

Robot and Landmark Localization using Scene Planes and the 1D Trifocal Tensor

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

Abstract—This paper presents a method for robot and landmarks 2D localization, in man made environments, taking profit of scene planes. The method uses bearing-only measurements that are robustly matched in three views. In our experiments we obtain them from vertical lines corresponding to natural landmarks. With these three view line-matches a trifocal tensor can be computed. This tensor contains the three views geometry and is used to estimate the aforementioned localization. As it is very usual to find a planar surface, we use the homography corresponding to that plane to obtain the tensor with one match less than the general case method. This implies lower computational complexity, mainly when trying a robust estimation, where we see a reduction in the number of iterations needed. Another advantage of obtaining an homography during the process is that it can help to automatically detect singular situations, such us totally planar scenes. It is shown that our proposal performs similarly to the general case method in a general scenario and better in case that we have some dominant plane in the scene. This paper includes simulated results proving this, as well as examples with real images from conventional and omnidirectional cameras.

Index Terms - 1D trifocal tensor, scene planes, bearing-only data, localization, SLAM initialization.

I. INTRODUCTION

Most robot autonomous tasks can not be accomplished just using the odometry information, due to its well-known limitations. Laser range and vision sensors are mostly used to provide the robot with scene information to carry out those autonomous tasks. Mobile robots work many times on planar surfaces. To define the scene or the situation in this case, i.e. to be localized, three motion parameters for a robot location and two more for each feature are needed. This is the task we are dealing with in this work (Fig.1).

In the last years, many Simultaneous Localization and Mapping (SLAM) algorithms have been proposed as a good method to achieve those tasks in unknown scenes, using different sensors, e.g. [1]. We are going to focus in the case of using bearing-only data. In this case, the SLAM methods can be seen as an iterative process that need to be initialized somehow, because just with one bearing-measurements acquisition we can not directly estimate the distance for the observed landmarks. In case of planar motion this initialization can be done with linear methods using three different initial acquisitions, as explained in [2]. When working with images, the projection in a 1D virtual retina of vertical landmarks in

the scene can be treated as bearing-only data. Similarly in omnidirectional images for the radial lines, which came from projected scene vertical landmarks [3]. As it is known, typically three acquisitions from different positions of the bearing-sensor are needed to recover robot and landmark localization. The trifocal tensor gives a closed formulation to relate those three views. Recently a work has appeared proposing a way to avoid the need of these three first acquisitions. It is based in multi-hypotesis ideas [4], with good performance for the SLAM, but it increases a little the complexity of the problem and they still need several acquisitions until they have a defined estimation of landmarks position.

The 1D trifocal tensor was previously presented [5], together with an algorithm to compute motion from it in a closed form. The 1D trifocal tensor has also been used for calibration of 1D cameras [6]. Also in its application to omnidirectional cameras we can find some previous related works, about localization [3] and about radial distortion correction [7]. This tensor can be computed linearly from seven matches although with calibrated cameras five matches are enough [5]. The use of constraints imposed by the scene can reduce the number of necessary matches. Here we study the situation with a plane available in the scene. These ideas have been used in case of general 3D scenes, following the well known two view plane-parallax constructions and extending it to more views, chapters 12 and 15 in [8]. The trifocal tensor and multi-view constraints based on homographies were studied in [9].

In this paper, we suppose a scene that contains at least one plane, more exactly three coplanar feature-matches, which is quite usual when a robot moves on man-made environments. We show how to estimate the 1D tensor from only four matches and evaluate the robot and landmarks localization obtained from that tensor. One goal of this work is to show a way to reduce the computational cost of previous methods aiming the same, by taking profit of the existence of a plane in the scene. This may be very useful in real time applications. Another advantage of obtaining an homography during the process is that it can help to automatically detect singular situations, such us totally planar scenes. We proof the good performance of the proposal with experiments both with simulated and real data. There are tests with pinhole cameras, where we use the projection of vertical landmarks in the scene, as well as with omnidirectional images, with the advantage in this case that the camera calibration is not needed.

II. PLANAR TRIFOCAL TENSOR

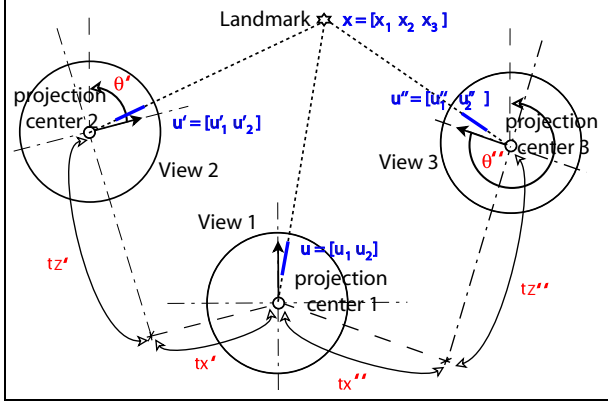


Fig. 1. The goal is to obtain the relative localization of the robot (θ' , $\mathbf{t}' = [t'_x, t'_z]$; θ'' , $\mathbf{t}'' = [t''_x, t''_z]$) and the position of the landmarks (\mathbf{x}), from the three view matches (\mathbf{u} \mathbf{u}' \mathbf{u}'') of natural landmarks.

