A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs

Yalin Bastanlar

Department of Computer Engineering, Izmir Institute of Technology, Izmir 35430, Turkey
Tel.: +90-232-750-7883; fax: +90-232-750-7862; e-mail: yalinbastanlar@iyte.edu.tr

A B S T R A C T

The external calibration of a camera system is essential for most of the applications that involve an omnidirectional and a pan-tilt-zoom (PTZ) camera. The methods in the literature fall into two major categories; 1) a complete external calibration of the system which allows all degrees of freedom but highly time consuming, 2) spatial mapping between the pixel coordinates in omnidirectional camera and pan/tilt angles of the PTZ camera instead of explicitly computing the rotation and translation. Most methods in this category make restrictive assumptions about the camera setup such as optical axes of the cameras coincide. We propose an external calibration method that is effective and practical. Using the two-view geometry principles and making reasonable assumptions about the camera setup, calibration is performed with just two scene points. We extract rotation using the point correspondences in images. Locating the PTZ camera in the omnidirectional image is used to find the translation parameters and the real distance between the two scene points lets us compute the translation in correct scale. Results of the simulated and real image experiments show that our method works effectively in real world cases and its accuracy is comparable to the state-of-the-art methods.

1. Introduction

Hybrid camera systems consisting of an omnidirectional camera and a pan-tilt-zoom (PTZ) camera are widely used especially in surveillance applications. An omnidirectional camera provides 360° horizontal field of view with a low resolution whereas a PTZ camera provides high resolution images viewing a certain direction. A hybrid system combines the powerful aspects of both camera types and aims wide-angle high resolution surveillance. A typical task is to detect a moving object via omnidirectional camera and directing the PTZ camera towards the position of the moving object [1].

The external calibration of a hybrid system, i.e. estimation of camera poses with respect to each other, is fundamental for a cooperative use. Previously proposed calibration methods are either not practical enough to effectively determine the extrinsic parameters or they make restrictive assumptions which limit the applicability. This is the main motivation of our study.

We propose a practical calibration method that is based on two-view geometry principles and makes reasonable assumptions about the camera setup. Our method firstly extracts rotation using only two scene points and their pixel coordinates in the hybrid image pair. Afterwards, PTZ camera is located in the omnidirectional image and its pixel coordinates are used to find the translation parameters. Finally, the real 3D distance between the two scene points lets us compute the translation in correct scale, which is the distance between camera centers. Intrinsic calibrations of both cameras are obtained a priori.

The organization of the paper is as follows. In Section 2, we summarize the related work and explain the difference in our method. In Section 3, we explain the steps of the proposed two-point calibration method. The results of our experiments are presented in Section 4. The average accuracy obtained with both synthetic and real images are given together with a discussion comparing our results with other state-of-the-art methods. Section 5 summarizes the conclusions of our study.

2. Related work

A significant portion of the previously proposed methods perform a complete external calibration of the hybrid system without restricting the rotation and translation between the hybrid camera pair. Although these methods provide accurate results, the calibration procedures are time consuming due to extracting required number of point correspondences. Moreover, in most cases these methods are computationally expensive. For instance, a large pattern on the floor is required for the method in [2]. Following the internal calibration of the omnidirectional camera at the ceiling, the geometric relationship between omnidirectional and perspective camera is derived using point correspondences in both camera images. Then the perspective cameras are calibrated. For the method in [3], a calibration pattern is required to be captured at different spatial positions. In [4], the extrinsic parameters (rotation and translation between the cameras) of the hybrid system are extracted via 3D Euclidean reconstruction of scene points following projective reconstruction by factorization which is computationally involved and also expensive due to using non-linear minimization techniques such as Levenberg–

Marquardt. In another method falling into this category [5], external calibration is performed by solving the epipolar geometry between the two cameras. The relative position of the active camera with respect to the static camera is solved together with the parameters associated with active camera's pan and tilt mechanism. In [6], authors focus on multi-view structure-from-motion and perform external calibration with essential matrix estimation using many points. The omnidirectional and perspective cameras are moving freely in 3D space with any orientation.

Another major group of approaches, that use PTZ camera in a hybrid system, do not solve for extrinsic parameters explicitly. Instead, they compute a spatial mapping between omnidirectional and pan-tilt parameters of the PTZ camera. In other words, they estimate the corresponding pan and tilt angles of the PTZ camera for a given pixel coordinate in omnidirectional image. It is assumed that the pan and tilt angles of a PTZ camera are highly correlated with the corresponding pixel coordinates in the omnidirectional image. For some methods, this mapping is based on data collection and fitting (interpolation) where no geometric information about the camera setup is used ([7], [8]). Other methods of spatial mapping (such as [9], [10], [11], [12], [13]) make assumptions to be able to use geometric constraints. One of the most common assumptions is that the optical axes of the two cameras coincide [12] (i.e. one is on top of the other). In some studies, this assumption is even further extended. For instance, in [10] authors assume that the camera origins are at the same location which is not possible in a real camera setup. They relate the distance between a pixel and the center of the omnidirectional image to the vertical angle (tilt) that is used to move the PTZ camera.

Another common restriction/assumption for a camera setup is that the relative position and orientation of the cameras are known. Examples are given in [9], [11] and [14]. Instead of employing an external calibration method, such as ours, they use the manually measured distance between the cameras. In [11], the height of the omnidirectional camera is fixed as well to be able to measure the distance to the object of interest.

