SO(3)-invariant asymptotic observers for dense depth field estimation based on visual data and known camera motion

Nadège Zarrouati, Emanuel Aldea, Pierre Rouchon

SO(3)-invariant asymptotic observers for dense depth field estimation based on visual data and known camera motion

Nadège Zarrouati * Emanuel Aldea † Pierre Rouchon ‡

September 30, 2011

Abstract

In this paper, we use known camera motion associated to a video sequence of a static scene in order to estimate and incrementally refine the surrounding depth field. We exploit the SO(3)-invariance of brightness and depth fields dynamics to customize standard image processing techniques. Inspired by the Horn-Schunck method, we propose a SO(3)-invariant cost to estimate the depth field. At each time step, this provides a diffusion equation on the unit Riemannian sphere that is numerically solved to obtain a real time depth field estimation of the entire field of view. Two asymptotic observers are derived from the governing equations of dynamics, respectively based on optical flow and depth estimations: implemented on noisy sequences of synthetic images as well as on real data, they perform a more robust and accurate depth estimation. This approach is complementary to most methods employing state observers for range estimation, which uniquely concern single or isolated feature points.

1 Introduction

Many vision applications are aimed at assisting in interacting with the environment. In military as well as in civilian applications, moving in an environment requires topographical knowledge: either in order to avoid obstacles or to engage targets. Since this information is often inaccessible in advance, the real-time computation of a 3D map is a goal that has kept the research community busy

*Nadège Zarrouati is with DGA, 7-9 rue des Mathurins, 92220 Bagneux, France, PHD candidate at Mines-ParisTech, Centre Automatique et Systèmes, Unité Mathématiques et Systèmes 60, boulevard Saint-Michel 75272 Paris Cedex, France nadege.zarrouati@mines-paristech.fr

†Emanuel Aldea is with SYSNAV, 57 rue de Montigny, 27200 Vernon, France emanuel.aldea@sysnav.fr

‡Pierre Rouchon is with Mines-ParisTech, Centre Automatique et Systèmes, Unité Mathématiques et Systèmes 60, boulevard Saint-Michel 75272 Paris Cedex, France pierre.rouchon@mines-paristech.fr

for many years. For example, environment reconstruction is tightly related to the SLAM problem [1], which is addressed by nonlinear filtering of observed key feature locations (e.g. [2, 3]), or by bundle adjustment [4, 5]. However, estimating a sparse point cloud is often insufficient, yet the transition between a discrete local distribution of 3D locations to a continuous depth estimation of the surroundings is an ongoing research topic. Dynamical systems provide interesting means for incrementally estimating depth information based on the output of vision sensors, since only the current estimates are required, and image batch processing is avoided.

For our work, we are interested in recovering in real-time the depth field around the carrier under the assumptions of known camera motion and known projection model for the onboard monocular camera. The problem of designing an observer to estimate the depth of *single or isolated keypoints* has raised a lot of interest, specifically in the case where the relative motion of the carrier is described by constant known [6, 7], constant unknown [8] or time-varying known [9, 10, 11, 12, 13] affine dynamics. From a different perspective, the seminal paper of [14] performs incremental depth field refining for the whole field of view via *iconic* (pixel-wise) Kalman filtering. Video systems, typically found on autonomous vehicles, have successfully used this approach for refining the disparity values obtained by stereo cameras in order to estimate the free space ahead [15]. Average optical flow estimations over planar surfaces have also been used for terrain following, in order to stabilize the carrier at a certain pseudo-distance [16]. Yet, none of these methods provide an accurate dense depth estimation in a general setting concerning the environment and the camera dynamics.

We propose a novel frame of methods relying on a system of partial differential equations describing the $SO(3)$-invariant dynamics of the brightness perceived by the camera and of the depth field of the environment. Based on this invariant kinematic model and the knowledge of the camera motion, these methods provide dense estimations of the depth field at each time-step and exploit such $SO(3)$-invariance.

The present paper is structured as follows. The invariant equations governing the dynamics of the brightness and depth fields are recalled in section 2 and their formulation in pinhole coordinates is given. In section 3, we adapt the Horn-Schunck algorithm to a variational method providing depth estimation. In section 4, we propose two asymptotic observers for depth field estimations: the first one is based on standard optical flow measures and the second one enables the refinement of rough or inaccurate depth estimations; we prove their convergence under geometric assumptions concerning the camera dynamics and the environment. In section 5, we test these methods on synthetic data and compare their accuracy, their robustness to noise and their convergence rate; tested on real data, this approach gives promising results.

2 The $SO(3)$-invariant model

2.1 The partial differential system on $\mathbb{S}^2$

The model is based on geometric assumptions introduced in [17]. We consider a spherical camera, whose motion is known. Linear and angular velocities $v(t)$ and $\omega(t)$ are expressed in the camera frame. Position of the optical center in the reference frame $\mathcal{R}$ is denoted by $C(t)$. Orientation versus $\mathcal{R}$ is given by the quaternion $q(t)$: any vector ς in the camera frame corresponds to the vector $q\varsigma q^*$ in the reference frame $\mathcal{R}$ using the identification of vectors as imaginary quaternions. We have thus: $\frac{d}{dt}q = \frac{1}{2}q\omega$. A pixel is labeled by the unit vector η in the camera frame: η belongs to the sphere $\mathbb{S}^2$ and receives the brightness $y(t,\eta)$. Thus at each time t, the image produced by the camera is described by the scalar field $\mathbb{S}^2 \ni \eta \mapsto y(t,\eta) \in \mathbb{R}$.

The scene is modeled as a closed, C^1 and convex surface Σ of $\mathbb{R}^3$, diffeomorphic to $\mathbb{S}^2$. The camera is inside the domain $\Omega \subset \mathbb{R}^3$ delimited by $\Sigma = \partial\Omega$. To a point $M \in \Sigma$ corresponds one and only one camera pixel: if the points of Σ are labeled by $s \in \mathbb{S}^2$, for each time t, a continuous and invertible transformation $\mathbb{S}^2 \ni s \mapsto \phi(t,s) \in \mathbb{S}^2$ enables to express η as a function of s: $\eta = \phi(t,s)$.

The density of light emitted by a point $M(s) \in \Sigma$ does not depend on the direction of emission (Σ is a Lambertian surface) and is independent of t (the scene is static). This means that $y(t,\eta)$ depends only on s: thus y can be seen either as a function of (t,η) or, via the transformation ϕ, as a function of s. The distance $C(t)M(s)$ between the optical center and the object seen in the direction $\eta = \phi(t,s)$ is denoted by $D(t,\eta)$, and its inverse by $\Gamma = 1/D$. Fig.1 illustrates the model and the notations. We assume that $s \mapsto y(s)$ is a C^1 function. For each t, $s \mapsto D(t,s)$ is C^1 since Σ is a C^1 surface of $\mathbb{R}^3$.

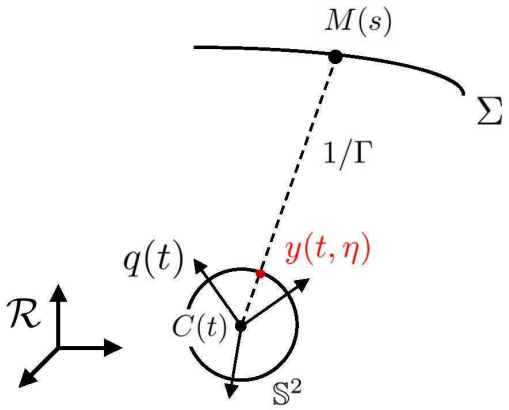

Figure 1: Model and notations of a spherical camera in a static environment.

