Constrained Multiple Planar Template Tracking for Central Catadioptric Cameras

Christopher Mei Selim Benhimane Ezio Malis
Patrick Rives
INRIA Sophia Antipolis - ICARE Project team
firstname.lastname@sophia.inria.fr

Abstract

This paper presents a homography-based approach for tracking multiple planar templates with central catadioptric cameras (which include perspective cameras). We extend the standard notion of homography to this wider range of devices through the unified projection model on the sphere. To enforce the same movement of the camera for the different planes, we parametrise the homography by the Lie algebra of the special euclidean group $\mathbb{SE}(3)$ and estimate the normal and depth for each plane. With this model, we use a minimisation technique to obtain a close to second-order convergence rate with a complexity similar to a first-order approach. The proposed method takes into account the non-uniform resolution of the sensor and proved robust to poor initial values for the plane normals. To assess the precision of the approach, the developed algorithm was tested on the estimation of the displacement of a mobile robot in a real application. We compare the results when the planes are tracked independently to the constrained case. We show that the proposed minimisation approach leads to better results in terms of speed and precision than current tracking algorithms.

1 Introduction

The combination of a convex mirror with a camera can enhance the field of view and improve tasks such as motion estimation, autonomous navigation and localisation. However to apply the results obtained in the field of projective geometry, a single viewpoint is needed. In [2], Baker and Nayar derive all the combinations of a single mirror and a camera that have this desired property. Geyer [10] and Barreto [4] have developed a unified projection model for these devices using a sphere (as opposed to a plane) as a projection surface. Under certain conditions, this model can be used for fish-eye cameras [22].

The focus of research in omnidirectional cameras is currently egomotion estimation [21, 11] and visual servoing [15, 12]. Visual tracking, which is a fundamental step for various computer vision applications, has seen very few articles for catadioptric systems, [15, 19] being exceptions. In [19], the authors propose an approach for tracking single planes by parameterising the homographies with $\mathbb{SL}(3)$.

The apparent difficulty of tracking with these devices comes from the non-linear projection model resulting in changes of shape in the image that makes the direct use of

methods such as KLT [17] nearly impossible. Parametric models [13, 20, 1] such as the homography-based approach presented in this article are well adapted to this problem. Previous related work using homography-based tracking for perspective cameras include [5] and [6] which extend the work proposed by Hager [13]. Homographies have also been used for visual servoing with central catadioptric cameras [12] and share with this article the notion of homographies for points belonging to the sphere of the unified projection model. The single viewpoint property means it would be possible to track in an unwarped perspective view. This is however undesirable for the following reasons : 1) it introduces a discontinuity in the Jacobian (at least two planes are needed to represent the 360^o field of view), 2) the non-uniform resolution is not taken into account and 3) the approach is inefficient (in terms of speed and memory usage).

The motivation for imposing the same camera euclidean motion for each tracked template is to improve the motion estimation and simultaneously the tracking. [7] imposes this constraint but assumes that the plane equations have been precomputed. Molton and Davison [9] estimate the normals of planes to improve the tracking of planar templates but estimate the motion independently ("de-centralised" approach). In [16], the authors linearise the equations to apply a Kalman filter. In this article, we propose to estimate the normals and depths on-line and impose the euclidean constraints directly during the tracking. We also detail how to adapt the efficient second-order minimisation algorithm proposed by Malis [18] in order to improve the convergence domain and speed of standard first-order minimisation algorithms [17, 13, 20, 1]. We show how to take into account the non-linear resolution and the distortion of the sensor. We also discuss the initialisation of the plane normals. In the last section, we compare constrained to unconstrained tracking with the use of the odometry of a mobile robot as ground truth.

