2012 American Control Conference
Fairmont Queen Elizabeth, Montréal, Canada
June 27-June 29, 2012

Curvilinear velocity estimation using low-quality stereo-vision systems and a gyrometer

Nadège Zarrouati, Mathieu Hillion and Nicolas Petit

Abstract— **We propose in this paper a method to estimate the velocity of a rigid body, using a novel stereo-vision principle. It is presented and applied in a laboratory test case which is representative of low-cost navigation for ground vehicles. The method exploits the dynamics of a scalar field obtained by weighting and averaging the brightness perceived by two embedded neighboring cameras. To be more specific, the cameras are complemented with a gyrometer to retrieve the curvilinear velocity of the moving rigid body. The proposed method is first tested on synthetic data, then on real data, and shows robustness to poor quality of image data. Significant levels of noise and blur are tested; in addition, this method does not require high resolution images, as opposed to any existing methods based on triangulation and tracking of keypoints.**

I. INTRODUCTION

In many ground transportations or land navigation applications where the path of a vehicle is known in advance, the current position of this vehicle can be derived from the distance traveled from the starting point (the curvilinear abscissa along the path). By extension, a similar approach is still valid when, instead of the path, the current heading of the vehicle is accessible (via embedded gyroscopes for example, or GPS compass [1]), in the spirit of *dead reckoning* [2], [3]. The traveled distance can be rapidly obtained by direct measurements, such as odometry (equipping most of wheeled robots and experimental vehicles), which rely on wheel motion to obtain position and orientation. Unfortunately, this mechanical device is very sensitive to uncertainties produced by wheel sliding on the ground, or soil imperfections. It is thus required to regularly register the robot, i.e. get absolute or relative positions estimates, with respect to known positions obtained by other means [4], [5]. Another widely considered possibility consists in integrating the velocity of the vehicle, which is itself obtained by integration of the acceleration provided by an Inertial Measurement Unit (IMU) [6], [7]. The cumulated error of the double integration can only be kept relatively small for high quality low-bias systems such as the ones used for missile guidance and control [8]: these systems are unfortunately too expensive to be considered for mass-markets applications. IMU-based low-cost solutions are typically coupled with GPS sparse measurements, only available in outdoors and in low buildings urban environments [9], [10], [11]. As a result, there is still a need for novel velocity estimation techniques provided that they can be low-cost and robust. This is the problem we consider, using embedded cameras and low-cost IMUs [12].

Since the 1980', research in image processing and computer vision has tackled the problem of pose estimation [13], [14]. Miniaturization and low-cost commercialization of vision systems has made this solution eligible for large-scale use, but the robustness, accuracy and real-time compatibility of these methods are not fully assessed yet. Two types of vision systems have been extensively studied, namely monocular and binocular (or stereoscopic) systems. The first one relies on optical flow analysis [15], [16], [2] or image point correspondences [17], [3]. Without any additional sensor output, it intrinsically lacks a scaling factor: velocity at any time t can only be known relatively to another (e.g. initial) velocity [18]. On the contrary, the distance between the optical centers of the cameras (the baseline) in a stereo-vision system directly provides the missing scaling factor for the employed triangulation method. Triangulation matches keypoints (or geometric primitives) in the image to estimate the 3-dimensional coordinates of the underlying physical objects under consideration [19]. The image-referenced dynamics of the projection of these landmarks is then filtered to obtain the pose estimate of the vision system [20], [21], [22]. These techniques are rapidly evolving, but the stereo preprocessing (rectification, prefiltering and correlation), which is necessary to robustify the correspondence step, significantly weighs their computational cost down.

The approach we propose bypasses the discussed correspondence problem to directly estimate the curvilinear velocity from the image. Our approach is inspired by the techniques of distributed magnetometry navigation first presented in [23] and developed since then [24]: in a static environment, given a measured field B only depending on the state of the measuring apparatus (namely the position x and orientation θ of the stereo-vision system), any temporal variation of the sensed value of B is directly related to variations of the state (x,θ). Here, B is simply built from the brightness of the perceived images. From this seemingly basic principle, we derive a method to estimate the curvilinear velocity. We show that this method is robust to noise and to blurring effects, can be used with low-cost material, and is compatible with real-time applications (as data can be significantly down-sampled which alleviates the

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

Mathieu Hillion is with SYSNAV, 57 rue de Montigny, 27200 Vernon, France emanuel.aldea@sysnav.fr

Nicolas Petit is with MINES-ParisTech, Centre Automatique et Systèmes, Unité Mathématiques et Systèmes 60, boulevard Saint-Michel 75272 Paris Cedex, France nicolas.petit@mines-paristech.fr

978-1-4577-1096-4/12/$26.00 ©2012 AACC

computational burden). It is thus robust to poor quality image data. To our knowledge, there are few attempts to treat the image perceived by a vision system as a physical signal, just as temperature or magnetic field: the variations of this signal only depend on the configuration of the measuring system.

