The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Self-location from monocular uncalibrated vision using reference omniviews

L. Puig, J.J. Guerrero

Abstract— **In this paper we present a novel approach to perform indoor self-localization using reference omnidirectional images. We only need one omnidirectional image of the whole scene stored in the robot memory and a conventional uncalibrated on-board camera. We match the omnidirectional image and the conventional images captured by the on-board camera and compute the hybrid epipolar geometry using lifted coordinates and robust techniques. We map the epipole in the reference omnidirectional image to a ground plane through a homography in lifted coordinates also, giving the position of the robot in the planar ground, and its uncertainty. We perform experiments with simulated and real data to show the feasibility of this new self-localization approach.**

I. INTRODUCTION

In recent years the use of catadioptric omnidirectional systems in mobile robot localization and navigation tasks and in visual surveillance applications has increased considerably. The main advantages of such systems is their wide field of view and the central single view point property. These characteristics allow overcoming the visibility constraint and help the geometrical analysis of the information captured by catadioptric cameras. Many authors have studied such systems, from their geometric properties [1] to the relation between two or more images [2], [3]. Others deal with the problem of mixing such systems with perspective cameras and establish geometrical relations such as hybrid fundamental matrices and tensors [4], [5], [6] to help the image matching. A comparative between some of these methods using lifted coordinates is presented in [7]. Recently Sturm and Barreto [8] developed a complete model for all central catadioptric systems.

Catadioptric systems are usually used in tasks of robot navigation and self-localization. Ishiguro and Tsuji [9] describe a method for robot localization from memorized omnidirectional views, which are stored using Fourier coefficients; similarly, Pajdla and Hlaváč [10] use the image phase of a panoramic view for robot localization. Cauchois et al. [11] present a method for robot localization by correlating real and synthesized omnidirectional images, but they can only handle small viewpoint changes. Matsumoto et al. [12] present a similar method based on simply comparing cylindrical gray-level images. In [13] they propose their own representation which is called *fast wide baseline feature matching* and the use of a topological map instead of a metric one. In [14] a new method for navigation and localization using a collection of 1D omnidirectional images formed by averaging the center scanlines of a cylindrical view is proposed. More recently some authors assume they have a sequence of omnidirectional images stored in memory, [15], [16] from which they find the closest to the current frame. In [15] a fisheye lens is used. The essential matrix is computed from the current frame and the next in the visual memory, which is used to feed a control law to guide the robot to the next position. In [16] a catadioptric system is used to perform a hierarchical localization. They go from topological localization using features matching to metric localization using the trifocal tensor.

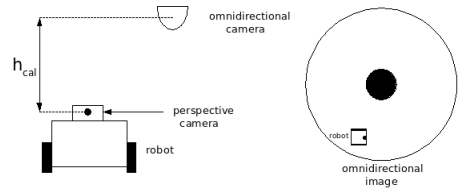


Fig. 1. Scenario where an omnidirectional image is taken from a camera on the ceiling and a conventional camera on the robot is used for self-localization.

As we have seen many authors try to deal with the problem of navigation and robot localization using omnidirectional images acquired by dioptric (fisheye) or catadioptric system mounted on the robot. Many times this type of sensors are expensive and not easily available as perspective cameras. Some other times the omnidirectional sensors are in a fixed position (surveillance) or they are used to explore a certain environment as in the Street View System[1]. In these cases we could use a different type of sensor to acquire the current image and to perform a matching between them. A similar scenario is depicted in Fig. 1 where, a *virtual* surveillance omnidirectional camera is located on the ceiling and a mobile robot, with a perspective camera on-board, share the field of view (FOV). We say *virtual* because the omnidirectional images of the scene are stored in the robot memory and there is neither required of a physical omnidirectional camera nor the communication with the robot. In this work we propose to use a combination of omnidirectional and perspective images to perform the robot localization in a scene. To perform the self-localization the robot matches the current perspective image acquired by the perspective camera with the omnidirectional image of the scene stored in memory. Once we have enough correspondences we compute the epipolar geometry encapsulated in the hybrid fundamental matrix. From this fundamental matrix we compute the omnidirectional epipole

L. Puig and J.J. Guerrero are with the DIIS-I3A, Universidad de Zaragoza, Zaragoza, Spain {lpuig, jguerrer}@unizar.es

This research has been funded by the Spanish projects VISPA DPI2009-14664 and UZ2007-TEC05.

