DIIS - I3A
C/ María de Luna num. 1
E-50018 Zaragoza
Spain

Internal Report: 2006-V07

Localization with Omnidirectional Images using the Radial Trifocal Tensor

C. Sagüés, A.C. Murillo, J.J. Guerrero, T. Goedemé, T. Tuytelaars, L. Van Gool

If you want to cite this report, please use the following reference instead:
Localization with Omnidirectional Images using the Radial Trifocal Tensor, C. Sagüés, A.C. Murillo, J.J. Guerrero, T. Goedemé, T. Tuytelaars, L. Van Gool *IEEE Int. Conference on Robotics and Automation*, pages 551-556, Orlando - Florida, May 2006.

This work was supported by project MCYT/FEDER - DPI2003 07986 and DGA(CONSI+D)/CAI - IT17/05

Localization with Omnidirectional Images using the Radial Trifocal Tensor

C. Sagüés, A.C. Murillo, J.J. Guerrero
DIIS - I3A, University of Zaragoza, Spain
{csagues, acm, jguerrer}@unizar.es

T. Goedemé, T. Tuytelaars, L. Van Gool
PSI-VISICS, University of Leuven, Belgium
{tgoedeme, Tinne.Tuytelaars, Luc.VanGool}@esat.kuleuven.be

***Abstract*— In this paper we present a technique to linearly recover 2D structure and motion in man made environments from three uncalibrated omnidirectional views. We use vertical lines from the scene which are projected as radial lines in the images and are automatically matched. The algorithm is based on a 1D radial trifocal tensor which encodes the relations of the three views and the projected lines. We include experiments with real images, which demonstrate the good performance of the method and its application to robotic tasks, such as robot localization based in a database of reference images or to obtain the initial values of robot and landmarks localization in SLAM algorithms.**

I. Introduction

Omnidirectional images have received increasing interest from robotics researchers lately [1], [2], [3]. This kind of images give a 360° view of the environment in a single image and the landmarks stay longer in the field of view than in conventional cameras. This benefits especially robotics applications that need to estimate the motion. For example for navigation (when we need to correct the trajectory due for instance to odometry errors) or for localization (where the current image is matched to a database of reference images taken at known positions).

For motion detection in omnidirectional images, some authors [4] have proposed to use known ego-motion algorithms once the relationship between the spherical projection model and the catadioptric omnidirectional model has been established. Others have matched points to compute structure and motion through a catadioptric fundamental matrix estimated with a non-linear method [5].

In this work we propose a method for 2D localization based on automatic matching of vertical lines and the 1D radial trifocal tensor. The results of the accuracy with real images are shown in the experimental section. Cobzas *et al.* [6] have previously worked on 2D localization with 3 images, with good results in localization but without simultaneously solving the line matching problem and using a different kind of panoramic images (cylindrical ones formed as a mosaic).

Most man-made natural environments contain a lot of vertical lines from natural landmarks (for example, doors or corridors when moving indoors, and lampposts or buildings when moving outdoors). Such lines are useful in our application because of several reasons. First, lines are robust against partial occlusions. Besides, as we work with an omnidirectional camera with vertical axis, these lines are the only lines that keep their straightness being always projected as radial lines in the image. Therefore, they can be easily used to linearly compute structure and motion and also to automatically estimate the center of projection in the image, which is a very important calibration parameter in omnidirectional images.

Matching features in omnidirectional images is more difficult than in conventional ones, mainly due to the deformations and lower resolution of the images. We designed an algorithm for matching lines in two omnidirectional images, which deals with these problems reasonably well [7]. This is the method used as first step in this work.

At least three views are necessary to solve the structure and motion problem from lines. Therefore, we have extended our matching algorithm in order to get triple matches. Once we have radial lines matched in 3 images, we can estimate a 1D radial trifocal tensor. This term was introduced in an application for calibrating radial distortion [8]. It is necessary to use a robust method (*ransac* in our case) to check the matches, due to the inevitable chance for outliers. We do simultaneously the robust matching and the computation of the trifocal tensor. The 1D trifocal tensor encodes the three-view relation between images [9], and it can be used to linearly compute both the position of the robot and the localization of the natural landmarks. In robotics, it was previously used for SLAM algorithm from bearing-only data but using conventional cameras and artificial landmarks on a plane [10].

II. 1D Radial Trifocal Tensor

A vertical line $\mathbf{v}$ in the world always gives a radial line $\mathbf{r}$ when projected in an omnidirectional image taken with vertical optical axis (Fig. 1). A radial line is related to its image points in such a way that if we name $(x_i, y_i, 1)$ to any point along the radial line, then the direction of that line will be $\mathbf{r_i} = (x_i, y_i)$.

