Fusion of inertial and visual : a geometrical observer-based approach

S. Bonnabel * **P. Rouchon** **

* *Centre de Robotique, Ecole des Mines de Paris, France (silvere.bonnabel@mines-paristech.fr)*
** *Centre Automatique et Systèmes, Ecole des Mines de Paris, France (rouchon@cas.ensmp.fr)*

Abstract: The problem of combination between inertial sensors and CCD cameras is of paramount importance in various applications in robotics and autonomous navigation. In this paper we develop a totally geometric model for analysis of this problem, independently from a camera model and from the structure of the scene (landmarks etc.). This formulation can be used for data fusion in several inertial navigation problems. The estimation is then decoupled from the structure of the scene. We use it in the particular case of the estimation of the gyroscopes bias and we build a nonlinear observer which is easy to compute, provides an estimation of the biais, filters the image, and is by construction very robust to noise.

Keywords: Non-linear observer, geometrical methods, inertial vision.

1. INTRODUCTION

IN dynamic vision the inverse problem of recovering information from a sequence of images is studied (see e.g [10]). The main goal is to estimate the motion of the camera and the structure of the scene. In the monocular case, without any information about the scene, the camera translation can only be estimated up to a scale factor [2]. Combining the output of the camera with inertial sensors can give additionnal information and make the problem observable [11, 8]. One application of this type of data fusion is the field of inertial navigation. Indeed in low-cost navigation systems, position and attitude are usually estimated using the measurement of the relatively inaccurate gyroscopes and accelerometers on the one hand, and velocity measurements (given by an air-data system or a Doppler radar), magnetic sensors, and/or CCD cameras on the other. The various measurements are fused according to the (flat-Earth) motion equations of the aircraft, usually by a gain-scheduled observer or an extended Kalman filter [5].

The central problem is to extract information from a sequence of images. The main body of research has been devoted to feature-based methods. They assume the point correspondences in all the images of the sequence are available (or line correspondances [11] etc.). They require to select good features in the image. This is almost always done thanks to human intervention (see [9]). Even when good features are to be found, they still are subject to noise and occlusion. From a practical point of view, the features can also be "lost" when the image is moving too fast. In this paper we formulate the problem independently from the structure (shape) of the scene, and the choice of any features. We compute the dynamics (1) of the output signal (the image from the camera) and we use it directly in the estimation problem. Even if we only estimate the gyro bias in this paper, the model could be used for pose estimation in particular cases (planar motion, additional velocity sensor etc). It is an alternative to the usual inertial-vision fusion methods in velocity, attitude, and bias estimation for UAV (see e.g. [4]).

In section II and III, we consider a flying body equipped with an inertial measurement unit (accelerometers and gyrometers) and a spherical video camera [1] . We give a geometrical modelisation of the problem. In section IV we assume that the environment is far, and we build an observer (2) to estimate the bias of the gyroscopes, independently from the structure of the environment. The observer provides a filter for the image, the signal being integrated over space and time. In particular, no spatio-temporal derivatives of the image are required. Moreover our observer-based method requires fewer computations than an extended Kalman Filter, or any estimator using feature-based visual methods. Even if the design is based on the use of a spherical camera and the invariance by rotation, it can be adapted to a variety of camera models including the standard pinhole model, as proved in section IV-C. For the problem of vision and inertial sensor cooperation see e.g. [4, 11, 8].

2. CAMERA MOTION : ASSUMPTIONS AND NOTATIONS

The model is based on simple assumptions which are common in this aera of research (see [9, 12]). We consider a flying object equipped with a spherical camera, which "sees" any direction of the space.

Geometry of the scene

The scene (the environment) is modeled as a surface $\mathcal{S} \subset \mathbb{R}^3$ surrounding the object and diffeomorphic to the sphere $\mathbb{S}^2$. We consider that the geometry of the environment is such that the camera can "see" any point of the environment at any time (for instance $\mathcal{S}$ is the boundary of a convex volume). It means that any point inside the scene is the origin of a ray which intersects $\mathcal{S}$ only once.

