

Tracking planes in omnidirectional stereovision

Guillaume Caron, E. Marchand, E. Mouaddib

▶ **To cite this version:**

Guillaume Caron, E. Marchand, E. Mouaddib. Tracking planes in omnidirectional stereovision. IEEE Int. Conf. on Robotics and Automation, ICRA'11, May 2011, Shanghai, China. pp.6306-6311. hal-00639646

HAL Id: hal-00639646
https://inria.hal.science/hal-00639646

Submitted on 9 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tracking planes in omnidirectional stereovision

Guillaume Caron, Eric Marchand and El Mustapha Mouaddib

***Abstract*— Omnidirectional cameras allow direct tracking and motion estimation of planar regions in images during a long period of time. However, using only one camera leads to plane and trajectory reconstruction up to a scale factor. We propose to develop dense plane tracking based on omnidirectional stereovision to answer this issue. The presented method estimates simultaneously the parameters of several 3D planes along with the camera motion in a spherical model formulation. Results show the efficiency of the approach.**

I. Introduction

Motion estimation remains one of the key issue in the development of vision systems. Although the omnidirectional camera geometrical model (central projection model [3]) is now well known, few researches have been done to solve visual tracking problems. Omnidirectional cameras offer a very wide field of view, allowing the perception of landmarks during a long period of time in image sequences, which is an important asset to the precision and robustness of an object tracking process.

Available tracking techniques uses either a 3D model leading mainly to a pose estimation process (eg, [8] in the perspective case and [14] for central catadioptric image) or 2D features, such as geometrical primitives (point, segments, circles, etc.) or object contours (such as active contours). One can also consider 2D models which can be a reference image (or a template). The goal is to estimate the motion (or warp) between the current image and the reference template. Examples of such approaches are differential tracking methods such as the KLT [1], [4], [9], [12].

In this paper, we consider a such approach within an omnidirectional stereovision sensor.

The warp to be estimated is a homography. This 2D motion model has the advantage to be exact for a plane whatever the camera motions are. This motion model was first defined for perspective camera model but is in fact valid for all central projection cameras [15], [16] of which some omnidirectional ones [2].

The central projection model have already been used for plane tracking in omnidirectional monocular images (eg [15]). This approach is similar to the one proposed in [4] although the registration between reference template and current image is done directly on the omnidirectional images. Nevertheless, the image gradients, needed to link image intensities to the variation of the motion, are computed in the image plane. However, it has been shown [6] that the filtering of omnidirectional images should be done on the spherical image rather than on the omnidirectional planar image. Indeed, the spherical geometry, for these images, is adapted to keep the properties of the convolution product used to compute gradients since it allows a regular sampling of the image whereas this is not possible in the image plane.

This work proposes to formulate the simultaneous tracking of image planar templates, constrained by homographies, directly within the spherical image space. Furthermore, rather than estimate a homography for each plane, we propose to estimate the camera pose, up to a scale factor, and recover the parameters of 3D planes. Based on this new spherical formulation, we then propose to use a stereoscopic catadioptric sensor allowing to perform the complete stereoscopic plane reconstruction, including scale factor, and tracking of spherical image planar regions. These three aspects together are the contribution of this work.

Thanks to the unified spherical projection model, the proposed stereoscopic plane reconstruction and tracking approach is generic and apply to all systems with multiple central cameras. But practically, stereoscopic systems may suffer from synchronization problems and even chromatic inconsistency between cameras. Hopefully, catadioptric omnidirectional stereovision can be achieved using a unique camera and several mirrors which solves the two latter problems. The used stereo catadioptric system is composed by an orthographic camera and four parabolic mirrors. This configuration (Fig. 1) has been proven to be optimal [7] and the image resolution in stereo tracking results is taken from this work.

The reminder of the paper presents the spherical tracking of a plane in monocular central vision, recalling the camera



(a) (b) (c)

Fig. 1. (a) Mirror part of a stereo sensor composed by (b) an orthographic camera and four coplanar paraboloidal mirrors [7] ; (c) the sensor on the mobile robot (reversed for stability reasons).

Guillaume Caron and El Mustapha Mouaddib are with Universit de Picardie Jules Verne, MIS laboratory, Amiens, FRANCE; e-mail: {guillaume.caron, mouaddib}@u-picardie.fr

Eric Marchand is with Université de Rennes 1, IRISA, INRIA Lagadic, Rennes, France; e-mail: eric.marchand@irisa.fr

projection model and the projective transformation of a 3D plane motion before developing the spherical tracking itself. Then, the stereoscopic spherical plane tracking is presented. Finally, tracking results using a single camera and a stereo omnidirectional camera are presented, in static and mobile camera conditions. A conclusion highlighting the contributions of this work closes the paper.

II. Spherical plane tracking

A. Camera model

According to the unified model [3], a central projection camera can be modeled by a first projection on a sphere, with a unit radius and a center $(0,0,\xi)^\mathsf{T}$ expressed in the camera frame, followed by a perspective projection on the image plane. Such a model can be defined using a unique parameter ξ which depends intrinsically on the central camera type.