In [13], Tan *et al*. propose a method to calculate the relative position of the optical centers of the two cameras based on parameters extracted from two sample scene points. Then, they use this relative distance information as an input to spatial mapping. For relative position estimation, they assume that the sample points (also target object to be tracked) are on a 2D plane and optical axes of both omnidirectional and the PTZ cameras are perpendicular to this plane.

The methods in this second group are more practical than the ones in the first group (complete external calibration), however the assumptions they make can be too restrictive due to several reasons: 1) Optical axes of the two cameras may not coincide. This assumption can be satisfied only for the setups where one camera is exactly on top of each other. 2) The distance between the cameras may not be measured manually. One may not be able to put a measuring tape (or a measuring laser) between them due to an obstacle. Putting a great effort to manually measure the distance is not practical since it is to be repeated when any of the cameras is moved. (e.g. cameras in a parking lot are moved for a better or a different view [13]). This requires a practical method for re-calibration. 3) The point correspondences used for calibration may not be on a 2D plane, or this plane may not be perpendicular to the optical axes. Such an assumption restricts the method to use a suitable (or prepared) surface with planar feature points on it.

Our method does not make the assumptions made by the previously proposed practical solutions, therefore the restrictions mentioned in the previous paragraph do not exist in our approach. Only assumption we make is mounting the cameras to the surfaces that are parallel to the ground, which can be satisfied by using man-made surfaces such as ceilings (indoor or outdoor). Our method is similar to the relative position estimation method in [13] but without the restriction of point correspondences should lie on a 2D plane which is perpendicular to the optical axes of the cameras. Here, we do not propose a spatial mapping method (as done in [9], [10], [11], [12], [13]), although the parameters estimated with our method can be used for any spatial mapping method.

3. Our method

Our method is based on the principles of two-view camera geometry where a 3x3 matrix, called the fundamental matrix, encompasses the geometric relation (the translation and rotation) between the two cameras or views from two different positions of a camera. With the standard method, eight point correspondences between the two views are required to compute the fundamental matrix [15], or 7 point correspondences if the rank constraint is used.

In case of calibrated cameras, another matrix, called the essential matrix, can be computed with the point correspondences. Rotation and translation parameters can be extracted from the essential matrix easily. It is possible to compute the essential matrix with as few as five point correspondences [16], instead of eight, however the algorithm is computationally involved as it requires Gröbner basis solver to find the roots of a tenth degree polynomial.

Reasonable assumptions about the camera setup let us perform the external calibration with easier procedures and less number of point correspondences. For instance, when the optical axes of the two cameras coincide (e.g.[10],[12]) there is only two degrees of freedom: the translation in the vertical direction and the rotation around the common optical axis. As mentioned previously, we do not assume a certain camera setup and we aim to develop a practical external calibration method that can be used when the cameras are moved. However, one reasonable assumption we make is mounting the cameras to the surfaces that are parallel to the ground such as the ceiling. This makes the optical axis of the omnidirectional camera perpendicular to the ground. As for the PTZ camera, there are two more rotational degrees of freedom which are the rotation between the cameras around the optical axis and the tilt angle. We set the PTZ camera to the docking reference (i.e. zero pan and tilt) however the angle between the zero pan and the coordinate system of the second camera is still to be estimated. We denote this angle with β in the rest of the paper. Tilt angle of the PTZ camera, on the other hand, is assumed to be zero relying on the docking reference. In the following, we explain how the remaining extrinsic parameters (β angle and the translation vector) of such setup are solved using only two point correspondences.

3.1. Intrinsic calibration

Our method employs an omnidirectional and a PTZ camera for which internal calibrations are obtained a priori. This is not a strong assumption since there are readily available toolboxes and it has to be done just once. Only the zoom parameter of PTZ camera affects internal calibration, which can be fixed to a certain value during the calibration task. Once the internal and external calibration tasks are over, zooming in and out does not affect extrinsic parameters.

For PTZ camera calibration, we use the method proposed in [17] and implementation is available as a MATLAB Toolbox [18]. For the omnidirectional camera, we use the sphere camera model [19] which is able to cover both catadioptric (mirrored) omnidirectional cameras and fisheye cameras. There are a few calibration methods proposed for the calibration using the sphere model [20][21], we preferred to employ [20] since a MATLAB toolbox is provided.

3.2. Two-view geometry

Let us briefly review the geometry between two cameras (or views). Coordinates of points in two views have the following relationship

$$\boldsymbol{x}'^T \boldsymbol{F} \boldsymbol{x} = 0 \tag{1}$$

where $\boldsymbol{F}$ is the 3x3 fundamental matrix, and $\boldsymbol{x}'$ and $\boldsymbol{x}$ are the corresponding points in the two images in homogeneous coordinates. When the cameras are calibrated, fundamental matrix can be formulized using the calibration matrices of the two cameras, $\boldsymbol{K}'$ and $\boldsymbol{K}$, as follows

$$\boldsymbol{F} = \boldsymbol{K}'^{-T} \boldsymbol{E} \boldsymbol{K}^{-1} \tag{2}$$

where $\boldsymbol{E}$ is the 3x3 essential matrix satisfying the following equation

$$\widehat{\boldsymbol{x}}'^T \boldsymbol{E} \widehat{\boldsymbol{x}} = 0 \tag{3}$$

This looks similar to the fundamental matrix relationship with the difference that $\widehat{\boldsymbol{x}}'$ and $\widehat{\boldsymbol{x}}$ are normalized 3D rays. These are the vectors representing rays outgoing from camera center and passing through the image point. Essential matrix actually encompasses the rotation and translation information between the camera views and can be written as

$$\boldsymbol{E} = [\boldsymbol{t}]_{\times} \boldsymbol{R} \tag{4}$$

where $[\boldsymbol{t}]_{\times}$ is the cross product in matrix form obtained with the translation vector $\boldsymbol{t}$, and $\boldsymbol{R}$ is the 3x3 rotation matrix.