2 Lie-group homography parameterisation

2.1 Notations

Let $\mathbf{R} \in \mathbb{SO}(3)$ be the rotation of the camera and $\mathbf{t} \in \mathbb{R}^3$ its translation. $\mathbf{R}$ can be written as $\mathbf{R} = \exp([\mathbf{r}]_\times)$ where $\mathbf{r} = \mathbf{u}\theta$ ($\mathbf{u}$ and θ being the axis and the angle of rotation respectively). The camera displacement between two views can be represented by a (4×4) matrix $\mathbf{T} \in \mathbb{SE}(3)$ (the Special Euclidean Group) :

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

The standard planar homography matrix $\mathbf{H}$ is defined up to a scale factor : $\mathbf{H}(\mathbf{T}, \mathbf{n}) \sim \mathbf{R} + \mathbf{t}\mathbf{n}_d^{*\top}$ where $\mathbf{n}_d^* = \mathbf{n}^*/d^*$ is the ratio between the normal vector to the plane $\mathbf{n}^*$ (a unit vector) and the distance d^* of the plane to the origin of the reference frame.

2.2 The Lie Algebra of $\mathbb{SE}(3)$ and the exponential map

Let $\mathbf{A}_i$, with $i \in \{1, 2, ..., 6\}$, be a basis of the Lie Algebra $\mathfrak{se}(3)$ (i.e. the dimension of the Lie Algebra $\mathfrak{se}(3)$ is 6). Any matrix $\mathbf{A} \in \mathfrak{se}(3)$ can be written as a linear combination of the matrices $\mathbf{A}_i$:

$$\mathbf{A}(\mathbf{x}) = \sum_{i=1}^{6} x_i \mathbf{A}_i \tag{2}$$

where $\mathbf{x} = (x_1,\ x_2,\ ...,\ x_6)$ is a (6×1) vector and x_i is the i-th element of the base field.

Let the (3×1) vectors $\mathbf{b}_x = (1,0,0)$, $\mathbf{b}_y = (0,1,0)$ and $\mathbf{b}_z = (0,0,1)$ be the natural orthonormal basis of $\mathbb{R}^3$. Knowing that the dimension of the matrices $\mathbf{A}_i$ is (4×4), the generators for the translation $(\mathbf{A}_1,...,\mathbf{A}_3)$ and rotation $(\mathbf{A}_4,...,\mathbf{A}_6)$ are :

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{b}_x \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{b}_y \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_3 = \begin{bmatrix} \mathbf{0} & \mathbf{b}_z \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_4 = \begin{bmatrix} [\mathbf{b}_x]_\times & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_5 = \begin{bmatrix} [\mathbf{b}_y]_\times & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_6 = \begin{bmatrix} [\mathbf{b}_z]_\times & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \tag{3}$$

where $[\mathbf{b}_i]_\times$ is the antisymmetric matrix associated to the vector $\mathbf{b}_i$ (i.e. $[\mathbf{b}_i]_\times \in \mathfrak{so}(3)$). The exponential map links the Lie Algebra to the Lie Group : $\exp : \mathfrak{se}(3) \rightarrow \mathbb{SE}(3)$

There exist an open cube v about $\mathbf{0}$ in $\mathfrak{se}(3)$ and an open neighbourhood U of the identity matrix $\mathbf{I}$ in $\mathbb{SE}(3)$ such that $\exp : v \rightarrow U$ is smooth and one-to-one onto, with a smooth inverse. The neighbourhood U of $\mathbf{I}$ is very large (i.e. the rotation angle must be less than π). Consequently, a homography matrix $\mathbf{H}$ is a function of $\mathbf{T}$ that can be locally parameterised as :

$$\mathbf{H}(\mathbf{T}(\mathbf{x})) = \mathbf{H}\left(\exp\left(\sum_{i=1}^{6} x_i \mathbf{A}_i\right)\right) \tag{4}$$

3 Generalised homography-based tracking

3.1 Unified projection model

For sake of completeness, we present here a slightly modified version of the projection model of Geyer [10] and Barreto [4] (see Fig. 1). The projection of 3D points can be done in the following steps (the values for the parameters ξ and η are detailed Table 1) :

1. world points in the mirror frame are projected onto the unit sphere, $(\mathscr{X})_{\mathscr{F}_m} \longrightarrow (\mathscr{X}_s)_{\mathscr{F}_m} = \frac{\mathscr{X}}{\|\mathscr{X}\|} = (X_s, Y_s, Z_s)$

2. the points are then changed to a new reference frame centered in $\mathscr{C}_p = (0,0,\xi)$, $(\mathscr{X}_s)_{\mathscr{F}_m} \longrightarrow (\mathscr{X}_s)_{\mathscr{F}_p} = (X_s, Y_s, Z_s - \xi)$