A vertical line in the 3D space can be represented by three homogeneous coordinates in $\mathcal{P}^2$ as $\mathbf{v} = [v_{(1)}, v_{(2)}, v_{(3)}]^T$. Its projection in the 2D image plane (a radial line) can be represented in $\mathcal{P}^1$ as a 1D homogeneous vector $\mathbf{r} = (r_{(1)}, r_{(2)})^T$. This mapping from $\mathcal{P}^2$ to $\mathcal{P}^1$ can be represented by a 2×3 matrix $\mathbf{P}$.

For three images (O_1, O_2, O_3), we have the following set

This work was supported by project MCYT/FEDER - DPI2003 07986 and DGA(CONSI+D)/CAI - IT17/05

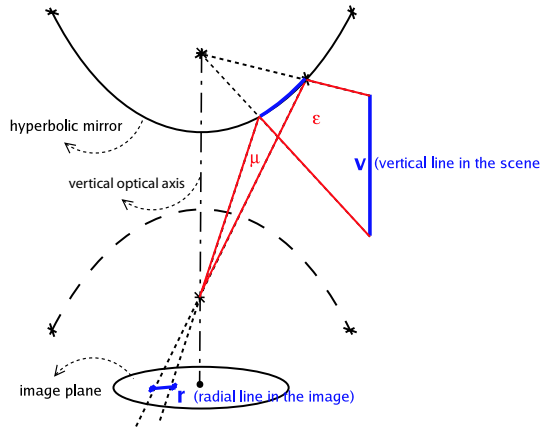

Fig. 1. Vertical line (**v**), its corresponding projection planes (ε, μ), which are coincident, and the projected radial line (**r**) in the image.

of equations:

$$\lambda_1 \mathbf{r}_1 = \mathbf{P}_1 \mathbf{v}; \qquad \lambda_2 \mathbf{r}_2 = \mathbf{P}_2 \mathbf{v}; \qquad \lambda_3 \mathbf{r}_3 = \mathbf{P}_3 \mathbf{v} \tag{1}$$

where λ_1, λ_2 and λ_3 are scale factors.

These projection equations (1) in the three images can be written in the following way

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{r}_1 & 0 & 0 \\ \mathbf{P}_2 & 0 & \mathbf{r}_2 & 0 \\ \mathbf{P}_3 & 0 & 0 & \mathbf{r}_3 \end{bmatrix} [\mathbf{v}, -\lambda_1, -\lambda_2, -\lambda_3]^T = 0. \tag{2}$$

Since a solution exists and neither **v** nor the scale factors can be null, then the right null space of the measurement matrix should have non-zero dimension which results in

$$\begin{vmatrix} \mathbf{P}_1 & \mathbf{r}_1 & 0 & 0 \\ \mathbf{P}_2 & 0 & \mathbf{r}_2 & 0 \\ \mathbf{P}_3 & 0 & 0 & \mathbf{r}_3 \end{vmatrix} = 0 \tag{3}$$

being $|\ |$ the determinant. This produces a trilinear constraint for the corresponding image radial lines in the three views [11]

$$\Sigma_{i=1}^{2} \Sigma_{j=1}^{2} \Sigma_{k=1}^{2} T_{ijk} \mathbf{r}_{1(i)} \mathbf{r}_{2(j)} \mathbf{r}_{3(k)} = 0, \tag{4}$$

where subindex (i, j, k) indicates the component of the vector and T_{ijk} $(i, j, k = 1, 2)$ are eight elements of the $2 \times 2 \times 2$ trifocal tensor of the three omnidirectional cameras. The elements of this tensor are the 3×3 minors of the 6×3 matrix $[\mathbf{P}_1\, \mathbf{P}_2\, \mathbf{P}_3]^T$ in such a way that to obtain $T_{i\underline{jk}} = [\overline{i}\ \overline{j}\ \overline{k}]$ we take the $\overline{i}th$ row of $\mathbf{P}_1$, the $\overline{j}th$ row of $\mathbf{P}_2$ and the kth row of $\mathbf{P}_3$, with $(\overline{.})$ meaning a mapping from [1,2] to [2,-1] (i.e., $\overline{1}$ means 2nd row and $\overline{2}$ means 1st row with the sign changed).

Let w be the number of views from the different robot locations, and l the number of landmarks seen by the robot (assuming for simplicity that all landmarks are visible in all views). As we get one equation from each landmark in each view, we have wl equations. We have 3 motion parameters to compute for each robot position (except for the first one, because we locate it in the origin) and 2 parameters for each landmark. If the number of views is 2, the problem is unsolvable, even with infinite number of landmarks, $wl < 3w - 3 + 2l - 1$, (-1 because the results are up to scale factor). We can see that the minimum number of views necessary to solve this problem is 3, with at least 5 landmarks.