The paper is organized as follows. In Section II, the notations of the problem are introduced, the field B is defined, and the governing dynamics equations are derived. In Section III, the method used to implement the equations in a real application is described and an observer for curvilinear velocity is presented. In Section IV, technical details for implementation of this method are considered, and results of simulations and experiments are given. In Section V, the robustness of the method is tested. Finally, some future directions are discussed in Section VI

II. Preliminaries

In this section, we state the problem notations involving one preliminarily idealized camera and the equations governing the weighted-averaged brightness during translation and rotation motion.

A. Notations and frameworks

The model of spherical camera we consider is based on geometric assumptions introduced in [25]. Linear and angular velocities $v(t)$ and $\omega(t)$ with respect to an inertial frame of reference $\mathcal{R}$ are expressed in the (body) camera frame $\mathcal{R}_b$. Position of the optical center of the camera with respect to $\mathcal{R}$ is denoted by $C(t) = x(t)$. Orientation of $\mathcal{R}_b$ with respect to $\mathcal{R}$ is given by the quaternion $q(t)$: any vector ζ in the camera frame corresponds to the vector $q\zeta q^*$ in the reference frame $\mathcal{R}$ using the identification of vectors as imaginary quaternions, where * denotes conjugation. A pixel is labeled by the unit vector η in the camera frame: η belongs to the sphere $\mathbb{S}^2$ and receives the intensity y (for grayscale images).

The scene is modeled as a closed, C^1 and convex surface Σ of $\mathbb{R}^3$, diffeomorphic to $\mathbb{S}^2$. The camera lies 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. The density of light emitted by M does not depend on the direction of emission (Σ is a Lambertian surface) and is independent of t (the scene is static and the lighting is constant). By the identity $q\eta q^* = \frac{C(t)M}{\|C(t)M\|}$, where η is identified to an imaginary quaternion, and under all previous assumptions, y is a function of η, $x(t)$ and $q(t)$.

At each instant t during the motion, the image produced by the camera is described by the scalar field $\mathbb{S}^2 \ni \eta \mapsto y(\eta, x(t), q(t)) \in \mathbb{R}$.

The distance $\|C(t)M\|$ between the optical center and the object seen in the direction η is denoted by D, and its inverse by $\Gamma = 1/D$. Fig.1 illustrates the model and the notations.

Under these assumptions, during the motion at velocity v and rate of turn ω, the intensity $y(\eta, x(t), q(t))$ satisfies the following partial differential equation:

$$\frac{\partial y}{\partial t} = -\nabla y \cdot (\eta \times (\omega + \Gamma \eta \times v)) \tag{1}$$

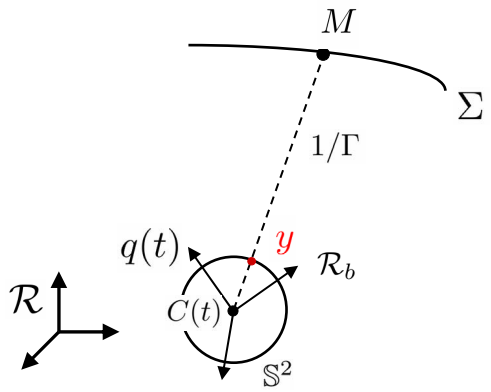

Fig. 1. Model and notations of a spherical camera in a static environment.

where ∇y is the gradient of y with respect to the Riemannian metric on $\mathbb{S}^2$. The value of ∇y at $\eta \in \mathbb{S}^2$ is identified with a vector of $\mathbb{R}^3$ tangent to the sphere at the point η, itself being identified to a unitary vector of $\mathbb{R}^3$ in the camera moving frame. In (1), 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 $a \times b$. More details on how to obtain this equation can be found in [25].

Remark 1: Equation (1) is the sum of two influences, the rotational and translational displacements, respectively. The depth Γ scales the translational velocity.

B. Governing equations

From the (scalar) intensity field y, one can build another scalar field, noted B, defined as:

$$B(x, q) = \int_{\mathbb{S}^2} \varphi(-\eta) y(\eta, x, q)\, d\sigma_\eta \tag{2}$$

where φ is a smooth (differentiable) convolution kernel, defined on the unit sphere $\mathbb{S}^2$, and $d\sigma_\eta$ denotes the infinitesimal surface element on $\mathbb{S}^2$. Clearly, B does not directly depend on time since the environment is static. The variations of B are directly related to the camera motion. The chain rule applied to $B(x(t), q(t))$ yields:

$$\frac{dB}{dt} = V^T \nabla_x B + \Omega^T \nabla_\theta B \tag{3}$$

where ∇_x denotes the gradient with respect to the position x of the optical center, ∇_θ denotes the gradient with respect to the 3-dimensional orientation of the optical axis of the camera, V and Ω are the 3-dimensional linear and angular velocities of the camera with respect to the inertial frame $\mathcal{R}$. For the sake of exploiting this equation, it has to be projected onto the reference frame $\mathcal{R}_b$ to which the sensors are attached.

III. Exploitation of weighted-averaged intensities for velocity estimation

A. Problem statement

Consider a rigid body in translation and rotation motion with respect to an inertial frame of reference $\mathcal{R}$ equipped

with sensors aligned with a body-attached frame of reference $\mathcal{R}_b$. The motion occurs in a plane (say the plane defined by x_b and y_b-axis). It is desired to estimate the projection of the translational velocity V_{x_b} along one of the axis of $\mathcal{R}_b$ (say the x_b-axis).