3. we then project the point onto the normalised plane, $\mathbf{m} = (\frac{X_s}{Z_s-\xi}, \frac{Y_s}{Z_s-\xi}, 1) = \hbar(\mathscr{X}_s)$

4. the final projection involves a generalised camera projection matrix $\mathbf{K}$ (with γ the generalised focal length, (u_0, v_0) the principal point, s the skew and r the aspect ratio)
$$\mathbf{p} = \mathbf{K}\mathbf{m} = \begin{bmatrix} \gamma & \gamma s & u_0 \\ 0 & \gamma r & v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{m} = k(\mathbf{m})$$

The function $\hbar$ is bijective and $\hbar^{-1}(\mathbf{m}) = \begin{bmatrix} \frac{-\xi - \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} x \\ \frac{-\xi - \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} y \\ \frac{-\xi - \sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} + \xi \end{bmatrix}$

We will call lifting the calculation of the point $\mathscr{X}_s$ corresponding to a given point $\mathbf{m}$ (or $\mathbf{p}$ according to the context).

Table 1: Unified model parameters

	ξ	γ
Parabola	1	$-2pf$
Hyperbola	$\frac{df}{\sqrt{d^2+4p^2}}$	$\frac{-2pf}{\sqrt{d^2+4p^2}}$
Ellipse	$\frac{df}{\sqrt{d^2+4p^2}}$	$\frac{2pf}{\sqrt{d^2+4p^2}}$
Planar	0	-f
Perspective	0	f
d : distance between focal points $4p$: latus rectum f : camera focal length		

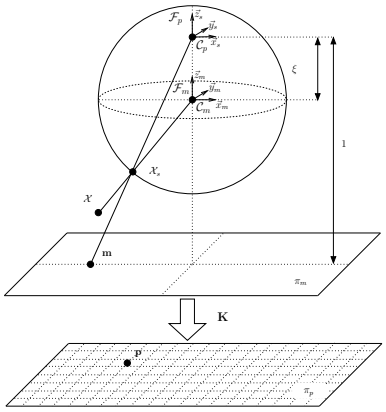

Figure 1: Unified image formation

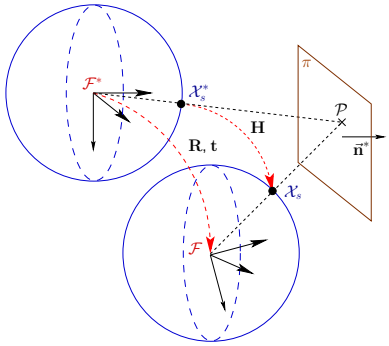

Figure 2: Homography between points on the sphere

3.2 Minimisation problem

Let $\mathscr{I}^*$ be the reference image. We will call reference template j, a region of size q_j of $\mathscr{I}^*$ corresponding to the projection of a planar region j of the scene.

A homography is a projective transformation and is thus valid for all central catadioptric devices (this is an exact projective property and not an approximation). This can be shown by considering two points on a plane related by a planar homography $\mathbf{H}$: $\mathscr{X} = \mathbf{H}\mathscr{X}^*$. Projecting these points to the unit sphere leads to : $\rho\mathscr{X}_s = \rho^*\mathbf{H}\mathscr{X}_s^*$.

The following scheme, illustrated by figure 2, shows the relationship between image points in two different frames :

$$\begin{array}{ccccccc} & & \mathbf{p}^* & \xrightarrow{k^{-1}} & \mathbf{m} & \xrightarrow{\hbar^{-1}} & \mathscr{X}_s^* \\ & & & & & & \downarrow \mathbf{H} \\ \mathbf{p} & \xleftarrow{k} & \mathbf{m} & \xleftarrow{\hbar} & \mathscr{X}_s & \xleftarrow{\frac{\mathscr{X}}{\|\mathscr{X}\|}} & \mathscr{X} \end{array} \tag{5}$$

Let $\mathbf{w}$ be the application of $\mathbf{H}$ followed by the normalisation and let $\Pi = k \circ \hbar$ be the

transformation between the sphere and the image plane : $\mathbf{p} = \Pi(\mathscr{X}_s)$.

