Robust Auto-Calibration of a PTZ Camera with Non-overlapping FOV

Nazim Ashraf and Hassan Foroosh
Computational Imaging Lab, University of Central Florida, Orlando, FL 32816
{*nazim, foroosh*}*@eecs.ucf.edu*

Abstract

We consider the problem of auto-calibration of cameras, which are fixed in location but are free to rotate while changing their internal parameters by zooming. Our method is based on line correspondences between two views, which may have non-overlapping field of view. Camera calibration from images having non-overlapping field of view is the basic motivation behind this research. The key observation is that the planes formed by the optic center and the line correspondences are really the same plane. We use this fact together with the orthonormality constraint of the rotation matrix to estimate the unknown camera parameters. We show experimental results on synthetic and real data, and analyze the accuracy and stability of our method.

1. Introduction

With the development of state-of-the-art systems, which can accurately detect, recognize, and track objects as they move through a scene, observation of *activity* from stationary cameras has become an important topic of research. But a fundamental issue is that due to perspective projection, the measurements made from the images do not reflect metric data as such; they are projectively distorted. To make real world measurements, we need to calibrate the camera. Therefore, camera calibration has been a subject of keen interest for researchers.

Original work on camera calibration started with Faugeras et al. [6], who considered a freely moving camera with unknown but constant internal parameters. Since then, several methods have been proposed [8, 15, 16, 12, 21], some of which consider special camera motions such as pure translation [18] or pure rotation [9]. More recent methods also consider auto-calibration under varying internal parameters [10, 12, 13, 19, 14].

The most related work to our paper is the auto-calibration method for rotating and zooming cameras by Agapito et al [1], who used the mapping of the image of the absolute conic (IAC) between two images by the infinite homography to impose constraints on camera internal parameters for a pair of images. However, we propose a new robust auto-calibration method to estimate camera intrinsic and extrinsic parameters given line correspondences between frames. We use the fact that the planes formed by the optic center and the line correspondences are really the same plane. This coupled with the orthonormality constraint of the rotation matrix is used to solve for the camera parameters. The main advantage of our method over Agapito's work is that in our method, the images used for calibration may have non-overlapping field of view.

The rest of the paper is organized as follows. We begin with a discussion of preliminary material (section 2), which includes the geometry of a basic pinhole projective camera, and a brief discussion of the particular case of zero translation. We then describe our method in section 3. Finally, we discuss experimental results on synthetic and real data (section 4), and conclude (section 5).

2. Background Material

2.1. Pinhole Camera Model

For a pinhole camera model, a 3D point $\mathbf{M} = [X \;\; Y \;\; Z \;\; 1]^T$ and its corresponding image projection $\mathbf{m} = [u \; v \; 1]^T$ are related via a 3×4 matrix $\mathbf{P}$ by

$$\mathbf{m} \sim \underbrace{\mathbf{K}[\mathbf{r}_1 \; \mathbf{r}_2 \; \mathbf{r}_3 \; \mathbf{t}]}_{\mathbf{P}} \mathbf{M}, \quad \mathbf{K} = \begin{pmatrix} \lambda f & \gamma & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1}$$

where $\sim$ indicates equality up to multiplication by a non-zero scale factor, $\mathbf{r}_i$ are the columns of the rotation matrix $\mathbf{R}$, which represents the orientation of the camera, $\mathbf{t}$ is the translation vector, and $\mathbf{K}$ is a nonsingular 3×3 upper triangular matrix known as the camera calibration matrix including five parameters, i.e. the focal

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

length f, the skew γ, the aspect ratio λ and the principal point at $(u_0,\ v_0)$. As argued by [1, 19], it is safe to assume $\lambda = 1$, and $\gamma = 0$. Without loss of generality, we choose to locate the origin of the camera's coordinate system at the optic center. Hence, the projection matrix for each view i can be written as

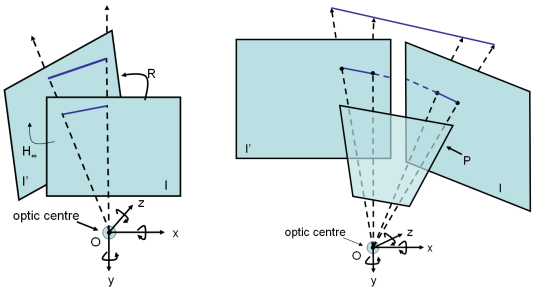
$$\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i|\mathbf{0}] \tag{2}$$