Under these assumptions, we first have:

$$\left.\frac{\partial y}{\partial t}\right|_s = 0, \quad \left.\frac{\partial \Gamma}{\partial t}\right|_s = \Gamma^2 v \cdot \eta \tag{1}$$

then

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

where h is any scalar field defined on $\mathbb{S}^2$ and ∇h its gradient with respect to the Riemannian metric on $\mathbb{S}^2$. The value of ∇h at $\eta \in \mathbb{S}^2$ is identified with a vector of $\mathbb{R}^3$ tangent to the sphere at the point η also identified to a unitary vector of $\mathbb{R}^3$ in the camera moving frame. The Euclidean scalar product of two vectors a and b in $\mathbb{R}^3$ is denoted by $a \cdot b$ and their wedge product by $a \times b$. By differentiation, the identity $q\eta q^* = \frac{\overrightarrow{C(t)M(s)}}{||C(t)M(s)||}$, where * denotes conjugation and η is identified to an imaginary quaternion, yields

$$\left.\frac{\partial \eta}{\partial t}\right|_s = \eta \times (\omega + \Gamma \eta \times v) \tag{2}$$

since the vector $\eta \times \omega$ corresponds to the imaginary quaternion $(\omega\eta - \eta\omega)/2$. Therefore, the intensity $y(t, \eta)$ and the inverse depth $\Gamma(t, \eta)$ satisfy the following equations:

$$\frac{\partial y}{\partial t} = -\nabla y \cdot (\eta \times (\omega + \Gamma \eta \times v)) \tag{3}$$

$$\frac{\partial \Gamma}{\partial t} = -\nabla \Gamma \cdot (\eta \times (\omega + \Gamma \eta \times v)) + \Gamma^2 v \cdot \eta \tag{4}$$

Equations (3) and (4) are $SO(3)$-invariant: they remain unchanged by any rotation described by the quaternion σ and changing (η, ω, v) to $(\sigma\eta\sigma^*, \sigma\omega\sigma^*, \sigma v\sigma^*)$. Equation (3) is the well-known optical flow equation that can be found under different forms in numerous papers (see [18] or [13] for example), while (4) is less standard (see e.g., [17]).

2.2 The system in pinhole coordinates

To use this model with camera data, one needs to write the invariant equations (3) and (4) with local coordinates on $\mathbb{S}^2$ corresponding to a rectangular grid of pixels. One popular solution is to use the pinhole camera model, where the pixel of coordinates (z_1, z_2) corresponds to the unit vector $\eta \in \mathbb{S}^2$ of coordinates in $\mathbb{R}^3$: $\left(1 + z_1^2 + z_2^2\right)^{-1/2} (z_1, z_2, 1)^T$. The optical camera axis (pixel $(z_1, z_2) = (0, 0)$) corresponds here to the direction z_3. Directions 1 and 2 correspond respectively to the horizontal axis from left to right and to the vertical axis from top to bottom on the image frame.

The gradients ∇y and $\nabla \Gamma$ must be expressed with respect to z_1 and z_2. Let us detail this derivation for y. Firstly, ∇y is tangent to $\mathbb{S}^2$, thus $\nabla y \cdot \eta = 0$.

Secondly, the differential dy corresponds to $\nabla y \cdot d\eta$ and to $\frac{\partial y}{\partial z_1}dz_1 + \frac{\partial y}{\partial z_2}dz_2$. By identification, we get the Cartesian coordinates of ∇y in $\mathbb{R}^3$. Similarly we get the three coordinates of $\nabla\Gamma$. Injecting these expressions in (3) and (4), we get the following partial differential equations (PDE) corresponding to (3) and (4) in local pinhole coordinates:

$$\begin{aligned}
\frac{\partial y}{\partial t} =& -\frac{\partial y}{\partial z_1}\begin{bmatrix} z_1 z_2\omega_1 - (1+{z_1}^2)\omega_2 + z_2\omega_3 \\ +\Gamma\sqrt{1+z_1^2+z_2^2}(-v_1+z_1v_3)\end{bmatrix} \\
& -\frac{\partial y}{\partial z_2}\begin{bmatrix} (1+{z_2}^2)\omega_1 - z_1 z_2\omega_2 - z_1\omega_3 \\ +\Gamma\sqrt{1+z_1^2+z_2^2}(-v_2+z_2v_3)\end{bmatrix} \\
\frac{\partial \Gamma}{\partial t} =& -\frac{\partial \Gamma}{\partial z_1}\begin{bmatrix} z_1 z_2\omega_1 - (1+{z_1}^2)\omega_2 + z_2\omega_3 \\ +\Gamma\sqrt{1+z_1^2+z_2^2}(-v_1+z_1v_3)\end{bmatrix} \\
& -\frac{\partial \Gamma}{\partial z_2}\begin{bmatrix} (1+{z_2}^2)\omega_1 - z_1 z_2\omega_2 - z_1\omega_3 \\ +\Gamma\sqrt{1+z_1^2+z_2^2}(-v_2+z_2v_3)\end{bmatrix} \\
& +\Gamma^2(z_1v_1+z_2v_2+v_3)
\end{aligned}$$

where v_1, v_2, v_3, ω_1, ω_2, ω_3 are the components of linear and angular velocities in the camera frame.

3 Depth estimation inspired by Horn-Schunck method

3.1 The Horn-Schunck variational method

In [19], Horn and Schunck described a method to compute the optical flow, defined as "the distribution of apparent velocities of movement of brightness patterns in an image". The entire method is based on the optical flow constraint written in the compact form

$$\frac{\partial y}{\partial t} + V_1\frac{\partial y}{\partial z_1} + V_2\frac{\partial y}{\partial z_2} = 0 \tag{5}$$

Identification with (5) yields

$$\begin{aligned}
V_1(t,z) &= f_1(z,\omega(t)) + \Gamma(t,z)g_1(z,v(t)) \\
V_2(t,z) &= f_2(z,\omega(t)) + \Gamma(t,z)g_2(z,v(t))
\end{aligned} \tag{6}$$

with

$$\begin{aligned}
f_1(z,\omega) &= z_1 z_2\omega_1 - (1+{z_1}^2)\omega_2 + z_2\omega_3 \\
g_1(z,v) &= \sqrt{1+z_1^2+z_2^2}(-v_1+z_1v_3) \\
f_2(z,\omega) &= (1+{z_2}^2)\omega_1 - z_1 z_2\omega_2 - z_1\omega_3 \\
g_2(z,v) &= \sqrt{1+z_1^2+z_2^2}(-v_2+z_2v_3).
\end{aligned} \tag{7}$$

For each time t, the apparent velocity field $V = V_1 \frac{\partial}{\partial z_1} + V_2 \frac{\partial}{\partial z_2}$ is then estimated by minimizing versus $W = W_1 \frac{\partial}{\partial z_1} + W_2 \frac{\partial}{\partial z_2}$ the following cost (the image $\mathcal{I}$ is a rectangle of $\mathbb{R}^2$ here)

$$I(W) = \iint_{\mathcal{I}} \Big(\big(\frac{\partial y}{\partial t} + W_1 \frac{\partial y}{\partial z_1} + W_2 \frac{\partial y}{\partial z_2} \big)^2 + \alpha^2 (\nabla {W_1}^2 + \nabla {W_2}^2) \Big) dz_1 dz_2 \quad (8)$$

where ∇ is the gradient operator in the Euclidian plane (z_1, z_2), $\alpha > 0$ is a regularization parameter and the partial derivatives $\frac{\partial y}{\partial t}$, $\frac{\partial y}{\partial z_1}$ and $\frac{\partial y}{\partial z_2}$ are assumed to be known.