B. Proposed solution

In our proposed solution, we consider a set of sensors, refered to as "apparatus" in the following. The sensors consist of two cameras of the type studied in Section II and one gyrometer measuring the rate of turn $\dot{\theta}_c$ of the rigid body (identified to $\mathcal{R}_b$) with respect to $\mathcal{R}$. The two cameras are located along the x_b-axis. As will appear from (3), the velocity V_{x_b} satisfies the equation

$$\frac{dB}{dt} = V_{x_b}\frac{\partial B}{\partial x_b} + \dot{\theta}_c\frac{\partial B}{\partial \theta_c} \tag{4}$$

where B is defined as in (2). Each term of (4), except V_{x_b}, can be measured with the proposed apparatus.

Remark 2: Note that in this notation, $\theta_c(t)$ replaces $q(t)$, and designs the orientation of the optical axis of a camera with respect to $\mathcal{R}$. This notation stands when orientation is reduced to one angle as in this specific case of motion.

C. Determining the derivatives

We first establish (4) from (3). Note that the operators ∇_θ and $-\nabla_\eta$ can be readily identified:

$$\nabla_\theta = -\nabla_\eta \tag{5}$$

Indeed, a change of direction of the optical axis is strictly equivalent to an opposite change of indexation of the pixels, as long as any object of the scene remains in the field of view of the camera (which is the case for the spherical camera under consideration). By integration by parts, this yields

$$\begin{aligned}\nabla_\theta B &= \int_{\mathbb{S}^2} \varphi(-\eta)\nabla_\theta \left(y\left(\eta, x, q\right)\right) d\sigma_\eta \\ &= -\int_{\mathbb{S}^2} \nabla_\eta \left(\varphi(-\eta)\right) y\left(\eta, x, q\right) d\sigma_\eta \end{aligned} \tag{6}$$

This means that for any given convolution kernel φ, $\nabla_\theta B$ can be as easily computed as B from any image y, i.e. from any field $\eta \mapsto y(\eta, x, q)$.

However, as underlined in Remark 1, $\nabla_x B$ can not be retrieved from a single camera without any notion of the depth field Γ: one can not predict the position of an object in an image after a non-purely rotational motion if the relative position of the object is unknown. This forces to get $\nabla_x B$ by some spatial differentiation scheme, thus to use at least one additional camera. For each camera i, with optical center $C^i(t)$ and optical axis direction $q^i(t)$ with respect to $\mathcal{R}$, a scalar field B^i is computed (the same convolution kernel is used for all cameras). The B^i variables actually correspond to $B(x^i, q^i)$. If the cameras are close enough, the first order relation between any B^i and B^j is:

$$B^j \approx B^i + C^iC^j \cdot \nabla_x B + \Delta\theta_{ij} \cdot \nabla_\theta B \tag{7}$$

where $\Delta\theta_{ij}$ is the rotation between q^i and q^j. In this apparatus, the cameras (as well as the gyrometer) are rigidly attached to the solid body. Assuming that their relative (constant) position and orientation C^iC^j and $\Delta\theta_{ij}$ are known, e.g. thanks to preliminary offline measurements or identification procedure, (7) together with (6) enables to determine the projection of $\nabla_x B$ on C^iC^j. In other words, with two cameras aligned along the x_b-axis of $\mathcal{R}_b$, if the distance between the optical centers of the cameras h is sufficiently small, this yields

$$\frac{\partial B}{\partial x_b} = \frac{B^2 - B^1 - \Delta\theta_{ij} \cdot \nabla_\theta B}{h} \tag{8}$$

Formally, through (4), this shows the observability of V_{x_b} using measurements from the proposed apparatus consisting of two cameras and a gyrometer.

D. An observer to reconstruct the velocity

Considering $\frac{\partial B}{\partial x_b}$ as a measurement obtained from (8), we propose the following observer based on the dynamics (4) to estimate the velocity V_{x_b} modeled as a constant for simplicity (extension to low-pass filters driven by white noise, as usually considered for performance improvement, can be readily addressed):

$$\begin{cases} \dot{\widehat{B}} = \widehat{V}\frac{\partial B}{\partial x_b} + \dot{\theta}_c\frac{\partial B}{\partial \theta_c} - (\widehat{B} - B)L_1\frac{\partial B}{\partial x_b} \\ \dot{\widehat{V}} = -l_2(\widehat{B} - B)\frac{\partial B}{\partial x_b} \end{cases} \tag{9}$$

where $\widehat{B}$ and $\widehat{V}$ are the estimates of B and V_{x_b}, respectively; B, $\frac{\partial B}{\partial \theta_c}$ and $\frac{\partial B}{\partial x_b}$ are provided by the cameras as explained in III-C; $\dot{\theta}_c$ is directly provided by an embedded gyrometer (which can be filtered for sake of noise reduction); $l_2 > 0$ and L_1 is a scalar defined below. The error dynamics with respect to

$$\begin{cases} \frac{dB}{dt} = V_{x_b}\frac{\partial B}{\partial x_b} + \dot{\theta}_c\frac{\partial B}{\partial \theta_c} \\ \dot{V}_{x_b} = 0 \end{cases} \tag{10}$$

is

$$\begin{cases} \dot{\tilde{B}} = \tilde{V}\frac{\partial B}{\partial x_b} - \tilde{B}L_1\frac{\partial B}{\partial x_b} \\ \dot{\tilde{V}} = -l_2\tilde{B}\frac{\partial B}{\partial x_b} \end{cases} \tag{11}$$