The projection of a scene point can now be expressed as

$$\mathbf{x} = \mathbf{K}_i[\mathbf{R}_i|\mathbf{0}] \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \\ \mathbf{1} \end{bmatrix} = \mathbf{K}_i\mathbf{R}_i \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} \tag{3}$$

This implies that for rotating cameras, the mapping of 3D rays to image points is encoded by the 3×3 invertible projective transformation:

$$\mathbf{P}_i = \mathbf{K}_i\mathbf{R}_i \tag{4}$$



(a) (b)

Figure 1. (a) Corresponding image points are related by the infinite homography. (b) For a line correspondence between two views, the planes defined by the optic center and the line segments are one and the same plane (marked as P).

3. Robust Auto-Calibration

The problem of auto-calibration a rotating and zooming camera is determining the calibration matrices of each frame. As mentioned earlier, we use line correspondences to calibrate the camera which is free to vary its intrinsic parameters by zooming in/out. However, we assume $\lambda = 1$, and $\gamma = 0$.

The key observation behind our method is that given a line correspondence between two views, the planes defined by the optic center and each of the line segments, are one and the same. To see this, refer to figure 1 (b). In this figure, a line segment present in the world is projected onto frames, I and I' for a rotating and zooming camera. The noteworthy feature is that the plane defined by the optic center, O, and the *imaged* line segments remains the same (marked as P). Indeed, even if the two views have non-overlapping field of view, this particular trait of line correspondences still holds. In fact, figure 1 (a) shows just that very case. Let us now see how we can use this for calibration:

Assume the line correspondences are between line $\mathbf{l}_i \sim [{l_x}^i, {l_y}^i, {l_w}^i]$, and line $\mathbf{l}_j \sim [{l_x}^j, {l_y}^j, {l_w}^j]$. Using the fact that the plane remains the same, we can write:

$$\mathbf{P}_i^T\mathbf{l}_i \sim \mathbf{P}_j^T\mathbf{l}_j \tag{5}$$

Substituting eq 4 in eq 5, we get:

$$\mathbf{R}_i^T\mathbf{K}_i^T\mathbf{l}_i \sim \mathbf{R}_j^T\mathbf{K}_j^T\mathbf{l}_j \tag{6}$$

Re-arranging this equation, we get:

$$\mathbf{l}_j \sim \mathbf{K}_j^{-T}\mathbf{R}_j^{-T}\mathbf{R}_i^T\mathbf{K}_i^T\mathbf{l}_i \sim \mathbf{K}_j^{-T}\mathbf{R}_j\mathbf{R}_i^T\mathbf{K}_i^T\mathbf{l}_i \sim \mathbf{K}_j^{-T}\mathbf{R}_{ij}^T\mathbf{K}_i^T\mathbf{l}_i \tag{7}$$

This equation relates the *known* line correspondences with the 12 unknown variables, namely f_i, f_j, u_o, v_o, and the 8 parameters of $\mathbf{R}_{ij}$. Each line correspondence provides us with two such equations. We require at least four line correspondences, which would provide us with 8 equations. Given these 8 equations, we can solve for the parameters of $\mathbf{R}_{ij}$ in terms of f_i, f_j, u_o, and v_o. If, however, we have more than 4 line correspondences, we have an over-determined system, which can still be used to serve the same purpose. Then, we can use the orthonormality constraint of $\mathbf{R}_{ij}$ to extract 5 equations involving only the intrinsic parameters of the camera. We found that 4 out of these 5 equations were independent, which can be solved for f_i, f_j, u_o, and v_o. Since these equations might be too complex to solve simultaneously, a more practical method is to express one or more variable(s) in terms of the other variables and minimize the resulting equation(s) using the Levenberg-Marquardt minimization method [20].