Such Horn-Schunk estimation of V at time t is denoted by

$$V_{\text{HS}}(t, z) = V_{\text{HS}1}(t, z) \frac{\partial}{\partial z_1} + V_{\text{HS}2}(t, z) \frac{\partial}{\partial z_2}.$$

For each time t, usual calculus of variation yields the following PDE's for V_{HS}:

$$\left(\frac{\partial y}{\partial z_1} \right)^2 V_{\text{HS}1} + \frac{\partial y}{\partial z_1} \frac{\partial y}{\partial z_2} V_{\text{HS}2} = \alpha^2 \Delta V_{\text{HS}1} - \frac{\partial y}{\partial z_1} \frac{\partial y}{\partial t}$$
$$\frac{\partial y}{\partial z_1} \frac{\partial y}{\partial z_2} V_{\text{HS}1} + \left(\frac{\partial y}{\partial z_2} \right)^2 V_{\text{HS}2} = \alpha^2 \Delta V_{\text{HS}2} - \frac{\partial y}{\partial z_2} \frac{\partial y}{\partial t}$$

with boundary conditions $\frac{\partial V_{\text{HS}1}}{\partial n} = \frac{\partial V_{\text{HS}2}}{\partial n} = 0$ (n the normal to $\partial \mathcal{I}$). Here Δ is the Laplacian operator in the Euclidian space (z_1, z_2). The numerical resolution is usually based on

- computations of $\frac{\partial y}{\partial z_1}$, $\frac{\partial y}{\partial z_2}$ and $\frac{\partial y}{\partial t}$ via differentiation filters (Sobel filtering) directly from the image data at different times around t.
- approximation of $\Delta V_{\text{HS}1}$ and $\Delta V_{\text{HS}2}$ by the difference between the weighted mean $\bar{V_{\text{HS}1}}$ and $\bar{V_{\text{HS}2}}$ of $V_{\text{HS}1}$ and $V_{\text{HS}2}$ on the neighboring pixels and their values at the current pixel;
- iterative resolution (Jacobi scheme) of the resulting linear system in $V_{\text{HS}1}$ and $V_{\text{HS}2}$.

The convergence of this numerical method of resolution was proven in [20]. Three parameters have a direct impact on the speed of convergence and on the precision: the regularization parameter α, the number of iterations for the Jacobi scheme and the initial values of $V_{\text{HS}1}$ and $V_{\text{HS}2}$ at the beginning of this iteration step. To be specific, α should neither be too small in order to filter noise appearing in differentiation filters applied on y, nor too large in order to have V_{HS} close to V when $\nabla V \neq 0$.

3.2 Adaptation to depth estimation

Instead of minimizing the cost I given by (8) with respect to any W_1 and W_2, let us define a new invariant cost J,

$$J(\Upsilon) = \iint_{\mathcal{J}} \Bigg(\left(\frac{\partial y}{\partial t} + \nabla y \cdot (\eta \times (\omega + \Upsilon \eta \times v)) \right)^2 + \alpha^2 \nabla \Upsilon^2 \Bigg) d\sigma_\eta \quad (9)$$

and minimize it with respect to any depth profile $\mathcal{J} \ni \eta \mapsto \Upsilon(t, \eta) \in \mathbb{R}$. The time t is fixed here and $d\sigma_\eta$ is the Riemannian infinitesimal surface element on $\mathbb{S}^2$. $\mathcal{J} \subset \mathbb{S}^2$ is the domain where y is measured and $\alpha > 0$ is the regularization parameter.

The first order stationary condition of J with respect to any variation of Υ yields the following invariant PDE characterizing the resulting estimation Γ_{HS} of Γ:

$$\alpha^2 \Delta \Gamma_{\text{HS}} = \left(\frac{\partial y}{\partial t} + \nabla y \cdot (\eta \times (\omega + \Gamma_{\text{HS}} \eta \times v)) \right) \ldots \ldots (\nabla y \cdot (\eta \times (\eta \times v))) \qquad \text{on } \mathcal{J} \quad (10)$$

$$\frac{\partial \Gamma_{\text{HS}}}{\partial n} = 0 \qquad \text{on } \partial \mathcal{J} \quad (11)$$

where $\Delta \Gamma_{\text{HS}}$ is the Laplacian of Γ_{HS} on the Riemannian sphere $\mathbb{S}^2$ and $\partial \mathcal{J}$ is the boundary of $\mathcal{J}$, assumed to be piece-wise smooth and with unit normal vector n.

In pinhole coordinates (z_1, z_2), we have

$$\begin{aligned} d\sigma_\eta &= \left(1 + z_1^2 + z_2^2\right)^{-3/2} dz_1 dz_2 \\ \nabla \Upsilon^2 &= \left(1 + z_1^2 + z_2^2\right) \left(\frac{\partial \Upsilon}{\partial z_1}^2 + \frac{\partial \Upsilon}{\partial z_2}^2 + (z_1 \frac{\partial \Upsilon}{\partial z_1} + z_2 \frac{\partial \Upsilon}{\partial z_2})^2 \right) \\ \left(\frac{\partial y}{\partial t} + \nabla y \cdot (\eta \times (\omega + \Upsilon \eta \times v)) \right)^2 &= (F + \Upsilon G)^2 \end{aligned}$$

where

$$\begin{aligned} F &= \frac{\partial y}{\partial t} + f_1(z, \omega) \frac{\partial y}{\partial z_1} + f_2(z, \omega) \frac{\partial y}{\partial z_2} \\ G &= g_1(z, v) \frac{\partial y}{\partial z_1} + g_2(z, v) \frac{\partial y}{\partial z_2}. \end{aligned} \quad (12)$$

Consequently, the first order stationary condition (10) reads in (z_1, z_2) coordi-

nates:

$$\begin{aligned}\Gamma_{\text{HS}}G^2 + FG =& \alpha^2 \left[\frac{\partial}{z_1} \left(\frac{1+z_1^2}{\sqrt{1+z_1^2+z_2^2}} \frac{\partial \Gamma_{\text{HS}}}{\partial z_1} \right) \right. \\ &+ \frac{\partial}{z_2} \left(\frac{1+z_2^2}{\sqrt{1+z_1^2+z_2^2}} \frac{\partial \Gamma_{\text{HS}}}{\partial z_2} \right) \\ &+ \frac{\partial}{z_2} \left(\frac{z_1 z_2}{\sqrt{1+z_1^2+z_2^2}} \frac{\partial \Gamma_{\text{HS}}}{\partial z_1} \right) \\ &\left. + \frac{\partial}{z_1} \left(\frac{z_1 z_2}{\sqrt{1+z_1^2+z_2^2}} \frac{\partial \Gamma_{\text{HS}}}{\partial z_2} \right) \right]\end{aligned} \tag{13}$$

on the rectangular domain $\mathcal{I} = [-\bar{z}_1, \bar{z}_1] \times [-\bar{z}_2, \bar{z}_2]$ ($\bar{z}_1, \bar{z}_2 > 0$ with $\bar{z}_1^2 + \bar{z}_2^2 < 1$).

The right term of (13) corresponds to the Laplacian operator on the Riemannian sphere $\mathbb{S}^2$ in pinhole coordinates. The numerical resolution of this scalar diffusion providing the estimation Γ_{HS} of Γ is similar to the one used for the Horn-Schunck estimation V_{HS} of V. The functional $I(W)$ defined in (8) is minimized with respect to two varying parameters W_1 and W_2 while there is really only one unknown function in this problem: the depth field Γ. On the contrary, the functional $J(\Upsilon)$ takes full advantage of the knowledge of the camera dynamics since the only varying parameter here is Υ.