Knowing intrinsic parameters $\gamma = \{p_x, p_y, u_0, v_0, \xi\}$, a 3D point $\mathbf{X} = (X,Y,Z)^\mathsf{T}$ is first projected as $\mathbf{X}_\mathcal{S}$ on the unitary sphere:

$$\mathbf{X}_\mathcal{S} = (X_\mathcal{S}, Y_\mathcal{S}, Z_\mathcal{S})^\mathsf{T} = pr_\mathcal{S}(\mathbf{X}) = \left(\frac{X}{\rho}, \frac{Y}{\rho}, \frac{Z}{\rho}\right)^\mathsf{T} \quad (1)$$

where $\rho = \sqrt{X^2+Y^2+Z^2}$. Then this point is projected in the image plane as $\mathbf{x} = (x, y, 1)$. The relationship between $\mathbf{X}$ and $\mathbf{x}$ is expressed as:

$$\mathbf{x} = pr_\xi(\mathbf{X}) \quad \text{with} \quad x = \frac{X}{Z+\xi\rho} \quad \text{and} \quad y = \frac{Y}{Z+\xi\rho} \ . \quad (2)$$

$\mathbf{x}$ is the point on the normalized image plane and the image point in pixelic coordinates is obtained by the application of intrinsic parameters matrix $\mathbf{u} = \mathbf{K}\mathbf{x}$.

To back-project $\mathbf{x}$ to the equivalent sphere, the function $pr_\xi^{-1}(.)$ is used, knowing $\beta = \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1}$: $\mathbf{X}_\mathcal{S} = pr_\xi^{-1}(\mathbf{x}) = (\beta x, \beta y, \beta - \xi)^\mathsf{T}$.

B. Projective transformation associated to a 3D plane

The coordinates of a point on the 3D plane in two views (for c_1 and c_2) are related by a homography. Considering the rotation ${}^{c_2}\mathbf{R}_{c_1}$ and the translation ${}^{c_2}\mathbf{t}_{c_1}$ between the two camera positions over time, the homography $\mathbf{H}$, defined up to a scale factor [15], is given by:

$$\mathbf{H} \propto {}^{c_2}\mathbf{R}_{c_1} - \frac{{}^{c_2}\mathbf{t}_{c_1}\,{}^{c_1}\mathbf{n}^\mathsf{T}}{{}^{c_1}d}, \ {}^{c_2}\mathbf{t}_{c_1} \in \mathbb{R}^3 \text{ and } {}^{c_2}\mathbf{R}_{c_1} \in \mathbf{SO}(3) \quad (3)$$

where ${}^{c_1}\mathbf{n}$ is the normal to the plane of interest expressed in the first camera frame and ${}^{c_1}d$ the orthogonal distance from the origin of this coordinates system to the plane. The homography is a 3×3 matrix and has 9 parameters, $h_{1,1}$ to $h_{3,3}$ and 8 degrees of freedom. To fix the scale, $h_{3,3}$ is forced to be equal to 1.

According to the unified model, a central camera is modeled by a sphere and the homography $\mathbf{H}$ links two corresponding points ${}^{c_1}\mathbf{X}_\mathcal{S}$ and ${}^{c_2}\mathbf{X}_\mathcal{S}$ on two spheres, up to a scale factor:

$${}^{c_2}\mathbf{X}'_\mathcal{S} = \alpha\,{}^{c_2}\mathbf{X}_\mathcal{S} = \mathbf{H}\,{}^{c_1}\mathbf{X}_\mathcal{S}. \quad (4)$$

Actually, a homography relates two corresponding view lines and since we are working on the equivalent sphere, the spherical point ${}^{c_2}\mathbf{X}_\mathcal{S}$ corresponding to ${}^{c_1}\mathbf{X}_\mathcal{S}$, after being transfered using $\mathbf{H}$, is:

$${}^{c_2}\mathbf{X}_\mathcal{S} = {}^{c_2}\mathbf{X}'_\mathcal{S} \ / \ ||{}^{c_2}\mathbf{X}'_\mathcal{S}||. \quad (5)$$

C. Plane tracking on the unit sphere

To achieve plane tracking, we propose to use a differential minimization method that aims to compute the homography allowing to transfer spherical points of the reference sphere to the current one minimizing the sum of square differences of spherical image intensities. Even if tracking a plane directly estimating the coefficients of the projective transformation matrix $\mathbf{H}$ leads to the minimal set of parameters to estimate, the tracking is here defined in the Euclidean space. Hence, the rotation, translation, plane normal and distance are explicitly estimated. Since the camera motion is the same for all the scene planes, this formulation allows to consider efficiently multi planes tracking (see section II-D).

The criterion to minimize for the plane tracking is then:

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{h}} \sum_{i=1}^{k} \left(\mathbf{I}_\mathcal{S}\left(\mathbf{H}\mathbf{X}^*_{\mathcal{S},i}\right) - \mathbf{I}_\mathcal{S}\left(\mathbf{X}^*_{\mathcal{S},i}\right)\right)^2 \quad (6)$$

where $\mathbf{X}^*_\mathcal{S}$ is a point of the reference region and the vector of parameters to estimate is $\mathbf{h} = (\mathbf{n}, d, \mathbf{r})^\mathsf{T}$. $\mathbf{r}$ is a vector representation of the relative pose and $\mathbf{H}$ is computed thanks to these parameters with the equation (3).