The 1D trifocal tensor has eight parameters up to a scale and it could be estimated from seven corresponding triplets. It is also possible to obtain it from only five three-view matches in calibrated cameras plus two additional conditions. It has been established that for a trifocal tensor, in case of calibrated cameras, the following additional constraints apply [9]:

$$\begin{aligned} -T_{111} + T_{122} + T_{212} + T_{221} = 0 \\ T_{112} + T_{121} + T_{211} - T_{222} = 0 \end{aligned} \tag{5}$$

which gives two additional linear equations to compute it.

We have shown, that using five matches and the calibration conditions is computationally more efficient and it gives better results in motion estimation than the classical seven degrees of freedom tensor computed from seven matches [12]. The computation of the trifocal tensor can be carried out as explained above using Singular Value Decomposition (SVD). With more matches, that procedure would give the least squares solution, which assumes that all the measures could be interpreted with the same model. This is very sensitive to outliers, so we need robust estimation methods to avoid them in the estimation. In our work we have chosen *ransac* [13], which makes a search in the space of solutions using random subsets of minimum number of matches.

III. From Lines to Motion Estimation

As explained before, our approach takes vertical lines from the scene as key features because they show several advantages when working with omnidirectional images. They are first matched based on descriptors of the lines. After this initial matching, the estimation of a 1D radial trifocal tensor and the computation of a more robust set of matches can be carried out simultaneously. The different steps of this algorithm are as follows,

- We automatically extract the image lines with their corresponding support region, that is the areas around the lines with similar gradient values [14].
- With the radial lines and using a *ransac* approach we robustly estimate the center of projection in the image, as it is not coincident with the center of the image. We need it to estimate the motion later. This step also rejects all the image lines that are not radial (i.e., non-vertical landmarks).
- Each line is described with a set of descriptors based on color and intensity signals of the image and on other line properties. Then, we match each couple of images (first and second, second and third) based on those descriptor differences. More details about this matching can be found in [7].
- Once we have two-view matches we consider as three-view ones the intersection of such sets. Based on the idea that all the neighboring lines should change in a similar way (rotate nearly the same), we can increase the number of matches until enough are found to be able to

estimate the trifocal tensor. We can make a first approach to reject some outliers, taking into account that all the triple matches should behave in a consistent way.

- From these matches we perform a robust estimation, using a *ransac* approach, of the radial trifocal tensor between the three images and simultaneously we reject the outliers (matches that do not fit the tensor). The accepted robust triplets of matches fit the geometry imposed by the tensor with an error under a chosen threshold.
- Finally, from the 1D radial trifocal tensor we can estimate the relative motion between views and the localization of the landmarks in the scene. This is explained in more detail in next section.

IV. Structure and Motion

As said before, the computed 1D radial trifocal tensor allows to estimate the relative positions of the robot and the landmarks.

A. Motion estimation

The motion can be computed from this radial trifocal tensor without any other kind of calibration of the camera than the assumption of squared pixels. Because the motion is restricted to a plane, four motion parameters, up to a scale factor for translation, should be computed, $\theta_{21}, \mathbf{t}_{21}, \theta_{31}, \mathbf{t}_{31}$ (Fig. 2).

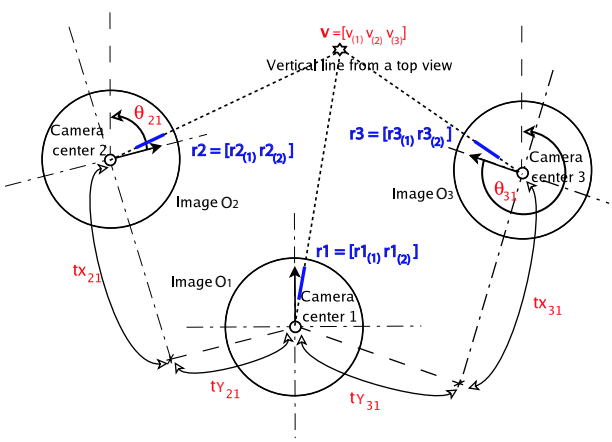

Fig. 2. The problem is to compute the relative position of the cameras $\theta_{21}, \mathbf{t}_{21}, \theta_{31}, \mathbf{t}_{31}$ and the localization of the landmarks $\mathbf{v}$

The method presented here is based on radial epipoles computed from homographies and it is the extension to omnidirectional images of the ideas presented by Dellaert [10].