4 Depth estimation via asymptotic observers

4.1 Asymptotic observer based on optical flow measures (V_{HS})

From any optical flow estimation, such as V_{HS}, it is reasonable to assume that we have access for each time t, to the components in pinhole coordinates of the vector field

$$\varpi_t : \mathbb{S}^2 \ni \eta \mapsto \varpi_t(\eta) = \eta \times (\omega + \Gamma \eta \times v) \in T_\eta \mathbb{S}^2 \tag{14}$$

appearing in (4). This vector field can be considered as a measured output for (4), expressed as $\varpi_t(\eta) = f_t(\eta) + \Gamma(t,\eta) g_t(\eta)$, where f_t and g_t are the vector fields

$$f_t : \mathbb{S}^2 \ni \eta \mapsto f_t(\eta) = \eta \times \omega \in T_\eta \mathbb{S}^2 \tag{15}$$

$$g_t : \mathbb{S}^2 \ni \eta \mapsto g_t(\eta) = \eta \times (\eta \times v) \in T_\eta \mathbb{S}^2. \tag{16}$$

This enables us to propose the following asymptotic observer for $D = 1/\Gamma$ since it obeys to $\frac{\partial D}{\partial t} = -\nabla D \cdot \varpi_t - v \cdot \eta$:

$$\frac{\partial \widehat{D}}{\partial t} = -\nabla \widehat{D} \cdot \varpi_t - v \cdot \eta + k g_t \cdot (\widehat{D} f_t + g_t - \widehat{D} \varpi_t) \tag{17}$$

where ϖ_t, f_t and g_t are known time-varying vector fields on $\mathbb{S}^2$ and $k > 0$ is a tuning parameter. This observer is trivially $SO(3)$ invariant and reads in pinhole coordinates:

$$\begin{aligned}\frac{\partial \widehat{D}}{\partial t} &= -\frac{\partial \widehat{D}}{\partial z_1}V_1 - \frac{\partial \widehat{D}}{\partial z_2}V_2 - (z_1 v_1 + z_2 v_2 + v_3) \\ &\quad + k\big(g_1(\widehat{D}f_1 + g_1 - \widehat{D}V_1) + g_2(\widehat{D}f_2 + g_2 - \widehat{D}V_2)\big)\end{aligned} \tag{18}$$

where V is given by any optical flow estimation and (f_1, f_2, g_1, g_2) are defined by (7).

As assumed in the first paragraphs of subsection 2.1, for each time t, there is a one to one smooth mapping between $\eta \in \mathbb{S}^2$ attached to the camera pixel and the scene point $M(s)$ corresponding to this pixel. This means that, for any $t \geq 0$, the flow $\phi(t, s)$ defined by

$$\left.\frac{\partial \phi}{\partial t}\right|_{(t,s)} = \varpi_t(\phi(t, s)), \quad \phi(0, s) = s \in \mathbb{S}^2 \tag{19}$$

defines a time varying diffeomorphism on $\mathbb{S}^2$. Let us denote by ϕ^{-1} the inverse diffeomorphism: $\phi(t, \phi^{-1}(t, \eta)) \equiv \eta$. Assume that $\Gamma(t, \eta) > 0$, $v(t)$ and $\omega(t)$ are uniformly bounded for $t \geq 0$ and $\eta \in \mathbb{S}^2$. This means that the trajectory of the camera center $C(t)$ remains strictly inside the convex surface Σ with minimal distance to Σ. These considerations motivate the assumptions used in the following theorem.

Theorem 1. *Consider $\Gamma(t, \eta)$ associated to the motion of the camera inside the domain Ω delimited by the scene Σ, a C^1, convex and closed surface as explained in sub-section 2.1. Assume that exist $\bar{v} > 0$, $\bar{\omega} > 0$, $\bar{\gamma} > 0$ and $\bar{\Gamma} > 0$ such that*

$$\forall t \geq 0,\ \forall \eta \in \mathbb{S}^2,\ |v(t)| \leq \bar{v},\ |\omega(t)| \leq \bar{\omega},\ \bar{\gamma} \leq \Gamma(t, \eta) \leq \bar{\Gamma}.$$

Then, for $t \geq 0$, $\Gamma(t, \eta)$ is a C^1 solution of (4). Consider the observer (17) with a C^1 initial condition versus η, $\widehat{D}(0, \eta)$. Then we have the following implications:

- *$\forall t \geq 0$, the solution $\widehat{D}(t, \eta)$ of (17) exists, is unique and remains C^1 versus η. Moreover*

$$t \mapsto \|\widehat{D}(t, \_) - D(t, \_)\|_{L^\infty} = \max_{\eta \in \mathbb{S}^2} \left|\widehat{D}(t, \eta) - D(t, \eta)\right|$$

  is decreasing (L^∞ stability).

- *if additionally for all $s \in \mathbb{S}^2$, $\int_0^{+\infty} \|g_\tau(\phi(\tau, s))\|^2 d\tau = +\infty$, then we have for all $p > 0$,*

$$\lim_{t \mapsto +\infty} \int_{\mathbb{S}^2} \left|\widehat{D}(t, \eta) - D(t, \eta)\right|^p d\sigma_\eta = 0$$

  (convergence in any L^p topology)

- *if additionally there is $\lambda > 0$ and $T > 0$ such that, for all $t \geq T$ and $s \in \mathbb{S}^2$, $\int_0^t \|g_\tau(\phi(\tau, s))\|^2 d\tau \geq \lambda t$, then we have, for all $t \geq T$,*

$$\|\widehat{D}(t,\cdot) - D(t,\cdot)\|_{L^\infty} \leq e^{-k\bar{\gamma}\lambda t}\|\widehat{D}(0,\cdot) - D(0,\cdot)\|_{L^\infty}$$

(exponential convergence in L^∞ topology).

Assumptions on $\int \|g_t(\phi(t,s))\|^2 dt$ can be seen as a condition of persistent excitations. It should be satisfied for generic motions of the camera.

Proof. The facts that v, ω and Γ are bounded and that the scene surface Σ is C^1, closed and convex, ensure that the mapping $\eta = \phi(t,s)$ and its inverse $s = \phi^{-1}(t,\eta)$ are C^1 diffeomorphism on $\mathbb{S}^2$ with bounded derivatives versus s and η for all time $t > 0$. Therefore, Γ is also a function of (t,s). Set $\overline{\Gamma}(t,s) = \Gamma(t,\phi(t,s))$: in the (t,s) independent variables the partial differential equation (4) becomes a set of ordinary differential equations indexed by s: $\left.\frac{\partial\overline{\Gamma}}{\partial t}\right|_s = \overline{\Gamma}^2 v(t)\cdot\phi(t,s)$ that reads also $\left.\frac{\partial(\overline{D})}{\partial t}\right|_s = -v(t)\cdot\phi(t,s)$ with $\overline{D} = 1/\overline{\Gamma}$. Thus $\overline{D}(t,s) - \overline{D}(0,s) = -\int_0^t v(\tau)\cdot\phi(\tau,s)\, d\tau$. Consequently, $\overline{\Gamma}$ is C^1 versus s and thus Γ is C^1 versus η.

Set $\overline{\widehat{D}}(t,s) = \widehat{D}(t,\phi(t,s))$. Then

$$\left.\frac{\partial\overline{\widehat{D}}}{\partial t}\right|_s = -v(t)\cdot\phi(t,s) + k\|g_t(\phi(t,s))\|^2\overline{\Gamma}(t,s)(\overline{D}(t,s) - \overline{\widehat{D}}(t,s)).$$

Set $E = \widehat{D} - D$ and $\overline{E} = \overline{\widehat{D}} - \overline{D}$. Then

$$\left.\frac{\partial\overline{E}}{\partial t}\right|_s = -k\overline{\Gamma}(t,s)\|g_t(\phi(t,s))\|^2\overline{E}(t,s) \tag{20}$$

Consequently, $\bar{E}$ is well defined for any $t > 0$ and C^1 versus s. Thus E and consequently $\widehat{D} = E + D$ are also well defined for all $t > 0$ and are C^1 versus η.