Considering a feature vector $\mathbf{I}_\mathcal{S}(\mathbf{h})$, the set of spherical image intensities of the region of interest, obtained after the transfer of the reference one using the homography, the error to be regulated is:

$$\mathbf{e} = \mathbf{I}_\mathcal{S}(\mathbf{h}) - \mathbf{I}^*_\mathcal{S}. \quad (7)$$

Imposing an exponential decrease of this error, we have $\dot{\mathbf{e}} = -\lambda\mathbf{e}$. The time variation of the intensity at a given point on the sphere is defined thanks to the Cartesian spherical optical flow constraint equation (OFCE). Assuming a temporal luminance consistency for a moving spherical image point:

$$\mathbf{I}_\mathcal{S}(\mathbf{X}_\mathcal{S} + \Delta\mathbf{X}_\mathcal{S}, t + \Delta t) = \mathbf{I}_\mathcal{S}(\mathbf{X}_\mathcal{S}, t). \quad (8)$$

and if $\Delta\mathbf{X}_\mathcal{S}$ is small enough, the OFCE is valid:

$$\nabla\mathbf{I}_\mathcal{S}^\mathsf{T}\dot{\mathbf{X}}_\mathcal{S} + \dot{\mathbf{I}}_\mathcal{S} = 0. \quad (9)$$

$\nabla\mathbf{I}_\mathcal{S}$ is the spatial Cartesian spherical gradient of $\mathbf{I}_\mathcal{S}(\mathbf{X}_\mathcal{S}, t)$. Hence, the time variation of $\mathbf{I}_\mathcal{S}$ is then expressed as:

$$\dot{\mathbf{I}}_\mathcal{S} = -\nabla\mathbf{I}_\mathcal{S}^\mathsf{T}\dot{\mathbf{X}}_\mathcal{S}. \quad (10)$$

$\dot{\mathbf{X}}_\mathcal{S}$ can be decomposed as:

$$\dot{\mathbf{X}}_\mathcal{S} = \frac{\mathrm{d}\mathbf{X}_\mathcal{S}}{\mathrm{d}t} = \frac{\partial\mathbf{X}_\mathcal{S}}{\partial\mathbf{h}}\frac{\mathrm{d}\mathbf{h}}{\mathrm{d}t} = \mathbf{G}_{\mathbf{X}_\mathcal{S}\mathbf{h}}\dot{\mathbf{h}} = \left[\mathbf{J}_{\mathbf{X}_\mathcal{S}} \mid \mathbf{L}_{\mathbf{X}_\mathcal{S}}\right] \begin{bmatrix}\dot{\mathbf{n}}\\ \dot{d}\\ \mathbf{v}\end{bmatrix} \quad (11)$$

where $\mathbf{v} = (\boldsymbol{\upsilon}, \boldsymbol{\omega})^\mathsf{T}$ is the camera relative pose velocity. Jacobian $\mathbf{J}_{\mathbf{X}_\mathcal{S}}$ expresses the variation of a spherical point w.r.t. the 3D plane (constrained normal and distance), and

$\mathbf{L}_{\mathbf{X}_\mathcal{S}}$ is known as the interaction matrix (for a relative pose, here) and is [10]:

$$\mathbf{L}_{\mathbf{X}_\mathcal{S}} = \left(\tfrac{1}{\rho}(\mathbf{X}_\mathcal{S}\mathbf{X}_\mathcal{S}^T - \mathbf{I}_{3\times3}) \qquad [\mathbf{X}_\mathcal{S}]_\times\right). \tag{12}$$

$\mathbf{L}_{\mathbf{X}_\mathcal{S}}$ mainly depends on the coordinates of a spherical point, apart from ρ, which is the distance of the 3D point in the current camera frame (eq. 1). This distance is obtained by intersecting a viewline and the 3D plane.

Equations (10) and (11) lead to the image Jacobian related to $\mathbf{I}_\mathcal{S}$ at spherical point $\mathbf{X}_\mathcal{S}$:

$$\mathbf{G}_{\mathbf{I}_\mathcal{S}}(\mathbf{X}_\mathcal{S}) = -\nabla\mathbf{I}_\mathcal{S}^\mathsf{T}\mathbf{G}_{\mathbf{X}_\mathcal{S}\mathbf{h}}. \tag{13}$$

D. Tracking of several planes

To deal with a moving camera and tracking multiple planes, the problem can be formulated to compute only one relative pose and several normal and distances. In such a case, intensities of a planar region are used to compute their corresponding plane whereas the intensities of all the planar regions are used to compute the relative pose. Hence, the Jacobian linking the variation of image intensities to the camera motion and parameters of p planes is:

$$\mathbf{G}_{\mathbf{I}_\mathcal{S}} = \begin{bmatrix} -\nabla\mathbf{I}_\mathcal{S}^{(1)} \\ -\nabla\mathbf{I}_\mathcal{S}^{(2)} \\ \cdots \\ -\nabla\mathbf{I}_\mathcal{S}^{(p)} \end{bmatrix}^\mathsf{T} \left[\begin{array}{cccc|c} \mathbf{J}_{\mathbf{X}_\mathcal{S}}^{(1)} & 0 & \dots & 0 & \mathbf{L}_{\mathbf{X}_\mathcal{S}}^{(1)} \\ 0 & \mathbf{J}_{\mathbf{X}_\mathcal{S}}^{(2)} & \dots & \vdots & \mathbf{L}_{\mathbf{X}_\mathcal{S}}^{(2)} \\ \vdots & \dots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & \mathbf{J}_{\mathbf{X}_\mathcal{S}}^{(p)} & \mathbf{L}_{\mathbf{X}_\mathcal{S}}^{(p)} \end{array}\right] \tag{14}$$

where (i) shows the plane number to which the Jacobian is related to and its corresponding feature set. Then, the error vectors corresponding to each plane are stacked together and are related, thanks to the latter multiple planes Jacobian, to the camera and planes parameters velocity:

$$\mathbf{V} = \begin{bmatrix} \dot{\mathbf{n}}^{(1)} & \dot{d}^{(1)} & \dot{\mathbf{n}}^{(2)} & \dot{d}^{(2)} & \dots & \dot{\mathbf{n}}^{(p)} & \dot{d}^{(p)} & \mathbf{v} \end{bmatrix}^\mathsf{T} \tag{15}$$

