CAMERA CALIBRATION WITHOUT METRIC INFORMATION USING 1D OBJECTS

Xiaochun Cao and Hassan Foroosh

School of Computer Science
University of Central Florida
Orlando, FL, 32816-3262
{xccao,foroosh}@cs.ucf.edu

ABSTRACT

This paper addresses the problem of calibrating a pin-hole camera from images of 1D objects. Assuming a unit aspect ratio and zero skew, we introduce a novel and simple approach that uses four observations of a 1D object and requires no information about the distances between the points on the object. This is in contrast to existing methods that use two images, but impose more restrictive configurations that require measured distances on the calibrating object. The key features of the proposed technique are its simplicity and ease of use due to the lack of need for any metric information. To demonstrate the effectiveness of the algorithm, we present the processing results on synthetic and real images.

1. INTRODUCTION

Calibration is the process of determining the intrinsic and the extrinsic camera parameters, in order to extract the Euclidean structure of the scene, and to determine the rigid motion of the camera with respect to the world coordinate frame. Classical techniques for camera calibration [1, 2, 3] require a so called calibration rig, with a set of correspondences between known points in the 3D space and their projections in the 2D image plane. Recent techniques propose more flexible plane-based calibration approaches [4, 5].

Zhang [6] has presented a method for camera calibration using a 1D object with one fixed point. However, this technique requires metric information about this 1D calibrating object, i.e. known position of other points along the line. This idea of camera calibration based on 1D constraints is formally analysed in [7]. Calibration techniques that do not require metric information [8, 9, 10] lead to solving more complex non-linear problems that are less tractable or ill-conditioned, and may not always be guarantied to converge to the optimal solution. Motivated by making camera calibration an off-the-shelf tool available to non-experts, we have developed an easy technique that uses a line (e.g. the border of a textbook), but does not require metric information. This versatile technique can also be applied to symmetric objects [11].

This work was partially supported by ONR grant #N00014-04-1-0512, and Sun Microsystem's grant #EDUD-7824-030482-US.

2. PRELIMINARIES

As is well known, for a pin-hole camera model, a 3D point $\mathbf{M}$ and its corresponding projection $\mathbf{m}$ in the image plane are related via

$$\mathbf{m} \sim \mathbf{A}[\mathbf{R}\ \mathbf{t}]\mathbf{M} \quad \text{where } \mathbf{A} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where $\sim$ indicates equality up to a non-zero scalar, $\mathbf{R}$ is the orthonormal rotation matrix, $\mathbf{t}$ is the translation vector, and $\mathbf{A}$ is the simplified camera intrinsic matrix with focal length f and $(u_0\ \ v_0)$ as the coordinates of the principal point $\mathbf{c}$. The simplified camera model assumes unit aspect ratio and zero camera skew, which typically is true for modern CCD cameras [12].

We now describe the configuration used herein for calibration. The 1D object can not be free moving as argued in [6]. One possibility is to move around a fixed pivoting point $\mathbf{P}$ as shown in Figure 1. Four images are taken as follows: one pair from the same viewpoint but with the 1D object at two distinct positions, and another pair from a different viewpoint but the same two positions of the object that were used in the first pair. Accordingly, two arbitrary points $\mathbf{M}_1$ and $\mathbf{M}_2$ with unknown locations on the 1D object and the corresponding points $\mathbf{M}'_1$ and $\mathbf{M}'_2$ after moving the 1D object to another position, yield two isosceles triangles $\mathbf{M}_1\mathbf{M}'_1\mathbf{P}$, and $\mathbf{M}_2\mathbf{M}'_2\mathbf{P}$. The two equal sides of these triangles are captured by two image pairs, where the images in each pair are obtained from the same viewpoint. Therefore, hereafter we refer to each image pair as a viewpoint. This configuration is depicted in Figure 1.

