmark is at the center of the camera frame: we assume that this solution is impossible. In consequence, we have $\forall t > 0$, $k'(t) = 0$.

We have shown that there exists an infinite number of possible solutions for the camera pose in the robot frame and the map which can produce the same outputs than the true solution. More particularly, any value of \widehat{t}_x and \widehat{t}_y can be chosen. However, there exists only one rotation matrix which is admissible for a translation value $\widehat{\mathbf{t}}$. Interestingly, the rotation matrix depends only on the true rotation matrix, the true camera translation, \widehat{t}_x , \widehat{t}_y and the ratio V_0/ω_0 (radius of curvature of the trajectory). It does not depend on the

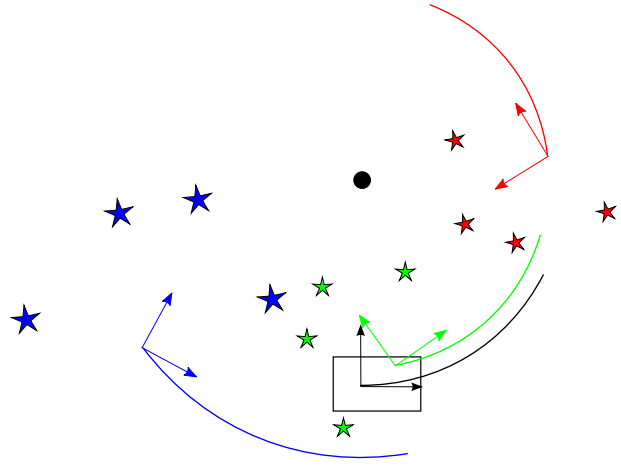


Fig. 2. Illustration of the observability issue when v/ω is constant

landmark or the pose of the robot. It can be seen as a kinematic constraint on the camera frame. Thus, **adding landmarks in the map will never eliminate any solution for \widehat{t}_x and \widehat{t}_y** . Finally, the new landmark position is defined uniquely by (18). In the case of several landmarks, eq. (18) would be applied for each landmark.

Finally, we show that the system is not observable when the inputs are constant. Moreover, any values of \widehat{t}_x and \widehat{t}_y can produce valid outputs. **Nevertheless, there is only one value possible of the camera rotation matrix and the map which satisfy the outputs when \widehat{t}_x and \widehat{t}_y are given.** This property will be used further to construct an observable trajectory.

3) Geometric interpretation:

In this paragraph a geometrical interpretation of the non observability properties in the case of constant inputs is proposed. In the following, let us consider Fig. 2:

- The robot trajectory during a time Δt is plotted in black: it is a circle.
- The true camera location is assumed to be the green frame. The real landmarks are thus the green stars. notice that there is a non zero γ angle.
- The red frame, landmarks and trajectory stands for an alternative solution for which we have shift the camera frame on the trajectory defined by the green circle. In that case, the camera motion during Δt is the red circle. If we apply a rotation to the new camera frame so that the angle with respect to the tangent of the circle is the same as in the true solution (the green one) and apply a rigid transformation on the green landmarks to ensure that their new coordinate in the red frame are preserved, the bearing measurements will be exactly the same. Moreover, it can be noted that the green and red solution can be superimposed. This is a first illustration of the non observability of this kind of trajectory.
- The blue frame corresponds to a solution in which the camera frame was moved on a trajectory with an higher radius (R') than the true trajectory (R). The transformation of the green landmarks to the blue ones