E. Parameters update

The time variation of planes and relative pose vector is computed thanks to a Levenberg-Marquart (LM) optimization method:

$$\mathbf{V} = -\lambda\left(\mathbf{Q} + \mu\,\mathrm{diag}(\mathbf{Q})\right)^{-1}\mathbf{G}_{\mathbf{I}_\mathcal{S}}^\mathsf{T}\begin{bmatrix} \mathbf{I}_\mathcal{S}(\mathbf{h}^{(1)}) - \mathbf{I}_\mathcal{S}^* \\ \vdots \\ \mathbf{I}_\mathcal{S}(\mathbf{h}^{(p)}) - \mathbf{I}_\mathcal{S}^* \end{bmatrix} \tag{16}$$

with $\mathbf{Q} = \mathbf{G}_{\mathbf{I}_\mathcal{S}}^\mathsf{T}\mathbf{G}_{\mathbf{I}_\mathcal{S}}$ and $\mathbf{h}^{(i)} = \left(\mathbf{n}^{(i)}, \mathbf{d}^{(i)}, \mathbf{r}\right)^\mathsf{T}$. If μ is very high, the optimization behaves like a steepest descent whereas a very low value for μ leads equation (16) to behave like a Gauss-Newton method. LM is often used since it ensures a better convergence, thanks to a controlled and efficient step.

The relative pose is updated at iteration it with the exponential map of $se(3)$ [13] using ${}^{c_2}\mathbf{M}_{c_1}^{it} = {}^{c_2}\mathbf{M}_{c_1}^{it-1}e^{[\mathbf{v}]}$. Plane parameters are updated by addition of their time variation.

F. Gradients computation on the sphere

Computing $\nabla\mathbf{I}_\mathcal{S}$ means to compute the spherical image derivation along $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ axes. Since the spherical image is a surface which can be minimally parameterized using azimuth and elevation angles, ϕ and θ, spherical image gradients are computed w.r.t. to these parameters and then transformed to get the "3D" image gradient.

Omnidirectional image filtering is only valid on the equivalent sphere where convolution is commutative with regular sampling. The technique to compute gradients is to use intensities from a spherical neighborhood for the convolution [6].

The two linear neighborhoods $\phi_\mathcal{N}$ and $\theta_\mathcal{N}$ of length l, considering $\Delta_{\phi,\theta} = \arccos(|\begin{pmatrix}0 & 0 & 1\end{pmatrix}.pr_\xi^{-1}(\begin{bmatrix}u_0+1 & v_0 & 1\end{bmatrix}^\mathsf{T})|)$, are expressed as:

$$\begin{cases} \phi_\mathcal{N} = \left\{\left(\theta,\ \phi + k\Delta_\phi\right)^\mathsf{T}, -\frac{l}{2} \le k \le \frac{l}{2}, k \neq 0\right\} \\ \theta_\mathcal{N} = \left\{\left(\theta + k\Delta_\theta,\ \phi\right)^\mathsf{T}, -\frac{l}{2} \le k \le \frac{l}{2}, k \neq 0\right\} \end{cases}. \tag{17}$$

Neighborhoods are then projected on the image plane to retrieve intensities. Nearest neighbor interpolation is done to make the process fast and to allow precomputation of spherical neighbors projected in the image plane for each pixel. These neighborhoods are used with an image derivative filter.

Finally, gradient values I_ϕ, I_θ are used to compute $I_{X_\mathcal{S}}$, $I_{Y_\mathcal{S}}$ and $I_{Z_\mathcal{S}}$, where $I_A = \frac{\partial I}{\partial A}$ knowing the Cartesian-polar coordinates transformation, $\phi = \arccos(Z_\mathcal{S})$ and $\theta = \arctan(Y_\mathcal{S}/X_\mathcal{S})$, and using partial derivatives $\frac{\partial\phi,\theta}{\partial\mathbf{X}_\mathcal{S}}$.

III. Stereoscopic spherical 3D plane tracking

Relative pose estimation in a monocular plane tracking, even with several planes, is estimated up to a scale factor. Dealing with a calibrated stereo rig allows to fix the scale and to have planes and relative position in meters.

Figure 2 shows the geometric relations between corresponding points in spherical views. Based on these considerations, the stereoscopic plane tracking has two main stages. First, the search for regions in the three views corresponding to the reference template is made, estimating the 3D plane parameters. Then, for any new stereo image, the relative pose is computed simultaneously to the plane parameters optimization. Therefore, we estimate only a minimal set of parameters.

A. Stereoscopic model

For a stereo rig of N cameras, each camera is modeled using the spherical model. Relative poses of $N-1$ cameras c_j, $j = 1..N$, of the rig, modeled by homogeneous matrices ${}^{c_j}\mathbf{M}_{c_1}$, are defined w.r.t. the reference one, c_1. In the present case, since the used stereo sensor for our experiments is composed of one orthographic camera and four parabolic mirrors, we consider a stereo rig of four central cameras ($N = 4$). Simultaneous intrinsic and extrinsic calibration has been done using an extension of a stereo omnidirectional virtual visual servoing approach [5].

