

INDOOR PTZ CAMERA CALIBRATION WITH CONCURRENT PT AXES

Jordi Sanchez-Riera, Jordi Salvador, Josep R. Casas
Image Processing Group, UPC – Technical University of Catalonia
Jordi Girona 1-3, edifici D5, 08034 Barcelona
{jsanchez,jordi,josep}@gps.tsc.upc.edu

Keywords: Pan Tilt Zoom Active Camera Calibration

Abstract: The introduction of active (pan-tilt-zoom or PTZ) cameras in Smart Rooms in addition to fixed static cameras allows to improve resolution in volumetric reconstruction, adding the capability to track smaller objects with higher precision in actual 3D world coordinates. To accomplish this goal, precise camera calibration data should be available for any pan, tilt, and zoom settings of each PTZ camera. The PTZ calibration method proposed in this paper introduces a novel solution to the problem of computing extrinsic and intrinsic parameters for active cameras. We first determine the rotation center of the camera expressed under an arbitrary world coordinate origin. Then, we obtain an equation relating any rotation of the camera with the movement of the principal point to define extrinsic parameters for any value of pan and tilt. Once this position is determined, we compute how intrinsic parameters change as a function of zoom. We validate our method by evaluating the re-projection error and its stability for points inside and outside the calibration set.

1 INTRODUCTION

Smart Rooms equipped with multiple calibrated cameras allow visual observation of the scene for application to multi-modal interfaces in human-computer interaction environments. Static cameras with wide-angle lenses may be used for far-field volumetric analysis of humans and objects in the room. Wide angle lenses might provide maximum room coverage, but the resolution of the images from these far-field static cameras is somewhat limited. The introduction of pan-tilt-zoom (PTZ) cameras permits a more detailed analysis of moving objects of interest by zooming in on the desired positions. In this context, calibration of the PTZ cameras for any pan, tilt, and zoom settings is fundamental to referring the images provided by these cameras to the working geometry of the actual 3D world, commonly employed by static cameras.

Our research group has built a room equipped with six wide-angle static cameras. The purpose of the static cameras is the computation of a 3D volumetric reconstruction of the foreground elements in the room by means of background subtraction techniques fol-

lowed by shape-from-silhouette. The volumetric data resulting from this sensor fusion process is exploited by different tracking and analysis algorithms working in the actual 3D world coordinates of the room. With wide angle lenses spatial room coverage from multiple cameras is almost complete, but the precision of the volumetric data is restricted by the resolution of the camera images. Feature extraction analysis operates at a coarse spatial scale, and cannot obtain detailed information, such as the positions of fingers and hands, which might correspond to convex volumetric blobs in the volumetric representation. The introduction of PTZ cameras in the Smart Room helps to provide a higher resolution volumetric reconstruction for hand and gesture recognition algorithms. Hand movements are closely followed by means of real-time adjustment of pan and tilt. The calibration of PTZ cameras is fundamental to referring their images to the common 3D geometry of the computed volumetric data.

Camera calibration methods can be divided into two groups. The first group makes use of a calibration pattern, such as a checkerboard, located at a known 3D world position. Calibration parameters are then

inferred from the detected pattern points in order to calibrate either a single camera (Tsai, 1987; Zhang, 2000; Heikkila, 2000), or several cameras simultaneously (Svoboda et al., 2005). The disadvantage of methods based on knowledge of the calibration pattern position is that they might not be convenient for large focal lengths.

The second group of methods, also known as auto-calibration or self-calibration, may also use a calibration pattern, but in any case the 3D world position is not known (Hartley and Zisserman, 2000; Agapito et al., 1999). These methods are more difficult to implement due to geometrical complexity problems, such as finding the absolute conic.

Most of the methods cited above have been developed for static cameras. Calibration of active cameras faces new challenges, such as computing variation of intrinsic and extrinsic parameters, respectively, as a function of the zoom value and the rotation (pan and tilt) angles. Effective calibration of the varying parameters for the PTZ cameras requires adaptation of the algorithms developed for static cameras.