Essential matrix is estimated with point correspondences from the two images. Each correspondence gives an equation that is equal to zero. Then, the rotation and translation parameters can be obtained from $\boldsymbol{E}$. In our method, as will be explained next, exploiting the constraints on the rotation and also extracting $\boldsymbol{t}$ from images, we are able to estimate $\boldsymbol{E}$ with as few as two point correspondences.

3.3. Estimating the unknowns in our method

Fig.1 shows the geometric relation between omnidirectional camera and the PTZ camera. Omnidirectional camera is represented as a perspective camera looking upwards to a mirror surface but any single-viewpoint omnidirectional camera complies with our approach. The Z axis of the PTZ camera is looking towards the scene points.
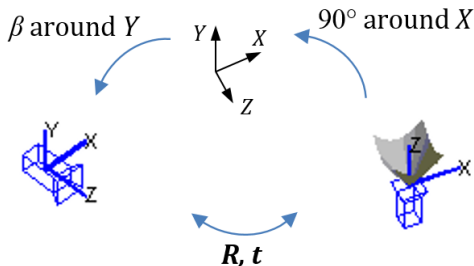

Fig. 1. Two view geometry (rotation and translation) between the omnidirectional and PTZ cameras. Omnidirectional camera is on the right, represented as a perspective camera looking upwards to a mirror. PTZ camera is represented as a perspective camera on the left.

We assume here the PTZ camera is in its docking reference, i.e. pan and tilt angles are zero. There is a rotation around the Y axis of the PTZ camera coordinate frame. This is the rotation between the docking reference of PTZ camera and the omnidirectional camera and it is denoted by β as mentioned earlier in this section. In addition, there is a 90° rotation around X axis. Since the cameras are mounted to the surfaces that are parallel to the ground, these two rotations are enough constitute the $\boldsymbol{R}$ between the cameras as follows:

$$\boldsymbol{R}_Y = \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix}, \boldsymbol{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(-90) & -\sin(-90) \\ 0 & \sin(-90) & \cos(-90) \end{bmatrix},$$

$$\boldsymbol{R} = \boldsymbol{R}_Y \boldsymbol{R}_X = \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 \\ 0 & 0 & 1 \\ -\sin(\beta) & -\cos(\beta) & 0 \end{bmatrix} \tag{5}$$

To make the estimation of $\boldsymbol{t}$ easier, we locate the PTZ camera in the omnidirectional image by manually marking the corresponding pixel in the omnidirectional image. This is not hard to achieve thanks to the wide FOV of the omnidirectional camera. In our experiments, for instance, we use a fisheye camera and it is enough not to place the PTZ camera at a higher level than the omnidirectional camera. An example of marking the PTZ camera center is given in Section 4.2. We will also present an analysis (in Section 4.1) to show how much our algorithm is robust to errors that occur while locating the PTZ camera.

The pixel coordinates of the marked PTZ camera location is converted to a 3D ray, $\boldsymbol{c}$, outgoing from the center of the omnidirectional camera. This is not the exact 3D location of the PTZ camera, however we know that the PTZ camera center is on vector $\boldsymbol{c}$. $\boldsymbol{R}$ and $\boldsymbol{c}$ together let us compute $\boldsymbol{t}$ up to a scale factor.

With the explained geometric relations, $\boldsymbol{E}$ is written as follows:

$$\boldsymbol{E}=[\boldsymbol{t}]_{\times}\boldsymbol{R}=\begin{bmatrix} -t_y \sin(\beta) & -t_y \cos(\beta) & -t_z \\ t_z \cos(\beta)+t_x \sin(\beta) & t_x \cos(\beta)-t_z \sin(\beta) & 0 \\ -t_y \cos(\beta) & t_y \sin(\beta) & t_x \end{bmatrix} \tag{6}$$

where $$[\boldsymbol{t}]_{\times}=\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{7}$$

and the components of the translation vector can be written as

$$\boldsymbol{t} = \boldsymbol{R} \cdot (-\boldsymbol{c}) = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 \\ 0 & 0 & 1 \\ -\sin(\beta) & -\cos(\beta) & 0 \end{bmatrix} \begin{bmatrix} -c_1 \\ -c_2 \\ -c_3 \end{bmatrix}$$

$$= \begin{bmatrix} -\cos(\beta)c_1 + \sin(\beta)c_2 \\ -c_3 \\ \sin(\beta)c_1 + \cos(\beta)c_2 \end{bmatrix} \tag{8}$$

Let $\widehat{\boldsymbol{x}}' = (x',y',1)$ and $\widehat{\boldsymbol{x}} = (x,y,1)$. Using equations (3) and (6), each point correspondence gives us an equation:

$$t_x(1+xy'\sin(\beta) + y'y\cos(\beta)) + t_y(y\sin(\beta) - x\cos(\beta)$$
$$- x'y\cos(\beta) - xx'\sin(\beta)) + t_z(xy'\cos(\beta) - x' - y'y\sin(\beta)) = 0 \tag{9}$$

This equation has one unknown, β, but two roots. Only one of these roots is true rotation angle around Y axis. To find the correct root we use two point correspondences, i.e. two equations, and select the common root. In real life cases, where the coordinates are noisy, the solutions that are close to each other are chosen.