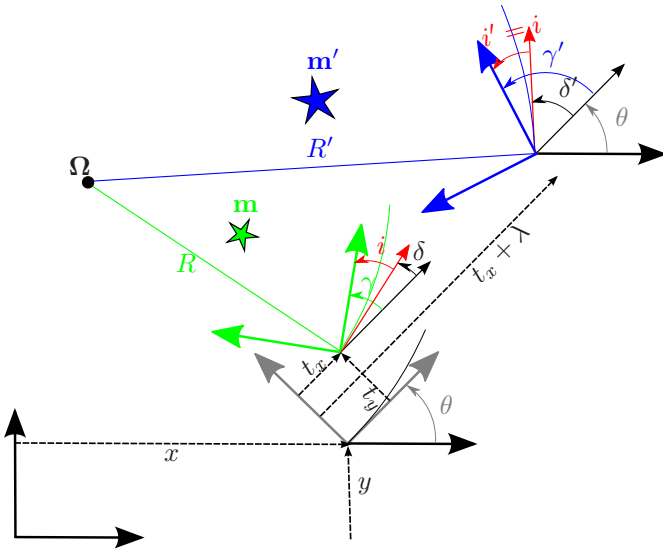


Fig. 3. Geometric transformation between two equivalent solutions in the case of a circular trajectory. We assume that the true solution is the green one. The blue one is one of the equivalent defined by translating the camera frame of λ in direction x .

corresponds to a similitude with an homothety factor of R'/R . Since a similitude conserves the angles and shape trajectory, the bearing measurements of the blue solution will be exactly the same than the one provided by the green solution. This is another example of the observability issue which introduce the problem of the **scale factor**.

Remark 4.1 (Case of the straight line):

The case of the straight line trajectory ($\omega = 0$) is also unobservable. It can be viewed as a limit case of the circular trajectory.

D. Case of multiple arcs of circle

Now, let us consider the case of a trajectory where v/ω is piecewise constant. It is equal to ρ_1 between t_0 and t_1 , equal to $\rho_2 \neq \rho_1$ between t_1 and t_2 , etc... Thus, the trajectory is composed by N connected arcs of circle with different curvatures. Let us denote t_x , t_y and γ the true camera pose and \mathbf{m} the true landmark position. As in the previous paragraph, we will denote with a prime an alternative solution that provides the same outputs. We can notice that:

- if we consider only the part of the trajectory between 0 and t_1 , we can choose any values for t'_x and t'_y and compute then γ' and \mathbf{m}' . Let \mathcal{S}_1 be the set of alternative solutions for this trajectory.
- if we consider only the part of the trajectory between t_1 and t_2 , we can choose any values for t'_x and t'_y and compute then γ' and \mathbf{m}' . Let \mathcal{S}_2 be the set of alternative solutions for this trajectory,
- etc...

If we have an alternative solution for the entire global trajectory, it has to belong to $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \dots \mathcal{S}_N$. A necessary

condition for \mathbf{m}' to belong to the intersection is that the solutions share the same homothetic factor. This factor is defined by the ratio between the true radius of curvature of the camera trajectory and the alternative one. The radius of curvature of the camera trajectory can be expressed as a function of its position and v/ω (see Fig. 3):

$$R = \sqrt{t_x^2 + (t_y - v/\omega)^2} \quad (21)$$

So, the problem can be stated as the proof of the existence of a solution for the following system:

$$\frac{\sqrt{t'_x{}^2 + (t'_y - \rho_1)^2}}{\sqrt{t_x^2 + (t_y - \rho_1)^2}} = \frac{\sqrt{t'_x{}^2 + (t'_y - \rho_2)^2}}{\sqrt{t_x^2 + (t_y - \rho_2)^2}} = \dots \quad (22)$$

that we can rewrite as

$$\frac{t'_x{}^2 + (t'_y - \rho_1)^2}{t_x^2 + (t_y - \rho_1)^2} = \frac{t'_x{}^2 + (t'_y - \rho_2)^2}{t_x^2 + (t_y - \rho_2)^2} = \dots \quad (23)$$

To solve this problem, we propose to study the function k defined by:

$$\rho \mapsto k(\rho) = \frac{t'_x{}^2 + (t'_y - \rho)^2}{t_x^2 + (t_y - \rho)^2} \quad (24)$$