To track the template j in the current image $\mathscr{I}$ is to find the transformation $\mathbf{H}(\overline{\mathbf{T}}, \overline{\mathbf{n}}_d^j)$ that warps the lifting of that region to the lifting of the reference template of $\mathscr{I}^*$:

$$\forall i, j \ : \mathscr{I}\left(\Pi(\mathbf{w}<\mathbf{H}(\overline{\mathbf{T}}, \overline{\mathbf{n}}_d^j)><\mathscr{X}_s^{ij*}>)\right) = \mathscr{I}^*(\mathbf{p}_{ij}) \tag{6}$$

In other words, knowing an approximation $\widehat{\mathbf{T}}$ of $\overline{\mathbf{T}}$ and $\widehat{\mathbf{n}}_d^j$ of $\overline{\mathbf{n}}_d^j$, the problem is to find the incremental transformation $\mathbf{T}(\mathbf{x})$ and $\mathbf{n}_d^j(\mathbf{x})$ that minimises the sum of squared differences (SSD) over all the pixels and over the m planes ($\mathbf{x}$ contains the 6 transformation parameters and the $3 \times m$ parameters for the normals and depths) :

$$\begin{cases} F(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{q_j} \|\mathbf{f}_{ij}\|^2 \\ \mathbf{f}_{ij} = \mathscr{I}\left(\Pi(\mathbf{w}<\mathbf{H}(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \widehat{\mathbf{n}}_d^j + \mathbf{n}_d^j(\mathbf{x}))><\mathscr{X}_s^{ij*}>)\right) - \mathscr{I}^*(\mathbf{p}_{ij}) \end{cases} \tag{7}$$

The minimal number of parameters in equation (7) is in fact $6+3\times m-1$ because the first homography has only 8 degrees of freedom. However the extra degree of freedom empirically gave better results probably due to the better conditioning of the Jacobian (all the values for the normals have the same amplitude).

3.3 Obtaining fast convergence

Equation (7) can be solved using a first-order approach [17, 13, 20, 1], however Benhimane and Malis showed in [5] that the cost function can be written as :

$$F(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{0}) + \frac{\mathbf{J}(\mathbf{0}) + \mathbf{J}(\mathbf{x})}{2}\mathbf{x}\right\|^2 + \mathscr{R}(\|\mathbf{x}\|^3) \tag{8}$$

which leads to the following second-order local minimiser (with $^+$ indicating the pseudo-inverse) :

$$\widehat{\mathbf{x}} = -\left(\frac{\mathbf{J}(\mathbf{0}) + \mathbf{J}(\mathbf{x}_0)}{2}\right)^+ \mathbf{f}(\mathbf{0}) \tag{9}$$

This does not however solve the problem as $\mathbf{J}(\mathbf{x}_0)$ generally depends on $\mathbf{H}(\overline{\mathbf{T}}, \overline{\mathbf{n}}_d^j)$ (or equivalently on the unknown $\mathbf{x}_0$). We will now show how we can avoid this dependency and calculate the second-order increment for central catadioptric sensors.

The Jacobians $\mathbf{J}(\mathbf{0})$ and $\mathbf{J}(\mathbf{x}_0)$, that correspond respectively to the current and the reference Jacobians, can be written as :

$$\mathbf{J}(\mathbf{0}) = \mathbf{J}_{\mathscr{I}}\,\mathbf{J}_{\Pi}\,[\mathbf{J}_{H_T}\mathbf{J}_T(\mathbf{0}) \quad \mathbf{J}_{H_n}\mathbf{J}_n(\mathbf{0})] \tag{10}$$

$$\mathbf{J}(\mathbf{x}_0) = \mathbf{J}_{\mathscr{I}^*}\,\mathbf{J}_{\Pi}\,[\mathbf{J}_{H_T^*}\mathbf{J}_{T^*}(\mathbf{x}_0) \quad \mathbf{J}_{H_n^*}\mathbf{J}_{n^*}(\mathbf{x}_0)] \tag{11}$$

with H_T the homography seen as a function of the transformation $\mathbf{T}$ and H_n the homography seen as a function of the plane normal $\mathbf{n}_d$.