[1] The use of a spherical camera is a theoretical hypothesis. In practice panoramic or wide angle cameras with suitable image transformation are used. Our work shows how to use only partial views associated to this problem in the final section.

Choice of the variables

Let us parameterize the environment $\mathcal{S}$ by a variable $s \in \mathbb{S}^2$ such that to any s corresponds one and only one point $M(s) \in \mathcal{S} \subset \mathbb{R}^3$. Let $C \in \mathbb{R}^3$ be the position of the center of gravity of the object. The implementation of the model equations are strongly simplified when the body orientation is described by a quaternion of length 1 (rather than by Euler angles or a rotation matrix). The use of quaternions is standard in aeronautics. We define the convention of reading any vector of $\mathbb{R}^3$ as a pure imaginary quaternion, as explained in a recap on quaternions, section 6.1. The orientation of the object is the quaternion of unitary norm $q \in \mathbb{H}_1$. It corresponds to the rotation which maps the earth-fixed frame to the body-fixed frame. Let v be the velocity expressed in the body-fixed frame. By definition $v(t) = q^{-1} * \frac{d}{dt}C * q$ since $\frac{d}{dt}C$ is the velocity in the earth-fixed frame. Let $\eta \in \mathbb{S}^2$ be the unitary vector, expressed in the body-fixed frame, pointing in a certain direction of the space. We identify (see section 6.1) vectors of $\mathbb{R}^3$ with quaternions whose first coordinate is equal to zero. The structure of the scene implies that, to one direction in space η, there is one and only one corresponding point of the scene $M(s) \in \mathcal{S}$ for all t. Thus by definition

$$D(t,s)\; q * \eta * q^{-1} = \overrightarrow{C(t)M(s)}$$

with

$$D(t,s) = \|\overrightarrow{C(t)M(s)}\|$$

and

$$s = \varphi(t,\eta)$$

where φ is a bijective function with respect to η for all t. Let its inverse be $\eta = \psi(t,s)$.

Radiance of the scene

According to the complete paper [9], it is a common assumption to assume the changes of the content and lighting of the scene to be small over the time. Thus we assume that every point $M(s)$ of the environment has a radiant flux (total power of light emitted) $I(s)$ which is constant over the time. Even if we could make the simple assumption (see e.g. [12]) that the light received from the scene does not depend on the distance to the scene, we are going to take into account the variation of luminosity with the distance the following way. In the standard lambertian reflectance model (see e.g. [7]) of surfaces, the light perceived depends on the angle between the surface normal and the specified direction, to take into account the perspective. We relax this assumption assuming the light is emitted (reflected) isotropically. We thus consider a point source model: the amount of light received by the object (whose position is C) coming from $M(s)$ is proportionnal to the radiant flux, as well as to the solid angle that the object subtends at the point $M(s)$, and it writes $K\frac{I(s)}{D^2(t,s)}$ where $K > 0$ is a normalisation factor. The output of the system is the amount of light received by the camera in any direction:

$$y(t,\eta) = K\frac{I(s)}{D^2(t,s)} \quad \text{with} \quad s = \varphi(t,\eta)$$

3. KINEMATIC MODEL

Let v be the velocity vector of the center of mass in the body-fixed frame. Let $\omega(t)$ be the instantaneous angular velocity vector in the body-fixed frame. It is a known input since it is measured by the gyroscopes. Let $a(t)$ be the specific acceleration vector, i.e, the aerodynamics forces divided by the body mass. It is measured by the accelerometers. Let $\mathbf{A}_{grav}$ be the gravity vector in the earth-fixed frame (we take the same notations as [3]).