Intrinsics Calibration for PTZs: One approach for zoom calibration considers each zoom position as a static camera and then calibrates the PTZ as multiple static cameras. (Willson, 1994) provides a study on zoom calibration describing the changes in focal length, principal point, and focus based on Tsai calibration (Tsai, 1987). The large number of possible combinations for zoom and focus values makes the computation of intrinsic parameters ineffective. Other approaches limit the re-projection error as described in (Chen et al., 2001), which proposes the calibration of extreme settings followed by the comparison of the re-projection error with an arbitrary threshold. Positions with errors above the threshold are re-calibrated with a mid point and the extreme. The process is repeated until the range of zoom and focus settings is below the re-projection error threshold. A quartic function is finally found to best match the obtained parameters. However, the strategy of multiple static cameras presents problems with increasing zoom values, that reduce the constraints needed to solve the equations making results inconsistent (Ruiz et al., 2002). Huang (Huang et al., 2007) proposes an algorithm for cameras with telephoto lenses to face this problem.

An alternative approach for calibrating intrinsic parameters under varying zoom values is the separate computation of focal length and principal point although, for the latter, (Li and Lavest, 1996) claim that calibration results are not significantly affected if the principal point is assumed to be constant.

Extrinsics Calibration for PTZs: Extrinsic parameters are important for the correct application of intrinsic calibration results, but the methods mentioned so far focus on the computation of intrinsic parameters without describing how to compute extrinsics in much detail. (Sinha and Pollefeys, 2004) proposes computing both intrinsics and extrinsics in an outdoor environment using an homography between images acquired at different pan, tilt and zoom. (Davis and Chen, 2003) compute extrinsics by finding the camera rotation axis. Further references for computing both extrinsics and intrinsics for surveillance applications are (Kim et al., 2006; Senior et al., 2005; Lim et al., 2003).

Proposal: To overcome the problems of the different methods mentioned above for the calibration of active cameras, we propose to find the rotation center of the camera. Once the rotation center is known, extrinsic parameters can be found by simply applying a geometric formula, and then intrinsic parameters can be determined by a simple bundle adjustment.

In the following, we first review the specifics of the active camera model, then describe the proposed calibration method in Section 3, and finally we present the experimental results in Section 4 and draw conclusions to close the paper.

2 THE PTZ CAMERA MODEL

Camera calibration is important to the relation of the physical world with the images captured by the camera. This relation is defined by a mathematical model. The pin-hole camera model is based on the perspective transform (a 3×3 matrix K) explaining the relation from the 3D world to the 2D image plane. Unfortunately, real camera optics introduce distortion that must be modeled and corrected. The camera model is not complete without positioning the camera in the physical world according to arbitrarily chosen world coordinates. For this we need to know the rotation (a 3×3 matrix R) and translation (a 3×1 vector T) of the camera coordinates with respect to the world coordinates.

The equation resulting from the above description relates the 3D homogeneous points X in world coordinates with 2D homogeneous points x in the image plane as follows:

$$x = K[R|T]X \quad (1)$$

2.1 Extrinsic Parameters

For PTZ cameras, R and T depend on how the principal point changes with pan and tilt movements. If the principal point is different from the rotation center of the camera (e.g. by a shift r in the optical axis), then any change in the pan (or tilt) angle of α degrees will produce a rotation and a translation of the principal point, which, due to the mechanics of the PTZ camera, will describe a sphere of radius r centered on the rotation center. Considering R_r and T_r , the rotation and translation describing the position of the rotation center as reference position, any pan (or tilt) movement will produce an additional rotation R_a and translation T_a describing the position of the principal point.