It can be shown that $\frac{dk}{d\rho}$ is always of the sign of the following polynomial:

$$P(\rho) = (t_y - t'_y)\rho^2 + (t_x^2 - t_x'^2 + t_y^2 - t_y'^2)\rho + (t'_y - t_y)(t_y^2 + t_x^2) \quad (25)$$

Three cases have to be considered depending on the degree of the polynomial $P(\rho)$:

- 1) $t_y \neq t'_y$. In this case, $P(\rho)$ is a parabola and two real roots exist.⁴ In consequence, the sign of $\frac{dk}{d\rho}$ changes two times when ρ is describing \mathbb{R} . This means that for a value ρ_1 , it may exist at the maximum two other values ρ_2 and ρ_3 that can satisfy the constrain. So, as we have 3 constraints to satisfy with different values of ρ (which implies $k(\rho_1) = k(\rho_2)$, $k(\rho_1) = k(\rho_3)$, $k(\rho_2) = k(\rho_3)$), one of the constraints will be always impossible to satisfy.
- 2) $t_y = t'_y$ but $t_x \neq t'_x$. In this case $P(\rho)$ is a line and has one real root. In consequence, the sign of $\frac{dk}{d\rho}$ changes one time when ρ is describing \mathbb{R} . This means that for a value ρ_1 , it may exist at the maximum one other value ρ_2 that can satisfy the constraint. So, one constraint will be always impossible to satisfy if the radius of curvature change two times.
- 3) $t_y = t'_y$ but $t_x = t'_x$. In that case, $\frac{dk}{d\rho}$ is always equal to zero. So, all the constraints are always satisfied. This corresponds to the case where the alternative solution **is the true solution**.

Finally, we have shown a sufficient condition of observability in the case of a constant piecewise command: **if the ratio v/ω takes at least four different values, there is only one value possible for the camera position**. This property is stronger than classical observability properties since it is

⁴It is easy to check that the discriminant of the parabola is positive.

global and not limited in a neighbourhood of the solution. At the limit case, a continuous variation of v/ω keeps the system always observable.

E. Conclusion of the analysis

To conclude the observability analysis, three main results have been demonstrated:

- 1) First, we showed that the system of SLAM augmented with the camera position in the robot frame is observable. So, it exists a set of inputs that makes the true solution distinguishable in a neighbourhood given.
- 2) Then, we showed a case where the system can not be observed: the radius of curvature of the trajectory is constant. This implies a necessary condition on the inputs: the ratio v/ω has to change.
- 3) Finally, we showed a sufficient condition on the inputs to make the system observable. The value of v/ω has to take four different values. This property provides us with a global observability. However, we guess that one changement of radius can make the system locally observable but we did not prove it. Nevertheless, the condition of four different values for v/ω is very simple to satisfy and provides strong properties.

V. IMPLEMENTATION

A. Filtering algorithm

The implemented method for solving the SLAM problem is based on a Smoothing And Mapping (SAM) approach ([2], [6]). It is a filtering algorithm which computes the entire trajectory of the robot and the map using all the measurement and inputs available from the starting instant. So, at time t , the state to estimate is $[\mathbf{x}_{1:t}^T, \mathbf{x}_C^T, \mathbf{m}^T]^T$. At each step, we assume that we have a prediction of the trajectory ($\mu^{\mathbf{x}_{1:t}}$), the camera frame (μ^C) and the landmarks ($\mu^{\mathbf{m}}$). These predictions are used to linearize all the equations of the system. With the Gaussian assumptions on the errors (see the paragraph III-B) and the linearization, the final probability density associated to $[\mathbf{x}_{1:t}^T, \mathbf{x}_C^T, \mathbf{m}^T]^T$ is a Gaussian which information parameters are computed from the predicted values, the different Jacobian and the covariances matrices \mathbf{Q}_t and \mathbf{R}_t that were introduced in paragraph III-B (see [2] for example to understand how these parameters are computed).