Proposition 1: The error dynamics (11), where $L_1 \triangleq l_1\frac{\partial B}{\partial x_b}$, with $l_1 > 0$ and $l_2 > 0$ is globally asymptotically stable ($\lim_{t\to+\infty}\tilde{B} = 0$ and $\lim_{t\to+\infty}\tilde{V} = 0$) provided that $\frac{\partial B}{\partial x_b}$ is non zero.

Proof: Consider the Lyapunov function W

$$W = |\widehat{B} - B|^2 + \frac{1}{l_2}|\widehat{V} - V_{x_b}|^2 \tag{12}$$

Using (11), one gets

$$\dot{W} = -2\tilde{B}^2 L_1\frac{\partial B}{\partial x_b} \tag{13}$$

Thus, choosing $L_1 = l_1\frac{\partial B}{\partial x_b}$, with $l_1 > 0$ yields $\dot{W} \leq 0$. If the set of trajectories $\left\{(\tilde{B}, \tilde{V}) \text{ s.t. } \dot{W} = 0\right\}$ only contains the origin $(\tilde{B}, \tilde{V}) = (0, 0)$, then LaSalle's invariance principle [26] enables to conclude that the trivial trajectory is globally asymptotically stable (see also [27] for a proof in non autonomous cases).

Trajectories lying in the set $\dot{W} = 0$ satisfy

$$\tilde{B}\frac{\partial B}{\partial x_b} = 0 \tag{14}$$

Assuming that $\frac{\partial B}{\partial x_b}$ is non zero, i.e. the image is not uniform, this yields $\tilde{B} = 0$. In this set, the dynamics (4) yields

$$\dot{\tilde{B}} = \tilde{V}\frac{\partial B}{\partial x_b} - l_1\tilde{B}\left(\frac{\partial B}{\partial x_b}\right)^2 \tag{15}$$

and assures that $\tilde{V} = 0$, which concludes the proof.

■

Remark 3: With four cameras located at the vertices of a non-singular tetrahedron, $\nabla_x B$ would be fully characterized by (7). Using three independent kernels for each camera gives a vector B with three components, and $\nabla_x B$ is then a 3x3 matrix. Assuming that three gyrometers give measurements of the angular velocity of the rigid body, observer (9) is generalized into

$$\begin{cases} \dot{\widehat{B}} = \widehat{V}^T\nabla_x B + \Omega^T\nabla_\theta B - (\widehat{B}^T - B^T)L_1\nabla_x B \\ \dot{\widehat{V}} = -l_2(\widehat{B}^T - B^T)\nabla_x B \end{cases} \tag{16}$$

With $L_1 = l_1\nabla_x B^T$ and assuming that $\nabla_x B$ is non singular, the previous proposition is still valid, and observer (16) estimates the three components of the translational speed V.

IV. Implementation, simulations and experimental results

A. Choice of the kernel

Realistic cameras have limited fields of view on contrary of spherical cameras. Thus it is essential to understand the importance of the boundary effects.

Firstly, when stating the governing partial differential equation (3), one assumes that the value of B at time $t + dt$ is predictable from the knowledge of the value B at time t and the motion of the camera. This is obviously not true when objects appear or disappear in the image from the sides of the image frame. Fortunately, it is possible to minimize the impact of these boundary source terms. A solution is to decrease the weight of these objects in B, e.g. by choosing small values for the convolution kernel φ on sufficiently large strips bordering the sides of the image frame (where objects are likely to appear or disappear). For planar trajectories, this only concerns the left and right sides of both cameras. The width Λ of these "zero" strips is chosen to ensure that it covers any new object entering the image frame. In the following, a possible value for Λ depending on the motion parameters (and sampling time) of the camera and on the environment is given.

Considering again that translation and rotation occur in a plane, and that the linear velocity is orthogonal to the optical axis of the camera (to ease the computation), we model the image as the intersection of the unit sphere $\mathbb{S}^2$ and the plane of motion of the camera. Fig.2 illustrates this viewpoint.

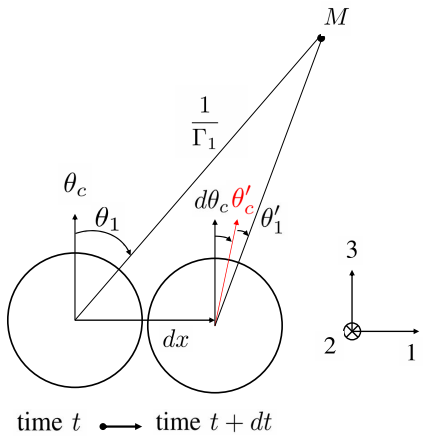

Fig. 2. Infinitesimal translation and rotation of a spherical camera tracking a fixed point of the environment.