The bearing-only data obtained by a robot moving on a planar surface can be converted to measurements in 1D perspective cameras using a projective formulation. Thus, we can easily convert a bearing measurement α from a scene feature to its projective formulation in a 1D virtual retina as $\mathbf{u} = (\tan \alpha, 1)^T$ or $\mathbf{u} = (\sin \alpha, \cos \alpha)^T$, which are projectively equivalent. In our case, the bearing-only data is particularized to vertical lines detected in images. We can consider only the x line coordinate in the image is relevant. Therefore they can be treated as elements of the \mathcal{P}^1 projective space and so we are in the same 1D case. With three views from different positions a trifocal tensor can be linearly computed, and robot and landmarks localization can be obtained (Fig. 1).

Let us name the homogeneous representation of a feature in \mathcal{P}^2 space as $\mathbf{x} = [x_1, x_2, x_3]^T$ and its homogeneous representation in the \mathcal{P}^1 projective space as $\mathbf{u} = [u_1, u_2]^T$. This projection to \mathcal{P}^1 projective space can be expressed by a 2×3 matrix \mathbf{M} in each image, in such a way that

$$\lambda \mathbf{u} = \mathbf{M}\mathbf{x}; \quad \lambda' \mathbf{u}' = \mathbf{M}'\mathbf{x}; \quad \lambda'' \mathbf{u}'' = \mathbf{M}''\mathbf{x} \quad (1)$$

where λ , λ' and λ'' are the respective scale factors.

Let us suppose all the scene features in a common reference frame placed in the first robot location. Then, the projection matrixes relating the observed features in the scene and in the corresponding image are $\mathbf{M} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$, $\mathbf{M}' = \mathbf{K}[\mathbf{R}'|\mathbf{t}']$ and $\mathbf{M}'' = \mathbf{K}[\mathbf{R}''|\mathbf{t}'']$, for the first, second and third robot locations respectively. These matrixes are composed by internal and external parameters. The internal ones are enclosed in the calibration matrix $\mathbf{K} = \begin{bmatrix} f & c_0 \\ 0 & 1 \end{bmatrix}$, where f is the focal length in pixels and c_0 is the position of the principal point. In case of omnidirectional images, the calibration matrix used is the identity. Supposing squared pixels, the only parameter to calibrate is the center of projection, what can be automatically done from the radial lines.

The external parameters are the translations $\mathbf{t}' = [t'_x, t'_z]^T$, $\mathbf{t}'' = [t''_x, t''_z]^T$ and rotations $\mathbf{R}' = \begin{bmatrix} \cos \theta' & \sin \theta' \\ -\sin \theta' & \cos \theta' \end{bmatrix}$,

$\mathbf{R}'' = \begin{bmatrix} \cos \theta'' & \sin \theta'' \\ -\sin \theta'' & \cos \theta'' \end{bmatrix}$ made by the sensor from the second and third position in relation to the first (Fig. 1).

The projection equations (1) from the three locations can be written in the following way

$$\begin{bmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{bmatrix} [\mathbf{x}, -\lambda, -\lambda', -\lambda'']^T = 0. \quad (2)$$

As neither \mathbf{x} nor the scale factors can be null, it originates

$$\begin{vmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{vmatrix} = 0 \quad (3)$$

that can be written as the trifocal constraint for 3 views [6]:

$$\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 \mathbf{T}_{ijk} u_i u'_j u''_k = 0. \quad (4)$$

This can be developed as

$$\begin{aligned} & T_{111} u_1 u'_1 u''_1 + T_{112} u_1 u'_1 u''_2 + T_{121} u_1 u'_2 u''_1 + T_{122} u_1 u'_2 u''_2 + \\ & T_{211} u_2 u'_1 u''_1 + T_{212} u_2 u'_1 u''_2 + T_{221} u_2 u'_2 u''_1 + T_{222} u_2 u'_2 u''_2 = 0, \end{aligned} \quad (5)$$

where T_{ijk} ($i, j, k = 1, 2$) are the eight elements of the $2 \times 2 \times 2$ trifocal tensor whose components are the 3×3 minors of the 6×3 matrix $[\mathbf{M} \mathbf{M}' \mathbf{M}'']^T$, in such a way that to obtain $T_{ijk} = [\bar{i} \bar{j} \bar{k}]$ we take the $\bar{i}th$ row of \mathbf{M} , the $\bar{j}th$ row of \mathbf{M}' and the $\bar{k}th$ row of \mathbf{M}'' , meaning $\bar{\cdot}$ a mapping from $[1, 2]$ to $[2, -1]$ ($\bar{1}$ would mean 2^{nd} row and $\bar{2}$ would mean 1^{st} row with sign changed).