Despite the Jacobian $\mathbf{J}(\mathbf{x}_0)$ generally depending on the unknown $\mathbf{x}_0$, thanks to the left-invariance property of the vector fields on $\mathbb{SE}(3)$, the following identity can be proven :

$$\mathbf{J}_{T^*}(\mathbf{x}_0)\mathbf{x}_0 = \mathbf{J}_T(\mathbf{0})\mathbf{x}_0 \tag{12}$$

we also have : $\mathbf{J}_{n^*}(\mathbf{x}_0)\mathbf{x}_0 = \mathbf{J}_n(\mathbf{0})\mathbf{x}_0$.

If we assume that $\mathbf{H}(\overline{\mathbf{T}},\overline{\mathbf{n}}) \approx \mathbf{H}(\widehat{\mathbf{T}},\widehat{\mathbf{n}})$, in equation (9), we can use $\mathbf{J}(\mathbf{0})\mathbf{x}_0$ instead of $\mathbf{J}(\mathbf{x}_0)\mathbf{x}_0$. The update $\widehat{\mathbf{x}}$ of the second-order minimisation algorithm can then be computed as follows :

$$\widehat{\mathbf{x}} = -\left(\left(\frac{\mathbf{J}_{\mathscr{I}} + \mathbf{J}_{\mathscr{I}^*}}{2}\right)\mathbf{J}_\Pi \begin{bmatrix}\mathbf{J}_{H_T T}(\mathbf{0}) & \mathbf{J}_{H_n n}(\mathbf{0})\end{bmatrix}\right)^+ \mathbf{f}(\mathbf{0}) \tag{13}$$

The computational complexity is almost the same as for a first-order algorithm because the reference Jacobian $\mathbf{J}_{\mathscr{I}^*}$ only needs to be calculated once and then added to the current jacobian (the division by 2 can be taken out of the pseudo-inverse).

It is important to note that $\mathbf{J}_{\mathscr{I}^*}$ and $\mathbf{J}_{\mathscr{I}}$ are the Jacobians taken in the images (that can be approximated with for example a Sobel filter). The non-linear properties of the central catadioptric sensors are taken into account with the Jacobian of the projection function $\mathbf{J}_\Pi$. We thus avoid transforming the image into a perspective view.

We may also note that the matrix is sparse so the algorithm can make the most of sparse linear algebra and avoid the full inversion of the Jacobian matrix [14].

In comparison, the inverse compositional [1] update would be ($\mathbf{J}_{\mathscr{I}^*}$ is constant) :

$$\widehat{\mathbf{x}} = -\left(\mathbf{J}_{\mathscr{I}^*}\mathbf{J}_\Pi \begin{bmatrix}\mathbf{J}_{H_T T}(\mathbf{0}) & \mathbf{J}_{H_n n}(\mathbf{0})\end{bmatrix}\right)^+ \mathbf{f}(\mathbf{0}) \tag{14}$$

with the pseudo-inverse calculated *once and for all*. The normals and depths are estimated at each step so $\mathbf{J}_{H_n n}$ is *not constant* and this update step is incorrect (this is explained more in detail in [3]). If we re-calculate the pseudo-inverse, the algorithm becomes correct but loses its advantage of low computational cost. The proposed algorithm is then more adequate : for a low computational overhead (the computation of $\mathbf{J}_{\mathscr{I}}$) it offers a close to second-order convergence rate and a larger convergence domain.

4 Experimental results : constrained and unconstrained tracking and comparison of different minimisation approaches

For the experiments, we used a central catadioptric camera[1] comprised of a S80 parabolic mirror from RemoteReality with a telecentric lens and perspective camera with an image resolution of 1024×768. The camera was mounted on a mobile robot with precise odometry that we will consider as ground truth.

Figure 3 shows the templates tracked in the experiments. They are numbered in counterclockwise order from 1 to 3 starting top left. To fix the scale factor, we measured the distance from the camera to the third plane (0.5 m) (the plane that proved the stablest while tracking). The sizes of the rectangular patches englobing the selected regions from 1 to 3 where respectively 94×69, 106×93 and 102×69.