A homography is a mapping between projective coordinates of two radial lines (from now on, whenever we are talking about lines, we refer to radial lines). The 1D trifocal tensor can be decomposed in several intrinsic homographies as explained in [15]. So if we take the implicit homography $\mathbf{H}_{23}$ corresponding to an image line in the first image O_1, it gives a relationship between the second O_2 and third image O_3.

$$\mathbf{r}_2 = \mathbf{H}_{23}\mathbf{r}_3$$

The homographies $\mathbf{H}^X_{23}$ and $\mathbf{H}^Y_{23}$, corresponding to the lines $\mathbf{r}_{1x}$ and $\mathbf{r}_{1y}$ respectively, can be obtained from the trifocal

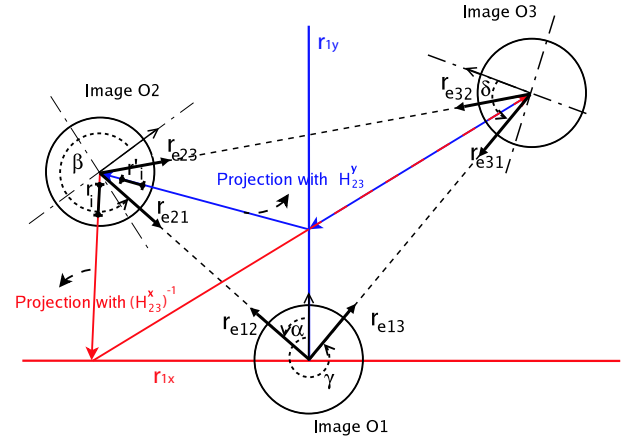

Fig. 3. Three omnidirectional images (O_1, O_2 and O_3). Radial epipoles $\mathbf{r}_{\mathbf{e}12}$, $\mathbf{r}_{\mathbf{e}21}$, $\mathbf{r}_{\mathbf{e}13}$, $\mathbf{r}_{\mathbf{e}31}$ and their orientations in the image α, β, γ, δ respectively. The radial epipoles are invariant to the reprojection by any homology, while the other image radial lines are not. For example, a generic line $\mathbf{r}_i$ goes through $(\mathbf{H}^X_{23})^{-1}\mathbf{H}^Y_{23}$ to $\mathbf{r}'_i$.

tensor (Fig. 3), and similarly for the homographies of the other images. We obtain,

$$\mathbf{H}^X_{12} = \begin{bmatrix} -T_{211} & -T_{221} \\ T_{111} & T_{121} \end{bmatrix} \quad \mathbf{H}^Y_{12} = \begin{bmatrix} -T_{212} & -T_{222} \\ T_{112} & T_{122} \end{bmatrix}$$

$$\mathbf{H}^X_{23} = \begin{bmatrix} -T_{121} & -T_{122} \\ T_{111} & T_{112} \end{bmatrix} \quad \mathbf{H}^Y_{23} = \begin{bmatrix} -T_{221} & -T_{222} \\ T_{211} & T_{212} \end{bmatrix}$$

$$\mathbf{H}^X_{31} = \begin{bmatrix} -T_{112} & -T_{212} \\ T_{111} & T_{211} \end{bmatrix} \quad \mathbf{H}^Y_{31} = \begin{bmatrix} -T_{122} & -T_{222} \\ T_{121} & T_{221} \end{bmatrix}$$

With those homographies the corresponding homologies for the three images will be $\mathbf{H}_1 = \mathbf{H}^X_{12}(\mathbf{H}^Y_{12})^{-1}$, $\mathbf{H}_2 = \mathbf{H}^X_{23}(\mathbf{H}^Y_{23})^{-1}$ and $\mathbf{H}_3 = \mathbf{H}^X_{31}(\mathbf{H}^Y_{31})^{-1}$.

The radial epipoles are the only radial lines mapped to themselves through the homology ($\mathbf{r}_\mathbf{e} = \mathbf{H}\mathbf{r}_\mathbf{e}$). So, the eigenvectors of the homologies are the radial epipoles of the images. From the eigenvectors of $\mathbf{H}_1$ we obtain $(\mathbf{r}_{\mathbf{e}12}, \mathbf{r}_{\mathbf{e}13})$ corresponding to the radial epipoles of O_2 and O_3 in the first image O_1.

Similarly the radial epipoles $(\mathbf{r}_{\mathbf{e}21}, \mathbf{r}_{\mathbf{e}23})$ and $(\mathbf{r}_{e31}, \mathbf{r}_{\mathbf{e}32})$ from $\mathbf{H}_2$ and $\mathbf{H}_3$ respectively can be obtained.