We consider an object M being at time t at the boundary of the image frame: we want to ensure that at time $t+\delta t$, it belongs to the "zero" strips of width Λ. The direction of the optical axis with respect to $\mathcal{R}$ is denoted θ_c at time t and θ'_c at time $t + \delta t$ where δt is a small time increment, with $\theta'_c - \theta_c = \delta\theta_c$. The position x of the optical center undergoes an infinitesimal translation δx between t and $t + \delta t$. θ_1 is the direction at t in the field of view of the camera where the fixed point M of the environment, distant of $1/\Gamma_1$ is perceived, taken from the optical axis, and θ'_1 is the direction of perception of the same point at time $t + \delta t$. In this case, the infinitesimal motion leads to:

$$\theta_1^{'} = \arctan\left(\tan\theta_1 - \delta x\frac{\Gamma_1}{\cos\theta_1}\right) - \delta\theta_c \tag{17}$$

$$= \theta_1 - \delta x\Gamma_1\cos\theta_1 - \delta\theta_c \tag{18}$$

at first order in δx and $\delta\theta_c$.

This yields the choice for Λ:

$$\Lambda \geq \left(V_{x_b}^{\max}\Gamma^{\max} + \dot{\theta_c^{\max}}\right)\delta t \geq |\theta_1^{'} - \theta_1| \tag{19}$$

Secondly, in (6), the boundary term in the integration by parts is null by integration on the closed domain $\mathbb{S}^2$, but this is no longer the case with a partial area of $\mathbb{S}^2$. However, the choice of small values of φ on the sides of the frame ensures that the boundary term is canceled (for $\delta t \to 0$), and (6) can still be a valid model.

Finally, φ is chosen as a one-dimensional smooth function, repeated along the vertical direction, which is orthogonal to the plane where the trajectory lies: a choice of appropriately scaled Gaussian derivatives enables to easily control the width Λ and to have an analytic expression of any spatial derivatives of the kernel; φ is chosen to enhance the contrasts of the image and thus to better discriminate the relative motion of the objects in the image. This can be achieved by tuning the derivative order of the considered Gaussian function.

B. Simulations

The asymptotic observer (9) proposed in III-D is tested on a sequence of synthetic images described by the following characteristics:

- *virtual cameras*: 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 cameras*: the trajectory of the mobile to which the cameras are fixed is planar: first linear, then circular, with radius 15 m; the tangential velocity is continuous and slowly variable, from 0.5 m/sec to 0.6 m/sec; the trajectory lasts 16 sec;
- *virtual scene*: it consists of a plane orthogonal 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;
- *stereo-vision system*: the distance between the optical centers is $h = 8$ cm; the segment joining the centers is tangent to the trajectory at its middle point to ensure that the linear velocity of the mobile is tangent to the trajectory; the optical axis directions form an angle $\delta\theta_c = 0$ or 1.6 deg.

Fig.3 summarizes the virtual setup used to generate the sequence of images. Notice that, contrary to the model, the tangential velocity is slowly variable: the following simulations intend to show that the proposed method can overcome the limitations of the model.

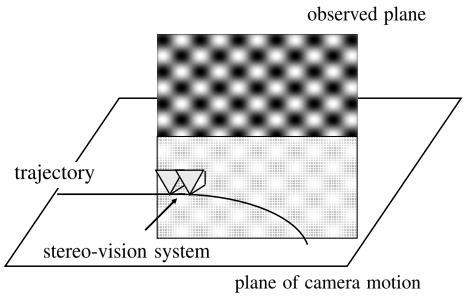

Fig. 3. Virtual setup used to generate the synthetic sequence of images used in the simulations.

Fig.4.a) shows the velocity profile.

In the first simulation, the optical axis directions of the two cameras are perfectly aligned. In this case, (8) simply writes

$$\frac{dB}{dt} = V_{x_b}\frac{B^2 - B^1}{h} + \dot{\theta}_c\frac{\partial B}{\partial \theta_c} \qquad (20)$$

The observer (9) is tested on that sequence and gives a velocity estimate V_{est}: the error $V_{\text{est}} - V_{x_b}$ is represented in Fig.4.b). After a rapid convergence step, the error stays below 0.002 $m.s^{-1}$.

In the second simulation, the optical axis are slightly misaligned, as they most certainly are in real systems: $\delta\theta_c =$ 1.6 deg, corresponding to a shift of 20 pixels in an image that needs to be compensated as in (8). The observers yields a velocity estimate whose difference to the real velocity is represented in Fig.4.c). After a rapid convergence step, the error stays below 0.015 $m.s^{-1}$.