Remark 5.1 (Gaussian representation):

Information parameters are the dual representation of the classical mean vector and covariance matrix. Let Ω and ξ be the information matrix and vector associated to a Gaussian density and Σ and μ the covariance matrix and mean. They are linked by:

$$\begin{cases} \Omega = \Sigma^{-1} \\ \xi = \Omega\mu \end{cases} \quad (26)$$

The advantage of the information representation is the **sparsity** of the information matrix whereas the covariance matrix is almost full. So, efficient methods can be used to compute an estimation by $\Omega^{-1}\xi$. In comparison to the Extended Kalman Filter which uses also a linearization method (but only on the current pose), the SAM algorithm has a

similar computational complexity⁵ but a better consistence since it can correct past estimation errors.

It remains also possible to extract some covariance parameters, like the covariance matrix associated to the current pose or the camera parameters⁶

Finally, our estimation algorithm can be summarized up as follow:

- 1) First, we assume that we have an estimation at time t : $\mu_{1:t}^{\mathbf{x}}$, μ^C and $\mu^{\mathbf{m}}$
- 2) Then, we use the odometry measurement to make a prediction of \mathbf{x}_{t+1} from $\mu_t^{\mathbf{x}}$ (we denote $\hat{\mathbf{x}}_{t+1}$ this prediction). We concatenate this vector with $\mu_{1:t}^{\mathbf{x}}$ to have a linearization trajectory:
$$\mu_{1:t}^{\mathbf{x}^T} = [\mu_{1:t}^{\mathbf{x}^T}, \hat{\mathbf{x}}_{t+1}^T]^T \quad (27)$$
- 3) We apply the SAM equations with $\mu_{1:t}^{\mathbf{x}}$, μ^C and $\mu^{\mathbf{m}}$ and all the measurements to compute the information parameters of the full system (Ω and ξ)
- 4) The estimation of all the states μ is given by $\Omega^{-1}\xi$.
- 5) Finally, new values for $\mu_{1:t}^{\mathbf{x}}$, μ^C and $\mu^{\mathbf{m}}$ are extracted from μ .

B. Initializations

1) Robot first pose:

We recall in the observability analysis that the overall system of SLAM is observable up to a global translation and rotation. To overpass this problem, the initial robot pose is fixed to **zero without uncertainty**. In practice, we can do it by adding $\frac{1}{\epsilon}\mathbf{I}_{3 \times 3}$ to $\Omega_{1:3,1:3}$; where ϵ is a positive value very close to zero.⁷ Obviously, $\mu_{1:3}^{\mathbf{x}}$ is set to zero.

2) Landmark initialization:

Landmark initialization is not obvious. Indeed, we are in the bearing-only case: landmarks cannot be initialized when they are seen for the first time. At least two observation with parallax are necessary. However, if the parallax is too small, landmark initialization can be ill-conditioned. Such initialization can make lead the SAM algorithm to diverge since it assumes that the linearizations are valid.

In consequence, the initialization is delayed until it exists parallax enough to guarantee a consistent linearization. Then, the first and last observations is used to compute an initial value for the landmark. This value will be used in the next step to linearize all the equations in which the landmark appears.

3) Camera frame initialization:

The initialization of the camera frame parameters have to satisfy linearity constraints. It is important to notice that we cannot set the initial information parameters to zero. Such situation implies that we do not have any information about the camera position. In practice, it leads to the impossibility to get an estimation of the landmarks parameters when the

⁵This is due to the fact that the EKF works with a covariance matrix which size is proportional to the size of the map and which is not sparse.

⁶Obviously, we never extract the whole parameters which is computationally expensive and not very useful.

⁷It will assume an initialization with a standard deviation of $\sqrt{\epsilon}$ meter.

observability criterion is not satisfied: the information matrix is not invertible. This makes the algorithm to diverge.