The sequence is composed of 120 images. The mobile robot covered a distance of about 2 m. The initial values given for the normals with depths was $[1;0;0]$ (the same results were obtained for values $[0;1;0]$, $[0;0;1]$, $[0;0;1000]$ ...). These initial values are far from the "real" values that can be deduced from Fig. 11 and Fig. 12 (with $d_3 = 0.5$ m) :

[1] the camera was calibrated using a open-source toolbox available on http://www-sop.inria.fr/icare/personnel/Christopher.Mei/Toolbox.html

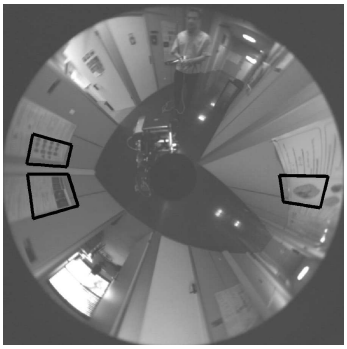

Figure 3: Tracked templates

$[-0.38;-0.31;0]$, $[-0.4;0.6;0]$ and $[1.2;-1.6;0]$. The algorithm proved to be relatively insensitive to the initial values when an extra degree of freedom was given (as explained in Section 3.2). This can be explained from the normals appearing in the homography as a product with the translation. A more detailed study would be needed to determine the region of convergence.

The motion for the planes tracked independently was obtained by applying the median over the rotation and translation recovered from the homographies, we will call this algorithm **SPT**.

The algorithm for the constrained tracking was tested with a forward compositional minimisation (**MPT_FC**) and with the proposed algorithm (**MPT_ESM**). The inverse compositional was also tested but failed after two iterations (which was to be expected as the initial values for the normals are far from correct). Applying only one update step as appears in the Kalman filter [16] also failed after two iterations even with the correct normals.

Figures 4 to 9 compare the odometry (in full lines) to the motion estimation using the **SPT**, **MPT_FC** and **MPT_ESM** algorithms (lines with symbols). The number of iterations needed to converge appears on the figure for the **MPT_FC** and **MPT_ESM** algorithms in a black dotted line with the number of iterations indicated on the right Y-axis. Figure 10 shows the templates at different stages in the tracking (only for **MPT_ESM**). The normals estimated on-line are represented in Figure 11. The distances estimated for planes 1 and 2 are detailed in Figure 12.

MPT_ESM which is a close to second-order approach gave slightly more precise results than **MPT_FC** (first-order approach) and in less iterations : 7 iterations were needed for **MPT_ESM** compared to 13 for **MPT_FC** (median value over the first 60 images). We will now compare **MPT_ESM** to **SPT**.

The motion estimation was precise except between iterations 75 and 100 where a reflection (that can be seen on the tracked templates in Fig. 10) on the poster of template 3 generated errors in the normal estimates but also in the distance estimates. We used a non-robust minimisation approach which is not able to cope adequately with illumination errors. However the tracking was able to recover after iteration 110. When the templates where tracked independently (i.e. the motion and normal estimates were extracted directly from the homography), the patches did not give a sufficient amount of information

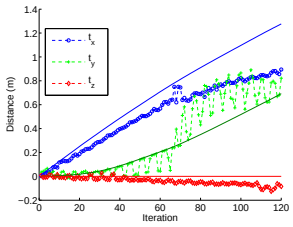

Figure 4: Estimation of the robot's translation (**SPT**)

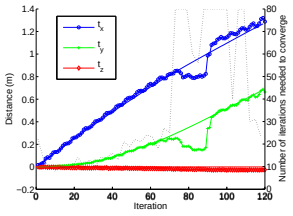

Figure 5: Estimation of the robot's translation (**MPT_FC**)

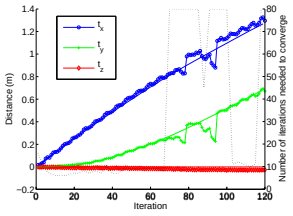

Figure 6: Estimation of the robot's translation (**MPT_ESM**)

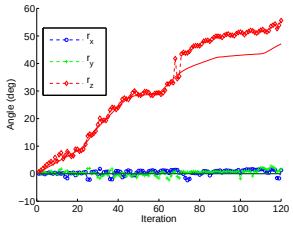

Figure 7: Estimation of the robot's rotation (**SPT**)