3.1. Maximum Likelihood Estimation

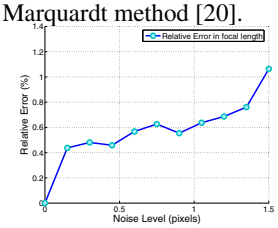
We refine our estimate by using bundle adjustment. Assuming we have noise w on the measured lines $\hat{l}$ is additive and described by Gaussian distribution with mean zero and standard deviation, σ, then the measured $\hat{l}$ is related to true location by:

$$\hat{\mathbf{l}} = \mathbf{l} + \mathbf{w} = \mathbf{K}\mathbf{R}\bar{\mathbf{L}} + \mathbf{w} \tag{8}$$

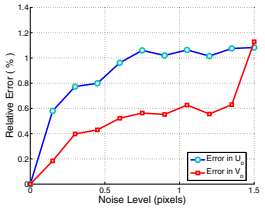
It is straightforward to prove that the Maximum Likelihood Estimate is given by:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{D}(\hat{l}_{ij}, \mathbf{K}_i\mathbf{R}_i\bar{\mathbf{L}}_j)^2 \tag{9}$$

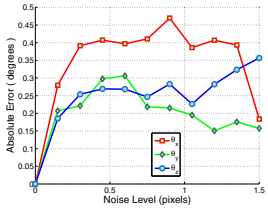
where $\mathbf{D}(l_i, l_j)$ is defined as $\mathbf{D} = \|l_i \times l_j\|$ In other words, we are estimating the $\mathbf{K_i}$, $\mathbf{R_i}$, and $\bar{\mathbf{L}}_j$ for which the squared distance of the measured lines to the true image lines for all lines across all views would be minimum. The minimization of this estimate is a non-linear problem and can be accomplished using Levenberg-Marquardt method [20].



(a) Relative error in f.



(b) Relative error in u_o, and v_o.



(c) Absolute Error in θ_x,θ_y, and θ_z.

Figure 2. Performance vs. Noise Level on synthetic data.

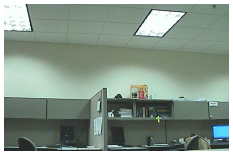
4. Experimental Results

In order to test the accuracy of the proposed solution, we experimented with synthetic and real data.

4.1. Synthetic Data

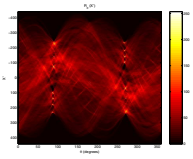
To evaluate the performance of our method, we carried out detailed experimentation with synthetic data. The good feature of using synthetic data is that the ground truth is known and the level of noise can be controlled. These can be used to test the accuracy and the stability of a given method.

The camera was simulated having the following parameters: $f = 1000, \lambda = 1.5, u_o = 512, v_o = 384, \gamma = 0$. 1000 lines were generated and Gaussian noise with zero mean and standard deviation of $\sigma \leq 1.5$ were added to these lines. For each noise level - which varied from 0.01 to 1.5 in our experiments - we performed 1000 independent trials. It must be noted that in the case of intrinsic parameters, the *relative difference* with respect to the parameter in question is a more meaningful error measure geometrically than simple absolute error. Therefore, we measured the relative error of estimated f, $\mathbf{u}_o$, and $\mathbf{v}_o$ with respect to the true values of these parameters. The results are shown in figure 2. It can be observed from these results that for intrinsic parameters, the relative error increases almost linearly with respect to the noise level. For a minimum noise level of 1.5 pixels, the error for f, u_o, and v_o is observed to be less than 1.1%, 1.2%, and 1.1% respectively. For extrinsic parameters, the absolute error remains well below 0.48%. These results demonstrates that our method is accurate and stable to noise.

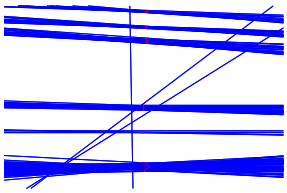


(a) A sample image.

(b) After applying Canny Edge Detection.



(c) Rendering of Hough Transform results.



(d) After line detection using Hough Transform.

Figure 3. (a)-(d) Steps in line detection.

4.2. Real Data

We use a SONY® SNC-RZ30N PTZ camera with an image resolution of 736×480 to obtain the real data, where the ground truth rotation angles are known. We deliberately do not change the intrinsic parameters so that we can estimate the accuracy of our parameters in the absence of ground truth.

We use the method of Devernay[4] to correct lens distortion. Then Canny edge detection [2] is used to extract the edges. Finally, we apply the Hough Transform [5] to detect lines. The process of line detection for one sample image is shown in 3 (a)-(d).