[1] http://maps.google.com

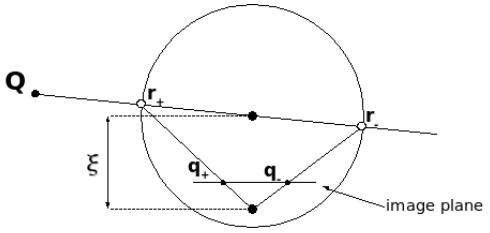


Fig. 2. Projection of a 3D point using the sphere camera model.

representing the perspective camera in the scene, in this case the robot. Once the robot is located in the omnidirectional image we can obtain some geometric information using the hybrid homography. This homography relates the position of the robot in the omnidirectional image with the ground plane giving us the relative position of the robot in meters with respect to the origin of the ground plane. We also studied the uncertainty of the location of a point in the ground plane adapting the methodology developed for conventional cameras [17] to omnidirectional ones.

The rest of the paper is organized as follows. In Section II, we present the generic camera model and the lifted coordinates used. Section III explains the method used to compute the position of the robot on a ground plane using a hybrid homography. Section IV explains the method used to compute the hybrid fundamental matrix and how to extract from it the position of the robot in the image. In Section V some experiments with synthetic and real images are presented. Finally conclusions are stated in Section VI.

II. Generic Camera Model

In this section we introduce briefly the sphere camera model which can explain all central catadioptric systems and the perspective cameras with radial distortion. The earliest work was developed by Geyer and Daniilidis [18] where they show the equivalence of the image geometries obtained by the catadioptric projection and the composition of projections of a sphere. In [5] Barreto and Daniilidis extend the model to consider perspective cameras with distortion. Recently in [8], Sturm and Barreto propose the generic projection model which considers any central catadioptric system. This model assumes all central catadioptric cameras can be modeled by a unit sphere and a perspective camera, such that the projection of 3D points can be performed in two steps (Fig. 2). First, one projects the scene onto the sphere, obtaining the intersection of the unit sphere and the line joining its center and a 3D point. There are two such intersection points, $\mathbf{r}_+$ and $\mathbf{r}_-$. These points are then projected into a perspective image, resulting in two image points, $\mathbf{q}_+$ and $\mathbf{q}_-$, one of which is physically true. This model covers all central catadioptric cameras, encoded by ξ, which is the distance between the center of the second projection and the center of the sphere. If $0 < \xi < 1$ we have a hyper-catadioptric system, $\xi = 1$ a para-catadioptric and $\xi = 0$ the classical pin-hole model.

The two intersection points of the sphere $r_\pm$ and the line joining its center and the 3D point $\mathbf{Q}$ are $(Q_1, Q_2, Q_3 \pm \sqrt{Q_1^2, Q_2^2, Q_3^2})^\mathsf{T}$. Their images in the perspective camera are

$$\mathbf{q}_\pm \sim \mathsf{K}\mathbf{r}_\pm \sim \mathsf{K} \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \pm \xi\sqrt{Q_1^2, Q_2^2, Q_3^2} \end{pmatrix} \quad (1)$$

where K represents the intrinsic and extrinsic parameters of the perspective camera. A more detailed explanation can be found in [8].

A. Lifted coordinates

The derivation of (multi-) linear relations for catadioptric imagery requires the use of lifted coordinates. They allow to generalize the transformations and multiview tensors from conventional perspective images to catadioptric systems, where the projective invariant entities are conics instead of lines. The Veronese map $V_{n,d}$ of degree d maps points of $\wp^n$ into points of an m dimensional projective space $\wp^m$, with $m = \begin{pmatrix} n+d \\ d \end{pmatrix} - 1$. Consider the second order Veronese map $V_{2,2}$, that embeds the projective plane into the 5D projective space, by lifting the coordinates of point $\mathbf{q} = (q_1, q_2, q_3)$ to

$$\hat{\mathbf{q}} = (q_1^2, q_1q_2, q_2^2, q_1q_3, q_2q_3, q_3^2) \quad (2)$$

III. From Omniview to Ground Location

In this section we explain the process to map a point in the omnidirectional image to a position in the ground plane. We assume a planar robot motion. To perform this task we calibrate the omnidirectional image with respect to the ground plane using a homography with lifted coordinates. We have tested two models to compute the hybrid homography. The generic 6×6 homography model H_{cata} and a simplified 3×4 model H_{sim}.