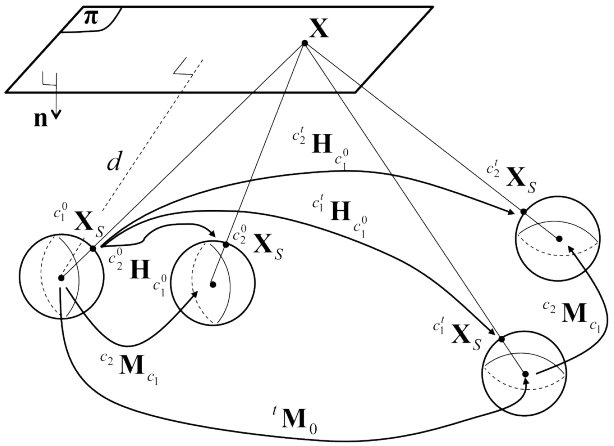


Fig. 2. Geometric relations between views. For brevity, a two views stereo rig is presented.

B. Stereoscopic spherical 3D plane tracking

Tackling planar region tracking in stereovision images, the reference region is defined in the part of the image corresponding to the first camera c_1. Before tracking the plane over time, dense correspondence is needed in the three other views. It is simultaneously done with the 3D plane reconstruction in a tracking process, thanks to the short baseline of the used stereo sensor.

1) Stereoscopic initialization: For the stereo initialization step, only the plane normal and its orthogonal distance to the reference camera center is computed. Indeed, from homogeneous transformations $^{c_j}\mathbf{M}_{c_1}$, rotation and translation from a camera to the reference one are known. Hence, the intensities of the three other views are related to the image of c_1, merging together the information of the four views to compute one 3D plane expressed in the c_1 frame.

A spherical point of the first camera is transfered to the camera j using equation (4) and $^{c_j}\mathbf{M}_{c_1}$ obtained from the stereo calibration. Merging together the four spherical images in a LM minimization method, leads to the following Jacobian and error vector:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{I_{\mathcal{S}_2}}(\mathbf{X}_\mathcal{S}) \\ \mathbf{J}_{I_{\mathcal{S}_3}}(\mathbf{X}_\mathcal{S}) \\ \mathbf{J}_{I_{\mathcal{S}_4}}(\mathbf{X}_\mathcal{S}) \end{bmatrix} \text{ and } \mathbf{e} = \begin{bmatrix} \mathbf{I}_{\mathcal{S}_2}\left(^{c_2}\mathbf{H}_{c_1}\mathbf{X}^*_\mathcal{S}\right) - \mathbf{I}_{\mathcal{S}_1}\left(\mathbf{X}^*_\mathcal{S}\right) \\ \mathbf{I}_{\mathcal{S}_3}\left(^{c_3}\mathbf{H}_{c_1}\mathbf{X}^*_\mathcal{S}\right) - \mathbf{I}_{\mathcal{S}_1}\left(\mathbf{X}^*_\mathcal{S}\right) \\ \mathbf{I}_{\mathcal{S}_4}\left(^{c_4}\mathbf{H}_{c_1}\mathbf{X}^*_\mathcal{S}\right) - \mathbf{I}_{\mathcal{S}_1}\left(\mathbf{X}^*_\mathcal{S}\right) \end{bmatrix} \tag{18}$$

where $\mathbf{J}_{I_{\mathcal{S}_j}}(\mathbf{X}_\mathcal{S})$ is the Jacobian related to intensity of the j-th spherical image at point $\mathbf{X}_\mathcal{S}$ (eq. (11)):

$$\mathbf{J}_{I_{\mathcal{S}_j}}(\mathbf{X}_\mathcal{S}) = -\nabla\mathbf{I}^\mathsf{T}_{\mathcal{S}_j}\begin{bmatrix}\frac{\partial\mathbf{X}_\mathcal{S}}{\partial\mathbf{N}} & \frac{\partial\mathbf{X}_\mathcal{S}}{\partial d}\end{bmatrix}. \tag{19}$$

Then, the normal and the distance are updated adding their time variation to their previous values and homographies $^{c_j}\mathbf{H}_{c_1}$ are recomputed after this update using $^{c_j}\mathbf{R}_{c_1}$ and $^{c_j}\mathbf{t}_{c_1}$ obtained by calibration, similarly to equation (3).

2) Eight views stereoscopic tracking of several planes: After the stereoscopic initialization step of each plane, their tracking over time is led, estimating the relative pose $^{t}\mathbf{M}_0$ between two stereo rig poses. This relative pose is expressed between the current position at time t and the reference position at time 0 of the reference camera, c_1^0 and c_1^t, of the stereo rig. Hence, at time t, image features of cameras c_2, c_3, c_4 are expressed w.r.t. c_1. Then, the information at time t are linked to the image of camera c_1 at time 0. The information of the four cameras at time 0 are still used to constrain the parameters of the 3D plane.

So, to sum up, the image region of c_1^0, corresponding to the plane of interest, forms the desired features and the images of c_2^0, c_3^0, c_4^0 are used to optimize $^{c_1}\mathbf{n}$ and d. c_1^t, c_2^t, c_3^t, c_4^t are used to optimize both the 3D plane parameters and the relative pose of the stereo rig.

Considering the pose $^{c_j^t}\mathbf{M}_{c_1^0}$, the pose Jacobian $\mathbf{L}_{\mathbf{I}_\mathcal{S}}(\mathbf{X}_\mathcal{S})$ related to $\mathbf{I}_\mathcal{S}$ at point $\mathbf{X}_\mathcal{S}$ is:

$$\mathbf{L}_{\mathbf{I}_{\mathcal{S}_j}}(\mathbf{X}_\mathcal{S}) = -\nabla\mathbf{I}^\mathsf{T}_{\mathcal{S}_j}\mathbf{L}_{\mathbf{X}_{\mathcal{S}_j}}. \tag{20}$$

$\mathbf{L}_{\mathbf{X}_\mathcal{S}}$ is defined in equation (12).