Alternatively, we can obtain the rest of the epipoles by projecting the first computed ones $(\mathbf{r}_{\mathbf{e}12}, \mathbf{r}_{\mathbf{e}13})$, to the other two images through any of the corresponding homographies like this

$$\begin{aligned} \mathbf{r}_{e21} &= (\mathbf{H}^X_{12})^{-1}\mathbf{r}_{e12} & \mathbf{r}_{e23} &= (\mathbf{H}^X_{12})^{-1}\mathbf{r}_{e13} \\ \mathbf{r}_{e31} &= \mathbf{H}^X_{31}\mathbf{r}_{e13} & \mathbf{r}_{e32} &= \mathbf{H}^X_{31}\mathbf{r}_{e12} \end{aligned}$$

If we call α, β the relative orientations of the radial epipoles $\mathbf{r}_{\mathbf{e}12}$ and $\mathbf{r}_{\mathbf{e}21}$ in the images O_1 and O_2 (Fig. 3), then the relative orientation (we get two symmetric solutions) between both cameras is,

$$\begin{aligned} \theta_{21} &= \beta - \alpha \\ \theta_{21} &= \beta - \alpha + \pi \end{aligned} \tag{6}$$

The translation $\mathbf{t}_{21} = (t_{x21}, t_{y21})^T$ is given, up to a scale, by the direction of the corresponding radial epipole, $\mathbf{t}_{21} = \mathbf{r}_{\mathbf{e}21}$.

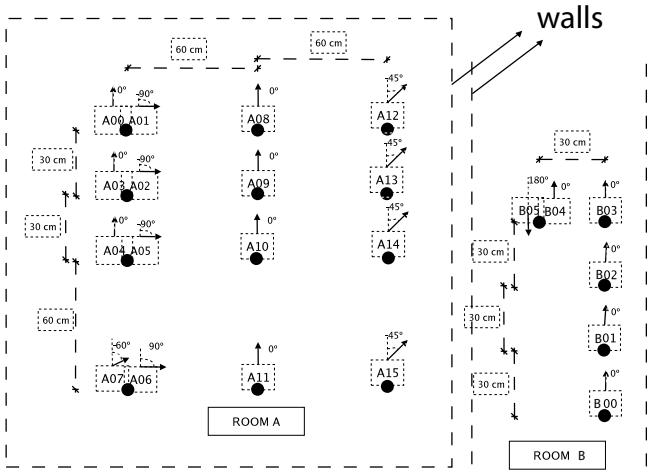

Fig. 4. Schema of positions of the images in the rooms of the experiments.

Similarly if we call γ and δ the relative orientations of the radial epipoles $\mathbf{r_e}_{13}$ and $\mathbf{r_e}_{31}$ in the images O_1 and O_3 (Fig. 3), then the relative orientation between both cameras is (we also get two symmetric solutions)

$$\begin{aligned} \theta_{31} &= \delta - \gamma \\ \theta_{31} &= \delta - \gamma + \pi \end{aligned} \tag{7}$$

The translation $\mathbf{t}_{31} = (t_{x31}, t_{y31})^T$ is given, up to a scale, by the direction of the corresponding image radial epipole, $\mathbf{t}_{31} = \mathbf{r_e}_{31}$.

B. Recovering the position of the landmarks

The vertical lines obtained from natural landmarks, usually correspond with doors, walls or other important features of the scene.

An arbitrary vertical line, $\mathbf{v} = (v_{(1)}, v_{(2)}, v_{(3)})^T$, gives rise to an image measurement in each of the three views: $\mathbf{r}_1 = [\mathbf{0}|\mathbf{I}]\mathbf{v}$; $\mathbf{r}_2 = [\mathbf{R}_{21}|\mathbf{t}_{21}]\mathbf{v}$ and $\mathbf{r}_3 = [\mathbf{R}_{31}|\mathbf{t}_{31}]\mathbf{v}$ respectively. Where $\mathbf{R}_{21}$, $\mathbf{R}_{31}$ denote the rotation matrices between the corresponding camera positions. They can be written as

$$\mathbf{R}_{21} = \begin{bmatrix} \cos\theta_{21} & \sin\theta_{21} \\ -\sin\theta_{21} & \cos\theta_{21} \end{bmatrix}, \mathbf{R}_{31} = \begin{bmatrix} \cos\theta_{31} & \sin\theta_{31} \\ -\sin\theta_{31} & \cos\theta_{31} \end{bmatrix} \tag{8}$$

From those we can write

$$\begin{aligned} \mathbf{r}_1 \times [\mathbf{I}|\mathbf{0}]\mathbf{v} &= 0 \\ \mathbf{r}_2 \times [\mathbf{R}_{21}|\mathbf{t}_{21}]\mathbf{v} &= 0 \\ \mathbf{r}_3 \times [\mathbf{R}_{31}|\mathbf{t}_{31}]\mathbf{v} &= 0 \end{aligned} \tag{9}$$

where $\times$ indicates the cross product.

They can be explicitly developed to give three equations, although two could be enough to solve the position of the natural landmark $\mathbf{v}$.

V. Experimental results

We have performed several experiments with real images (640x480) of different indoor scenes. We have a database of 70 images sorted in different rooms and with known relative location. They were acquired with an omnidirectional vision system, consisting of a color camera, with vertical optical axis, looking at a hyperbolic mirror.