A. Generic Homography

This homography maps a point in the omnidirectional image in lifted coordinates $\hat{\mathbf{q}}_c$ to a degenerate dual conic containing the point in the ground plane $\Omega \sim \mathbf{q}_{p+}\mathbf{q}_{p-}^\mathsf{T} + \mathbf{q}_{p-}\mathbf{q}_{p+}^\mathsf{T}$.

$$\Omega = \mathsf{H}_{cata}\ \hat{\mathbf{q}}_c \quad (3)$$

To compute this homography we use a DLT-like (Direct Linear Transformation) approach. As in the perspective case we need correspondences $\mathbf{q}_p^i \leftrightarrow \mathbf{q}_c^i$ between points in the ground plane $\mathbf{q}_p^i$ and omnidirectional image points $\mathbf{q}_c^i$. From Eq. 3 we obtain

$$\widehat{[\mathbf{q}_c]}_\times\ \mathsf{H}_{cata}\ \hat{\mathbf{q}}_p = \mathbf{0} \quad (4)$$

If the j-th row of the matrix H_{cata} is denoted by $\mathbf{h}_j^\mathsf{T}$ and arranging Eq. 4 we have

$$\widehat{[\mathbf{q}_c]}_\times \otimes \hat{\mathbf{q}}_p \begin{pmatrix} \mathbf{h}_1^\mathsf{T} \\ \mathbf{h}_2^\mathsf{T} \\ \mathbf{h}_3^\mathsf{T} \\ \mathbf{h}_4^\mathsf{T} \\ \mathbf{h}_5^\mathsf{T} \\ \mathbf{h}_6^\mathsf{T} \end{pmatrix} = \mathbf{0} \quad (5)$$

These equations have the form $\mathsf{A}^i\mathbf{h} = \mathbf{0}$, where A^i is a 36×6 matrix, and $\mathbf{h} = (\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \mathbf{h}_3^\mathsf{T}, \mathbf{h}_4^\mathsf{T}, \mathbf{h}_5^\mathsf{T}, \mathbf{h}_6^\mathsf{T})^\mathsf{T}$ is a 36-vector made up of the entries of matrix H_{cata}. The matrix A^i has the form

$$\mathsf{A}^i = \begin{pmatrix} 0 & 0 & q_3^2\hat{\mathbf{q}}_p & \cdots \\ 0 & -q_3^2\hat{\mathbf{q}}_p & 0 & \cdots \\ q_3^2\hat{\mathbf{q}}_p & 0 & 0 & \cdots \\ 0 & q_3q_2\hat{\mathbf{q}}_p & -q_3q_1\hat{\mathbf{q}}_p & \cdots \\ -q_3q_2\hat{\mathbf{q}}_p & q_3q_1\hat{\mathbf{q}}_p & 0 & \cdots \\ q_2^2\hat{\mathbf{q}}_p & -2q_2q_1\hat{\mathbf{q}}_p & q_1^2\hat{\mathbf{q}}_p & \cdots \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} 0 & -2q_3q_2\hat{\mathbf{q}}_p & q_2^2\hat{\mathbf{q}}_p \\ q_3q_2\hat{\mathbf{q}}_p & q_3q_1\hat{\mathbf{q}}_p & -q_2q_1\hat{\mathbf{q}}_p \\ -2q_3q_1\hat{\mathbf{q}}_p & 0 & q_1^2\hat{\mathbf{q}}_p \\ -q_2^2\hat{\mathbf{q}}_p & q_2q_1\hat{\mathbf{q}}_p & 0 \\ q_2q_1\hat{\mathbf{q}}_p & -q_1^2\hat{\mathbf{q}}_p & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and is rank 3, so each correspondence gives 3 equations. Thus we need at least 12 correspondences to compute H_{cata}.

To get the two points encapsulated in the conic Ω we have to verify the equation $(q_1, q_2, 0)\Omega(q_1, q_2, 0)^\mathsf{T} = 0$ and solve the resultant quadratic equation for q_1, q_2.