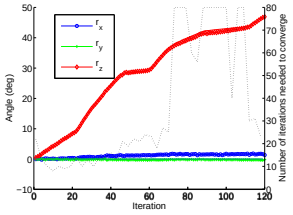

Figure 8: Estimation of the robot's rotation (**MPT_FC**)

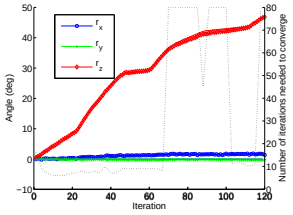

Figure 9: Estimation of the robot's rotation (**MPT_ESM**)

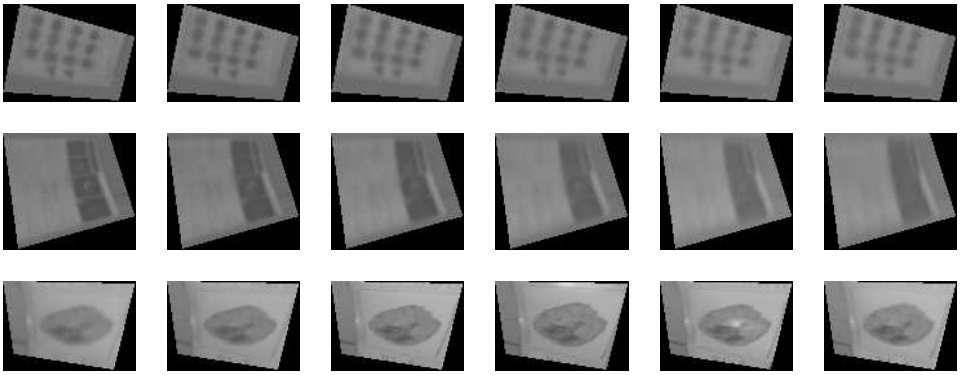

Figure 10: Reprojection of the templates for iterations 0,25,50,75,100,120 in the reference image using the estimated homography (**MPT_ESM**)

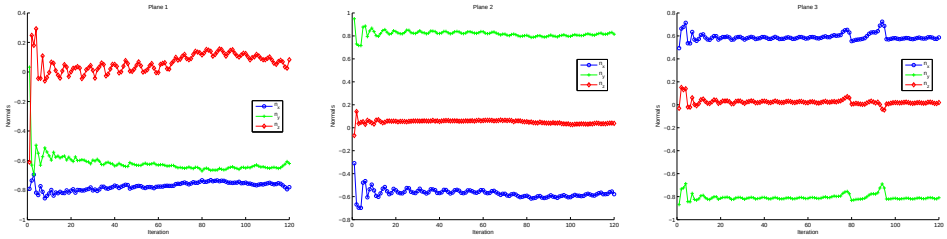

Figure 11: Normals estimated for planes 1 to 3 (**MPT_ESM**)

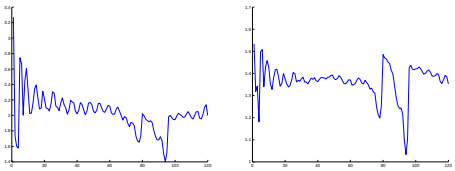

Figure 12: Estimation of the plane distances for planes 1 and 2 (**MPT_ESM**)

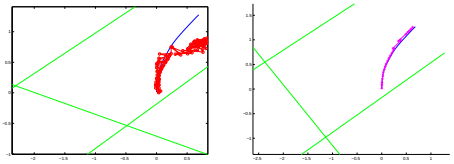

Figure 13: Robot's motion in the XY-plane for **SPT** and **MPT_ESM**

to enable a good estimate of the motion. With the illumination problem arising on the *stablest* estimated plane, the motion estimation becomes erratic (Fig. 4 and Fig. 7).

The **MPT_ESM** gave a translation estimate with a maximum error of [16,23,3] cm for [x,y,z] and an absolute mean error of [2.6, 2.0, 1.6] cm, for the rotation the maximum error was of [1.85, 0.43, 0.64] deg over the [x,y,z] rotation axis with a mean error of [1.14, 0.22, 0.22] deg. Estimating the distance proved sensitive to small errors, the variance over the sequence was respectively of $\sigma = 23.6$ cm and $\sigma = 7.33$ cm for planes 1 and 2.