Using two correspondences also gives us a chance to compute the actual scale of the translation, since our estimated translation vector was correct up to a scale factor. The two points used for rotation estimation can be reconstructed in 3D space after the camera matrices are obtained. This is done by triangulation [22]. The real distance between these two points can then be divided to the distance between estimated (reconstructed) points to obtain the scale ratio. This ratio is then used to correct the estimated translation, t.

3.4. The two-point calibration algorithm

The complete procedure of the proposed external calibration method can be enumerated as follows:

Step 1: Read the image coordinates of two sample points in both omnidirectional and PTZ camera images. Calculate the outgoing 3D rays of these sample points (using intrinsic parameters).

Step 2: Locate the PTZ camera in the omnidirectional image by manually marking the corresponding pixel. Calculate the 3D ray of that pixel, obtain c. This is the direction of PTZ camera center when looked from the omnidirectional camera center.

Step 3: Estimate β using the two-view geometry relations. Repeat this for two different point correspondences. Each correspondence gives two solutions, select the common one. If both solutions are common (points with same pan angle) choose another two-point pair.

Step 4: Estimate t using β (Step 3) and c (Step 2). The estimated translation at this point is correct up to a scale factor.

Step 5: Compute E with the estimated β and t. Calculate the 3D locations of the two sample points by triangulation. Also compute the distance between these reconstructed 3D points.

Step 6: Measure the distance between the two 3D points' real locations manually. The difference between the measured and the estimated (Step 5) distances is the scale ratio. Use this ratio to correct the scale of the t estimated in Step 4. The t obtained here is the final estimate of translation.

3.5. The rotation parameters when the PTZ camera is turned after calibration

During the calibration procedure, pan and tilt angles of the PTZ camera are set to the docking references (zero angle). Once the calibration is over, certainly the camera within the PTZ dome is rotated. This does not affect the estimated translation t since the camera centers are not moved; however applied pan/tilt angles change the rotations around X and Y axes and matrices given in Eq. 5 are no more valid. Let us explain how to update the rotation parameters.

With the external calibration, the angle between the omnidirectional camera and the zero pan of the PTZ camera is estimated (denoted by β). When the PTZ camera is rotated within its dome, the actual rotation between the current camera orientations can be found by adding the 'within-dome' rotation of PTZ camera (let us denote by ϕ) to the estimated β. As a result,

referring to Fig. 1, the rotation around Y axis anytime during the operation is equal to $\phi + \beta$.

Another concern is the tilt applied to the PTZ camera. The external calibration is performed when the tilt is zero (docking reference) and at that point there is a rotation of 90° around X axis (cf. R_X in Eq. 5 and Fig.1). After moving the PTZ camera within the dome with an amount of γ, total rotation around X axis becomes $\gamma+90°$.

4. Experiments

4.1. Experiments with simulated images

First, we analyze the proposed method in a simulated environment. Fig. 2 shows the simulated setup for the experiment. We place our omnidirectional camera to the origin. It is represented as a perspective camera looking upwards to a mirror. Simulated PTZ camera is located exactly at (-0.8, 0.2, 0). 200 points were randomly generated and distributed in a 3D volume to which perspective camera is directed. Fig. 3 shows sample omnidirectional and perspective images. The PTZ image is one megapixel, whereas the omnidirectional image is five megapixels. These resolution values are in accordance with the real cameras we possess which are used for the real image experiments.
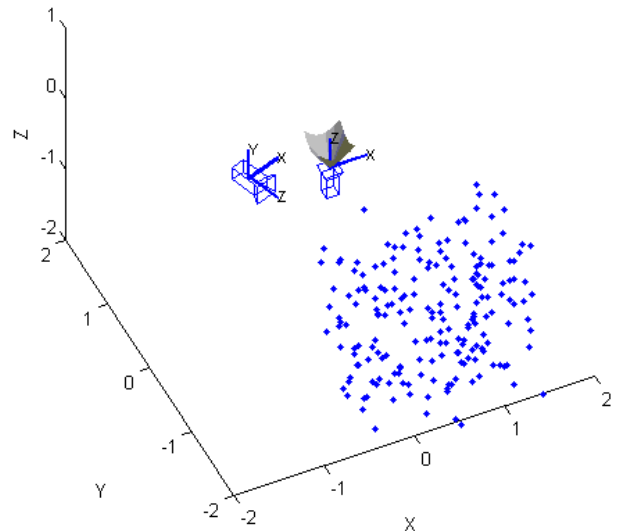