B. Simplified Homography

The second approach is derived from the hybrid epipolar geometry approach used in [7]. In this work, a 3×4 hybrid fundamental matrix was tested which is the theoretical model for a para-catadioptric systems, being a good approximation for other central catadioptric systems if the camera has square pixels. From this example we compute a 3×4 homography H_{sim} which maps a 4-vector lifted coordinates of a point in the omnidirectional image $\hat{\mathbf{q}}_c$ to a point in the ground plane $\mathbf{q}_p$ in homogeneous coordinates.

$$\mathbf{q}_p = \mathsf{H}_{sim}\hat{\mathbf{q}}_c \quad (7)$$

The lifting from 3-vector to 4-vector is $\hat{\mathbf{q}}_c \sim (q_1^2 + q_2^2, q_1q_3, q_2q_3, q_3^2)^\mathsf{T}$ [4]. To compute this model we need at least 6 correspondences since each correspondence gives two equations and in H_{sim} we have 11 degrees of freedom.

C. Uncertainty

This homography transformation produces a non-homogeneous uncertainty distribution. So in order to have an estimation of the uncertainty of the robot location we must consider the uncertainty transformation from the image to the ground plane. We adapt the approach proposed by [17] to omnidirectional images using lifted coordinates. We assume uncertain image points $\mathbf{q}$ with $\sigma_{q_1} = \sigma_{q_2} = \sigma$ the covariance in the image coordinates and we consider an exact H_{sim}.

As we are dealing with lifted coordinates in the image plane we use the Jacobian J of this lifting to translate the error into the lifted coordinates

$$\Lambda_{\hat{\mathbf{q}}} = \mathsf{J}\Sigma\mathsf{J}^\mathsf{T} \quad (8)$$

Now we propagate the error from the points in the image to the points in the plane by

$$\Lambda_{\mathbf{Q}} = \mathsf{H}_{sim}\Lambda_{\hat{\mathbf{q}}}{\mathsf{H}_{sim}}^\mathsf{T} \quad (9)$$

As result we have a covariance matrix for homogeneous coordinates of a point in the plane $\mathbf{Q} = (Q_1, Q_2, Q_3)$ which have to be converted into a 2×2 covariance matrix $\Lambda_{\mathbf{Q}}^{2\times2}$, which is performed as

$$\Lambda_{\mathbf{Q}}^{2\times2} = \nabla f \Lambda_{\mathbf{Q}} \nabla f^\mathsf{T} \quad (10)$$

where

$$\nabla f = 1/Q_3^2 \begin{pmatrix} Q_3 & 0 & -Q_1 \\ 0 & Q_3 & -Q_2 \end{pmatrix} \quad (11)$$

IV. Robot location in the omnidirectional image

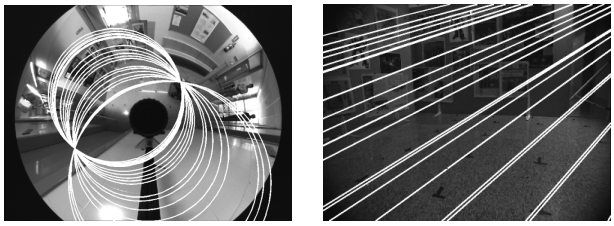
As it is well known, the epipoles are the location of the camera center in one view seeing by the other view. Extracting the epipole in the omnidirectional image we can locate the robot camera in a wide field of view. When the epipolar geometry is encapsulated by the fundamental matrix the computation of epipoles is reduced to extract the null vectors of such matrix. This problem increase its complexity when omnidirectional images are used, the null vector becomes a null space. Before computing the epipoles we have to compute the fundamental matrix between an omnidirectional image and a perspective one. In this work we use the 4×3 fundamental matrix F43 analyzed in [7]. This matrix is a useful simplification of the complete model, the 6×6 matrix F66. We choose the F43 matrix because as it is shown in [7], it requires less correspondences than the F66 and its performance is good enough with the mirror shape we use in this work. The following constraint describes the hybrid fundamental matrix:

$$\hat{\mathbf{q}}_c^\mathsf{T}\mathsf{F}_{cp}\mathbf{q}_{per} = 0 \quad (12)$$

where $\hat{\mathbf{q}}_c$ is the 4-vector lifted coordinates of a point in the catadioptric image and $\mathbf{q}_{per}$ represents a point in the perspective image in homogeneous coordinates.