Using simple geometry one can readily show that the axis of symmetry of these triangles is given by the line $\mathbf{SP}$, where $\mathbf{S}$ is the intersection of $\mathbf{M}_1\mathbf{M}'_2$ and $\mathbf{M}_2\mathbf{M}'_1$, projected to the point s in the image plane. Thus, we can choose the world coordinate system as follows: origin at $\mathbf{O}$, i.e. the intersection of $\mathbf{SP}$ and $\mathbf{M}_1\mathbf{M}'_1$, x-axis along $\mathbf{M}_1\mathbf{M}'_1$ with

0-7803-8554-3/04/$20.00 ©2004 IEEE.

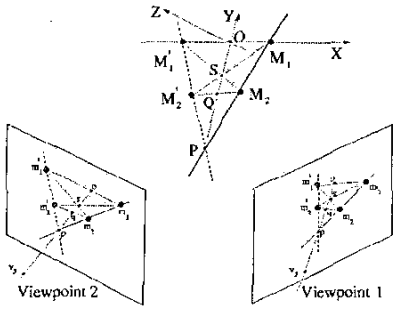


Fig. 1. Configuration of Calibration Using a 1D Object.

positive direction towards $\mathbf{M}_1$, y-axis along $\mathbf{PS}$ with positive direction towards $\mathbf{S}$, and z-axis given by the right-hand rule.

3. SOLVING CAMERA CALIBRATION

In the coordinate frame described above, all the four points $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_2'$, and $\mathbf{M}_1'$ are on the plan $Z = 0$. Therefore, the 3×4 projection matrix in equation (1) reduces to a 3×3 planar homograph $\mathbf{H}$ [13, 1]:

$$\mathbf{H} = [r_{31}\mathbf{v}_x \quad r_{32}\mathbf{v}_y \quad t_z\mathbf{o}] \tag{2}$$

where $\mathbf{v}_x = [v_{xx} \quad v_{xy} \quad 1]^T$ and $\mathbf{v}_y = [v_{yx} \quad v_{yy} \quad 1]^T$ are the x and y vanishing points in the image coordinate system, $\mathbf{o} = [o_x \quad o_y \quad 1]^T$ is the image point corresponding to the projection of $\mathbf{O}$, and r_{31} and r_{32} are the components of the rotation matrix $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$ given by

$$r_{31} = -\sin(\ _y) \tag{3}$$
$$r_{32} = \cos(\ _y)\sin(\ _x) \tag{4}$$

For our configuration, we obviously have

$$\mathbf{v}_x \sim (\mathbf{m}_1 \times \mathbf{m}_1') \times (\mathbf{m}_2 \times \mathbf{m}_2') \tag{5}$$
$$\mathbf{v}_y^T(\mathbf{s} \times \mathbf{p}) = 0 \tag{6}$$
$$\mathbf{o} \sim (\mathbf{m}_1 \times \mathbf{m}_1') \times (\mathbf{s} \times \mathbf{p}) \tag{7}$$

These are the five basic constraints on our homography $\mathbf{H}$, given four images of the 1D object. In general, a 2D homography has eight DOFs. This implies that we need three more independent constraints to solve the problem.

For a unit aspect ratio and zero skew, the principal point $\mathbf{c}$ is the ortho-center of the triangle with vertices at $\mathbf{v}_x$, $\mathbf{v}_y$, and $\mathbf{v}_z$ [12]. Thus,

$$(\mathbf{v}_x - \mathbf{c})^T(\mathbf{v}_y - \mathbf{c}) + f^2 = 0 \tag{8}$$

By combining equations (5),(6), and (8), we can show that $\mathbf{v}_y$ is as the form:

$$\begin{bmatrix} v_{xy} - v_0 - (f^2 - u_0v_{xx} + u_0^2 - v_0v_{xy} + v_0^2)l_2 \\ u_0 - v_{xx} + (f^2 - u_0v_{xx} + u_0^2 - v_0v_{xy} + v_0^2)l_1 \\ (v_{xx} - u_0)l_2 - (v_{xy} - v_0)l_1 \end{bmatrix} \tag{9}$$

where $(l_1, l_2, 1)$ is the line $\mathbf{s} \times \mathbf{p}$.