Fig. 2. Simulated environment with one omnidirectional camera, one PTZ (perspective) camera and a number of scene points.
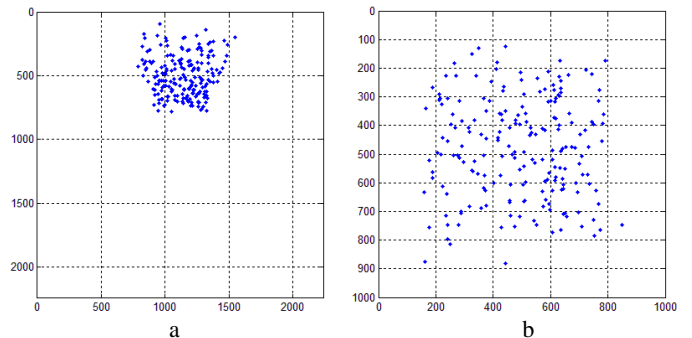

Fig. 3. Sample omnidirectional (a) and perspective (b) images captured in the simulated environment. Scene points cover the whole perspective image but only a part of the omnidirectional image.

When no noise is added to the point coordinates, the algorithm gives the correct values without any error. The only exception is that when both of the selected image points have the same pan angle, then two solutions are obtained, one of which is correct. In the experiments, we discard that point pair and pick a different one. In our simulations, we measured the percentage errors in estimated β and estimated t when different amounts and types of noise are added:

N1. Pixel coordinate noise: occurs while reading the corresponding point coordinates in the images. The standard deviation of the amount of the noise is incrementally increased from 0 to 1 pixel.

N2. PTZ camera location noise: occurs while locating the pixel coordinates of the PTZ camera within the omnidirectional image. The standard deviation of noise amount is incrementally increased from 0 to 5 pixels.

N3. Distance measurement noise: occurs while measuring (manually) the distance between the real 3D locations of the two selected point correspondences. The standard deviation of noise amount is incrementally increased from 0 to 5 percent.

N4. Tilt with respect to the ground noise: occurs when the zero tilt position of the PTZ camera is not perfectly parallel to the ground level. The standard deviation of noise amount is incrementally increased from 0 to 0.3 degrees.

For each different noise type and amount, the experiments were repeated 100 times and average error is recorded. In the experiment, results of which are given in Fig. 4, only pixel coordinate noise (N1) is added to the measurements obtained from the simulated environment. It can easily be inferred that an increase in coordinate reading noise increases the parameter estimation error. In the second experiment (Fig. 5), varying amount of PTZ camera location noise is added to a pixel coordinate noise of 0.4 pixels (N2+ fixed N1).

For the experiment in Fig. 6, varying amount of distance measurement noise is added to a pixel coordinate noise of 0.4 pixels and a PTZ camera location noise of 1.0 pixel (N3 + fixed N2 + fixed N1). Distance noise is given as a percentage of the original distance. E.g. 1% noise means 1 cm error in 1 m distance. We were expecting that only translation error is negatively affected by an increase in the distance measurement noise because the measured distance is used only for t estimation after β is estimated. The results are in accordance with this expectation.

For the experiment in Fig. 7, varying amount of tilt with respect to the ground noise is added to a pixel coordinate noise of 0.4 pixels and a PTZ camera location noise of 1.0 pixel (N4 + fixed N2 + fixed N1). Tilt noise is added as altered angles in both of the horizontal axes with a standard deviation of 0-0.3 degrees. We understand from the results that the estimations of translation vector and β angle are very sensitive to an existing tilt. More than 0.3 degree tilt severely disrupts the estimation accuracy. We will further discuss the effect of tilt while presenting real image experiments in the succeeding subsection.

4.2. Experiments with real images

Images of our experiments were captured with an Oncam Grandeye 360° omnidirectional camera with a fisheye sensor and a Samsung SNP-5300 PTZ camera pair. Both cameras were modelled with the sphere camera model and calibrated using the techniques given in Section 3.1. The cameras were fixed to tables that are supposed to be parallel to the ground. No extra effort was
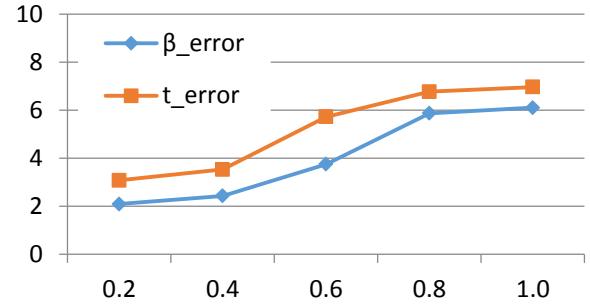

Fig. 4. Percentage errors in β and t estimates when varying amount of pixel coordinate noise is added. The std. dev. of the noise varies between 0.2 and 1 pixel. Original β is 0.35 radians (20°). Original $|t|$ is 0.8246 meters.
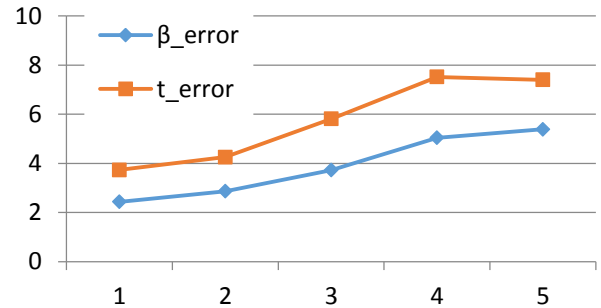