The reference is defined to be at pan and tilt zero, then the rotation R_a and translation T_a should be applied to the 3D homogeneous points X . Note that this transform is not a trivial operation. When pan and tilt displacement occur at the same time, two rotation matrices are needed and the order in which matrices are multiplied produces different results. Thus, the transform necessary to express the 3D homogeneous points X in correct principal point origin coordinates must be applied in the following order: Rotate the points R_r and translate T_r , then apply rotation R_a and after the translation T_a , finally translate the points back to the origin with transform T .

$$R_a = (R_t R_p) \quad (2)$$

$$R = R_r \quad (3)$$

$$T_a = T_p + T_t + [0, 0, r]^T \quad (4)$$

$$T = T_a + T_r \quad (5)$$

In the next subsections we describe the parameters in Equations (2-5) and the effects of pan and tilt independently. Our coordinate axis election is described considering z for the optical axis (in the depth direction) and x, y for the Cartesian coordinates on the image plane.

2.1.1 Pan and Tilt Movements

Figure 1 represents the displacement of the principal point when either pan or tilt changes. When pan changes by α degrees, it introduces a rotation $R_p(\alpha)$ and a translation $T_p(\alpha)$ parameterized by α :

$$R_p(\alpha) = \begin{bmatrix} \cos(\alpha) & 0 & \sin(\alpha) \\ 0 & 1 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix} \quad (6)$$

$$T_p(\alpha) = \begin{bmatrix} 0 \\ h \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ r \sin(\alpha) \\ -r(1 - \cos(\alpha)) \end{bmatrix} \quad (7)$$

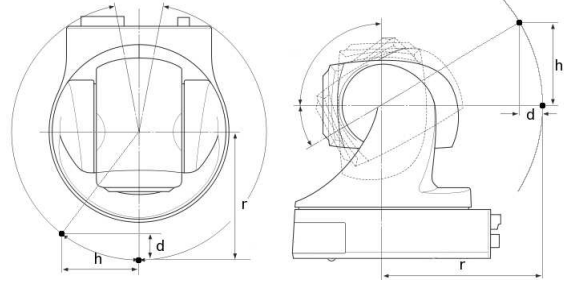


Figure 1: Principal point displacement (black dot) due to pan (left image) and tilt (right image).

When tilt changes by β degrees, it introduces a rotation $R_t(\beta)$ and a translation $T_t(\beta)$ parameterized by β :

$$R_t(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & -\sin(\beta) \\ 0 & \sin(\beta) & \cos(\beta) \end{bmatrix} \quad (8)$$

$$T_t(\beta) = \begin{bmatrix} h \\ 0 \\ d \end{bmatrix} = \begin{bmatrix} r \sin(\beta) \\ 0 \\ -r(1 - \cos(\beta)) \end{bmatrix} \quad (9)$$

2.2 Intrinsic Parameters

The perspective transform provides a 3x3 matrix, K , mapping the 3D space to the 2D camera plane assuming the pin-hole camera model. This mapping is characterized by the principal point and the focal length. The principal point (p_x, p_y) is located at the intersection of the optical axis with the image plane. Usually the camera is designed so that the principal point is at the center of the image, but it may shift depending on the camera optics and zoom settings. The focal length is the distance between the principal point and the focal point. Measured in pixels, the focal length is represented by two components (f_x, f_y) (the ‘‘aspect ratio’’ f_y/f_x is different from 1 only for non-square pixels). For PTZ cameras, we may assume that all the intrinsic parameters are a function of zoom:

$$K(z) = \begin{bmatrix} f_x(z) & sk(z) & p_x(z) \\ 0 & f_y(z) & p_y(z) \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

sk represents the skew coefficient defining the angle between the x and y image axes. Further, we must consider the distortion introduced by camera lenses,

represented by four parameters, two for radial and two for tangential distortion, as described in (Bouguet, 2007).

3 Calibration Method

The first goal of our calibration method is to find the rotation center of the camera and its distance to the principal point. Then, with just a geometric transform, we can obtain the extrinsic parameters for any given pan, tilt position. Once this is done, we can focus on the intrinsic parameters. Using the extrinsic parameters already found as the valid camera world position, we can obtain the intrinsic parameters considering that, when zoom changes, intrinsic parameters also change.