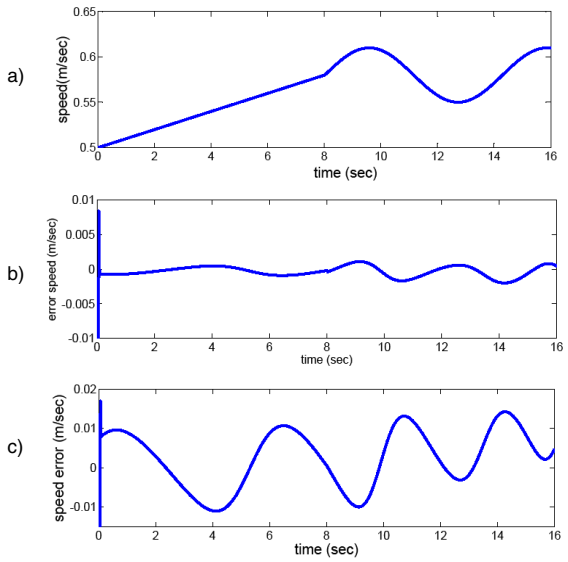

Fig. 4. a): Tangential velocity V_{x_b}: the trajectory is linear during the first 8 sec, then circular with radius 15 m. b): Error $V_{\text{est}} - V_{x_b}$ of estimation of the tangential velocity in the first simulation, where the optical axis are perfectly aligned. c): Error $V_{\text{est}} - V_{x_b}$ of estimation of the tangential velocity in the second simulation, where the optical axis form an angle of 1.6 deg.

Note that in both simulations, the error of estimation seems to be larger in the second phase of the trajectory (the circular trajectory), even though it stays bounded: this phenomenon is explained by the fact that the tangential velocity varies more rapidly in this phase.

C. Experimental results

To realize the experiments, a stereo-vision system was fixed on a motorized trolley traveling back and forth on a 2 meter-linear track pictured in Fig.5. The encoder of the motor enables to know the position of the trolley with a micrometric resolution. The stereo-vision system is composed of two Flea2 - Point Grey Research VGA video cameras (640 by 480 pixels) synchronized at 20.83 fps, with a Cinegon 1.8/4.8 C-mount lens, provideing an angular field of view of approximately 50 by 40 deg. The stereo-vision system is oriented orthogonally to the track.

The environment observed by the stereo-vision system is static, composed of random objects in an office space in a depth range of a few centimeters to 3.5 meters, figuring in Fig.6.

Note that the scene is uniquely 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 lighting of the room is variable, at a frequency that can not be easily ignored, which puts (1) and the governing equation (3) in

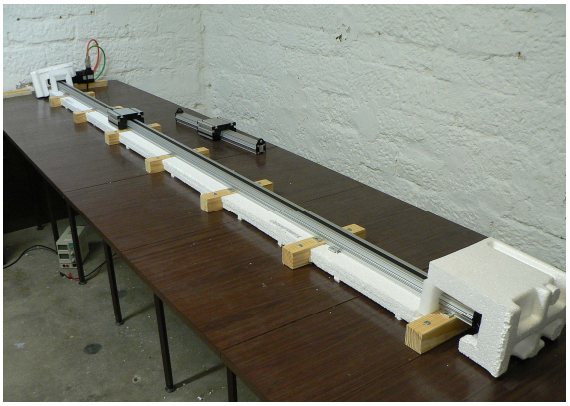

Fig. 5. Motorized trolley and track used to generate a linear motion for the experiments.

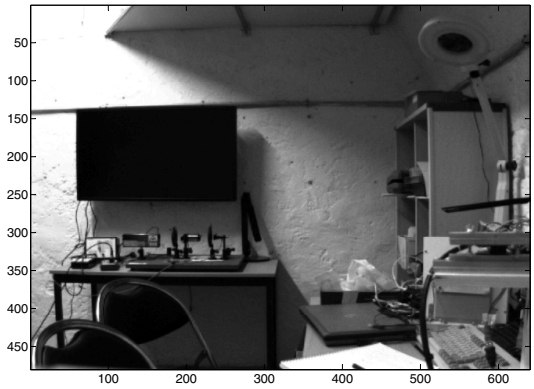

Fig. 6. Environment filmed by the stereo-vision system: an example of image processed for the experiment.

the wrong. However, the impact of this temporal dependance in the equations can be reduced by a normalization of the intensity of the images such as

$$y(x,y) = \frac{y(x,y)}{\bar{y}} \tag{21}$$

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 optical axis of the two cameras are slightly misaligned, forming an angle estimated to be 1.57 deg: this is taken into account in the approximation of $\frac{\partial B}{\partial x_b}$ as in the second simulation presented in IV-B.

With consideration of all these elements, the observer (9) provides an estimation V_{est} of the linear velocity of the trolley, which is on the other hand provided by the motor (V_{mes}) at a frequency of 500 Hz. Both are plotted in Fig.7.

Note that in spite of the rapid changes of sign of the velocity, the estimation quickly converges and follows every variation. In phases of constant velocity, the error does not reach zero, but noise remains on the same order of magnitude that the uncertainty of the measure (most of it due to lighting variations).

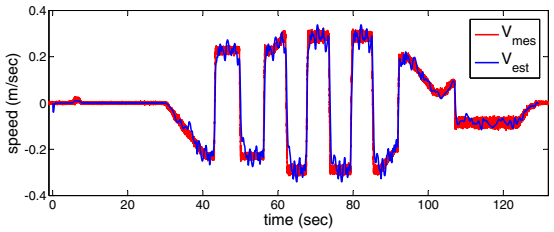

Fig. 7. Linear velocity: measured (in red) and estimated (in blue).