Being v the number of views, and l the number of bearing-only measurements we have vl equations. We have 3 motion parameters to compute for each robot location (except for the first one, because we locate it in the origin) and 2 parameters for each landmark. So the number of parameters to estimate is $3(v-1) + 2l - 1$ (-1 because we can only get the results up to a scale factor). If the number of images is 2, the problem is unsolvable, even with infinite number of landmarks ($vl \geq 3v - 3 + 2l - 1$). The minimum number of views necessary to solve this problem is 3, with at least 5 measurements.

The 1D trifocal tensor has 8 parameters up to a scale, so it can be estimated from 7 corresponding triplets. With calibrated cameras the following additional constraints apply [5]:

$$\begin{aligned} -T_{111} + T_{122} + T_{212} + T_{221} &= 0 \\ T_{112} + T_{121} + T_{211} - T_{222} &= 0, \end{aligned} \quad (6)$$

then only five three-view matches are needed.

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 [10]. The computation of the trifocal tensor can be carried out as explained above using Singular Value Decomposition (SVD). With more matches than the minimal case, 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 our work we have chosen *ransac* [11], which makes a search in the space of solutions using random subsets of minimum number of matches.

III. SCENE PLANE AND THE 1D TRIFOCAL TENSOR

When the robot moves in man-made environments, many times there are planes in the scene which can be used in the computation of the trifocal tensor, then the number of matches needed may be reduced. This idea has been applied in the case of three 3D views, with the 2D trifocal tensor, e.g. [12] [9]. In this section, we study that situation for the 1D trifocal tensor.

It is shown in the literature of multiple view geometry [13] that there exist a relation between the projections of a line in three images, the tensor \mathbf{T} defined between the three views and two homographies \mathbf{H}_{21} (from image 1 to 2) and \mathbf{H}_{31} (from image 1 to 3) corresponding to a transformation through the same plane but between different couple of images. There have been developed for line features in a 3D scene. We transfer those constraints to point features in a 2D scene with 1D projections. These new constraints are obtained as follows.

If we have a point projection in three views, \mathbf{u}, \mathbf{u}' and \mathbf{u}'' , and 2 homographies, \mathbf{H}_{21} and \mathbf{H}_{31} , the following relations are known for any point in the plane of the scene:

$$\mathbf{u}' = \mathbf{H}_{21}\mathbf{u}, \quad \mathbf{u}'' = \mathbf{H}_{31}\mathbf{u} \quad (7)$$

On the other hand, the constraint imposed by the 1D trifocal tensor (5) can be reordered as,

$$u_1(T_{111}u_1'u_1'' + T_{112}u_1'u_2'' + T_{121}u_2'u_1'' + T_{122}u_2'u_2'') + u_2(T_{211}u_1'u_1'' + T_{212}u_1'u_2'' + T_{221}u_2'u_1'' + T_{222}u_2'u_2'') = 0.$$

$$\text{If we name } \mathbf{T}_1 = \begin{bmatrix} T_{111} & T_{112} \\ T_{121} & T_{122} \end{bmatrix} \text{ and } \mathbf{T}_2 = \begin{bmatrix} T_{211} & T_{212} \\ T_{221} & T_{222} \end{bmatrix},$$

this equation can be written as $[u_1 u_2] \begin{bmatrix} \mathbf{u}'^T \mathbf{T}_1 \mathbf{u}'' \\ \mathbf{u}'^T \mathbf{T}_2 \mathbf{u}'' \end{bmatrix} = 0$.