Since for any s and $t_2 > t_1 \geq 0$ we have $|\overline{E}(t_2,s)| \leq |\overline{E}(t_1,s)|$, we have also

$$|\overline{E}(t_2,s)| \leq \max_\sigma |\overline{E}(t_1,\sigma)| = \|\widehat{D}(t_1,\cdot) - D(t_1,\cdot)\|_{L^\infty}.$$

Thus, taking the max versus s, we get

$$\|\widehat{D}(t_2,\cdot) - D(t_2,\cdot)\|_{L^\infty} \leq \|\widehat{D}(t_1,\cdot) - D(t_1,\cdot)\|_{L^\infty}.$$

Since

$$\overline{E}(t,s) = \overline{E}(0,s)e^{-k\int_0^t \left(\overline{\Gamma}(\tau,s)\|g_\tau(\phi(\tau,s))\|^2\right)d\tau}$$

we have ($\phi(0,s) \equiv s$).

$$|\overline{E}(t,s)| \leq |E(0,s)|e^{-k\bar{\gamma}\int_0^t \|g_\tau(\phi(\tau,s))\|^2\, d\tau}.$$

Take $p > 0$. Then

$$\int_{\mathbb{S}^2} \left|E(t,\eta)\right|^p d\sigma_\eta = \int_{\mathbb{S}^2} \left|\overline{E}(t,s)\right|^p \det\left(\frac{\partial \phi}{\partial s}(t,s)\right) d\sigma_s.$$

By assumption $\frac{\partial \phi}{\partial s}$ is bounded. Thus exists $C > 0$ such that

$$\|E(t,\_)\|_{L^p} = \int_{\mathbb{S}^2} \left|E(t,\eta)\right|^p d\sigma_\eta \leq C \int_{\mathbb{S}^2} \left|\overline{E}(t,s)\right|^p d\sigma_s.$$

When $\int_0^{+\infty} \|g_\tau(\phi(\tau,s))\|^2 d\tau = +\infty$, for each s we have $\lim_{t\mapsto+\infty} \overline{E}(t,s) = 0$. Moreover $|\overline{E}(t,s)|$ is uniformly bounded by the L^∞ function $E(0,s)$. By Lebesgue dominate convergence theorem $\lim_{t\mapsto+\infty} \|\overline{E}(t,\_)\|_{L^p} = 0$. Previous inequality leads to $\lim_{t\mapsto+\infty} \|E(t,\_)\|_{L^p} = 0$.

When, for $t > T$, $\int_0^T \|g_\tau(\phi(\tau,s))\|^2 d\tau \geq \lambda t$, we have, for all $s \in \mathbb{S}^2$, $|\overline{E}(t,s)| \leq |E(0,s)| e^{-k\bar{\gamma}\lambda t}$. Thus, for all $s \in \mathbb{S}^2$ we get $|\overline{E}(t,s)| \leq \|E(0,\_)\|_{L^\infty}$. Since $\eta \mapsto \phi^{-1}(t,\eta)$ is a diffeomorphism of $\mathbb{S}^2$, we get finally, for all $\eta \in \mathbb{S}^2$, $|E(t,\eta)| \leq \|E(0,\_)\|_{L^\infty} e^{-k\bar{\gamma}\lambda t}$. This proves $\|E(t,\_)\|_{L^\infty} \leq \|E(0,\_)\|_{L^\infty} e^{-k\bar{\gamma}\lambda t}$. □

4.2 Asymptotic observer based on rough depth estimation (Γ_{HS})

Instead of relying the observer on estimation V_{HS}, we can base it on Γ_{HS}. Then (17) becomes ($k > 0$)

$$\frac{\partial \widehat{D}}{\partial t} = -\nabla \widehat{D} \cdot (f_t + \Gamma_{\text{HS}} g_t) - v \cdot \eta + k(1 - \widehat{D}\Gamma_{\text{HS}}) \tag{21}$$

that reads in pinhole coordinates

$$\frac{\partial \widehat{D}}{\partial t} = -\frac{\partial \widehat{D}}{\partial z_1}(f_1 + \Gamma_{\text{HS}} g_1) - \frac{\partial \widehat{D}}{\partial z_2}(f_2 + \Gamma_{\text{HS}} g_2) - (z_1 v_1 + z_2 v_2 + v_3) + k(1 - \widehat{D}\Gamma_{\text{HS}}). \tag{22}$$

For this observer we have the following convergence result.

Theorem 2. *Take assumptions of theorem 1 concerning the scene surface Σ, $\Gamma = 1/D$, v and ω. Consider the observer (21) where Γ_{HS} coincides with Γ and where the initial condition is C^1 versus η. Then $\forall t \geq 0$, the solution $\widehat{D}(t,\eta)$ of (21) exists, is unique, remains C^1 versus η and*

$$\|\widehat{D}(t,\_) - D(t,\_)\|_{L^\infty} \leq e^{-k\bar{\gamma}t} \|\widehat{D}(0,\_) - D(0,\_)\|_{L^\infty}$$

(exponential convergence in L^∞ topology)

The proof, similar to the one of theorem 1, is left to the reader.

5 Simulations and numerical implementations

5.1 Sequence of synthetic images and method of comparison

The non-linear asymptotic observers described in section 4 are tested on a sequence of synthetic images characterized by the following:

- *virtual camera*: the size of each image is 640 by 480 pixels, the frame rate of the sequence is 60 Hz and the field of view is 50 deg by 40 deg;
- *motion of the virtual camera*: it consists of two combined translations in a vertical plane ($v_3 = \omega_1 = \omega_2 = \omega_3 = 0$), and the velocity profiles are sinusoids with magnitude 1 $m.s^{-1}$, and different pulsations (π for v_1 and 3π for v_2);
- *virtual scene*: it consists of a 4 m^2-plane placed at 3 m and tipped of an angle of 0.3 rad with respect to the plane of camera motion; the observed plane is virtually painted with a gray pattern, whose intensity varies in horizontal and vertical directions as a sinusoid function;
- *generation of the images*: each pixel of an image has an integer value varying from 1 to 256, directly depending on the intensity of the observed surface in the direction indexed by the pixel, to which a normally distributed noise varying with mean 0 and standard deviation σ is added.

The virtual setup used to generate the sequence of images is represented in Fig.2.

To compare the performances of both methods, we use the global error rate in the estimation of D, defined as

$$E = \int_{\mathcal{I}} \frac{|\widehat{D}(t,\eta) - D(t,\eta)|}{D(t,\eta)} d\sigma_\eta \; / \int_{\mathcal{I}} d\sigma_\eta \tag{23}$$

where D is the true value of the depth field, $\widehat{D}$ is the estimation computed by any of the proposed methods and $\mathcal{I}$ is the image frame.

5.2 Implementation of the depth estimation based on optical flow measures (V_{HS})

We test on the sequence described in 5.1 the depth estimation characterized by the partial differential equation (13). The optical flow input V_{HS} (V_1 and V_2 components) is computed by a classical Horn-Schunck method. Note that convergence theorem 1 assumes that the domain of definition of the image was the entire unit sphere $\mathbb{S}^2$. Here the field of view of our virtual camera limits this domain to a portion of the sphere $\mathcal{I}$. However, the motion of our virtual camera ensures that most of the points of the scene appearing in the first image stay in the field of view of the camera during the whole sequence. The convergence of