However, $\mathbf{L}_{\mathbf{X}_\mathcal{S}}$ of mirrors 2, 3 and 4 are transformed to be expressed in the c_1 frame [5] using the twist transformation matrix, computed thanks to $^{c_j}\mathbf{M}_{c_1}$, between velocity vectors of camera 1 and camera j. This leads to the Jacobian $\mathbf{G}$ for the stereoscopic tracking of several planes. For brevity, a Jacobian of two planes tracked by a stereo rig of two cameras is shown but it is extendable to any number of planes and cameras (four cameras and two planes in the results). The Jacobian relates spherical image intensities to the 3D planes and the relative pose between current and reference stereo rig positions:

$$\mathbf{G} = \left[\begin{array}{cc|c} \mathbf{J}^{(1)}_{\mathbf{I}_{\mathcal{S}_2^0}}(\mathbf{X}_\mathcal{S}) & 0 & 0 \\ 0 & \mathbf{J}^{(2)}_{\mathbf{I}_{\mathcal{S}_2^0}}(\mathbf{X}_\mathcal{S}) & 0 \\ \mathbf{J}^{(1)}_{\mathbf{I}_{\mathcal{S}_1^t}}(\mathbf{X}_\mathcal{S}) & 0 & \mathbf{L}^{(1)}_{\mathbf{I}_{\mathcal{S}_1^t}}(\mathbf{X}_\mathcal{S}) \\ \mathbf{J}^{(1)}_{\mathbf{I}_{\mathcal{S}_2^t}}(\mathbf{X}_\mathcal{S}) & 0 & \mathbf{L}^{(1)}_{\mathbf{I}_{\mathcal{S}_2^t}}(\mathbf{X}_\mathcal{S})^{c_2}\mathbf{V}_{c_1} \\ 0 & \mathbf{J}^{(2)}_{\mathbf{I}_{\mathcal{S}_1^t}}(\mathbf{X}_\mathcal{S}) & \mathbf{L}^{(2)}_{\mathbf{I}_{\mathcal{S}_1^t}}(\mathbf{X}_\mathcal{S}) \\ 0 & \mathbf{J}^{(2)}_{\mathbf{I}_{\mathcal{S}_2^t}}(\mathbf{X}_\mathcal{S}) & \mathbf{L}^{(2)}_{\mathbf{I}_{\mathcal{S}_2^t}}(\mathbf{X}_\mathcal{S})^{c_2}\mathbf{V}_{c_1} \end{array}\right]. \tag{21}$$

$^{c_j^t}\mathbf{H}_{c_1^0}$ is the homography, related to the plane of interest between c_1 at time 0 and c_j at time t, computed using estimated plane normal and distance and the rotation and translation extracted from $^{c_j^t}\mathbf{M}_{c_1^0} = {}^{c_j}\mathbf{M}_{c_1}{}^{t}\mathbf{M}_0$. Finally, Jacobians $\mathbf{J}_{\mathbf{I}_{\mathcal{S}_j^0}}$ are recomputed at each iteration since the 3D plane parameters are always updated.

For brevity, details of the LM optimization are not given here but the equation is really similar to equation (16). Finally, without giving much details, a robust estimator based on M-Estimation [11] is considered to reject outliers and improve robustness w.r.t. illumination changes, including specularity, and partial occlusion.

IV. Results

The initial plane selection is done manually by clicking four points to define a reference template in the first image. In stereovision images, the selection is only done in the top left mirror. Corresponding regions in other mirrors are automatically found.

The processing is currently real time for only one plane tracked by one camera but the implementation of the method clearly needs to be optimized as the processing time evolves as a function of the number of pixels in tracked regions.

A video presents these results in the research section of the website: http://mis.u-picardie.fr/∼g-caron.

A. Mobile omnidirectional camera

The tracking of several planes is experimented using a paracatadioptric camera acquiring images of 640×480 pixels (450×450 useful). Figure 3(a) shows the first image with selected planar regions of the processed sequence. The experiment is led with a mobile robot embedding the camera and making a loop, coming back near its original position and continuing its motion along a straight line in a corridor. In this sequence the motion of the planar regions in images is very important as well as perspective effects. Furthermore, each tracked plane is successively partially occluded (Fig. 3(b)) and some specularity appears. Finally, both tracked planes are visible in the major part of the sequence but not in the corridor. To deal with this kind of issue, when camera position is too close from the plane and if its image region surface is too small, the particular plane is no longer tracked.

The tracking succeeds all along the sequence and the estimated camera/robot trajectory (Fig. 3(e)) shows the stability of the estimation, even in presence of partial occlusions. The standard deviation of the Z estimates is 2.64 cm for the complete sequence, 1.5 cm when tracking two planes and 2.19 cm after stopping to track the plane $\mathbf{P}_1$. To compare, the tracking of plane $\mathbf{P}_0$ only leads to a Z standard deviation of 4 cm for the complete sequence (small instability when partially occluded). Variations of estimated normal parameters (always in the coordinates system of the first camera position) emphasizes also the estimation stability (Fig. 3(c) and 3(d)). Real planes are vertical as the camera axis and, hence, their estimated Z component is 0. X and Y components depends on the orientation of the plane in the first camera frame (Fig. 3(a)).

B. Static stereoscopic omnidirectional camera

In the static or mobile case, images are acquired at a resolution of 800×800 pixels, leading to a diameter of 400 pixels for each mirror, using the four mirrors catadioptric sensor (Fig. 1). In this experiment, the camera is static and a planar object is moved all around the camera (Fig. 4). The tracking is achieved even with important specular reflections on the real planar object (Fig. 4(b)) leading to differences between the current and the reference image template (Fig. 4(a)). The object does not makes a complete turn around the camera, that is why the final rotation angle around the vertical axis is not 0 (Fig. 4(c)).