The output of the system is $y(t,\eta)$: it is the light received by the pixel in the direction $\eta \in \mathbb{S}^2$ of the spherical camera. We are going to prove that the motion equations are given by (1). The motion of a flying rigid body (assuming the Earth is flat and defines an inertial frame) is described by

$$\begin{aligned}\frac{d}{dt}q &= \frac{1}{2}q * \omega \\ \frac{d}{dt}v &= v \times \omega + q^{-1} * \mathbf{A}_{grav} * q + a\end{aligned}$$

We use functions of three variables η, s, t but they only depend on two of them since we have $\eta = \psi(t,s)$. We are going to differentiate the variables above with respect to t with s held constant. In the sequel η represents the function of two variables $\psi(t,s)$, that we will differentiate with respect to t. Let $\frac{\partial}{\partial t}\big|_s$ denote differentation with respect to time t with s held constant. Let $\frac{\partial}{\partial t}\big|_\eta$ denote differentation with respect to time t with η held constant. We will use the “reduced” velocity

$$\xi(t,s) = \frac{v(t)}{D(t,s)}$$

where $v = q^{-1} * \frac{d}{dt}C * q$. Let us differentiate with respect to the time variable, the following structural equality which explicates the direction η in the earth-fixed frame (we forget t in the equations)

$$q * \eta * q^{-1} = \frac{\overrightarrow{C(t)M(s)}}{D(t,s)} = \frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|}$$

We have

$$\frac{\partial}{\partial t}\bigg|_s \left(\overrightarrow{CM}\right) = -\frac{d}{dt}C = -q * v * q^{-1}$$

and

$$\begin{aligned}\frac{\partial}{\partial t}\bigg|_s \|\overrightarrow{CM}(s)\| &= -\frac{d}{dt}C.\frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|} \\ &= -(q * v * q^{-1}).(q * \eta * q^{-1}) \\ &= -v.\eta\end{aligned}$$

where “.” is the scalar product, as the rotation preserves the scalar product. Gathering these last two results, along with the definitions of $D(t,s)$ and $\xi(t,s)$ gives

$$\frac{\partial}{\partial t}\bigg|_s \left(\frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|}\right) = q * (-\xi + (\eta.\xi)\eta) * q^{-1}$$

But we have using Leibniz differentiation rule $\frac{\partial}{\partial t}\big|_s \left(q * \eta * q^{-1}\right) = q * (\omega \times \eta) * q^{-1} + q * \frac{\partial \eta}{\partial t}\Big|_s * q^{-1}$. The first equation obtained concerns $\eta = \psi(t,s)$:

$$\frac{\partial \eta}{\partial t}\bigg|_s = -\omega \times \eta + (\eta.\xi)\eta - \xi$$

We also have

$$
\begin{aligned}
\left.\frac{\partial y}{\partial t}\right|_s &= \left.\frac{\partial}{\partial t}\right|_s \left(K \frac{I(s)}{|\overrightarrow{C(t)M(s)}|^2} \right) \\
&= -2K \frac{I(s)}{|\overrightarrow{CM}|^2} (q * \xi * q^{-1}.q * \eta * q^{-1}) \\
&= -2(\xi.\eta) y.
\end{aligned}
$$

as the rotation preserves the scalar product. The camera provides at any time the scalar field $\mathbb{S}^2 \ni \eta \mapsto y(t, \eta)$ which is the output of the system. Since

$$
\left.\frac{\partial y}{\partial t}\right|_s = \left.\frac{\partial y}{\partial t}\right|_\eta + \left.\frac{\partial y}{\partial \eta}\right|_t \left.\frac{\partial \eta}{\partial t}\right|_s
$$

The output scalar field obeys the following partial differential equation:

$$
\left.\frac{\partial y}{\partial t}\right|_\eta + \left.\frac{\partial y}{\partial \eta}\right|_t (-\omega \times \eta + (\eta.\xi)\eta - \xi) = -2(\xi.\eta) y.
$$

From $\left.\frac{\partial}{\partial t}\right|_s \|\overrightarrow{CM}(s)\| = -\frac{d}{dt} C . \frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|} = -v.\eta$ we deduce also ($D = \|\overrightarrow{CM}\|$)

$$
\left.\frac{\partial D}{\partial t}\right|_\eta + \left.\frac{\partial D}{\partial \eta}\right|_t (-\omega \times \eta + (\eta.\xi)\eta - \xi) = -v.\eta.
$$