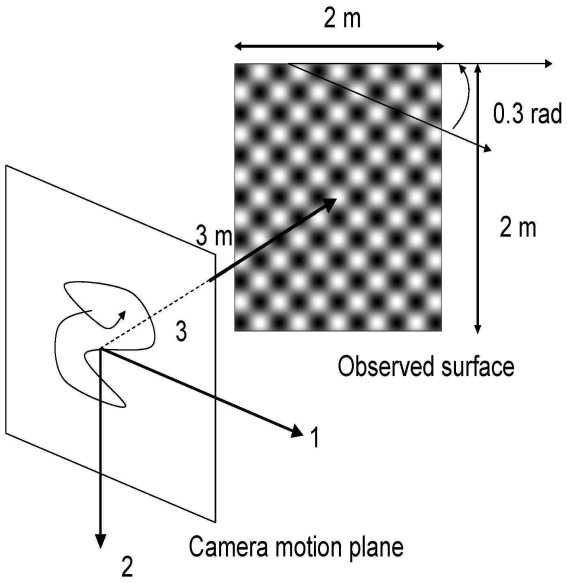

Figure 2: Virtual setup used to generate the sequence of images processed in 5

the method is only ensured for these points, and Neumann boundary conditions are chosen at the borders where optical flow points toward the inside of the image:

$$\frac{\partial \widehat{D}}{\partial n} = 0 \text{ if } n \cdot V_{\text{HS}} < 0 \tag{24}$$

The observer gain $k > 0$ is chosen in accordance with scaling considerations. Setting $k = 500\ s.m^{-2}$ provides a rapid convergence rate: we see on Fig. 3 that after a few frames, the initial relative error (blue curve) is reduced by 1/3. Setting $k = 50\ s.m^{-2}$ is more reasonable when dealing with noisy data: on Fig. 4 initial relative error is reduced by 1/3 after around 20 frames.

More precisely, the standard deviation σ of the noise added to the synthetic sequence of images is 1. The gains $k = 500$ and $k = 100$ are successively tested, and the associated error rates for $\widehat{D}$ are plotted in Fig. 3. As expected, the convergence is more rapid for a larger gain but at convergence, i.e., after 40 frames, the relative errors are similar and below 1.5%.

To test robustness when dealing with noisy data, the standard deviation σ is magnified by 20. The correction gain is tuned to $k = 50\ s.m^{-2}$. The converged errors after 40 frames significantly increases and yet stays between 12 and 14 %. Note that such permanent errors can not decrease since such noise level first affects the optical flow estimation V_{HS} that feeds the asymptotic observer. Compared to its true value V, the error level for V_{HS} is about 15 %. These results underline the fact that this approach is sensitive to input optical flow

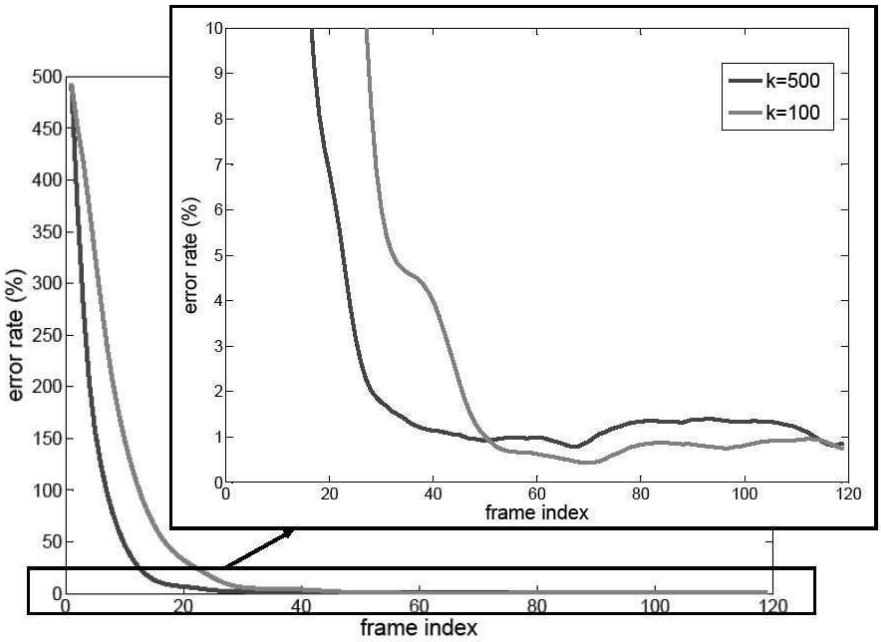

Figure 3: Relative estimation errors of the depth field D estimated by the asymptotic observer (18) filtering the optical flow input V_{HS} obtained by Horn-Schunck method, for different correction gains k. The noise corrupting the image data is normally distributed, with mean $\mu = 0$ and standard deviation $\sigma = 1$.

measures, but not directly to noise corrupting the image data.

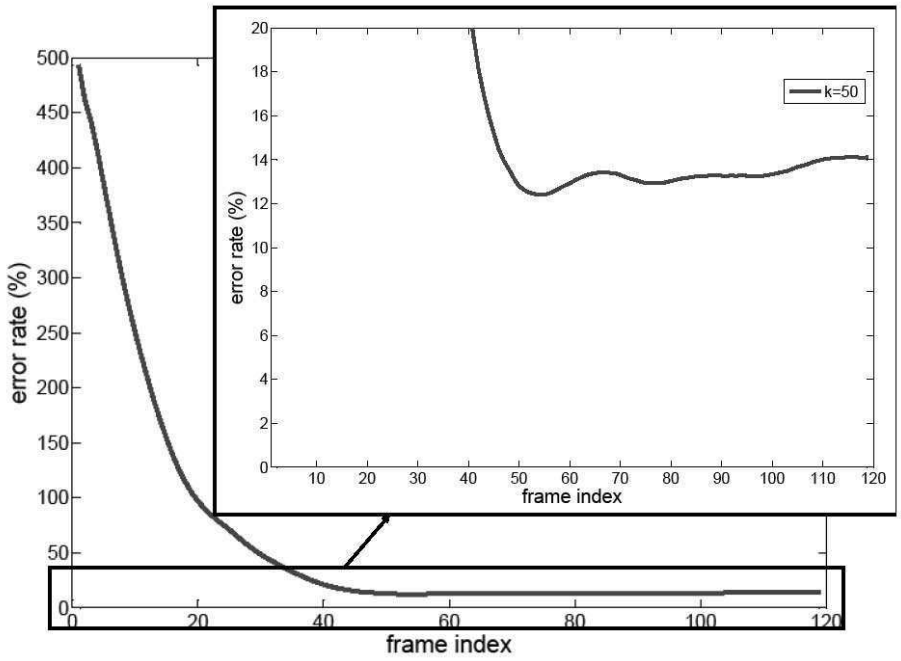

Figure 4: Relative estimation errors of the depth field D estimated by the asymptotic observer (18) filtering the optical flow input V_{HS} obtained by Horn-Schunck method.The noise corrupting the image data is normally distributed, with mean $\mu = 0$ and standard deviation $\sigma = 20$.

5.3 Implementation of the asymptotic observer based on rough depth estimation (Γ_{HS})

Subsequently, the observer described by (22) was applied to the same sequence. The input depth field Γ_{HS} is obtained as the output of (13). To adapt the numerical method to this model, we make the small-angle approximation by neglecting the second order terms z (we neglect the curvature of $\mathbb{S}^2$ and consider that the camera image corresponds to a small part of $\mathbb{S}^2$ that can be approximated by a small Euclidean rectangle; the error of this approximation is smaller than 3% for such rectangular image): (13) becomes

$$G^2\Gamma + FG = \alpha^2 \left(\frac{\partial^2 \Gamma}{\partial {z_1}^2} + \frac{\partial^2 \Gamma}{\partial {z_2}^2} \right) \tag{25}$$

F and G are computed using angular and linear velocities, ω and v, and differentiation (Sobel) filters directly applied on the image data $y(t, z_1, z_2)$. $\Delta\Gamma$ is

approximated by the difference between the weighted mean $\bar{\Gamma}$ of Γ on the neighboring pixels and its value at the current pixel. The resulting linear system in Γ is solved by the Jacobi iterative scheme, with an initialization provided by the previous estimation. The regularization parameter α is chosen accordingly to scaling considerations and taking into account the magnitude of noise: $\alpha = 300$ $m.s^{-1}$ provides a convergence in about 5 or 6 frames for relatively clean image data. As for the observer (22), the correction gain $k = 50$ $s.m^{-1}$ enables a convergence in around 20 frames.