TABLE I

TABLE OF ESTIMATION ERRORS FOR VARIOUS RESOLUTIONS

Resolution	Error
640x480	6.80%
320x240	6.82%
160x120	6.84%
80x60	6.82%
40x30	7.45%
20x15	9.22%

V. Robustness of the method

In this section, we demonstrate that the method is robust even when dealing with low resolution image data or with blurred images. The sequence of paired images processed was generated with the open source 3D computer graphics software Blender©. The scene is a simple 10 by 4-meters wall painted with a picture containing sharp contrasts. The cameras travel 14 meters on a line parallel to the wall, with a constant velocity; their optical axis are oriented orthogonally to the wall. The sequence lasts 4 seconds. To compare the performance of the method for the different sequences, we use the following error, expressed in percent:

$$E = \int_{t=0}^{T} \frac{|\widehat{V}(t) - V(t)|}{|V(t)|} dt \tag{22}$$

For the image sequence used as a reference , the resolution is 640x480 pixels, and the exposure time is 0 (in practice, negligible compared to the sampling time). The reference error is then 6.80 %, most of which being due to the convergence lag.

A. Robustness to low resolution

To show that high resolution is not required to provide a good estimation of the velocity, we will process different sequences corresponding to the same motion and different image resolutions. The field of view stays the same but each pixel corresponds to a larger angular part over which brightness is averaged. The errors associated to estimations for each resolution are gathered in Table I.

These results highlight that even for very low camera resolutions, the proposed method estimates the velocity with a very good accuracy.

B. Robustness to blur

Fast moving objects appear to be ”moving”, that is, blurred by their own motion in a movie frame if the distance traveled

TABLE II

TABLE OF ESTIMATION ERRORS FOR VARIOUS EXPOSURE TIMES

Exposure time	Error
0	6.80%
0.25	6.89%
0.5	6.88%
0.75	6.89%
1	6.91%
1.25	7.24%

in the field of view of the camera during the exposure time is significant. Thus, for a given speed, blur appears when increasing the exposure time. To show that this method is not sensitive to blur, i.e. to high speed, we process different sequences corresponding to the same motion and different exposure times, expressed as fractions of the sampling time which is constant. The errors associated to estimations for each exposure time are gathered in Table II.

These results highlight that even for significant blurring effects, the proposed method estimates the velocity with a very good accuracy.

VI. CONCLUSIONS AND FUTURE WORKS

In section II, the notations of the problem were introduced for the camera, the field B obtained by weighting and averaging of the brightness perceived by a camera was defined, and the governing dynamics equations for B were detailed. In section III, the method used to adapt the equations to estimate the projection V_{x_b} of the velocity on an axis were described: specifically, the approximations and techniques to measure the derivatives of B were detailed and the measuring apparatus was consequently deduced; an observer for V_{x_b} was proposed. In section IV, the choice of the convolution kernel φ was considered, and results of simulations and experiments proved the validity of the method. In V, we showed on synthetic sequences the robustness of this approach to blur and downsampling. In other words, this approach seems to provide even from low resolution image data an accurate estimation of a large range of velocities. It only requires two low-quality cameras combined with a gyroscope, while Structure-from-Motion methods either employ a high-resolution binocular system or a high-resolution monocular system combined with an additional sensor for scale factor estimation. The next step would be to actually embed the measuring apparatus on a ground vehicle and to compare the estimated velocity to a reference, which is by now the missing element.

REFERENCES

[1] E. Abbott and D. Powell, "Land-vehicle navigation using GPS," *Proceedings of the IEEE*, vol. 87, no. 1, pp. 145 –162, jan 1999.

[2] E. Bayramoglu, N. Andersen, N. Kjolstad Poulsen, J. Andersen, and O. Ravn, "Mobile robot navigation in a corridor using visual odometry," in *Advanced Robotics, 2009. ICAR 2009. International Conference on*, june 2009, pp. 1 –6.

[3] J. Shen, D. Tick, and N. Gans, "Localization through fusion of discrete and continuous epipolar geometry with wheel and IMU odometry," in *American Control Conference (ACC), 2011*, 29 2011-july 1 2011, pp. 1292 –1298.

[4] M. Tomono, "3D localization based on visual odometry and landmark recognition using image edge points," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, oct. 2010, pp. 5953 –5959.

[5] S.-B. Kim, S.-Y. Lee, T.-H. Hwang, and K.-H. Choi, "An advanced approach for navigation and image sensor integration for land vehicle navigation," in *IEEE VTS Fall 60th Vehicular Technology Conference*, vol. 7, 2004, pp. 4075–4078.

[6] I. Skog and P. Handel, "In-car positioning and navigation technologies; a survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 1, pp. 4 –21, march 2009.

[7] G. Dissanayake, S. Sukkarieh, E. Nebot, and H. Durrant-Whyte, "The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications," *Robotics and Automation, IEEE Transactions on*, vol. 17, no. 5, pp. 731–747, 2001.

[8] S. M. Bezick, A. J. Pue, and C. M. Patzelt, "Inertial navigation for guided missile systems," *Johns Hopkins APL technical digest*, vol. 28, pp. 331–342, 2010.

[9] C. Urmson and W. Whittaker, "Self-driving cars and the urban challenge," *Intelligent Systems, IEEE*, vol. 23, no. 2, pp. 66 –68, march-april 2008.