Let $\xi = v(t)/D$, let $\Lambda = 1/D$, they are both functions of (t, η). We have the following kinematic model (partial derivatives are with respect to the two independent variables (t, η)):

$$
\begin{aligned}
&\frac{d}{dt} q = \frac{1}{2} q * \omega \\
&\frac{d}{dt} v = v \times \omega + q^{-1} * \mathbf{A}_{grav} * q + a \\
&\frac{\partial \Lambda}{\partial t} = -\frac{\partial \Lambda}{\partial \eta} (\eta \times (\omega + \Lambda \eta \times v)) + \Lambda^2 v.\eta \\
&\frac{\partial y}{\partial t} = -\frac{\partial y}{\partial \eta} (\eta \times (\omega + \Lambda \eta \times v)) - 2\Lambda (v.\eta) y
\end{aligned} \tag{1}
$$

where

- $q(t) \in \mathbb{H}_1$ and $v(t) \in \mathbb{R}^3$ are the unmeasured part of the state, of finite dimension.
- $\mathbb{S}^2 \ni \eta \mapsto \Lambda(t, \eta) \in \mathbb{R}_*^+$ is the infinite dimensional part of the non-measured part of the state.
- $\mathbb{S}^2 \ni \eta \mapsto y(t, \eta) \in \mathbb{R}_*^+$ is the infinite dimensional part of the measured part of the state.
- The varying vectors $\omega(t), a(t) \in \mathbb{R}^3$ and the fixed vector $\mathbf{A}_{grav} \in \mathbb{R}^3 / \{0\}$ are known.

Note that this system is invariant under the action of the group $G = \mathbb{H}_1$. The action is defined by (right multiplication corresponding to a change of body-fixed frame)

$$
\begin{aligned}
(q, v, \eta, \omega, a, \mathbf{A}_{grav}) \mapsto (qg, &g^{-1} v g, g^{-1} \eta g, \ldots \\
&g^{-1} \omega g, g^{-1} a g, g^{-1} \mathbf{A}_{grav} g)
\end{aligned}
$$

for all $g \in G$. Λ and y are unchanged by the transformation since η is replaced with $g^{-1} \eta g$ (change of parameterization). It is also invariant under the other action of $G = \mathbb{H}_1$ (left multiplication, corresponding to a change of earth-fixed frame)

$$
(q, \mathbf{A}_{grav}) \mapsto (gq, g \mathbf{A}_{grav} g^{-1})
$$

where all the rest is unchanged. The model (1) is a totally intrinsic model based on the geometry of the problem. With some additional information (v is known, the motion is 2-dimensional...) it can be used to build observers in order to estimate quantities which are not directly measured, as in the sequel.

4. FAR ENVIRONMENT AND OBSERVER-BASED ESTIMATION OF THE GYROSCOPES BIAS

We are going to study a simple case. We assume that the environment (the scene) is far away (it can be the stars, the earth seen from a plane...), so that we can make the approximation $\Lambda = 1/\|CM\| \approx 0$ since $\|CM\|$ is very large. Thus (1) becomes

$$
\begin{aligned}
&\frac{d}{dt} q = \frac{1}{2} q * \omega \\
&\frac{d}{dt} v = v \times \omega + q^{-1} * \mathbf{A}_{grav} * q + a \\
&\frac{\partial y}{\partial t} = -\frac{\partial y}{\partial \eta} (\eta \times \omega)
\end{aligned}
$$

So we see that we can not estimate v and q using directly the signal y. But one can estimate a constant bias $c \in \mathbb{R}^3$ on the measurement ω_m of the gyroscope: $\omega = \omega_m + c$. Indeed in [13], an Extended Kalman Filter which corrects the gyro bias error using celestial observations from star trackers is built. But once again background knowledge about the scene is used, since 1183 bright stars are stored in the spacecraft computer, and the brighter ones are tracked. We are going to correct the bias without background knowledge on the scene as follows: the evolution in time of the output map y now writes

$$
\frac{\partial y}{\partial t} + \frac{\partial y}{\partial \eta} (\eta \times \omega) = 0
$$

which can be written

$$
\frac{\partial y}{\partial t} + \frac{\partial y}{\partial \eta} (\eta \times \omega_m) + \frac{\partial y}{\partial \eta} (\eta \times c) = 0
$$

At that stage, several solutions can be imagined to estimate $c \in \mathbb{R}^3$ from the measurement $\mathbb{S}^2 \ni \eta \mapsto y(t, \eta)$. Let us give one of the Lyapunov type.