Fig. 5. Percentage errors in β and t estimates when varying amount of PTZ camera location noise is added. The std. dev. of the noise varies between 1 and 5 pixels. There is also a fixed amount of pixel coordinate noise (0.4 pixels). Original β is 0.35 radians (20°). Original $|t|$ is 0.8246 meters.
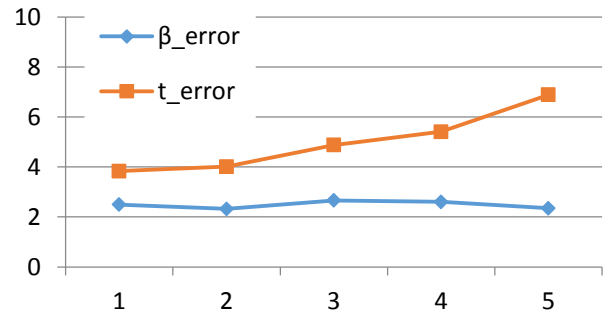

Fig. 6. Percentage errors in β and t estimates when varying amount of distance measurement noise (1%-5%) is added. There is also a fixed amount of pixel coordinate noise (0.4 pixels) and PTZ camera location noise (1 pixel). Original β is 0.35 radians (20°). Original $|t|$ is 0.8246 m.
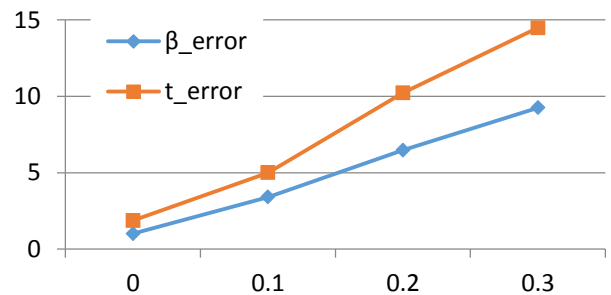

Fig. 7. Percentage errors in β and t estimates when varying amount of tilt w.r.t. the ground noise is added. The std. dev. of the noise varies between 0 and 0.3 degrees in both horizontal axes. There is also a fixed amount of pixel coordinate noise (0.4 pixels) and PTZ camera location noise (1 pixel). Original β is 0.35 radians (20°). Original $|t|$ is 0.8246 m.

spent to make them perfectly parallel. The true distance between the camera centers were 105 cm (measured manually). A hybrid image pair can be seen in Fig. 8.

As summarized in Section 3.4, our algorithm starts with selecting two point correspondences in the hybrid image pair. To be able to repeat the experiment many times and record the average performance, we selected more than 10 point correspondences for each of the hybrid image pair used. An example set of correspondences are also marked in Fig. 8. The next step is locating the PTZ camera in the omnidirectional image. The location of the PTZ camera is marked with a large red circle in the omnidirectional image. The center of that circle, marked with a red dot, is assumed to be the center of the PTZ camera, coordinates of which is used in the rest of the algorithm. After β and t are estimated (Steps 3 and 4), the 3D locations of the two sample points are computed by triangulation (Step 5). In Fig. 9, we show a top-view of the 12 reconstructed 3D scene points, two of which are used at each trial of our method.
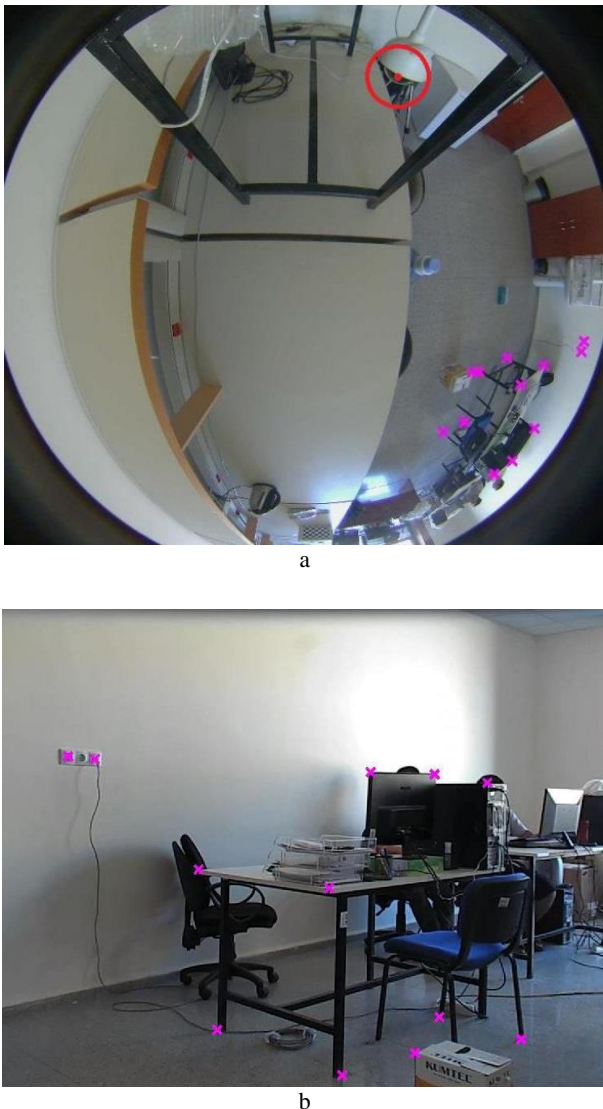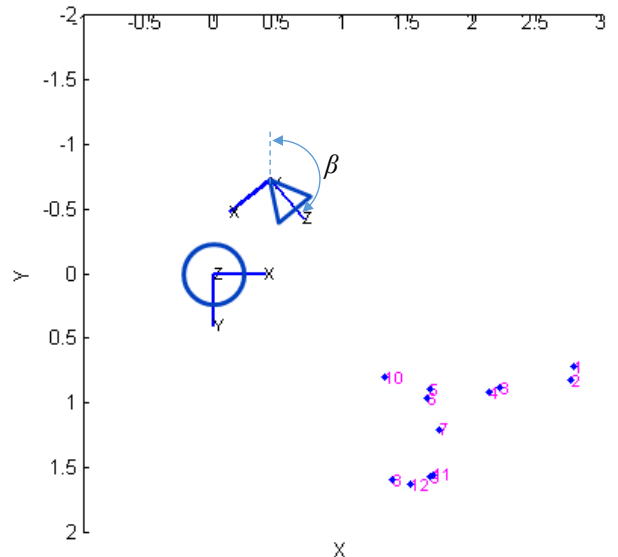