To compute the rotation center and its distance to the principal point, first we place the calibration pattern in a known and measured position in the room respect our world origin, and we take some images at several known pan and tilt positions at zoom zero (widest lens). One of the images is taken as reference and the others are described by its relative pan and tilt position with respect to the reference. Then a bundle adjustment, shown in Equation (11) is applied to minimize the reprojection error.

$$\sum_{i=1}^n \sum_{j=1}^m \|(x_{ij} - \hat{x}(R_i, R_{pi}, R_{ti}, T_i, T_{pi}, T_{ti}, X_j, kc, a_p, a_t))\|^2 \quad (11)$$

where $\hat{x}(R_i, \dots)$ is the projection of point X_j in image i according to Equation (1) with modifications explained in Section 2.1 and, using four parameters for lens distortion kc . R_i, T_i are the rotation and translation matrices for the reference image. R_{pi}, R_{ti} and T_{pi}, T_{ti} are the additional rotation and translation matrices due to the relative pan and tilt movements, described in Equations (6-9). Our formulation assumes there are n images with m points in each image. Depending on the camera position, it will not be possible to take images of the calibration pattern at pan and tilt zero, thus a_p and a_t describe the pan and tilt of the reference image respect the zero position.

In order to reach convergence, it is important to start from a good initial guess of the parameters' values. In our case the initial guess for the intrinsic and extrinsic parameters is obtained with (Bouguet, 2007), although any other method, such as (Tsai, 1987), could be equally valid.

With the same principle, we can consider the values of rotation center and radius just found after the previous error minimization as an initial guess for a second

bundle adjustment. In this case, we aim at minimizing error with respect to all the parameters: Intrinsic and extrinsic (rotation center and radius). This minimization is computed on the functional shown in Equation (12).

$$\sum_{i=1}^n \sum_{j=1}^m \|(x_{ij} - \hat{x}(K, R_i, R_{pi}, R_{ti}, T_i, T_{pi}, T_{ti}, X_j, kc, a_p, a_t))\|^2 \quad (12)$$

Note that the intrinsic parameters K , are not present in the Equation (11). With a calibrated PT (pan-tilt) camera after this second minimization, we need to calibrate the intrinsic parameters as a function of zoom. For this purpose, we take some images at different zoom values of the calibration pattern in the previous reference position. As we increase the zoom, the calibration pattern ends up outside of the image, therefore, a smaller calibration pattern is used. This second pattern must be aligned with the previous one to keep it in the same measured position. In this case, the minimized functional is shown in Equation (13).

$$\sum_{j=1}^m \|(x_j - \hat{x}(K(z), X_j))\|^2 \quad (13)$$

Note that only one image is used each time and only intrinsic parameters are minimized. In our particular case, the camera has 1000 step positions for zoom. This is a large number of values and it would not be a good idea to perform the minimization for all the possible values. Therefore, we take the images sampling the whole zoom range at equal intervals. With this, we obtain the intrinsic parameters for some of the values of zoom and then, we can fit a function. This function is as simple as a polynomial of n degrees and $n + 1$ coefficients.

4 Results

The proposed algorithm has been implemented in Matlab as an extension of the (Bouguet, 2007) toolbox. In order to test the correctness and limits of the method, we have studied how pan and tilt range affects the re-projection error and have determined which is the optimal number of degrees to fit a function and the optimal sampling of the zoom range. Another test performed to ensure that the correct parameters are found consists in the evaluation of the reprojection error of a known 3D position different from those used to calibrate the camera, in the Smart Room. For all recordings we have used two different calibration patterns consisting of black and white squares. The images have been grabbed using a Sony-EVI D70

camera, controlled with the Evi-Lib library (EVI-Lib, 2006) for adjustment of pan, tilt, and zoom.