4.1 *The observer*

We are going to build a non-linear observer for this infinite dimensional problem. The mean of y on the orbit spanned by the action of rotations on the sphere $\int_{\mathbb{S}^2} y(t, \eta) d\sigma_\eta$ where $d\sigma$ is the aera element of the unit sphere $\mathbb{S}^2$, is a scalar invariant, since the group action is an isometry. It is independent from the time since it is the total amount of light received from the scene, which is constant over the time according to our assumptions applied to a far environment.

Let ∇y denote the gradient of the scalar function $\mathbb{S}^2 \ni \eta \mapsto y(t, \eta)$. So ∇y can be seen as a vector of $\mathbb{R}^3$, tangent to the sphere at η, and thus $\nabla y \cdot \eta = 0$. We have

$$
\frac{\partial y}{\partial \eta} (\eta \times \omega_m) = \nabla y \cdot (\eta \times \omega_m) = (\nabla y \times \eta) \cdot \omega_m
$$

and the dynamics writes:

$$
\frac{\partial y}{\partial t} = -(\nabla y \times \eta) \cdot \omega = -(\nabla y \times \eta) \cdot (\omega_m + c)
$$

Let $k_y, k_c > 0$ be two constant gains. Consider the following asymptotic observer:

$$
\begin{aligned}
&\frac{\partial \hat{y}}{\partial t} = -(\nabla \hat{y} \times \eta) \cdot (\omega_m + \hat{c}) - k_y (\hat{y} - y) \\
&\frac{d}{dt} \hat{c} = k_c \int_{\mathbb{S}^2} (\hat{y} - y)(\nabla \hat{y} \times \eta) \, d\sigma
\end{aligned} \tag{2}
$$

At this point, in time the authors are unable to provied a rigorous proof of convergence of the observer, however, we offer the following discussion to justify its form. The discussion relies on the following Lyapunov function:

$$V = \frac{1}{2}\int_{\mathbb{S}^2} (\hat{y} - y)^2 d\sigma + \frac{1}{2k_c}\|\hat{c} - c\|^2.$$

For all q, $\eta \mapsto q^{-1} * \eta * q$ is an isometry of $\mathbb{S}^2$. Thus $d\sigma_\eta = d\sigma_{q^{-1}*\eta*q}$ and

$$\int_{\mathbb{S}^2} (\hat{y}(t,\eta) - y(t,\eta))^2 d\sigma_\eta = \int_{\mathbb{S}^2} (\hat{y}(t, q^{-1} * \eta * q) - y(t, q^{-1} * \eta * q))^2 d\sigma_\eta$$

Note that one can view η in the second member of the equation above as a constant vector of the earth-fixed frame. $q^{-1} * \eta * q$ is this vector expressed in the moving frame, thus $y(t, q^{-1} * \eta * q)$ is the light coming from a fixed point ("s=cste") of the scene, since it corresponds to the direction η. It does not depend on the time. We have

$$\frac{d}{dt}\left(\int_{\mathbb{S}^2} (\hat{y}(t,\eta) - y(t,\eta))^2 d\sigma_\eta\right) = \int_{\mathbb{S}^2} \frac{d}{dt}\left(\hat{y}(t, q^{-1} * \eta * q) - y(t, q^{-1} * \eta * q)\right)^2 d\sigma_\eta$$

Using the equality $\frac{d}{dt}q = \frac{1}{2}q * \omega$, and $\frac{\partial(\hat{y}-y)}{\partial\eta}(\eta \times \omega) = (\nabla(\hat{y} - y) \times \eta) \cdot \omega$ we have

$$\frac{d}{dt}\left(\int_{\mathbb{S}^2} (\hat{y} - y)^2 d\sigma\right) = \int_{\mathbb{S}^2} 2(\hat{y} - y)\left(\frac{\partial(\hat{y} - y)}{\partial t} + (\nabla(\hat{y} - y) \times \eta) \cdot \omega\right) d\sigma$$