C. Mobile stereoscopic omnidirectional camera

To evaluate the behavior of the proposed several planes tracking approach, the four view stereo sensor (Fig. 1) is used to recover the motion of the robot embedding it in a pure translation (Fig. 5).

The pure translation experiment aims to show the efficiency of the tracking method w.r.t. important motion between the first (Fig. 5(a)) and the last (Fig. 5(b)) image of the sequence. 230 images are acquired for a total motion of 2.40 m. The estimated trajectory is really near the desired one (Fig. 5(c)) with a final position error, at the farthest from the observed planes, of around 10 cm. This result is obtained without adding a manually measured distance or any other information. The precision of course depends on the calibration precision and the achievement of the tracking which is efficient, here, since it succeeds despite size and orientation variations of tracked regions in images. The tracking also shows its robustness to illumination variations, specially for specularities. Finally, one can note the stability of the estimation which is also emphasized by the stability of the two plane normal vectors (Fig. 5(d)) when they have reached their optimum. This optimum can only be reached when the distance between the current and the reference view is wide enough to reduce uncertainties.

V. Conclusion

The tracking of 3D planes in omnidirectional stereoscopic images has been successfully led. The spherical formulation of the problem is theoretically consistent and gives precise results for the tracking as well as for the stereo system motion estimation. A particular omnidirectional stereo camera has been used with properties of perfect synchronization and chromatic similarity. Results also show the robustness of the approach to occlusions.

Future work will be focused on a fully automated plane detection and tracking method based on this work.

References

[1] S. Baker and I. Matthews, *Equivalence and Efficiency of Image Alignment Algorithms*, IEEE Int. Conf. on Computer Vision and Pattern Recognition, December 2001.

[2] S. Baker and S. K. Nayar, *A Theory of Single-Viewpoint Catadioptric Image Formation*, Int. Journal on Computer Vision **35** (1999), no. 2.

[3] J. P. Barreto and H. Araujo, *Issues on the Geometry of Central Catadioptric Imaging*, IEEE Int. Conf. on Computer Vision and Pattern Recognition (Hawaii, USA), December 2001.

[4] S. Benhimane and E. Malis, *Homography-based 2D Visual Tracking and Servoing*, Int. Journal of Robotics Research **26** (2007), no. 7.

[5] G. Caron, E. Marchand, and E. Mouaddib, *3D Model Based Pose Estimation For Omnidirectional Stereovision*, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (St. Louis, USA), October 2009.

[6] C. Demonceaux and P. Vasseur, *Omnidirectional image processing using geodesic metric*, Int. Conf. on Image Processing (Cairo, Egypt), November 2009.

[7] L. Devendeville, G. Dequen, and E. Mouaddib, *On Omnidirectional Catadioptric Stereovision Design*, Iberian Conf. on Pattern Recognition and Image Analysis (Girona, Spain), 2007.

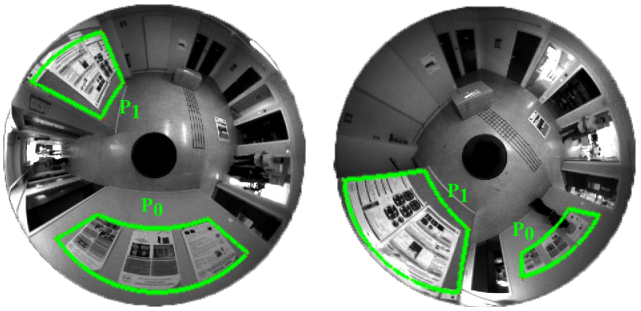
[8] T. Drummond and R. Cipolla, *Real-Time Visual Tracking of Complex Structures* , IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002), no. 7.

[9] G. Hager and P. Belhumeur, *Efficient Region Tracking With Parametric Models of Geometry and Illumination*, IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998), no. 10.

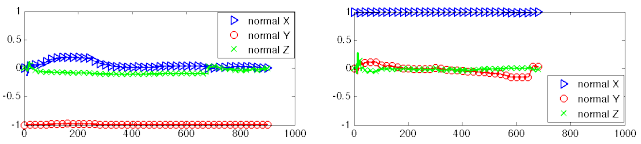
[10] T. Hamel and R. Mahony, *Visual servoing of an under-actuated dynamic rigid-body system: An image-based approach*, IEEE Trans. on Robotics and Automation **18** (2002), no. 2.

[11] P.-J. Huber, *Robust Statistics*, Wiler, New York, USA, 1981.