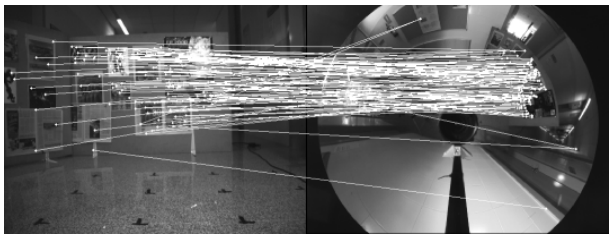
The automatic computing of the hybrid fundamental matrix can be found in [7]. In our current solution we avoid to do an unwarping in polar coordinates and use directly SIFT [19] points in the omnidirectional image. In this case a flip transformation is required to correct the reflexion transformation made by the mirror in the omnidirectional image. This transformation allows to match directly the SIFT points extracted from both images, and without this flip transformation the matching using SIFT points does not produce coherent results. Note that the SIFT descriptor is rotation and scale invariant but it is not projective invariant.

An example of the epipolar conics and lines can be seen in Fig. 3. The automatic matching before and after the robust hybrid fundamental matrix computation can be seen in Fig. 4(a) and Fig. 4(b).

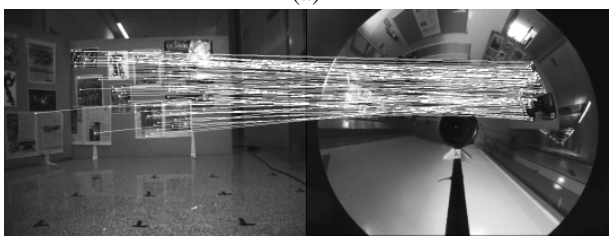


(a) (b)

Fig. 3. Example of epipolar conics (a) and epipolar lines (b) computed from the hybrid fundamental matrix F43.



(a)



(b)

Fig. 4. (a) Putative matching using just SIFT. (b) Matching using the hybrid fundamental matrix.

A. Computing the epipoles

The hybrid fundamental matrix F43 has a one-dimensional left null-space and the two epipoles of the catadioptric camera (Fig. 3(a)) are the left null-vectors that are valid lifted coordinates. This means that this 4-vectors must fit the following quadratic constraint:

$$\hat{\mathbf{q}}_c \sim \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix} \Leftrightarrow \hat{q}_{c1}\hat{q}_{c4} - \hat{q}_{c2}^2 - \hat{q}_{c3}^2 = 0 \qquad (13)$$

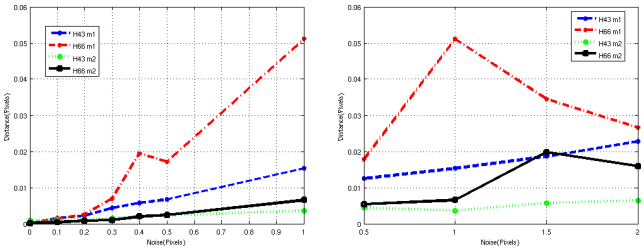
Once we obtain the 4-vectors which fit this constraint we use the same process explained in Section III-A to compute the two solutions for the epipole.

V. EXPERIMENTS

First we analyze the behavior of the two approaches to compute the homography described in section III. Once we define the model to compute the homography we perform experiments of the whole system with synthetical data to analyze its accuracy. Finally, to confirm the results we perform some experiments with real images going from a point A to a point B inside the same room with just one omnidirectional image stored in the robot memory and a perspective uncalibrated camera installed in the robot.

A. Generic Model (H_{cata}) vs. Simplified Model (H_{sim})

In this experiment we use a simulator which generates omnidirectional images coming from two different catadioptric systems using a hyperbolic mirror with $\xi = 0.9662$ (m1) and $\xi = 0.7054$ (m2) using the sphere model [5]. We take the parameters from two real hyperbolic mirrors designed by Neovision[2] and Accowle[3], respectively. The pattern to compute the homography is composed of a squared plane with 11×11 points and a distance between points of 40cm. The goal of the first experiment is to know the behavior of the two homography approaches in presence of noise. We add different amounts of Gaussian noise described by its standard deviation (σ) to the coordinates of the points in the omnidirectional image. The DLT algorithm is used to compute the homographies. For every σ we repeat the experiment 10 times in order to avoid particular cases due to random noise. The mean of these iterations is shown in Fig. 5(a). When the amount of noise is low both models shows a good performance, but when the amount of noise increases the performance of H_{cata} decreases and H_{sim} remains. This result shows that H_{cata} is more sensitive to noise than H_{sim}, this can be caused by the overparameterization of the general model. We also observe that the error corresponding to the mirror m2 is smaller than the one obtained with the mirror m1, this is explained because the area occupied by the pattern in the omnidirectional image is bigger using the mirror m2 than using the mirror m1.