As in section 5.2, we test the observer (22) for different levels of noise corrupting the image data. For $\sigma = 1$, the error rates associated to the input depth Γ_{HS} and to the estimated depth $\widehat{D}$ are plotted in Fig. 5. After only 6 images of the sequence, the error rate for Γ_{HS} is smaller than 4% and stays below this upper bound for the rest of the sequence. On the downside, the error rate stays larger than 2.5%. On the contrary, the error rate associated to $\widehat{D}$ keeps decreasing, and reaches the minimal value of 0.5 %.

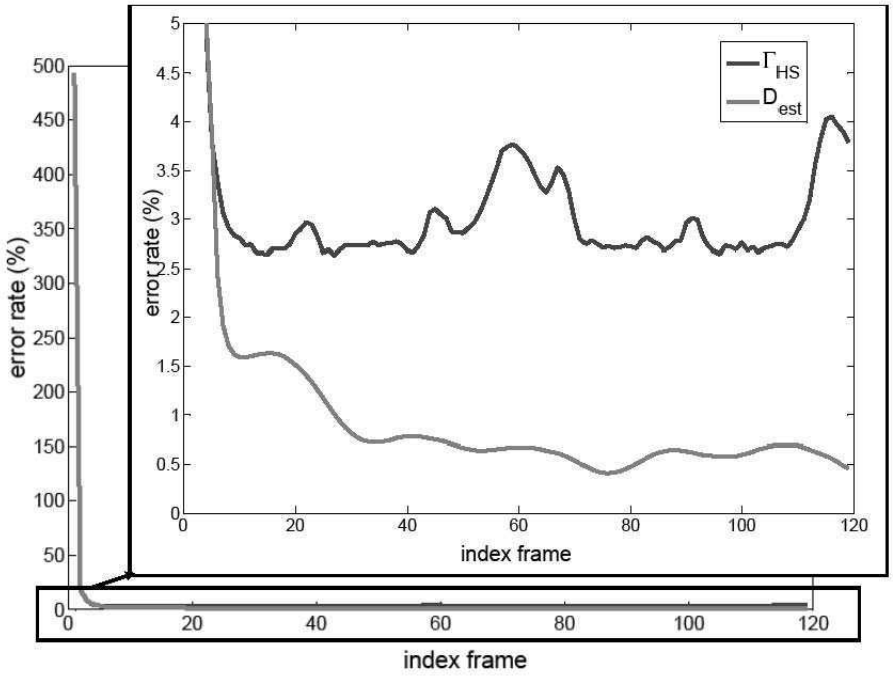

Figure 5: Relative estimation error of Γ_{HS} (blue), using the depth estimation inspired by Horn-Schunck, described in 3.2 and of D (red) estimated by the asymptotic observer (22) filtering Γ_{HS}.The noise corrupting the image data is normally distributed, with mean $\mu = 0$ and standard deviation $\sigma = 1$.

For $\sigma = 20$, the error rates associated to the input depth Γ_{HS} and to the estimated depth $\widehat{D}$ are plotted in Fig. 6. For the computation of Γ_{HS}, the

diffusion parameter α is increased to 1000 $m.s^{-1}$ to take into account such stronger noise. The observer filters the error associated to Γ_{HS} (between 4 and 8 %) to provide a 3 % accuracy. The results show a good robustness to noise for this observer.

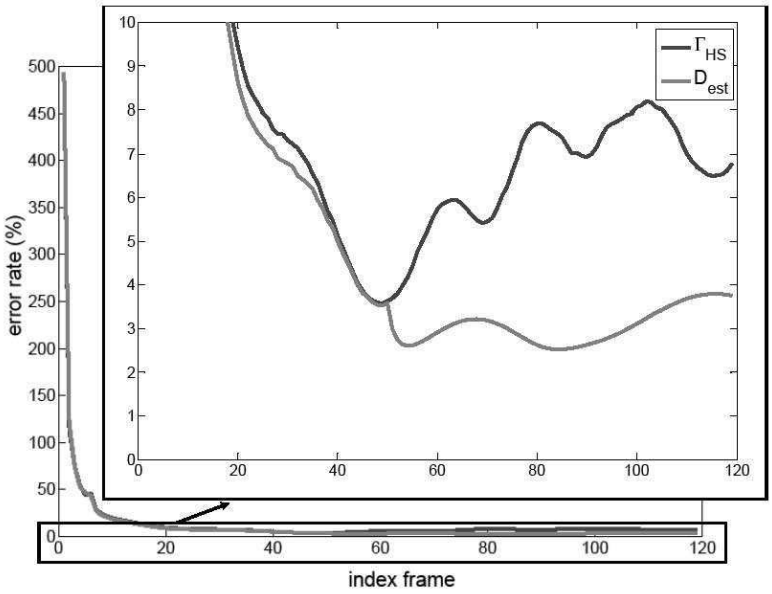

Figure 6: Relative estimation error of Γ_{HS} (blue), using the depth estimation inspired by Horn-Schunck, described in 3.2 and of D (red) estimated by the asymptotic observer (22) filtering Γ_{HS}.The noise corrupting the image data is normally distributed, with mean $\mu = 0$ and standard deviation $\sigma = 20$.

5.4 Experiment on real data

To realize the experiments, a camera was fixed on a motorized trolley traveling back and forth on a 2 meter-linear track in about 6 seconds. The resolution of the encoder of the motor enables to know the position and the speed of the trolley with a micrometric precision. The camera is a Flea2 - Point Grey Research VGA video cameras (640 by 480 pixels) acquiring data at 20.83 fps, with a Cinegon 1.8/4.8 C-mount lens, with an angular field of view of approximately 50 by 40 deg, and oriented orthogonally to the track. The scene is a static work environment, with desks, tables, chairs, lamps, lit up by electric light plugged on the mains, with frequency 50 Hz. The acquisition frame rate of the cameras produces an aliasing phenomenon on the video data at 4.17 Hz. In other words, the light intensity in the room is variable, at a frequency that can not be easily ignored, which does not comply with the initial hypothesis

(3). However, the impact of this temporal dependence in the equations can be reduced by a normalization of the intensity of the images such as $y(x,y) = \frac{y(x,y)}{\bar{y}}$ where x and y are the horizontal and vertical indexes of the pixels in the image, $y(x,y)$ is the intensity of this pixel and $\bar{y}$ is the mean intensity on the entire image.

The depth field was estimated via the asymptotic observer (18) based on optical flow measures. The components of the optical flow were computed using a high quality algorithm based on TV-L^1 method (see [21] for more details). The correction gain was tuned to $k = 100\ s.m^{-2}$. An example of image data is shown in Fig.7, and the depth estimate associated to that image at the same time $t*$ is shown in Fig.8. At that specific time $t^\star \approx 8$ s, the trolley already traveled once along the track and is on its way back toward its starting point. Some specific estimates are extracted from the whole depth field (two tables, two chairs, a screen, two walls) and highlighted in black; they are compared to real measures taken in the experimental room (in red): the estimate depth profile $\widehat{D}(t^\star, \cdot)$ exhibits a strong correlation with these seven punctual reference values of $D(t^\star, \cdot)$; the global appearance of the depth field looks very realistic.