There is a schema showing the images-grid from the part of the database corresponding to the rooms shown in the experiments (Fig. 4). The ground truth used to evaluate the motion estimation results was measured by hand with a tape measure and goniometer when taking the images.

A. Estimation of the projection center

First of all, we made some experiments to show the stability of the estimation of the projection center in the image. This parameter is very important because it determines the direction of the lines. This direction is used to estimate the tensor, that is very sensitive to noise and therefore we need to get the center coordinates as robust as possible.

We have observed that in one image the center estimated is stable. We made 50 different executions of the ransac-based estimation of the projection center in one image and its standard deviation was 0.8 and 1.4 pixels in the x and y coordinates respectively. Supposedly, the center should remain stable in all the images, but we have observed that in different images it changes till four pixels. This indicates that it is a good idea to obtain the projection center of every image separately. We can see a general example of an image with the extracted radial lines and the center of projection estimated from them (Fig. 5).

B. Robot and landmark localization

In this part of the experiments we want to show the performance of the process explained in this work to get the robot and landmark localization. First in *Test 1* with a common situation, where we automatically obtain the most similar images to the current one from a database. Next in *Test 2* with more complex situations, where we choose three more separated views from the database to test the scope of the method.

1) Test 1: In this experiment, we select one image from our database representing the current and unknown position of our robot, and we deduce its position based on other two images in the database. So, we compare the current image against all

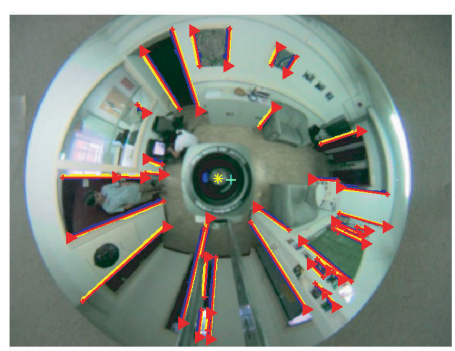

Fig. 5. All lines extracted in one image. The estimated center of projection with a yellow star ($*$) and the center of the image with blue cross (+).

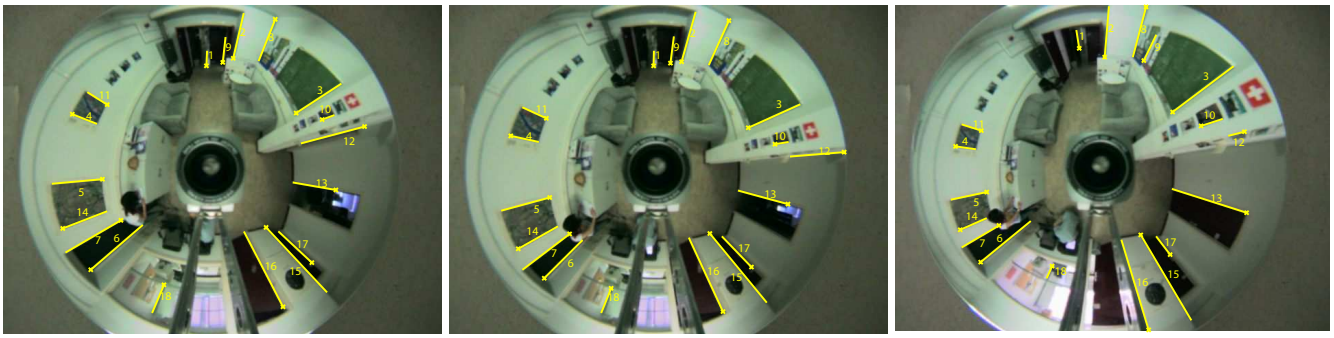

image A09 image A08 image A00

Fig. 6. Test 1. Triple robust matches obtained between current image and 2 reference ones from the database. There are 18 robust matches (1 wrong).

the others to find the most similar with a hierarchical method [7]. Next, we can apply to the current image and two of the database (the most similar found and one adjacent) the method to get structure and motion explained in this work. We show an example of this whole process. As current image we have *A*09, the image found as most similar is *A*08 and one adjacent to this second one is *A*00. We can see a typical solution of the robust matches obtained in these three images after the estimation of the radial trifocal tensor (Fig.6). The motion estimation between them is shown in Table I, where we can observe an accuracy to within one degree, both for the rotation and for the direction of translation. Notice this is close to the uncertainty on our ground-truth data. Moreover, the fully automatic method (including the line matching over the three images) was equally accurate than a method based on manual line matching. We also obtained the localization of scene landmarks as explained above (Section IV-B). We show a typical execution of this in Fig. 7. There we see the robot positions, the landmarks reconstruction with the ground truth motion, which was manually measured, and the reconstruction with the motion computed trough the trifocal tensor.