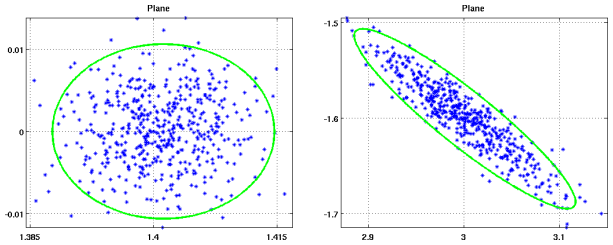


(a) (b)

Fig. 5. Comparison between the two approaches to compute the hybrid homography, $\mathsf{H}_{sim} = \mathsf{H}43$, $\mathsf{H}_{cata} = \mathsf{H}66$. (a) Performance with presence of noise. (b) Different heights test.

The goal of the following experiment is to show the behavior of the two approaches when the height between the omnidirectional camera and the plane pattern changes. We try 4 different heights, 0.5, 1.0, 1.5 and 2 meters. We also add $\sigma = 1$ pixel Gaussian noise to the image coordinates. The result of this experiment can be seen in Fig. 5(b). We observe that the error given by H_{sim} is smaller than the one coming from H_{cata} in both mirror cases. In the case of m1 this can be explained because this systems is close to a para-catadioptric system where H_{sim} is the exact model, but it also shows good results with a real hyperbolic mirror (m2). We also observe that the error using the mirror m2 is smaller than the one given by the mirror m1. With the last results we decided to choose H_{sim} to implement our system.

[2]http://www.neovision.cz
[3]http://www.accowle.com



(a) (b)

Fig. 6. Propagation error in the plane using H_{sim}. The theoretical ellipse of uncertainty is shown in green in (a) a point close to the center and (b) a point in the periphery. Units in the plane are meters.

TABLE I

ESTIMATIONS OF POSITION IN THE PLANE USING THE LIFTED HOMOGRAPHY

	Real position	Estimated H50	Estimated H100
Point 1	(154,-114)	(155.92,-112.49)	(154.90,-110.92)
Point 2	(260,108)	(266.07,106.22)	(259.70,106.39)
Point 3	(0.0,-471)	(0.02,-469.36)	(0.01,-467.62)

B. Uncertainty

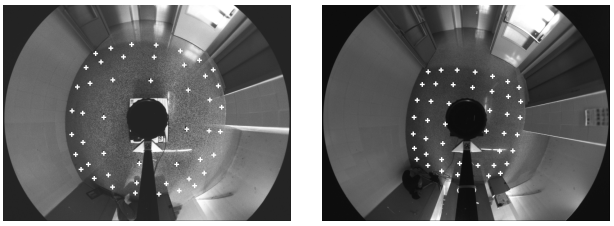
In this experiment we show the error propagation through H_{sim}. We select a point in the grid used to compute the homography. This point is used as the center for a Gaussian distribution with $\sigma = 1$ pixel. In Fig 6 we observe the error propagation in the plane corresponding to two different points in the omnidirectional image, one close to the center Fig. 6(a) with an error of $\pm 1.35cm$ and the other one close to the periphery Fig 6(b) with an error of $\pm 14.66cm$. We observe that the error varies depending on the position in omnidirectional image. This is because of the perspective effect, since there is high uncertainty approaching the line (conic) of the horizon of the ground plane, which corresponds with far away observations.

C. Homography using real images

Here we test different heights for the omnidirectional camera checking for location accuracy. The plane used has 7×7 points and the distance between them is 40cm giving a pattern of 240×240cm. The omnidirectional camera is located at the center of the scene. In Fig. 7 we can see two different omnidirectional images obtained from two different heights, 50cm (H50) and 100cm (H100) respectively. We also show the reprojection of the points used to compute the homography. To test the performance of the homographies we choose three different points which do not belong to the pattern. The goal is to check the accuracy when the points are far from the image center. Table I shows the location results of this experiment in centimeter units. This inaccuracy is caused by the perspective effect and by the non-homogeneous resolution of the omnidirectional image, i.e. a few pixels in the periphery of the image means more distance than the same pixels close to the image center. This situation validates the uncertainty analysis presented in Section V-B.