Therefore, we have the following equality up to a scale (\cong)

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \cong \begin{bmatrix} \mathbf{u}'^T \mathbf{T}_2 \mathbf{u}'' \\ -\mathbf{u}'^T \mathbf{T}_1 \mathbf{u}'' \end{bmatrix}, \quad \text{and substituting with (7) we get}$$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \cong \begin{bmatrix} \mathbf{u}^T \mathbf{H}_{21}^T \mathbf{T}_2 \mathbf{H}_{31} \mathbf{u} \\ -\mathbf{u}^T \mathbf{H}_{21}^T \mathbf{T}_1 \mathbf{H}_{31} \mathbf{u} \end{bmatrix}. \quad (8)$$

From (8) we obtain 4 additional constraints for the 1D tensor:

- First, equation (8) must be certain for any point \mathbf{u} . So let us consider that \mathbf{u} could be in the form $\mathbf{u} = [0 \ u_2]^T$ or $\mathbf{u} = [u_1 \ 0]^T$. Replacing in that equation with \mathbf{u} in these two special forms and developing the expressions we get two new constraints (to simplify the expressions, let us name $\mathbf{B}_1 = \mathbf{H}_{21}^T \mathbf{T}_2 \mathbf{H}_{31}$ and $\mathbf{B}_2 = -\mathbf{H}_{21}^T \mathbf{T}_1 \mathbf{H}_{31}$):

$$\begin{aligned} B_1(2,2) &= 0 \\ B_2(1,1) &= 0 \end{aligned} \quad (9)$$

where $B_n(a,b)$ means (row a , column b) of the matrix \mathbf{B}_n .

- Moreover the scale factor must be the same for both \mathbf{u} components in (8), therefore $\frac{\mathbf{u}^T \mathbf{H}_{21}^T \mathbf{T}_2 \mathbf{H}_{31} \mathbf{u}}{u_1} = \frac{-\mathbf{u}^T \mathbf{H}_{21}^T \mathbf{T}_1 \mathbf{H}_{31} \mathbf{u}}{u_2}$ must be true. Developing this expression we get the other two new constraints:

$$\begin{aligned} B_1(1,1) &= B_2(2,1) + B_2(1,2) \\ B_2(2,2) &= B_1(2,1) + B_1(1,2). \end{aligned} \quad (10)$$

It is known that 3 matched features at least are needed to compute 1D homographies from visual data in two 1D projections [14]. The corresponding coordinates in the projective space \mathcal{P}^1 of the matched features in first and second

images ($\mathbf{u} = [u_1, u_2]^T$ and $\mathbf{u}' = [u_1', u_2']^T$) are related through the homography \mathbf{H}_{21} : $\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} = \mathbf{H}_{21} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, with $\mathbf{H}_{21} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$, what provides one equation to solve \mathbf{H}_{21} :

$$\begin{bmatrix} u_1 u_2' & u_2 u_2' & u_1 u_1' & u_2 u_1' \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \end{bmatrix} = 0.$$

With the coordinates of at least three bearing-only measurements, in our case three vertical lines, we can construct a 3x4 \mathbf{A} matrix. The homography solution corresponds to the eigenvector associated to the least eigenvalue of the $\mathbf{A}^T \mathbf{A}$ matrix and it can be solved by singular value decomposition of matrix \mathbf{A} . Similarly to obtain \mathbf{H}_{31} .

The coplanarity condition reduces in one the minimum number of matched features needed to compute the tensor. Therefore, the tensor, in the calibrated case, can be computed from 4 matched features, three of them being coplanar in the scene. This tensor gives a general constraint for all observed landmarks from the three robot positions, independently of its location in the scene. This reduction of the minimum number of matches is specially convenient due to the robust technique used. In this case instead of doing a random search in a 5 degrees of freedom (d.o.f.) space of solutions, we have to do a search in a 3 d.o.f. space, to robustly estimate the homography and the features belonging to it, plus a second search in a 1 d.o.f. space of solutions, to estimate the tensor with the homography plus a feature match which is out of the plane.

For instance, let us suppose a situation with 40% of outliers. If we execute the algorithm for the 5 matches tensor, the ransac algorithm needs to perform 57 iterations to get a result with 99% probability of being correct. On the other hand, if we choose the 4 matches tensor estimation, the ransac algorithm will just need 19 (for homographies) + 6 (for tensor) iterations for the same level of confidence. To sum up, around twice more time required for the classical way of estimating the tensor. However, we should notice that this big difference is realistic only in the case that the plane is dominant in the scene. Otherwise we should consider higher level of outliers for the plane based method than for the general ones. Then, the outliers would be not only the wrong matches but also many matches which do not belong to the plane. If we suppose a 50% of outliers in that search, the number of iterations obtained (35+7 iterations) would still be lower than the 5 matches tensor, with the advantage that the estimation of the homography can give us some clue about singular situations (e.g. when all the scene is explained by it because it is totally planar).

IV. ROBOT AND LANDMARK LOCALIZATION

When the motion is performed on a plane, 6 parameters, up to a scale factor for translation, should be computed: $\theta', \mathbf{t}', \theta'', \mathbf{t}''$ (Fig. 1). The algorithm we use to compute motion recovers the epipoles with a technique proposed for the 3D case [15], also applied for the 2D case in [2]. We have also used it in a general scene with omnidirectional images [3].