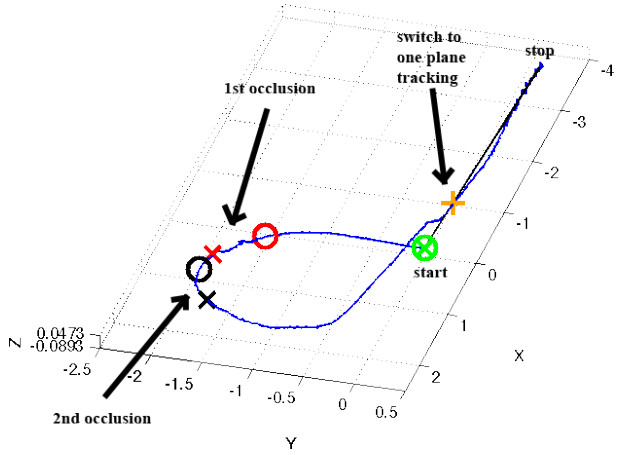
[12] B.D. Lucas and T. Kanade, *An Iterative Image Registration Technique with an Application to Stereo Vision*, Int. Joint Conf. on Artificial Intelligence, 1981.
[13] Y. Ma, S. Soatto, J. Košecká, and S. Sastry, *An invitation to 3-D vision*, Springer, 2004.
[14] E. Marchand and F. Chaumette, *Fitting 3D Models on Central Catadioptric Images*, IEEE Int. Conf. on Robotics and Automation (Roma, Italia), April 2007.
[15] C. Mei, S. Benhimane, E. Malis, and P. Rives, *Efficient Homography-Based Tracking and 3-D Reconstruction for Single-Viewpoint Sensors*, IEEE Trans. on Robotics **24** (2008), no. 6.
[16] A. Salazar-Garibay, E. Malis, and C. Mei, *Visual Tracking of Planes with an Uncalibrated Central Catadioptric Camera*, IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (Saint-Louis, USA), October 2009.



(a) Image 0 / 900 (b) Image 216 / 900 ($\mathbf{P}_0$ occluded)

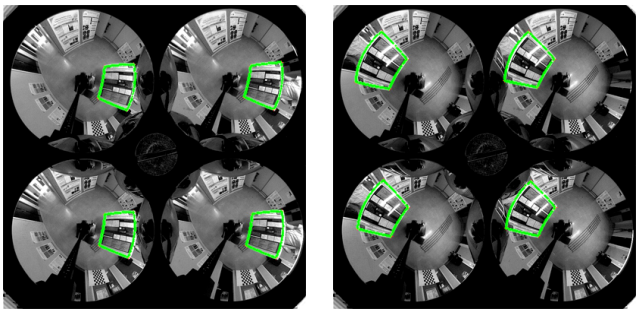


(c) Evolution of the $\mathbf{P}_0$ normal. (d) Evolution of the $\mathbf{P}_1$ normal.

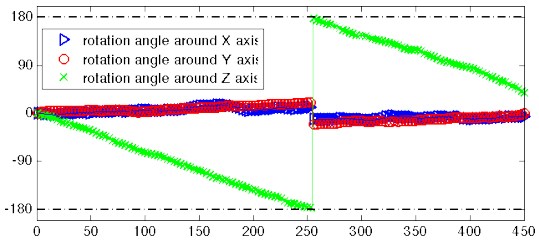


(e) Estimated trajectory with a circled green cross marking the starting position and two sets of a circle and a cross showing the trajectory parts when tracked planes are occluded (red for $\mathbf{P}_0$ and black for $\mathbf{P}_1$). The orange cross marks the end of the tracking of $\mathbf{P}_1$ and a straight black line links the start position to the end.

Fig. 3. Curve motion along 11 m with partial occlusion. Planes $\mathbf{P}_0$ and $\mathbf{P}_1$ are simultaneously tracked in a sequence of 900 images until $\mathbf{P}_1$ is no more visible. The rest of the estimation is done with the tracking of $\mathbf{P}_1$ (from the orange cross on trajectory (e)). To have the trajectory at the good scale in meter, the distance to $\mathbf{P}_0$ is manually measured from the initial camera position. The final position error is below 1 cm (unmeasurable) and the standard deviation of the estimation of the Z coordinate is about 2.64 cm for the complete estimated trajectory.

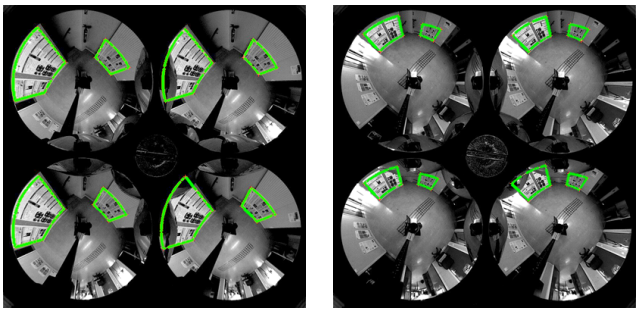


(a) Image 0 / 450 (b) Image 220 / 450

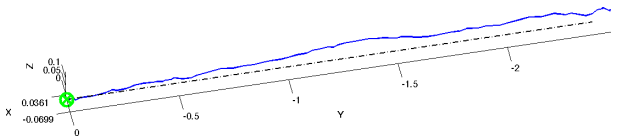


(c) Estimated angles of rotation w.r.t. the reference image.

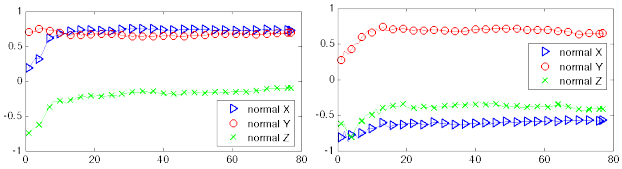
Fig. 4. Stereoscopic tracking of a plane all around the sensor. The rotation of the planar object is nearly pure around the camera axis (c).



(a) Image 0 / 230 (b) Image 230 / 230



(c) Estimated trajectory in blue and reference one in dashed black. The unit is the meter and the circled green cross shows the start position.



(d) Plane normal vectors evolution along the sequence.

Fig. 5. Straight motion along 2.40 m. The final position error is 10 cm in 3D leading to an error ratio over the path length of around 4%. Note that the smallest plane in images (a) and (b) is not vertical, that is why its Z component is not 0.