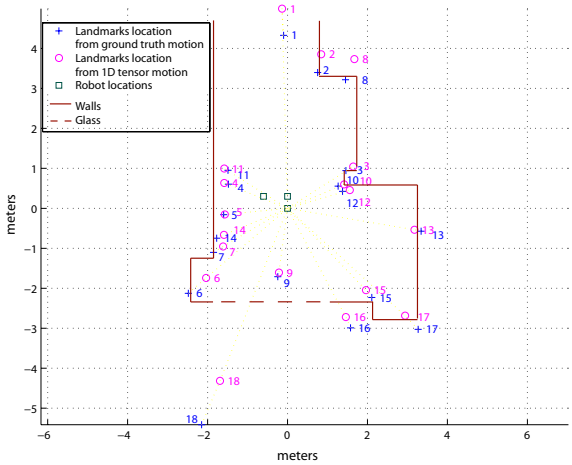

Fig. 7. Robot locations and landmarks reconstruction obtained with the ground truth motion (manually measured) and with the motion obtained through the 1D trifocal tensor.

Test 1	θ_{21}	θ_{31}	t_{21}	t_{31}
Reference	0°	0°	90°	153.44^o
Manual	-0.5^o	0.7^o	87°	152.8^o
Automatic	-0.7^o	0.7^o	91.1^o	153.8^o
(std)	(0.14^o)	(0.2^o)	(4°)	(0.84^o)

TABLE I

Results of motion estimation in *Test 1*. *Reference* is the reference motion of database of images (manually measured). *Manual* is the estimation with lines manually matched and *Automatic* the mean of the results in 50 executions with automatic matching (below its standard deviation).

Test 2		θ_{21}	θ_{31}	t_{21}	t_{31}
Case 1	*Reference*	0°	45°	26.5^o	26.5^o
	Manual	0.5^o	48.55^o	28.52^o	16°
	Automatic mean	4.03^o	46.46^o	15.26^o	-5.65^o
	(std)	(5.83^o)	(12.15^o)	(46.99^o)	(28.34^o)
	median	2.85^o	44.47^o	24.20^o	21.14^o
Case 2	*Reference*	0°	180°	90°	116.56^o
	Manual	3.08^o	176.31^o	92.5^o	112.55^o
	Automatic mean	3.47^o	174.52^o	94°	115.84^o
	(std)	(1.61^o)	(0.59^o)	(4.90^o)	(1.29^o)
	median	3.20^o	174.69^o	94.66^o	115.66^o

TABLE II

Results of motion estimation in *Test 2*. *Reference* shows the reference data (manually measured). *Manual* contains the estimation with manual matches and row *Automatic* the mean, its standard deviation and the median of 50 executions with automatic line matching.

2) Test 2: We have made some additional trials with images more separated, in order to try situations more useful because they allow to have smaller databases, but more difficult to get enough three-view matches in. Examples of the typical robust triple matches obtained in these cases are in Fig. 8. The results of the motion estimation in these situations are worse than in previous *Test 1*, mainly due to the lower quality of the obtained matches, but still acceptable. Table II shows averages of 50 executions of the estimation for each one. The worst behavior in *Case 1* could be due to the alignment of the three image centers. This could explain some bad executions that spoilt the average, because the median values were still acceptable. As future work we will study these special configurations.