If the distances between the collinear points were known, $\mathbf{v}_y$ could have been easily determined using the cross ratio $\mathbf{s}, \mathbf{v}_y; \mathbf{p}, \mathbf{o}$. However, in the absence of metric information, a natural solution can be obtained by equating the above cross-ratio across two viewpoints (i.e. two image pairs). In particular, since $\mathbf{v}_y$ is defined as a function of f and the principal point $\mathbf{c}$ in equation (9), a relation between f and the principal point $\mathbf{c}$ is obtained by the following equality between the two viewpoints:

$$\{\mathbf{s}, \mathbf{v}_y; \mathbf{p}, \mathbf{o}\}^1 = \{\mathbf{s}, \mathbf{v}_y; \mathbf{p}, \mathbf{o}\}^2 \tag{10}$$

where $\{\cdot, \cdot; \cdot, \cdot\}$ stands for the cross ratio of four points, and the superscripts indicate the viewpoints in which the cross-ratios are taken.

From (9) it follows that (10) is quadratic in f^2 and the coordinates of $\mathbf{c}$. Therefore one such equality provides two solutions for f^2 in terms of $\mathbf{c}$, of which only one is correct. However, the correct solution should minimize the symmetric transfer error of geometric distance as discussed later. We can compute $\mathbf{v}_y$ as a function of u_0 and v_0 using (9), from which we can find the three rotation angles as follows

$$\begin{matrix} j \\ z \end{matrix} = \tan^{-1}\left(\frac{v_{xy}^j - v_0^j}{v_{xx}^j - u_0^j}\right) \tag{11}$$

$$\begin{matrix} j \\ y \end{matrix} = -\tan^{-1}\left(\frac{f\sin(\ \begin{matrix} j \\ z \end{matrix})}{v_{xy}^j - v_0^j}\right) \tag{12}$$

$$\begin{matrix} j \\ x \end{matrix} = \tan^{-1}\left(\frac{f\cos(\begin{matrix} j \\ z \end{matrix})}{(v_{yy}^j - v_0^j)\cos(\ \begin{matrix} j \\ y \end{matrix}) - f\sin(\ \begin{matrix} j \\ z \end{matrix})\sin(\ \begin{matrix} j \\ y \end{matrix})}\right) \tag{13}$$

where j is the viewpoint number. Since in our case the world origin is not mapped to infinity, t_z can not be close to zero, and hence we can safely set the scalar t_z to 1. This yields the homography matrix in terms of the principal point up to a scalar t_z for the last column. As pointed out in [11], t_z doesn't affect the computation of camera internal parameters and rotation angles. We compute the ratio $\frac{t_z}{t_z'}$ by requiring both image points $\mathbf{m}$ and $\mathbf{m}'$ to be projected to the same 3D world point $\mathbf{M}$. Therefore, we can compute the principal point $\mathbf{c}$ by minimizing the following symmetric transfer error of geometric distance

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\Omega} \sum_i d(\mathbf{m}_i, \mathbf{H}^{-1}\mathbf{m}_i')^2 + d(\mathbf{m}_i', \mathbf{H}\mathbf{m}_i)^2 \tag{14}$$

where Ω is the 2D search space of u_0 and v_0 and $\mathbf{H} = \frac{t_z}{t_z'}\hat{\mathbf{H}}'\hat{\mathbf{H}}^{-1}$. Because the principal points of recent CCD cameras are very close to the center of the image, the searching space Ω can be narrowed down to a window around the image center. Experimentally, we found that better results are reached by searching for the optimal triplet (u_0, v_0, f) in a 3D searching space Ω' initializing around the image center and the estimated f from equations (10) and (14).