[10] K. Ohno, T. Tsubouchi, B. Shigematsu, S. Maeyama, and S. Yuta, "Outdoor navigation of a mobile robot between buildings based on DGPS and odometry data fusion," in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, vol. 2, sept. 2003, pp. 1978 – 1984 vol.2.

[11] D. A. Grejner-Brzezinska, Y. Yi, and C. K. Toth, "Bridging GPS gaps in urban canyons: The benefits of ZUPTs," *Navigation*, vol. 48, no. 4, pp. 217–225, 2001.

[12] P. Davidson, M. Vazquez, and R. Piche, "Uninterrupted portable car navigation system using GPS, map and inertial sensors data," in *IEEE 13th International Symposium on Consumer Electronics,*, 2009, pp. 836–840.

[13] R. Szeliski, *Computer vision, Algorithms and Applications*. Springer, 2010.

[14] S. Soatto, P. Perona, R. Frezza, and G. Picci, "Motion estimation via dynamic vision," in *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, vol. 4, dec 1994, pp. 3253 –3258 vol.4.

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

[16] A. Van den Hengel, W. Chojnacki, and M. J. Brooks, "Determining the translational speed of a camera from time-varying optical flow," in *Proceedings of the 1st international conference on Complex motion*, ser. IWCM'04, 2007, pp. 190–197.

[17] D. Nister, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756 –770, june 2004.

[18] A. Dani, N. Fischer, and W. Dixon, "Single camera structure and motion," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 238 –243, jan. 2012.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, June 1981.

[20] F. Le-Bras, T. Hamel, and R. Mahony, "Nonlinear observer-based visual control of a VTOL UAV," in *Proc. of the European Control Conf.*, 2007.

[21] Z. Zhang and O. D. Faugeras, "Three-dimensional motion computation and object segmentation in a long sequence of stereo frames," *International Journal of Computer Vision*, vol. 7, pp. 211–241.

[22] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, sept. 2008, pp. 3946 –3952.

[23] D. Vissière, A. Martin, and N. Petit, "Using spatially distributed magnetometers to increase IMU-based velocity estimation in perturbed areas," in *Proc. of the 46th IEEE Conf. on Decision and Control*, 2007.

[24] E. Dorveaux, T. Boudot, M. Hillion, and N. Petit, "Combining inertial measurements and distributed magnetometry for motion estimation," in *Proceedings of the 2011 American Control Conference*, San Francisco, États-Unis, Jul. 2011, pp. 4249–4256.

[25] N. Zarrouati, E. Aldea, and P. Rouchon, "SO(3)-invariant asymptotic observers for dense depth field estimation based on visual data and known camera motion," in *American Control Conference, Montreal*, 2012, to appear.

[26] H. K. Khalil, *Nonlinear Systems*. MacMillan, 1992.

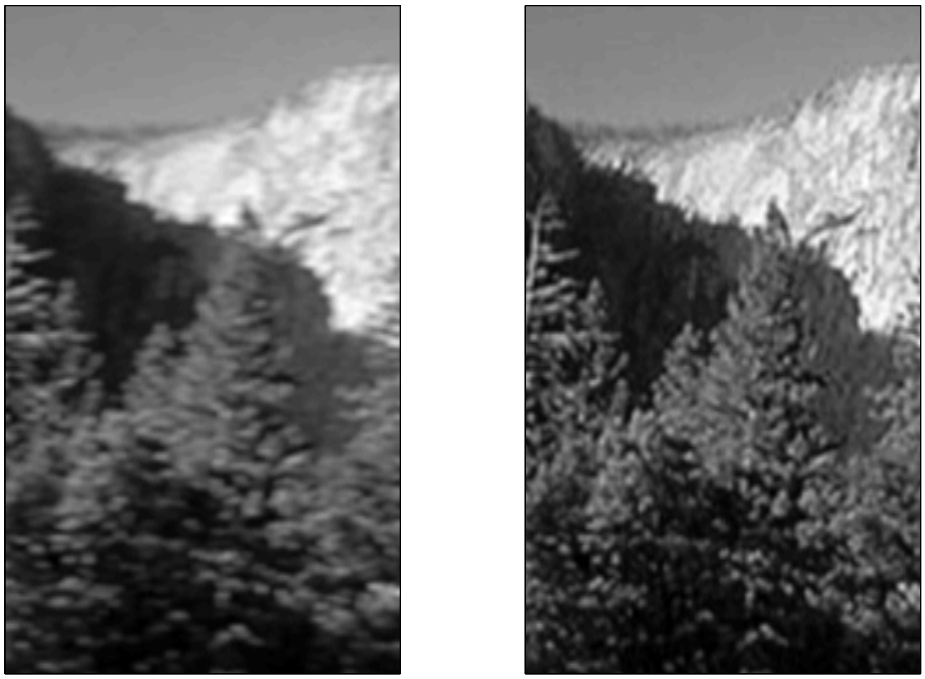

Fig. 8. Example of blurred image (left) versus original (right). Exposure time is 1.25, as reported in Table II.

[27] E. Dorveaux, "Magneto-inertial navigation: principles and application to an indoor pedometer," Ph.D. dissertation, École des Mines de Paris, 2011.