Consequently, an initialization corresponding to the knowledge that we have on these parameters is used. In practice, we choose to set the parameters to zero and to add the corresponding information matrix:

$$\Omega^C = \begin{bmatrix} 1\text{m}^{-2} & 0 & 0 \\ 0 & 1\text{m}^{-2} & 0 \\ 0 & 0 & \frac{1}{5^2}\text{deg}^{-2} \end{bmatrix} \quad (28)$$

This choice corresponds to a standard deviation of 1m on t_x and t_y and 5 degrees for γ . This is reasonable for our application.⁸

VI. EXPERIMENTATIONS

A. Experimental testbed

We tested our method on two different indoor real sequences:

- 1) The first one, (called *Borel sequence*) was acquired by a robot with undeformable wheels which provide a precise odometry. In this experiment, the robot is driven in corridors. So, the global coherence of the solution can be simply checked,
- 2) The second one (called *Kahn sequence*) was acquired by a robot equipped of **tires**. The contact surface is quite important and the tires are deformable. So, the odometry is quite imprecise. To check the global consistence of the solution, the end of the trajectory is exactly the begin of the trajectory. This gives an easy way to compute the global accumulated error.

Both robots were equipped with an omnidirectional camera to measure the environment. Interest points are selected and tracked using algorithms adapted to omnidirectional cameras ([3]). We converted images pixel coordinates in azimuth and elevation angles with the unified projection model ([8], [1]). The advantage of the omnidirectional cameras is that they provide a full 360deg field of view. The main drawback is that image processing is more complex and takes more time.

B. Results

1) Quality of our method:

The results of our method are provided in this paragraph. In both cases, the solution of our method was compared to a solution provided by an equivalent SLAM algorithm in which the camera parameters are assumed to be known. Both robot platforms were designed to place precisely different sensors. In consequence, a fair reference can be got by hand.⁹

Estimated values and the associated 3σ bounds (99% of probability) of the camera parameters are provided on table I. We can see that the parameters estimated are very close to the reference. It is important to notice that an initial standard deviation of one meter was used for the position parameters, which represent a poor initialisation in terms of information.

⁸The robots that we used have a length of about 1 meter. So, the 3σ bounds for t_x and t_y contains all the experimental platform.

⁹However, this reference is not as precise as a “real” ground truth. This is particularly true for the reference on γ which is difficult to evaluate.

		Reference	Estimation	3σ bounds
Borel	t_x (m)	0.55	0.5575	[0.5163, 0.5988]
	t_y (m)	-0.29	-0.2859	[-0.3939, -0.1780]
	γ (deg)	0	0.53	[0.2020, 0.8642]
Kahn	t_x (m)	-0.1955	-0.2046	[-0.2121, -0.1971]
	t_y (m)	0	-0.0227	[-0.0416, -0.0039]
	γ (deg)	0	-0.0198	[-0.3349, 0.2953]

TABLE I

NUMERIC RESULTS CONCERNING THE CAMERA PARAMETERS

It can be noticed that the estimation of t_x and t_y is quite conservative for the *Borel sequence* (the 3σ bounds are much larger than the one provided for the *Kahn sequence*). However, the center is very accurate. This may be due to a too conservative tuning of the covariances matrices \mathbf{R}_t and \mathbf{Q}_t for the *Borel sequence*. Finally, the angles γ are not perfectly consistent with the reference solution. Nevertheless, taking into account the uncertainty on the reference angle (about 1deg) makes the solution consistent.

Then, the SLAM results of our new solution were compared with the results provided by setting the camera parameters to the hand-made references and assuming they are perfectly known. Results are shown on Fig. 4. Concerning the *Borel sequence* our results (in red) are very close to the ones provided by using the reference parameters (in blue). Then, the results provided by our method for the *Kahn sequence* differ slightly from the reference solution. Our result seems more accurate since the end of the trajectory is a bit closer to the origin. Moreover, the *Kahn sequence* shows that very good results can be achieved with poor odometry. Finally, both results show that adding the camera parameters in the estimation does not change the good behavior of the SAM algorithm (moreover, it tends to enhance it).