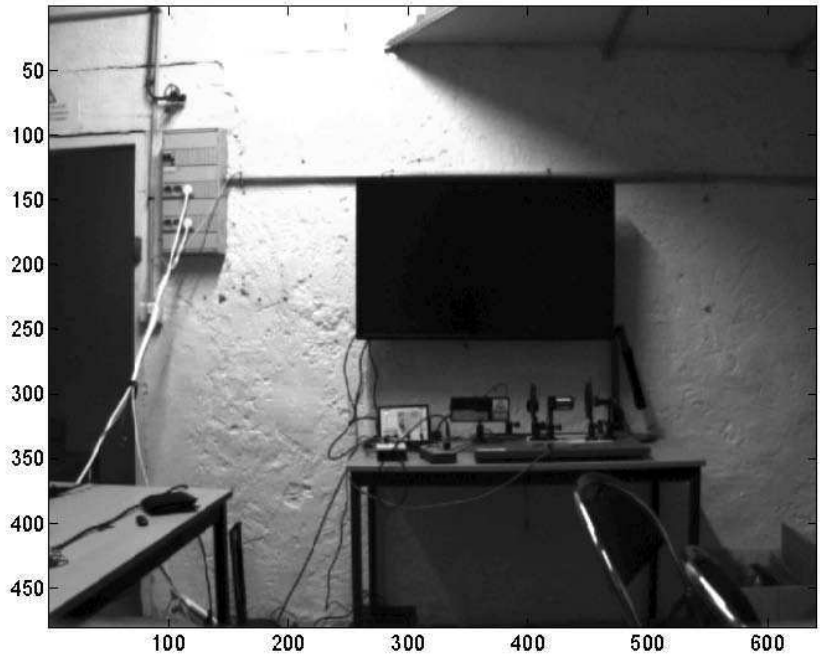

Figure 7: The static scene as seen by the camera

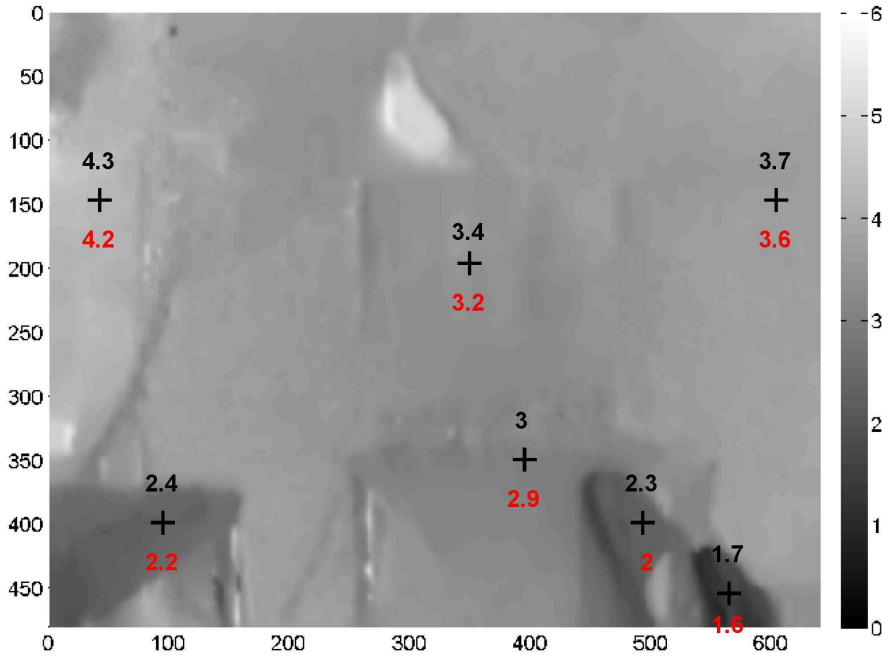

Figure 8: Estimation of the depth field associated with the image shown in Fig.7. Depth is associated to a gray level, whose scale in meters is on the right. Some estimates are extracted from the entire field (in black) and compared to real measures (in red).

6 Conclusions and future works

In section 2, we recalled a system of partial differential equations, describing the invariant dynamics of brightness and depth smooth fields under the assumptions of a static and Lambertian environment. We proposed in section 3 an adaptation of optical flow algorithms that take the best advantage of the $SO(3)$-invariance of these equations and the knowledge of camera dynamics: it yielded an $SO(3)$-invariant variational method to directly estimate the depth field. In section 4, we proposed two asymptotic observers, respectively based on optical flow and on depth estimates. We proved their convergence under geometric and persistent excitation assumptions. On synthetic images, we showed in section 5 that the variational method converges rapidly, but its performance is highly dependent of the noise level whereas both asymptotic observers filter this noise. These asymptotic obersevers based on image processing of the entire field of view of the camera seems to be an interesting tool to dense range estimation and a complement to methods based on feature tracking.

References

[1] R. Smith, M. Self, and P. Cheeseman, *Estimating uncertain spatial relationships in robotics*. New York, NY, USA: Springer-Verlag New York, Inc., 1990, pp. 167–193.

[2] J. Civera, A. Davison, and J. Montiel, "Inverse depth parametrization for monocular slam," *Robotics, IEEE Transactions on*, 2008.

[3] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *International Joint Conference on Artificial Intelligence IJCAI*, 2003.

[4] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Real-time monocular slam: Why filter?" in *ICRA*, 2010, pp. 2657–2664.

[5] H. Strasdat, J. M. M. Montiel, and A. Davison, "Scale drift-aware large scale monocular slam," in *Proceedings of Robotics: Science and Systems*, Zaragoza, Spain, June 2010.

[6] R. Abdursul, H. Inaba, and B. K. Ghosh, "Nonlinear observers for perspective time-invariant linear systems," *Automatica*.

[7] O. Dahl, Y. Wang, A. F. Lynch, and A. Heyden, "Observer forms for perspective systems," *Automatica*.

[8] A. Heyden and O. Dahl, "Provably convergent on-line structure and motion estimation for perspective systems," in *Computer Vision Workshops, 2009 IEEE 12th International Conference on*, 2009.

[9] X. Chen and H. Kano, "A new state observer for perspective systems," *Automatic Control, IEEE Transactions on*, 2002.

[10] W. Dixon, Y. Fang, D. Dawson, and T. Flynn, "Range identification for perspective vision systems," *Automatic Control, IEEE Transactions on*, vol. 48, no. 12, pp. 2232 – 2238, dec. 2003.

[11] D. Karagiannis and A. Astolfi, "A new solution to the problem of range identification in perspective vision systems," *Automatic Control, IEEE Transactions on*, vol. 50, no. 12, pp. 2074 – 2077, dec. 2005.

[12] A. D. Luca, G. Oriolo, and P. R. Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments." *I. J. Robotic Res.*, pp. 1093–1116, 2008.

[13] M. Sassano, D. Carnevale, and A. Astolfi, "Observer design for range and orientation identification," *Automatica*.

[14] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision*, vol. 3, no. 3, pp. 209–238, 1989.

[15] C. Hoilund, T. Moeslund, C. Madsen, and M. Trivedi, "Improving stereo camera depth measurements and benefiting from intermediate results," in *IEEE Intelligent Vehicles Symposium*, 2010, pp. 935 –940.

[16] B. Herisse, S. Oustrieres, T. Hamel, R. E. Mahony, and F.-X. Russotto, "A general optical flow based terrain-following strategy for a vtol uav using multiple views," in *ICRA*, 2010, pp. 3341–3348.

[17] S. Bonnabel and P. Rouchon, "Fusion of inertial and visual : a geometrical observer-based approach," *2nd Mediterranean Conference on Intelligent Systems and Automation (CISA'09)*, 2009.

[18] A. Censi, S. Han, S. B. Fuller, and R. M. Murray, "A bio-plausible design for visual attitude stabilization," in *CDC*, 2009, pp. 3513–3520.

[19] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[20] A. Mitiche and A. reza Mansouri, "On convergence of the horn and schunck optical-flow estimation method," *IEEE Transactions on Image Processing*, vol. 13, pp. 848–852, 2004.

[21] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, pp. 1–26, 2010.