Depending on the distance from the calibration pattern to the camera, the range of pan and tilt, such that the camera is able to grab an image of the whole pattern, will vary. In the selected pattern position, we are able to capture a maximum range of 35 degrees from one side to another of pan, and a maximum range of 20 degrees from one side to another of tilt. In Figure 2 we have represented the re-projection error (y-axis) as a function of pan range (x-axis) and tilt range (the lines drawn). Note that, if for any reason, we are not able to work with a tilt range larger than 10, the pan range needs to be at least 25 to keep re-projection error below 3 pixels. For a tilt range larger than 10, the algorithm is more flexible, allowing more variability for pan range and still keeping a re-projection error below 1.5 pixels.

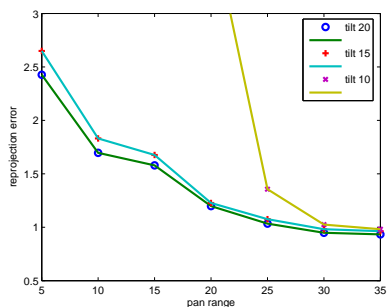


Figure 2: Re-projection error for extrinsic parameters as a function of pan and tilt

In Figure 3 we have represented the re-projection error (y-axis) as a function of the number of coefficients (x-axis) and intervals of zoom sampling (the lines drawn). For intervals of 25 or 50 steps, the results are very similar, except for coefficients larger than 14, when re-projection error of the line corresponding to 50 steps increases. This is because with 50 steps, over a range of 1000 possible steps, there is not enough data to fit a function with more than 14 coefficients. For similar reasons, when data is taken from intervals of 100 steps, it is not possible to fit a function with an arbitrary large number of coefficients. With a small number of coeffs. (e.g. 4), the error is larger than 5 pixels.

With camera parameters within correct re-projection error margins, it is also important to acknowledge to what extent the model is able to reproduce the same results for different positions than those used to calibrate the camera. For this purpose a second set of images has been acquired with the calibration pattern at different position. Figure 4 shows two exam-

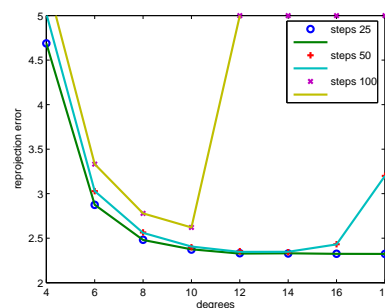


Figure 3: Re-projection error for intrinsic parameters for different polynomial degrees

ples of the re-projection error at different zoom values. Taking 40 different images at random positions with zoom zero (shortest focal length), we get a re-projection error of 1.20 pixels with a standard deviation of 0.56 pixels while taking 40 images at different zoom positions the re-projection error is 2.45 pixels with a standard deviation of 2.64 pixels. This proves that the re-projection error is very close to the values obtained during calibration.

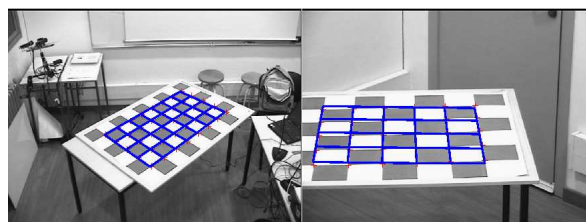


Figure 4: Grid projection. Left image zoom 0, error = 0.93 pixels. Right image zoom 550, error 1.83 pixels.

5 Conclusions

Calibration of active cameras has to overcome some difficulties that are not present for fixed cameras, such as solving calibration equations for telescopic lenses or large focal lengths and the fact that the rotation center of the PTZ camera is not co-located with the principal point. The goal of this work has been to obtain a calibration algorithm for PTZ cameras that improves the precision of volumetric reconstruction algorithms applied to a combination of fixed and PTZ camera images in our Smart Room. The principal features we wanted were real-time availability of calibration parameters for all range of pan, tilt and zoom values, with re-projection errors smaller than one pixel spac-

ing in order to have precise 3D volumetric reconstruction.