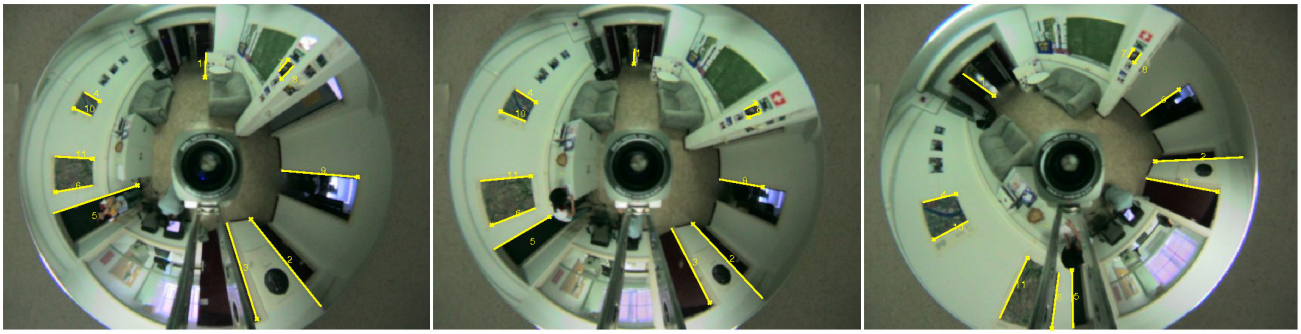

image A04 image A09 image A12

(a) Case 1

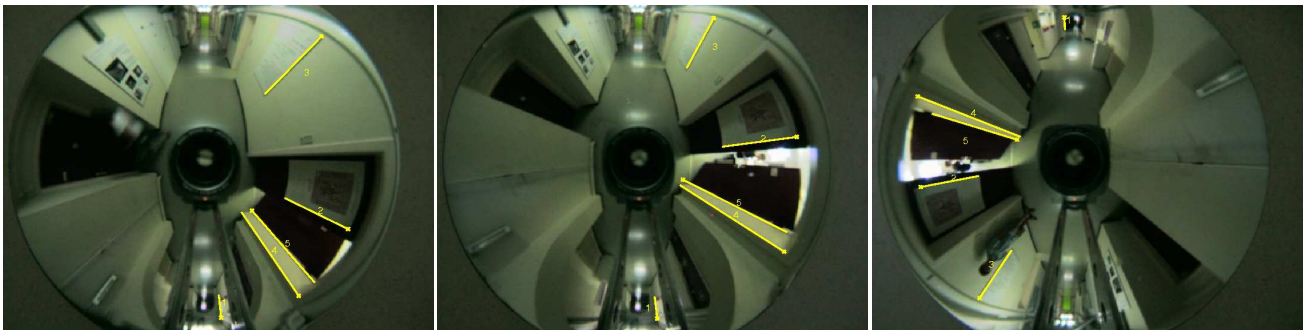

image B01 image B03 image B05

(b) Case 2

Test 2	Total matches	wrong	Robust matches	wrong
Case 1 - Fig.8(a)	15	4	11	1
Case 2 - Fig.8(b)	6	1	5	0

(c) Total matches before tensor estimation and Robust matches after the robust estimation of the tensor.

Fig. 8. Test 2. Triple robust matches obtained in several examples widely separated.

VI. Conclusions

We have presented a linear approach to recover robot and landmark localization with omnidirectional images through the 1D radial trifocal tensor. The features used are vertical lines, extracted from natural landmarks (doors, corners..). The radial lines in the images are matched in triplets based on line descriptors and their support regions. They are robustly filtered using the trilinear constraint imposed by the radial trifocal tensor. The experiments show high accuracy and stability for results in real situations. So they are suitable to be used in robotic tasks such as localization based on reference databases or to initialize SLAM algorithms.

References

[1] K. Imai, K. Tsuji, and M. Yachida, "Iconic memory-based omnidirectional route panorama navigation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 78–87, 2005.

[2] E. Menegatti, A. Pretto, and E. Pagello, "Testing omnidirectional vision-based monte-carlo localization under occlusion," in *Proc. of International Conference on Intelligent Robots and Systems, IROS 2004*, 2004.

[3] P. I. Corke, D. Strelow, and S. Singh, "Omnidirectional visual odometry for a planetary rover," in *Proc. of International Conference on Intelligent Robots and Systems, IROS 2004*, 2004.

[4] J. Gluckman and S. Nayar, "Ego-motion and omnidirectional cameras," in *IEEE International Conference on Computer Vision*, 1998.

[5] C. Geyer and K. Daniilidis, "Structure and motion from uncalibrated catadioptric views," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2001.

[6] D. Cobzas and H. Zhang, "2d robot localization with image-based panoramic models using vertical line features," in *Proc. of Vision Interface*, 2000, pp. 211–216.

[7] A. Murillo, C. Sagüés, J. Guerrero, T. Goedemé, T. Tuytelaars, and L. V. Gool, "Hierarchical localization by matching vertical lines in omnidirectional images," DIIS - I3A Universidad de Zaragoza," Technical report - 2005-V07, 2005.

[8] S. Thirthala and M. Pollefeys, "The radial trifocal tensor: A tool for calibrating the radial distortion of wide-angle cameras," in *Proc. of Computer Vision Pattern Recognition*, 2005.

[9] K. Åström and M. Oskarsson, "Solutions and ambiguities of the structure and motion problem for 1d retinal vision," *Journal of Mathematical Imaging and Vision*, vol. 12, pp. 121–135, 2000.

[10] F. Dellaert and A. Stroupe, "Linear 2d localization and mapping for single and multiple robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, May 2002.

[11] O. Faugeras, L. Quan, and P. Sturm, "Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1179–1185, 2000.

[12] J. Guerrero, C. Sagüés, and A. Murillo, "Localization and bearing-only matching using the planar trifocal tensor," DIIS - I3A Universidad de Zaragoza," Technical report - 2005-V06, 2005.

[13] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: John Wiley, 1987.

[14] J. Guerrero and C. Sagüés, "Robust line matching and estimate of homographies simultaneously," in *IbPRIA, Pattern Recognition and Image Analysis, LNCS 2652*, 2003, pp. 297–307.

[15] A. Shashua and M. Werman, "Trilinearity of three perspective views and its associate tensor," in *In Proceedings of the International Conference on Computer Vision (ICCV)*, June 1995, pp. 920–925.