Since $\frac{\partial(\hat{y}-y)}{\partial t} = -(\nabla(\hat{y}-y)\times\eta)\cdot\omega - k_y(\hat{y}-y) + (\nabla\hat{y}\times\eta)\cdot(c-\hat{c})$ et $\omega - c = \omega_m$, we have

$$\frac{d}{dt}\left(\int_{\mathbb{S}^2} (\hat{y} - y)^2 d\sigma\right) = 2\int_{\mathbb{S}^2} (\hat{y} - y)\left(-k_y(\hat{y} - y) - (\nabla\hat{y} \times \eta) \cdot (\hat{c} - c)\right) d\sigma$$

Thus

$$\frac{d}{dt}V = -2k_y \int_{\mathbb{S}^2} (\hat{y} - y)^2 d\sigma \le 0.$$

A more advanced convergence analysis requires the Lassalle invariance principle, which application to the infinite dimensional case is not so easy. Intuitively, we have asymptotically $\hat{y} = y$ so (which is not obvious here) $\nabla\hat{y} = \nabla y$. It implies $(\nabla y \times \eta) \cdot (\hat{c} - c) = 0$. So if the set of $\nabla y \times \eta$ spans all $\mathbb{R}^3$ when η takes values in all $\mathbb{S}^2$, one must have asymptotically $\hat{c} = c$. Even if these ideas are semi-rigorous, they allow to give a reasonnable condition of convergence: $\text{vect}\,(\nabla y \times \eta)_{\eta\in\mathbb{S}^2} = \mathbb{R}^3$, i.e., the image must have a contrast in all three directions.

4.2 Pinhole model

In this section, we show how the observer can be adapted if the camera used is modeled by the widespread standard pinhole camera model (see fig 1). In fact, we think that many results obtained with the totally symmetric model (1) (invariant by rotation using a spherical camera), can be extended to the non-symmetric case of an "usual" camera. This is a not-so-obvious feature since the convergence analysis of the observer (2) is based on the invariance by rotation of the integral over the whole sphere. To sum up, the use of spherical cameras allows us to enhance the geomery of the problem and the invariance by $SO(3)$ (that we identify here to $\mathbb{H}_1$, see the appendix) but this assumption can be relaxed, restricting the integrals to a portion of the sphere the following way: let $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ be a C^∞ function of only η, whose support corresponds to the camera "window" (or pinhole). It is equal to 0 everywhere except in the interior of the window, and equal to 1 inside a smaller window contained in the window of the camera. This smaller window must be chosen as large as possible. On the window, the measured output y is multiplied by ϕ to derive a new output $Y(t,\eta) = \phi(\eta)\; y(t,\eta)$. It obeys the following partial differential equation

$$\frac{\partial Y}{\partial t} + (\nabla Y \times \eta - y\nabla\phi \times \eta).\omega = 0$$

The observer is modified as follows

$$\frac{\partial\hat{Y}}{\partial t} = -(\nabla\hat{Y} \times \eta - y\nabla\phi \times \eta) \cdot (\omega_m + \hat{c}) - k_y(\hat{Y} - Y),$$
$$\frac{d}{dt}\hat{c} = k_c \int_{\mathbb{S}^2} (\hat{Y} - Y)(\nabla\hat{Y} \times \eta - y\nabla\phi \times \eta)\, d\sigma \tag{3}$$

We take the same Lyapunov function

$$V = \frac{1}{2}\int_{\mathbb{S}^2} (\hat{Y} - Y)^2 d\sigma + \frac{1}{2k_c}\|\hat{c} - c\|^2.$$

and we have now

$$\frac{\partial(\hat{Y} - Y)}{\partial t} = -\,(\nabla(\hat{Y} - Y) \times \eta) \cdot \omega - k_y(\hat{Y} - Y) - (\nabla\hat{Y} \times \eta - y\nabla\phi \times \eta) \cdot (\hat{c} - c)$$

and the first term of the right-hand expression vanishes when it is integrated over the whole sphere, as in the preceeding sections.