D. Epipole Estimation

In this experiment we analyze different positions for the perspective camera in the scene and compute the epipole



(a) height = 50cm (b) height = 100cm

Fig. 7. Omnidirectional images of the pattern at two different heights with the reprojection of the pattern points

TABLE II

ERROR IN THE EPIPOLE ESTIMATION DEPENDING ON THE CFOV, DEFINED BY THE BASELINE AND THE # OF MATCHES.

CFOV (rads)	Baseline(m)	# of Matches	Error(pixels)
1.83	2.91	246	0.57
1.13	2.50	219	0.84
1.04	2.12	200	1.08
0.97	1.80	176	1.25
0.75	1.58	149	2.86
0.54	1.50	119	4.96
0.26	1.58	90	25.83

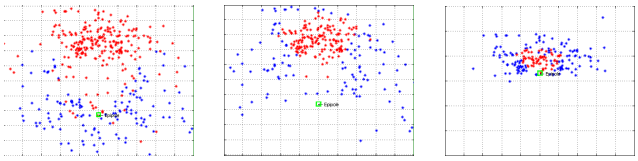


Fig. 8. Three different CFOVs, represented by the common observable points (red points) depending on the position of the perspective camera (green square) (See digital version).

in the omnidirectional image. We move the perspective camera along the $y-$axis at a height of 0.5m in the same x coordinate as the omnidirectional camera. Fig 8 presents three omnidirectional images showing the matched points in the scene (red points) being in the Common Field of View (CFOV) of both cameras. We naturally observe that moving the perspective camera decreases the CFOV causing a reduction in the number of matches and decreasing the accuracy of the epipole computed. We add Gaussian noise with $\sigma = 0.5$ to points in both images. Table II shows the estimation of the epipole location depending on the size of the CFOV. We observe that even when we have enough correspondences to compute the hybrid fundamental matrix the estimation of the epipole could be 25 pixels far from the real value. This is explained because these correspondences are localized in just a small part of the omnidirectional image.

E. Self-localization of the robot

In this experiment we will show the behavior of the whole self-localization system. We have a single omnidirectional image of the scene stored in the robot memory and we obtain a perspective image at different positions of the perspective camera. We match the perspective image with the omnidirectional one to compute the fundamental matrix. We have observed that the features used to perform this matching have to be distributed in a wide part of the omnidirectional

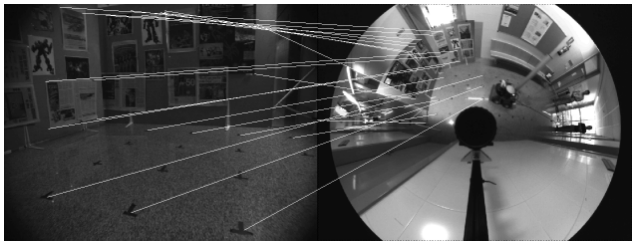
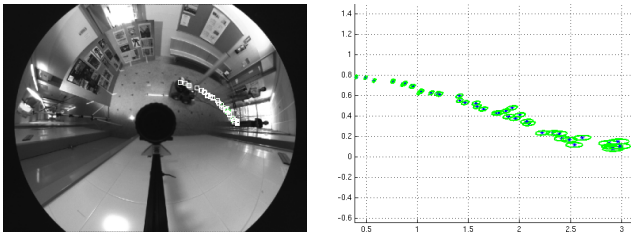


Fig. 9. Matching between the omnidirectional and the perspective image.



(a) (b)

Fig. 10. (a) Epipole trajectory with uncertainty. (b) Uncertainties of the epipoles in the ground plane, units are in meters.

image. If we do not fit this constraint the estimation of the epipole gives bad results. This is the reason why we give some correspondences manually. Once we have computed this matrix, we extract the epipole which gives the position of the perspective camera in the omnidirectional image. Previously, we calibrate the height of the omnidirectional camera with respect to the on-board perspective camera. We set the height between the two optical centers at the same height used to compute the homography. The robot follows a simple straight path, advancing for 3.5m. In Fig. 10 we can see the two phases of the whole approach, camera location on the omnidirectional image (matching and epipole computing) and the ground location through the homography. In Fig. 10(b) we show the uncertainty estimation. We observe that the estimation of the position is better when the robot is in the central area of the omnidirectional image. When the robot is in the periphery the location uncertainty increases as expected.