Here we explain a summary of this method to get the localization parameters from the 1D trifocal tensor:

- The directions of translation are given by the epipoles and the rotations between robot positions are obtained by trigonometric relations between the epipoles. To get the epipoles, we first obtain the intrinsic homographies of the tensor, corresponding to the X and Z axis of the images. From these homographies, we compute the corresponding homologies for the tree images. The epipoles are the eigenvectors mapped to themselves through these homologies.

- Once the rotation and translation between the cameras have been obtained, the landmark localization is computed solving the system of the 3 equations (two are enough) which project them in \mathcal{P}^1 (1) in the three images.

The method provides two solutions for the motion parameters, defined up to a scale for the translations and landmarks location. This scale and ambiguity problem can be solved easily with some extra information. It can be obtained for example from odometry or from other previous knowledge of the scene.

V. EXPERIMENTAL RESULTS

In this section we show several experiments with simulated data to show the trifocal tensor performance in motion computation, when estimated with 5 matches or with a scene plane and 4 matches. There are shown also experiments with different types of real images, to show the performance in robot localization and landmarks reconstruction.

A. Simulation Experiments

First, some tests were run with simulated data to establish the performance of motion estimation through the trifocal tensor. We implemented a simulator of 2D scenes which are projected into 1D virtual cameras with field of view of 53° and 1024 pixels. We present the results for two different simulated movements, MovA and MovB. The first one could fit a common multi-robot configuration, and the second one represents a typical situation with a mobile robot going forward. In Fig. 2 we can see the position of the cameras, its field of view and the localization of the features in each movement.

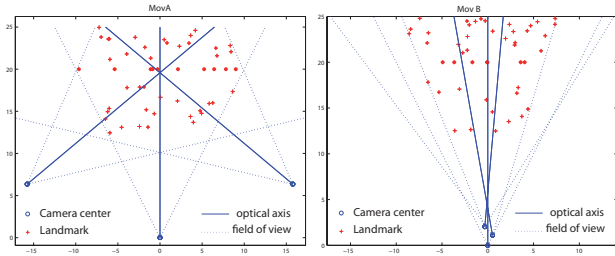


Fig. 2. *MovA and MovB*. Two simulated scenarios, showing landmarks and 3 camera positions with their corresponding field of view.

Measurement errors were simulated as gaussian random noise (of zero mean and standard deviations varying from 0 to 1 pixel) added to features image coordinates. Each experiment was repeated 100 times. The evaluation parameters shown for

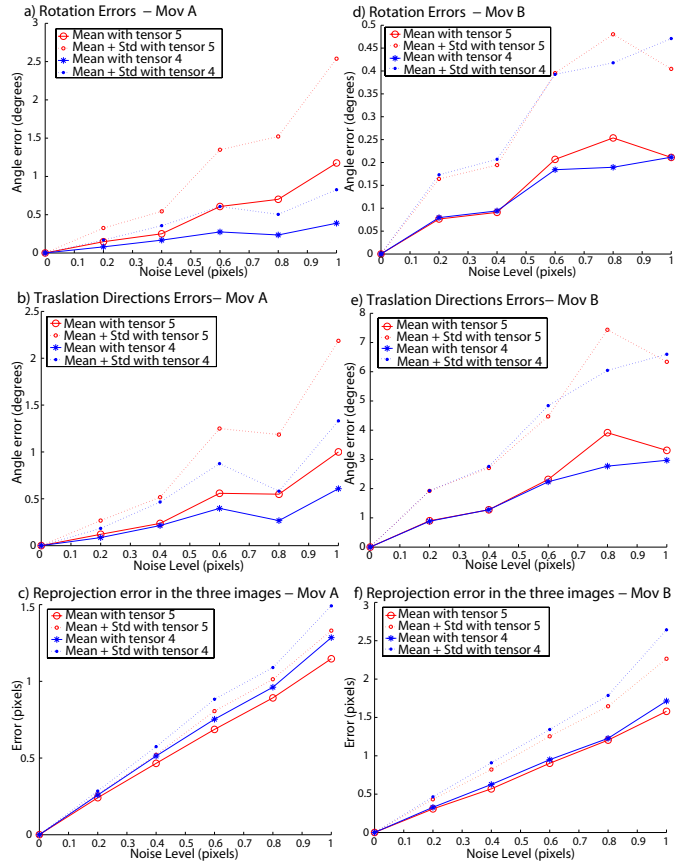


Fig. 3. *Dominant Plane case*: 20 matches in the plane and 10 out of it. Trifocal Tensor estimated with 5 and with 4 matches (100 executions for each case with different random matches). RMS error in rotation, translation direction and reprojection for MovA (left) and MovB(right) of Fig. 2.

each of them are: the RMS (root-mean-square) error in the computation of the rotation angles (θ' and θ''), the RMS error in the directions of translation (t' and t'') and the average RMS features reprojection error in the three images.