Fig. 9. A top-view of the reconstructed cameras and 12 reconstructed 3D scene points. The omnidirectional camera is indicated with a circle around it, Z axis of which is looking down. Z axis of the perspective camera (PTZ camera) is looking toward the reconstructed scene points. The distance between any two of these reconstructed points is compared with the measured real distance between the points to obtain the scale ratio.

Finally, the ratio between the measured distance between the two 3D points' real locations and the estimated distance (previous step) is assigned as the scale ratio. We use this ratio to refine the t estimate, magnitude of which is supposed to be equal to the actual distance measured between camera centers (105 cm). We have performed 30 experiments of estimating t using three different hybrid image pairs. The average of estimated $|t|$ is 105.5 cm where the standard deviation is 4.5 cm. Average percentage error is 3.1%, with a maximum error of 10.9% (estimated 93.5 cm). When we compare the average 3.1% distance error with the results of the simulated image experiments, we can conclude that the pixel coordinate noise, PTZ camera location noise and distance measurement noise are supposed to be below 0.4 pixels, 2 pixels and 2% respectively since these are the conditions that generate 4% error in $|t|$ estimation (cf. Figures 4, 5, and 6). Also having an average $|t|$ error of 3.1% indicates that tilt with respect to the ground is not more than 0.1° (cf. Fig.7).

Regarding the estimation of β, we do not have a ground truth value since the rotation between the camera axes cannot be perfectly set while fixing the cameras. However we are able to interpret the average and standard deviation of the estimates. The average of the β estimates for the image pair given in Figures 8 and 9 is 137.27° with a standard deviation of 0.31°. When the total of 30 estimates including different image pairs are considered, the standard deviation is 0.22°. This low variation in the estimates can be interpreted as an indication of high accuracy.

4.3. Comparison with other methods

We consider that the error in $|t|$ estimation is acceptable when compared with other methods in the literature. For the method given in [4], authors presented external calibration errors for varying amount of pixel coordinate noise (cf. simulated experiments results in Fig. 3 in [4]). When the noise is 0.4 pixels, the average rotation angle errors vary between 2°-5° and the translation error varies between 3% - 7% for each axis (would be larger for $|t|$). These error rates are higher than the ones we

Fig. 8. A pair of omnidirectional (a) and PTZ (b) camera images. 12 point correspondences which are used for the experiments are marked on both images.

obtained. The comparison given above made with first stage of the method in [4], which is solved linearly. We considered this stage as an alternative to ours since it is relatively simple, although it requires 9 to 36 point correspondences. In the second stage of the algorithm, the angle errors can be minimized to ~1° average and translation errors to ~1% average, however this requires applying projective reconstruction by factorization which is computationally expensive due to using non-linear minimization techniques such as Levenberg–Marquardt.

Among the previously proposed practical solutions for using an omnidirectional-PTZ camera pair, only of them actually estimates the relative position and orientation of the cameras, which is given in [13]. Therefore, we compared the accuracy of the parameter estimation of our method only with [13]. The authors of the study presented in [13] did not compare their distance estimation error with ground truth values. However they record their errors of pan and tilt angle estimates. Mean absolute errors were recorded as 2.15° and 0.77° for pan and tilt angles respectively. 2.15° makes an error of 11% when the actual value is 20°, which is the rotation value (β) set in our simulations. We generally obtained lower errors in β estimation (Figures 4-7). Moreover, in real image experiments we obtained a standard deviation of 0.22° in our estimates of β value.

Neglecting small rotation errors during setup makes our method very practical (external calibration with just two points). The effect of these small rotations would be significant if we were working on structure-from-motion and 3D reconstruction. However, in most of the applications including an omnidirectional-PTZ camera pair, the step after the external calibration is directing the active (pan-tilt) camera to the object located in the omnidirectional image. As we observe in the referred studies (e.g. [13]), the errors we face (namely <4° pan angle (β) error and 3% translation ($|t|$) error) do not prevent us from directing the PTZ camera successfully.

5. Conclusions

We have developed an effective and practical external calibration method for an omnidirectional-PTZ camera pair. The existing methods in the literature either solve the complete external calibration of the system allowing all degrees of freedom or they make restrictive assumptions about the camera setup such as camera optical axes should coincide or the selected scene points should lie on a 2D plane.