VI. Conclusions

In this paper we present a new approach to perform the robot self-localization, mixing one reference omnidirectional image and perspective images. We avoid the use of any catadioptric system by storing the omnidirectional images of the scene in the robot memory. The only sensor used is the on-board perspective camera installed on the robot. We propose a schema where the epipole, computed from the hybrid fundamental matrix, in the omnidirectional image is mapped to a ground plane by a hybrid homography previously computed. We observed that the accuracy of the epipole estimation depends on how well the correspondences are distributed in the perspective and omnidirectional images. The hybrid homography proposed uses a generic model that can relate a scene plane to any catadioptric image. We test our approach with simulated data and we obtained promising results using real images. We observed that the major drawbacks in practice are the hybrid matching and the computation of epipoles. In both cases we the features must be well distributed in both images which is difficult when the CFOV is small. If the matching is done between the whole perspective image but just a part of the omnidirectional one we can match the images but the epipole estimation could have big errors. We also observed a highly inhomogeneous uncertainty related to the position on the omnidirectional image, which is however well coded in our model. There is still work to do with this new self-localization approach, but it seems to be a good option since a set of omnidirectional reference images can describe easily the environment, and only a conventional uncalibrated camera is required on-board.

References

[1] Baker, S., Nayar, S.: A theory of single-viewpoint catadioptric image formation. Int. J. Comput. Vision **35(2)** (1999) 175–196

[2] Svoboda, T., Pajdla, T.: Epipolar geometry for central catadioptric cameras. Int. J. Comput. Vision **49** (2002) 23–37

[3] Geyer, C., Daniilidis, K.: Structure and motion from uncalibrated catadioptric views. Proc. IEEE Conference on Computer Vision and Pattern Recognition (2001)

[4] Sturm, P.: Mixing catadioptric and perspective cameras. In: Workshop on Omnidirectional Vision, Copenhagen, Denmark. (2002) 37–44

[5] Barreto, J.P., Daniilidis, K.: Epipolar geometry of central projection systems using veronese maps. In: Computer Vision and Pattern Recognition, Washington, DC, USA, IEEE Computer Society (2006) 1258–1265

[6] Barreto, J.: A unifying geometric representation for central projection systems. CVIU **103** (2006)

[7] Puig, L., Guerrero, J.J., Sturm, P.: Matching of omnidirectional and perspective images using the hybrid fundamental matrix. In: OMNIVIS. (2008)

[8] Sturm, P., Barreto, J.: General catadioptric imaging geometry. In: ECCV. (2008)

[9] Ishiguro, H., Tsuji, S.: Image-based memory of environment. In: in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems. (1996) 634–639

[10] Pajdla, T., Hlavác, V.: Zero phase representation of panoramic images for image vased localization. In: CAIP '99: Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns, London, UK, Springer-Verlag (1999) 550–557

[11] Cauchois, C., Brassart, E., Delahoche, L., Clerentin, A.: 3d localization with conical vision. In: in Omnivis 2003, the Fourth Workshop on Omnidirectional Vision. (2003)

[12] Matsumoto, Y., Ikeda, K., Inaba, M., Inoue, H.: Visual navigation using omnidirectional view sequence. Intelligent Robots and Systems. **1** (1999) 317–322

[13] Goedemé, T., Nuttin, M., Tuytelaars, T., Gool, L.V.: Omnidirectional vision based topological navigation. Int. J. Comput. Vision **74** (2007) 219–236

[14] Briggs, A., Li, Y., Scharstein, D., Wilder, M.: Robot navigation using 1d panoramic images. Robotics and Automation. ICRA. (2006) 2679–2685

[15] Courbon, J., Mezouar, Y., Eck, L., Martinet, P.: Efficient visual memory based navigation of indoor robot with a wide-field of view camera. In: Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on. (2008) 268–273

[16] Murillo, A.C., Sagues, C., Guerrero, J.J., Goedemé, T., Tuytelaars, T., Gool, L.V.: From omnidirectional images to hierarchical localization. Robotics and Autonomous Systems **55(5)** (2007) 372–382

[17] Criminisi, A., Reid, I., Zisserman, A.: A plane measuring device. Image and Vision Computing **17** (1999) 625–634

[18] Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical applications. In: ECCV (2). (2000) 445–461

[19] Lowe, D.: Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision. Volume 20. (2004) 91–110