We took into account that there is a plane in the scene, supposing the features that belong to the plane are known. In this situation, we can estimate the tensor with 1 match less than in a general case, as explained in Section III. We considered a plane parallel to the first image, placed 20 units ahead the origin, in both scenarios (Fig.2).

We evaluated the performance in the localization with the two ways to estimate the tensor, with 4 matches or with 5. There were different cases of study, depending how many matched features belong to the plane: most of them in the plane (dominant plane in the scene), equally distributed (no dominant plane) or all of them in the plane.

Simulating a general scenario (when there is no dominant plane, but still a plane exists), we obtained very similar results for both tensors. In these simulations we generated 10 matches on the plane and 20 or 30 out of it. In all the cases errors were similar between methods, an example is shown in Fig. 4. However, when the plane is the predominant element in the scene (most matched features are on it), we found some

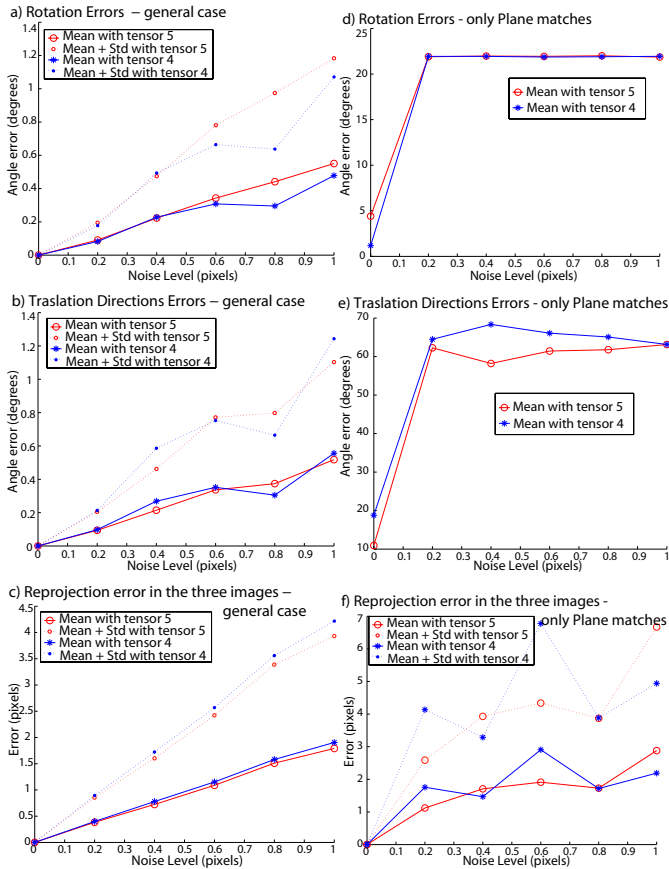


Fig. 4. Trifocal tensor estimated with 5 and with 4 matches (100 executions for each case with different random matches). RMS error in rotation, translation direction and reprojection for MovA of Fig. 2. Left: *General case*, 10 matches in the plane and 20 out of it (no dominant plane). Right: *All matches in the plane*, singular situation where the tensor does not exist.

advantages in the use of the 4 matches tensor. In Fig. 3 we can see the comparison between results from the tensor using plane constraints (with 4 matches, *tensor4*) and from the general tensor (with 5 matches, *tensor5*). In these simulations, we generated 20 random matches on the plane and 10 out of it. We can observe for MovA that *tensor4* behaves better than *tensor5*, specially when the noise increases. Computing this 4 matches tensor has another advantage, as the intermediate estimation of the homography can give us a clue about being in singular situations. For example, if the whole scene can be explained with the homography (all the matches fit it), we have a planar scene and then there is no sense to continue with the tensor estimation, as it does not exist in those cases. We also tried the localization estimation if all the matches belong to the plane, and the expected bad results can be seen in Fig. 4.

B. Real images Experiments

We show two examples with real images, one with conventional and other with omnidirectional cameras. In the first one we want to show the results with a scene where the plane is dominant. In the second one, when it is not. The scale factor was solved using data from the ground truth (only one known distance is necessary). The line matching is not the

subject of study here, we used methods developed in previous works, both for conventional images [10] and omnidirectional ones [3]. They are based in nearest neighbor search over the descriptors of the region around each line, and in checking some topological consistence similarly to the work in [16].