Around 10 images were captured while panning and tilting with $\theta_y = 1$ and $\theta_x = 1$ respectively; the intrinsic parameters remained constant. In order to validate the stability of the proposed system, we applied our method to every combinations of the 9 images out of the total

set. The results for the image sequence are shown in figure 4. The last three columns depict the values of the rotation angles, θ_x, θ_y, and θ_z as 1.41, 1.14, and 0.13 respectively, which are close to the ground truth angles. As mentioned earlier, camera zoom remained constant in the sequence, therefore, the value of column 4 and column 5, i.e. f_1 and f_2 are seen to be very close to each other. The results also demonstrate low median deviation for the estimated parameters.

Combination	u_o	v_o	f_1	f_2	θ_x	θ_y	θ_z
C_1	188.51	285.41	347.71	357.11	1.48	1.03	0.17
C_2	183.50	315.45	383.65	362.30	0.64	0.43	0.31
C_3	179.02	246.68	258.90	268.73	0.83	0.51	0.02
C_4	121.47	389.37	322.39	325.72	2.52	1.86	0.25
C_5	121.38	215.22	333.77	297.57	2.69	1.81	0.19
C_6	199.64	238.76	364.96	299.21	1.50	0.89	0.07
C_7	179.01	246.68	258.90	268.73	1.11	2.04	0.03
C_8	195.03	244.19	281.67	270.35	1.05	2.03	0.08
C_9	190.48	276.69	313.67	315.20	1.61	0.36	0.08
C_{10}	121.38	215.22	333.77	297.57	0.67	0.45	0.19
Mean	167.94	267.37	319.94	306.25	1.41	1.14	0.13
St. Dev.	32.74	52.97	42.43	34.14	0.71	0.71	0.10

Figure 4. Results obtained on real data.

5. Conclusion

In this paper, we have described a new robust technique for camera calibration given line correspondences. The method is based on the fact that the planes formed by the optic center and the line correspondences are really the same plane. This fact is used with the orthonormality constraint of the rotation matrix to estimate the unknown camera parameters. The advantage of the proposed method is that the images used for calibration may have non-overlapping field of view. We successfully demonstrate the proposed method on synthetic and real data.

References

[1] L. D. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *Int. J. Comput. Vision*, 45(2):107–127, 2001.

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[3] D. Capel and A. Zisserman. Automatic mosaicing with super-resolution zoom. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 885, Washington, DC, USA, 1998. IEEE Computer Society.

[4] F. Devernay and O. D. Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13(1):14–24, 2001.

[5] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, 1972.

[6] O. Faugeras, T. Luong, and S. Maybank. Camera self-calibration: theory and experiments. In *Proc. of ECCV*, pages 321–334, 1992.

[7] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartology. *Communications of the ACM*, 24(6):381–395, 1981.

[8] R. I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. ECCV*, pages 471–478, 1994.

[9] R. I. Hartley. Self-calibration of stationary cameras. *Int. J. Comput. Vision*, 22(1):5–23, 1997.

[10] R. I. Hartley, E. Hayman, L. D. Agapito, and I. Reid. Camera calibration and the search for infinity. In *Proc. IEEE ICCV*, pages 510–517, 1999.

[11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[12] A. Heyden and K. Astrom. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. IEEE CVPR*, pages 438–443, 1997.

[13] A. Heyden and K. Astrom. Flexible calibration: Minimal cases for auto-calibration. In *Proc. IEEE ICCV*, pages 350–355, 1999.

[14] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *J. Math. Imaging Vis.*, 13(2):131–146, 2000.

[15] J. Knight, A. Zisserman, and I. Reid. Linear auto-calibration for ground plane motion. In *Proc. IEEE CVPR*, pages 503–510. IEEE, 2003. Madison, Wisconsin.

[16] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. IEEE ICCV*, pages 293–300, 1999.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 6(2):91–110, 2004.

[18] T. Moons, L. Gool, M. Proesmans, and E. Pauwels. Affine reconstruction from perspective image pairs with a relative object-camera translation in between. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(1):77–83, 1996.

[19] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. J. Comput. Vision*, 32(1):7–25, 1999.

[20] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[21] B. Triggs. Autocalibration and the absolute quadric. In *Proc. IEEE CVPR*, pages 609–614, 1997.

[22] B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle adjustment — a modern synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–373, 1999.