Figure 13 shows the motion of the robot in the XY-plane for **MPT_ESM** with the odometry depicted in full lines and the estimated motion with connected crosses. The planes are also represented in the image from the estimates. The angle between the corridor walls were quite precisely estimated with 92.9 deg between planes 1 and 2 and 87.3 deg between planes 1 and 3. The results obtained using **SPT** depicted using connected circles in Fig. 13 did not give a satisfying estimate.

5 Conclusion

We have presented in this article an approach to tracking multiple templates to estimate the motion of the camera but also the normals and depths of the tracked planes. Through the unified projection model, we obtain a generic algorithm that also applies to the motion estimation of perspective cameras (and fish-eye lenses under certain conditions). This algorithm takes into account the non-uniform resolution and distortion of the sensor and avoids unwarping the images to perspective views. We have shown that linking the motion improves the quality of the estimates. Further improvements could be made to make the algorithm robust to illumination changes and partial occlusion [13]. Knowing when to estimate the normals would also improve the quality and robustness of the tracking.

References

[1] S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. In *CVPR*, pages 1090–1097, 2001. 2, 5, 6

[2] S. Baker and S. Nayar. A theory of catadioptric image formation. In *ICCV*, pages 422–427, 2001. 1

[3] S. Baker, R. Patil, K.M. Cheung, and I. Matthews. Lucas-kanade 20 years on: Part 5. Technical report, Robotics Institute, Carnegie Mellon University, 2004. 6

[4] Joao P. Barreto and Helder Araujo. Issues on the geometry of central catadioptric image formation. In *CVPR*, volume 2, pages 422–427, 2001. 1, 3

[5] S. Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In *IROS*, 2004. 2, 5

[6] J. M. Buenaposada and L. Baumela. Real-time tracking and estimation of planar pose. In *ICPR*, pages 697–700, 2002. 2

[7] Dana Cobzas and Peter Sturm. 3d ssd tracking with estimated 3d planes. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision, Victoria, Canada*, 2005. 2

[8] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 2(3):485–508, 1988.

[9] Locally Planar Patch Features for Real-Time Structure from Motion. Nicholas molton, andrew j. davison and ian reid. In *BMVC*, 2004. 2

[10] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *ECCV*, pages 445–461, 2000. 1, 3

[11] C. Geyer and K. Daniilidis. Mirrors in motion: Epipolar geometry and motion estimation. In *International Journal on Computer Vision*, pages 766–773, 2003. 1

[12] H. Hadj-Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet. 2 1/2 d visual servoing with central catadioptric cameras. In *IROS*, 2005. 1, 2

[13] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 1998. 2, 5, 9

[14] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition edition, 2003. 6

[15] F. Martin J. Barreto and R. Horaud. Visual servoing/tracking using central catadioptric cameras. In *Int. Symposium on Experimental Robotics*, Advanced Robotics Series, 2002. 1

[16] Hailin Jin, Paolo Favaro, and Stefano Soatto. A semi-direct approach to structure from motion. *The Visual Computer*, 19(6):377–394, 2003. 2, 7

[17] B. Lucas and T. Kanade. An iterative image registration technique with application to stereo vision. In *Int. Joint Conference on Artificial Intelligence*, pages 674–679, 1981. 2, 5

[18] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *IEEE Int. Conference on Robotics and Automation*, 2004. 2

[19] C. Mei, S. Benhimane, E. Malis, and P. Rives. Homography-based tracking for central catadioptric cameras. In *IROS*, October 2006. 1

[20] H. Y. Shum and R. Szeliski. Construction of panoramic image mosaics with global and local alignment. *International Journal on Computer Vision*, 16(1):63–84, 2000. 2, 5

[21] T. Svoboda and Pajdla. T. Epipolar geometry for central catadioptric cameras. *International Journal on Computer Vision*, 49(1):23–37, August 2002. 1

[22] Xianghua Ying and Zhanyi Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model? In *ECCV*, 2004. 1