1) *Using conventional cameras (1R)*: In this experiment we used a conventional camera with known calibration matrix. We automatically extracted and matched vertical lines in three views. These 3 images with the line matches (the ones used to estimate the plane are in red) and a scheme of these features reconstruction is shown in Fig. 5. In this case, as the plane is dominant in the scene, it is possible to estimate robustly which features belong to a plane. The tables at the bottom of the same figure contain the localization errors: rotations (θ' , θ'') and direction of translations (t' , t''), as well as the feature reconstruction errors: in the image reprojection and in the 2D reconstruction in the scene. The ground truth motion for this experiment was obtained with the aid of *Photomodeler* software, where a set of points are manually given to get a photogrammetric reconstruction. Similarly to the simulation results, we observe better performance in the results from the tensor computed through a homography (TT4). We made different tries using more or less matches from out of the plane. As could be expected, the more we decreased those set of matches which do not belong to the plane, the more the errors with the TT5 increased.

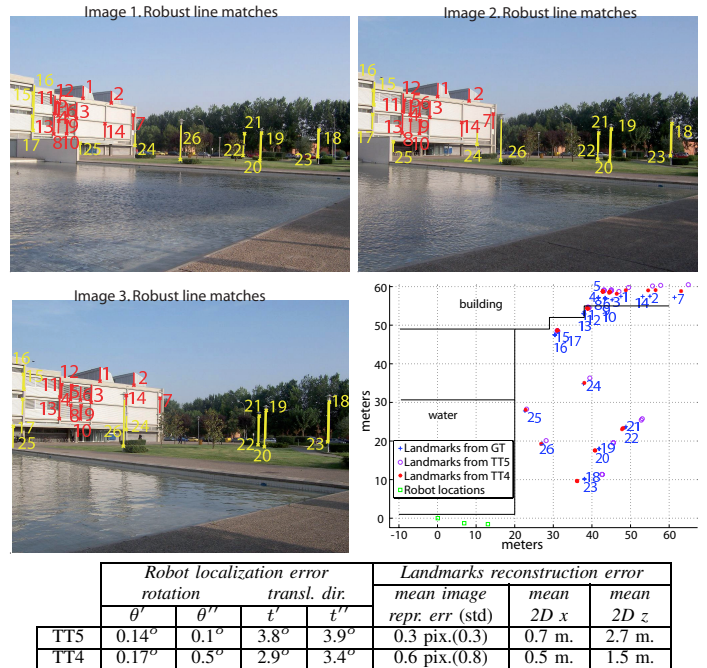


Fig. 5. *Experiment 1R*. Top & Middle-left: Outdoor real images with line robust matches. Coplanar lines [1..14] marked in red. Middle-right: Scene scheme with robot locations and landmarks reconstruction obtained through the classical tensor (TT5 in pink o), through the tensor with an homography (TT4 in red *) and landmarks location obtained from the ground truth motion (obtained with *Photomodeler*, in blue +). Bottom: robot and landmarks localization errors.

2) *Using omnidirectional cameras (2R)*: Next we show an example with omnidirectional images. In this case the

calibration of the camera is not necessary. If we suppose squared pixels, it is only required the center of projection. It does not coincide with the center of the image, but we estimate it automatically from the radial lines [3]. The trifocal tensor for this kind of image is also robustly computed from the projected radial lines (vertical landmarks of the scene). With this kind of images, the segmentation of the lines belonging to the same plane in the three views is a more difficult task. This is due to the wide field of view from the scene, what can make that many planes are visible all the time. Also the number of lines belonging to one specific plane may be too small, preventing from their automatic detection. In this cases, the homography/plane inliers are obtained using a priori knowledge about the scene. This problem has to be deeply studied in future works. In this experiment we selected them manually. In Fig. 6 the three views used are shown with the matched lines. The lines used to estimate an homography are marked in red. There we see also the scheme of the reconstruction, where the good performance of the proposal can be appreciated. Results from both tensors, as expected from the simulations, are quite similar, with the before mentioned advantages of the intermediate estimation of an homography. The errors in the rotation (θ' , θ'') and translation direction (t' , t'') estimation are shown in a table at the bottom of the figure, together with the features reconstruction errors: image reprojection and 2D reconstruction in the scene. Here the ground truth motion was obtained with a metric tape and a goniometer.