The observer is also given using the usual cartesian coordinates of the pinhole model in the appendix (section 6.2).

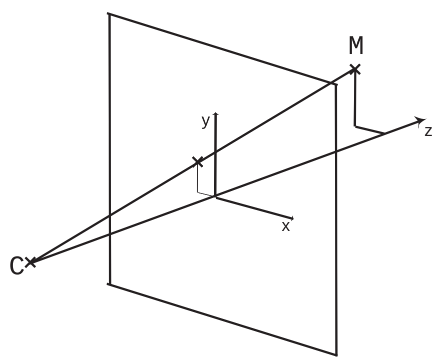

Fig. 1. Cartesian coordinates and "pinhole model". C is the optical center, Cz the optical axis, Cx and Cy are parrallel to the image plane, and (Cxyz) is an orthonormal frame. The point M is expressed in cartesian coordinates in the camera frame. We assume the focal length to be 1.

4.3 Comparison between our approach and optical-flow based methods

First of all we recall the observer (2) uses directly the output signal y without using any type of background knowledge of the scene. Moreover it provides a filter for the noisy output y, and computes a filtered image $\hat{y}$ in real time. Contrarily to featureless methods based on optical flow or spatio-temporal derivatives, which are standard when nothing is known about the environment (see [9]), the image described by y is never differentiated. It is even integrated over space and time when computing $\hat{c}$, which allows to very efficiently filter the high frequencies (noise). Integrations over space are standard in image processing (see e.g. [1]) and provide generally very efficient filters to noise, although we did not provide simulations in this article to support this claim.

The last advantage of this observer-based approach, is that the noise is never rectified. Indeed the general featureless methods almost always require a least squares fit. Thus the estimation uses the square of the output or its derivatives. For instance, if the measured output is noisy $y_m = y + w$ where w is a standard gaussian white noise, the mean of y_m^2 is not equal to the mean of y^2 since w^2 is not a white noise, and the estimation is biased.

5. CONCLUSION

The main contribution of this paper is to give a mathematical geometric formulation of the inertial-vision fusion problem, and to propose a simple observer to estimate the gyro bias. The method developped is different from the usual inertial-vision techniques. In the future this approach should be tested on real data, and extended to estimate other quantities (pose, attitude..) the following way. When Λ is small, one can neglect in (1) the second order terms in Λ and keep those of order one, including partial derivatives in η. We obtain the following approximated system

$$\begin{aligned}
\frac{d}{dt}q &= \frac{1}{2}q * \omega \\
\frac{d}{dt}v &= v \times \omega + q^{-1} * \mathbf{A}_{grav} * q + a \\
\frac{\partial \Lambda}{\partial t} &= -\frac{\partial \Lambda}{\partial \eta}(\eta \times \omega) \\
\frac{\partial y}{\partial t} &= -\frac{\partial y}{\partial \eta}(\eta \times (\omega + \Lambda\eta \times v)) - 2\Lambda(v.\eta)y
\end{aligned}$$

This approximated model can be used to do pose and attitude estimation in some particular cases.

6. APPENDIX

6.1 Quaternions

As in [6], we use the quaternion parameterization of SO(3) to derive filters for state estimation. The quaternions are a non commutative group. Any quaternion q can be written $q = q^0 + q^1 e_1 + q^2 e_2 + q^3 e_3$ with $(q^0, q^1, q^2, q^3) \in \mathbb{R}^4$, the multiplication $*$ is defined by

$e_1 * e_1 = -1$, $e_1 * e_2 = -e_2 * e_1 = e_3$ with circular permutations and the norm of q is $\sqrt{(q^0)^2 + (q^1)^2 + (q^2)^2 + (q^3)^2}$. Any vector $\boldsymbol{p} \in \mathbb{R}^3$ can be identified with the quaternion $p^1 e_1 + p^2 e_2 + p^3 e_3$. We make this identification systematically. Then one can associate to any quaternion whose norm is 1, a rotation matrix $R_q \in SO(3)$ thanks to the following equality: $q^{-1} * \boldsymbol{p} * q = R_q \boldsymbol{p}$ for all $\boldsymbol{p}$. The subgroup of quaternions whose norm is 1 is denoted by $\mathbb{H}_1$. Conversely, to any rotation R_q of $SO(3)$ are associated two quaternions $\pm q$ of length 1. Thus we will write the elements of $SO(3)$ as quaternions whose norm is 1 (denoted by $\mathbb{H}_1$) and the vectors of $\mathbb{R}^3$ as quaternions whose first coordinate is equal to 0. Numerically, quaternions are easier to manipulate and compute than matrices in $SO(3)$. The wedge product $v \times \omega$ of vectors of $\mathbb{R}^3$ writes for the associated quaternions: $(v * \omega - \omega * v)/2$.