2) Observability results:

Finally, the relevance of the observability condition was analysed. Results are presented on Fig. 5. They show that our implementation is consistent with respect to the observability analysis. Indeed, the 3σ bounds begin to decrease significantly (which means that the estimation really begins) when the radius of the trajectory shows significant variations. This validate both our implementation and the observability analysis presented in section IV.

VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed an accurate and consistent SLAM method in the context of non-calibrated fusion between image data and odometry. First, an original observability analysis was presented. Our study was not limited to the only notion of observability. The problem of characterising the “good inputs” was clearly addressed and solved: we showed that the radius of the trajectory has to change enough. To our knowledge, this is a new result: most studies limits the analysis to the observability condition (like the study in [7]). Then, our implementation to solve simultaneously the SLAM problem and the camera frame calibration was described. It is based on the SAM approach, which gives in general better results than the classical EKF.

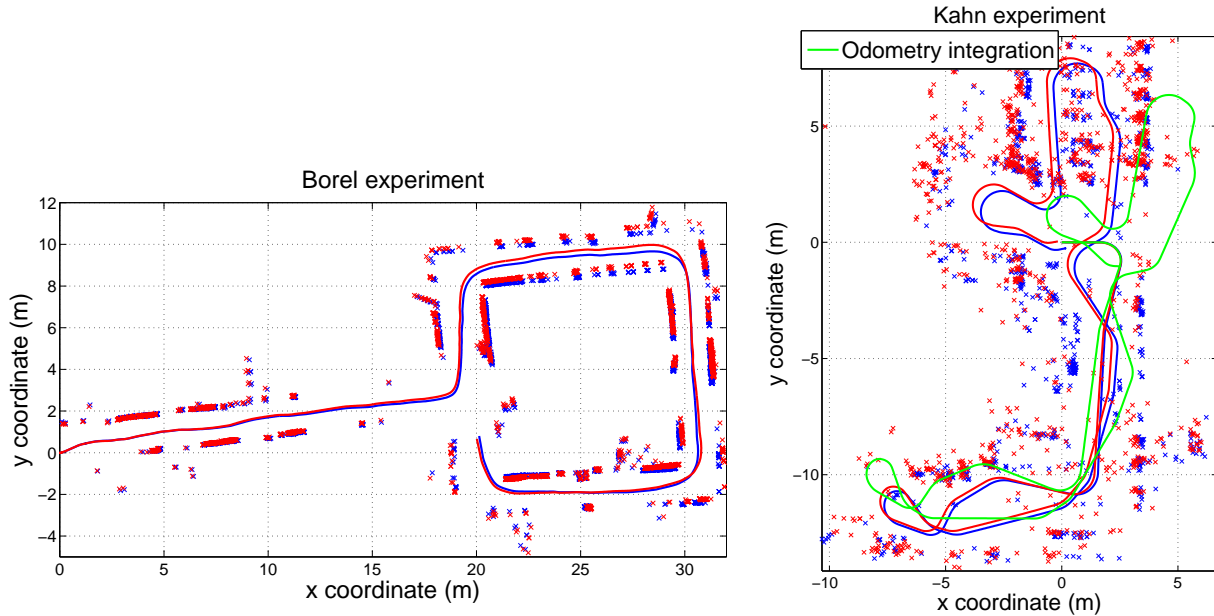


Fig. 4. Blue: SAM solution assuming the camera parameters are known. Red: solution provided by our algorithm — Left: Borel experimentation - Right: Kahn experimentation — We did not plot the odometry in the case of the Borel sequence since it is very close to the solutions red and blue.