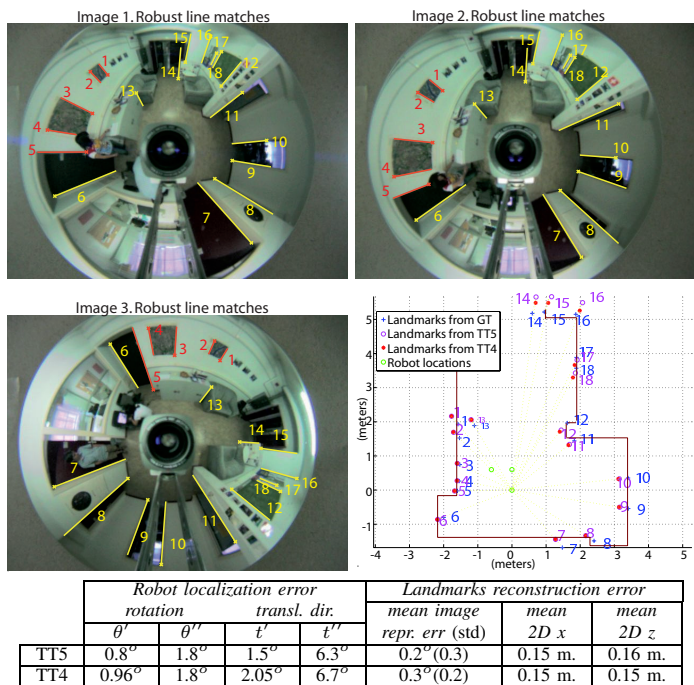


Fig. 6. Experiment 2R. Top & Middle-left: Indoor omnidirectional images with robust line matches. Coplanar lines [1..5] marked in red. Middle-right: Scene scheme with robot locations and landmarks reconstruction obtained through the classical tensor (TT5 in pink o), through the tensor with an homography (TT4 in red *) and landmarks location obtained from the ground truth motion (measured with a metric tape, in blue +). Bottom: robot and landmarks localization errors.

VI. CONCLUSIONS

In this paper we have presented a method to recover robot and landmark localization through a trifocal tensor. It is a low complexity (linear) method that takes profit of planes in the scene. It uses bearing-only measurements, e.g. obtained from conventional or omnidirectional images. An important advantage is that it can be estimated with only four matches, when three of them are located in a plane of the scene. This makes the method computationally less expensive than other similar ones and suitable for real time applications. There is no loss in performance in general cases, and it even gives better results when there is a dominant plane in the scene. Also notice the possibility of detect singular situations automatically in the intermediate step of homography estimation. The simulation and real images experiments show the good performance of our proposal, with quite low errors in localization and reconstruction, proving its suitability for robotic tasks, such as bearing only SLAM initialization or multi-robot localization.

REFERENCES

- [1] J.A. Castellanos, J. Neira, and J.D. Tardós. Multisensor fusion for simultaneous localization and map building. *IEEE Trans. Robotics and Automation*, 17(6):908–914, 2001.
- [2] F. Dellaert and A. Stroupe. Linear 2d localization and mapping for single and multiple robots. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*. IEEE, May 2002.
- [3] C. Sagues, A.C. Murillo, J.J. Guerrero, T. Goedemé, T. Tuytelaars, and L. Van Gool. Localization with omnidirectional images using the 1d radial trifocal tensor. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2006.
- [4] J. Sola, A. Monin, M. Devy, and T. Lemaire. Undelayed initialization in bearing only slam. In *IEEE/RSJ Int. conf. on Intelligent Robots and Systems*, 2005.
- [5] K. Åström and M. Oskarsson. Solutions and ambiguities of the structure and motion problem for 1d retinal vision. *Journal of Mathematical Imaging and Vision*, 12:121–135, 2000.
- [6] 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*, 22(10):1179–1185, 2000.
- [7] S. Thirithala 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 (CVPR-05)*, 2005.
- [8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.
- [9] L. Zelnik-Manor and M. Irani. Multiview constraints on homographies. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(2):214–223, 2002.
- [10] J.J. Guerrero, C. Sagüés, and A.C. Murillo. Localization and bearing-only data matching using the planar trifocal tensor. Technical report - 2005-v06, DIIS - I3A Universidad de Zaragoza, 2005.
- [11] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley, New York, 1987.
- [12] G. Cross, A. W. Fitzgibbon, and A. Zisserman. Parallax geometry of smooth surfaces in multiple views. In *Proc. of the 7th Int. Conference on Computer Vision*, pages 323–329, September 1999.
- [13] O. Faugeras, Quang-Tuan Luong, and T. Papadopolou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, 2001.
- [14] J.J. Guerrero, R. Martinez-Cantin, and C. Sagüés. Visual map-less navigation based on homographies. *Journal of Robotic Systems*, 22(10):569–581, 2005.
- [15] A. Shashua and M. Werman. Trilinearity of three perspective views and its associate tensor. In *Proc. of the International Conference on Computer Vision (ICCV)*, pages 920–925, June 1995.
- [16] H. Bay, V. Ferrari, and L. Van Gool. Wide-baseline stereo matching with line segments. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, June 2005.