6.2 Cartesian coordinates

In the classical camera pinhole model, each pixel of the image has two coordinates $(x, y) \in \mathbb{R}^2$. A point M of the scene has cartesian coordinates $(x, y, z) \in \mathbb{R}^3$ (see fig 1). Its projection on the sphere is $\mathbb{S}^2 \ni (x', y') = (\sin(\arctan x), \sin(\arctan y))$. We write $\eta = (0, x', y', \sqrt{1 - x'^2 - y'^2})^T$. Indeed $\eta \in \mathbb{S}^2$ is a quaternion of norm 1, whose first coordinate is 0. The output is now a function $h(t, x, y)$ (we let h denote the output map rather than y, not to be confused with the y-coordinate) and the observer (3) writes

$$\begin{aligned}
\frac{\partial \hat{Y}}{\partial t} &= -(\nabla \hat{Y} \times \eta - h(t, x, y)\nabla\phi \times \eta) \cdot (\omega_m + \hat{c}) \\
&\quad - k_y(\hat{Y} - Y), \\
\frac{d}{dt}\hat{c} &= k_c \int_{\substack{-L \le x \le L \\ -h \le y \le h}} (\hat{Y} - Y).. \\
&\quad ..(\nabla \hat{Y} \times \eta - h(t, x, y)\nabla\phi \times \eta)\, d\Sigma(x', y', z')
\end{aligned}$$

where $d\Sigma$ is the area element of the sphere using cartesian coordinates $z' = \sqrt{1 - x'^2 - y'^2}$.

REFERENCES

[1] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mech. Anal*, 123:199–257, 1993.

[2] S. Benhimane and E. malis. Homography-based 2d visual tracking and servoing. *The International Journal of Robotics Research*, page 661, July 2007.

[3] S. Bonnabel, Ph. Martin, and P. Rouchon. Symmetry-preserving observers. *IEEE Transactions on Automatic Control*, 53:2514 – 2526, 2008.

[4] T. Hamel R. Mahony Cheviron, T. and G. Baldwin. Robust nonlinear fusion of inertial and visual data for position, velocity and attitude estimation of uav. In *International Conference on Robotics and Automation*, 2007.

[5] John L. Crassidis, F. Landis Markley, and Yang Cheng. Nonlinear attitude filtering methods. *Journal of Guidance, Control,and Dynamics*, 30(1):12–28, 2007.

[6] G. Creamer. Spacecraft attitude determination using gyros and quaternion measurements. *Journal of Astronautical Sciences*, 44(3):357–371, 1996.

[7] O. Faugeras. *Three-Dimensional Computer Vision*. The MIT press, Cambridge, Massachussets, 1993.

[8] J. Lobo and J. Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12): 1597–1608, 2003.

[9] S. Mann and R. W. Picard. Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Trans. on Image Processing*, 6(9), 1997.

[10] A. Matveev, X. Hu, R. Frezza, and H. Rehbinder. Observers for systems with implicit output. *IEEE Trans. Automat. Contr.*, 45:168–173, 2000.

[11] H. Rehbinder and B.K. Ghosh. Pose estimation using line-based dynamic vision and inertial sensors. *IEEE Transactions on Automatic Control.*, 48(2):186–199, 2003.

[12] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41:393–413, 1996.

[13] A. Wu. Stellar inertial attitude determination for leo spacecraft. In *Proceedings of the 35th Conference on Decision and Control*, Kobe, Japan, December 1996.