State of the art methods did not provide a solution for these goals. Methods for static cameras use to fail for large focal lengths and require large tables to store calibration values for the pan-tilt-zoom range. Methods for active cameras were more adequate but provided a good solution for only a small range of pan and tilt.

The presented algorithm is flexible in the sense that it provides readily available calibration parameters for any combination of pan, tilt, and zoom, with a small re-projection error, even though it is slightly larger than one pixel.

Some sources of error have been identified. Probably, the most influential error source is the assumption of a nearly constant 3D position of the principal point of the camera with respect to the zoom when, in practice, the focal length in the zoom range of the camera can vary up to a few centimeters. The consideration of constant lens distortion for the zoom range is not true either, although its influence in the final results might be insignificant.

Another possible source of error is the position of the calibration pattern in the Smart Room. The Smart Room has some predefined points, precisely measured to put the calibration pattern on according to arbitrary coordinates. For each attempt to position the calibration pattern at predefined 3D world coordinates, there is an error of about 0.5cm in the physical location, which is small considering the actual dimensions of the Smart Room(4m × 5m), but still enough to affect camera calibration results.

ACKNOWLEDGMENTS

We want to thank to Jeff Mulligan for the ideas given in the algorithm definition and to Albert Gil for the support given for camera grabbing software.

REFERENCES

- Agapito, L. d., Hayman, E., and Reid, I. (1999). Self-calibration of a rotating camera with varying intrinsic parameters. In *Proc. 9th British Machine Vision Conference, Southampton.*, pages 105–114.
- Bouguet, J. Y. (2007). Camera calibration toolbox for matlab. (Available for download from <http://www.vision.caltech.edu/bouguetj/calib.doc>).
- Chen, Y.-S., Shih, S.-W., Hung, Y. P., and Fuh, C. S. (2001). Simple and efficient method of calibrating a motorized zoom lens. *Image and Vision Computing*, 19(14):1099–1110.
- Davis, J. and Chen, X. (2003). Calibrating pan-tilt cameras in wide-area surveillance networks. In *Ninth IEEE International Conference on Computer Vision, 2003. Proceedings*, number 1, pages 144–149.
- EVI-Lib (2006). C++ library for controlling the serial interface with sony color video cameras evi-d30, evi-d70, evi-d100 (<http://sourceforge.net/projects/evilib/>). <http://sourceforge.net/projects/evilib/>.
- Hartley, R. I. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049.
- Heikkila, J. (2000). Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077.
- Huang, X., Gao, J., and Yang, R. (2007). Calibrating pan-tilt cameras with telephoto lenses. In *Computer Vision - ACCV 2007, 8th Asian Conference on Computer Vision*, pages I: 127–137, Tokyo, Japan.
- Kim, N., Kim, I., and Kim, H. (2006). Video surveillance using dynamic configuration of multiple active cameras. In *IEEE International Conference on Image Processing (ICIP'06)*, pages 1761–1764.
- Li, M. and Lavest, J.-M. (1996). Some aspects of zoom lens camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1105–1110.
- Lim, S. N., Elgammal, A., and Davis, L. S. (2003). Image-based pan-tilt camera control in a multi-camera surveillance environment. In *Proceedings International Conference on Multimedia and Expo (ICME'03)*, volume 1, pages 1–645–8.
- Ruiz, A., López-de Teruel, P. E., and García-Mateos, G. (2002). A note on principal point estimability. In *16th International Conference on Pattern Recognition (ICPR'02)*, volume 2, pages 304–307.
- Senior, A., Hampapur, A., and Lu, M. (2005). Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration. In *Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)*, pages 433–438.
- Sinha, S. and Pollefeys, M. (2004). Towards calibrating a pan-tilt-zoom camera network. In *OMNIVIS 2004, ECCV Conference Workshop CD-rom proceedings*.
- Svoboda, T., Martinec, D., and Pajdla, T. (2005). A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422.
- Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, RA-3(4):323–344.
- Willson, R. G. (1994). *Modeling and Calibration of Automated Zoom Lenses*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.