With the method proposed in this paper, using the principles of two-view geometry and making reasonable assumptions about the camera setup, calibration can be performed with just two scene points. We assume the optical axes of both cameras to be perpendicular to the ground which can be satisfied by using surfaces like a table or the ceiling. We also fix the PTZ camera to its docking reference (zero pan and tilt angles). We extract rotation between the camera coordinate systems using the point correspondences in the hybrid image pair. Locating the PTZ camera in the omnidirectional image is used to find the translation parameters and the real 3D distance between the two scene points lets us compute the translation in correct scale.

What we propose is not a spatial calibration method, but the extrinsic parameters estimated with our method can be used for any spatial mapping method. This would make most of the existing methods more versatile as we do not make restrictive assumptions.

We performed simulated and real image experiments to analyze the applicability and accuracy of the proposed algorithm. Results show that our method works in real world cases and the accuracy is comparable to the state-of-the-art methods. Although the accuracy could be increased by using more point correspondences, this would deteriorate the practical side of our method.

References

[1] Yao, Y., Abidi, B., Abidi, M.: 'Fusion of Omnidirectional and PTZ Cameras for Accurate Cooperative Tracking', Proc. of the IEEE Int. Conf. on Video and Signal Based Surveillance (AVSS), 2006

[2] Chen, X., Yang, J., Waibel, A.: 'Calibration of a Hybrid Camera Network', Proc. of the Ninth IEEE International Conference on Computer Vision (ICCV), 2003

[3] He, B., Chen, Z., Li, Y.: 'Calibration method for a central catadioptric-perspective camera system', J. Opt. Soc. Am. A, 2012, 29, (11), pp. 2514-2524

[4] Deng, X., Wu, F., Wu, Y., et al.: 'Self-calibration of hybrid central catadioptric and perspective cameras', Computer Vision and Image Understanding, 2012, 116, pp. 715-729

[5] Horaud, R., Knossow, D., Michaelis, M.: 'Camera cooperation for achieving visual attention', Machine Vision and Applications, 2006, 16, (6), pp.331–342

[6] Bastanlar, Y., Temizel, A., Yardimci, Y. et al.: 'Multi-view Structure-from-Motion for Hybrid Camera Scenarios', Image and Vision Computing, 2012, 30, (8), pp. 557-572

[7] Chen, C.H., Yao, Y., Page, D., et al.: 'Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking', IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18, (8), pp. 1052-1063

[8] Senior, A.W., Hampapur, A., Lu, M.: 'Acquiring Multi-scale Images by Pan–Tilt–Zoom Control and Automatic Multi-camera Calibration', Proc. of Seventh IEEE Workshop on Application of Computer Vision, 2005, pp. 433–438

[9] Tarhan, M., Altug, E.: 'A Catadioptric and Pan-Tilt-Zoom Camera Pair Object Tracking System for UAVs', Journal of Intelligent Robot Systems, 2011, 61, pp. 119–134

[10] Scotti, G., Marcenaro, L., Coelho, C., et al.: 'Dual camera intelligent sensor for high definition 360 degrees surveillance', IEE Proceedings-Vision, Image and Signal Processing, 2005, 152, (2), pp. 250-257

[11] Cui, Y., Samarasekera, S., Huang, Q., et al.: 'Indoor Monitoring Via the Collaboration Between a Peripheral Sensor and a Foveal Sensor', IEEE Workshop on Visual Surveillance, 1998

[12] Liu, Y., Shi, H., Lai, S. et al.: 'A spatial calibration method for mater-slave surveillance system', Optik, 2014, 125, pp. 2479-2483

[13] Tan, S., Xia, Q., Basu, A., et al.: 'A Two-Point Spatial Mapping Method for Hybrid Vision Systems', Journal of Modern Optics, 2014, 61, (11), pp. 910-922

[14] Peixoto, P., Batista, J., Araujo, H.: 'Integration of information from several vision systems for a common task of surveillance', Robotics and Autonomous Systems, 2000, 31, pp. 99-108

[15] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, Cambridge Univ. Press, 2nd edn., 2004

[16] Stewenius, H., Engels, C., Nister, D.: 'Recent Developments on Direct Relative Orientation', ISPRS J. of Photogrammetry and Remote Sensing, 2006, 60, pp. 284–294

[17] Zhang, Z.: 'A flexible new technique for camera calibration', IEEE Trans. Pattern Analysis and Machine Intelligence, 2002, 22, pp.1330–1334

[18] Bouguet, J.Y.: 'Camera Calibration Toolbox for Matlab', http://www.vision.caltech.edu/bouguetj/calib_doc, accessed 30 September 2014

[19] Geyer, C., Daniilidis, K.: 'A unifying theory for central panoramic systems', European Conference on Computer Vision (ECCV), 2000

[20] Mei, C., Rives, P.: 'Single view point omnidirectional camera calibration from planar grids', IEEE Int. Conference on Robotics and Automation (ICRA), 2007, pp. 3945-3950

[21] Puig, L., Bastanlar, Y., Sturm, P. et al.: 'Calibration of Central Catadioptric Cameras Using a DLT-Like Approach', Int. Journal of Computer Vision (IJCV), 2011, 93, (1), pp.101–114

[22] Hartley, R., Sturm, P.: 'Triangulation', Computer Vision and Image Understanding, 1997, 68, (2), pp. 146–157