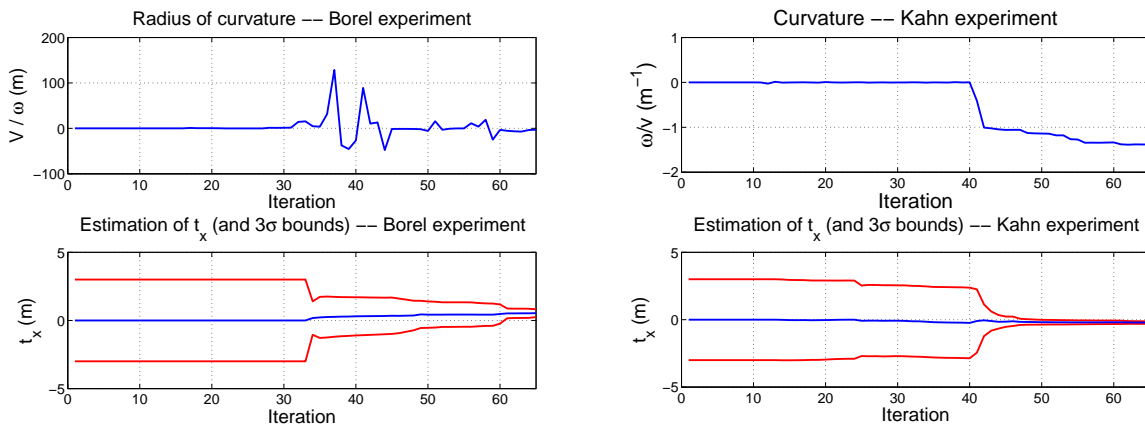


Fig. 5. Evolution of the radius or curvature (or curvature) and the estimation of t_x — Left: Borel experiment - Right:Kahn experiment

Our method was validated on real data. It gave very good results: it was possible to recover the robot trajectory, the map and the camera parameters in the case of a good odometry (*Borel sequence*) and in the case of a poor odometry (*Kahn sequence*). Finally, the experimentations also validated the observability analysis.

Future work will be focused on extending these results. Two new problems will be considered:

- 1) The z axis of the camera and the robot are not aligned. So 2 more rotation angles have to be calibrated,
- 2) The **intrinsic** parameters of the camera are not known and have to be calibrated by the SLAM algorithm

For each study, a complete observability analysis has to be done. It has to be validated on real data.

REFERENCES

- [1] João P. Barreto. *General Central Projection Systems, modeling, calibration and visual servoing*. PhD thesis, Department of electrical and computer engineering, 2003.
- [2] Frank Dellaert and Michael Kaess. Square Root SAM: Simultaneous Location and Mapping via Square Root Information Smoothing. *Journal of Robotics Research*, 2006.
- [3] H. Hadj-Abdelkader, E. Malis, and P. Rives. Spherical Image Processing for Accurate Odometry with Omnidirectional Cameras. In *The Eighth Workshop on Omnidirectional Vision*, October.
- [4] R. Hermann and A. J. Krener. Nonlinear controllability and observability. *IEEE Transactions On Automatic Control*, 22(5):728–740, 1977.
- [5] G.P. Huang, A.I. Mourikis, , and S.I. Roumeliotis. Analysis and improvement of the consistency of extended kalman filter based slam. In *Proceedings of IEEE ICRA08*, 2008.
- [6] Cyril Joly and Patrick Rives. Building Consistent Local Submaps with Omnidirectional SLAM. In *The Ninth Workshop on Omnidirectional Vision, Kyoto, Japan*, October 2009.
- [7] A. Martinelli, D. Scaramuzza, and R. Siegwart. Automatic self-calibration of a vision system during robot motion. may. 2006.
- [8] Christopher Mei. *Laser-Augmented Omnidirectional Vision for 3D Localisation and Mapping*. PhD thesis, Ecole des Mines de Paris, INRIA Sophia Antipolis, 2007.
- [9] J. Solà, M. Devy, A. Monin, and T. Lemaire. Undelayed initialization in bearing only slam. In *IEEE International Conference on Intelligent Robots and Systems*, 2005.