Point	Position1	Position2
$\mathbf{M}_1$	(-75 0 0)	(75 0 0)
$\mathbf{M}_2$	(-45 -150 0)	(45 -150 0)
$\mathbf{P}$	(0 -375 0)	(0 -375 0)

Table 1. Coordinates of 1D object at two positions

Viewpoint	x	y	z	t_x	t_y	t_z
1^{st}	10	-7	3.8	20	40	350
2^{nd}	12	6	-5	20	-24	350
3^{rd}	12	13	-12	10	70	360
4^{th}	12	30	-12	10	20	350
5^{th}	12	16	-5	10	-40	360
6^{th}	10	5	3.8	20	26	370
7^{th}	12	16	-15	10	-4	380

Table 2. External Parameters for seven different viewpoint.

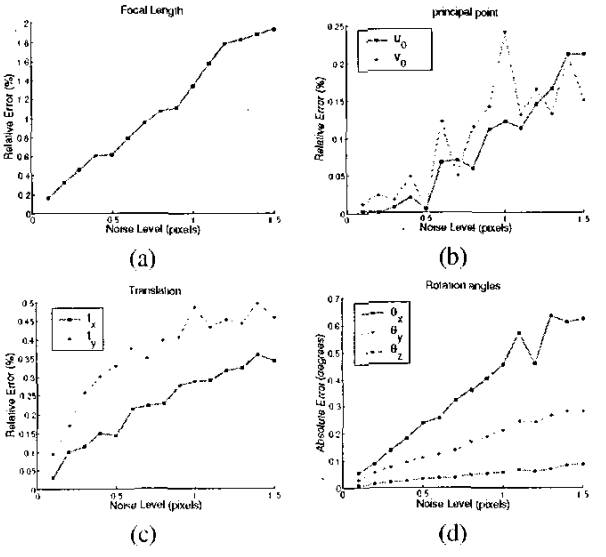


Fig. 2. Performance vs noise (in pixels) using $u_0 - v_0 - f$ search space, averaged over 100 independent trials: (a), (b) and (c) relative error for f, principal point and translation, (d) absolute errors in and rotation angles.

4. EXPERIMENTAL RESULTS

4.1. Computer Simulation

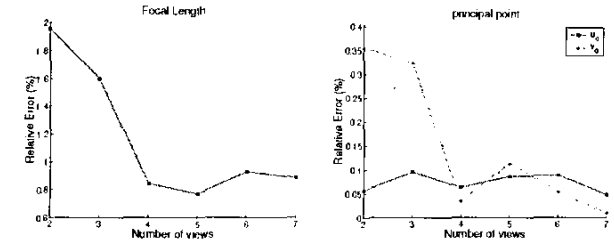
The simulated camera has a focal length of $f = 1020$, unit aspect ratio, zero skew, and the principal point at [316 243]. The image resolution is 640×480. In the experiments presented here, we observed the 1D objects randomly at seven positions listed in Table 2. For each observation, we switch the 1D object between two positions shown in Table 1. The 3D search space Ω' is $15 \times 15 \times 41$ with 1.0 pixel interval for f, u_0 and v_0.

Performance Versus Noise Level: In this experimentation, we used the first five image pairs in Table 2. Gaussian noise with zero mean and a standard deviation of $\sigma \leq 0.5$ was added to the projected image points. The estimated camera intrinsic and extrinsic parameters were then compared with the ground truth. For the rotation angles, we have shown the absolute errors of the 2^{nd} camera position in degrees. As argued by [14, 6] that the relative difference with respect to the focal length rather than the absolute error is a geometrically meaningful error measure. Therefore we measured the relative error of f and the principal point with respect to the focal length, and the translations with respect to t_z while varying the noise level from 0.1 pixels to 1.5 pixels. Results using 3D search space Ω' is shown in Figure 2. For $\sigma = 0.5$, which is larger than the typical noise in practical calibration, the relative error of focal length f is 1.92%. The maximum relative error of principal points is around 0.24% which is 2.44 pixels. Excellent performance is achieved for all extrinsic parameters, i.e. relative errors less that 0.36% for t_x and 0.50% for t_y, absolute errors less than 0.63 degree for $_x$, less than 0.28 degree for $_y$ and less than 0.086 degree for $_z$.

Performance Versus Number of Viewpoints: We also examined the performance with respect to the number of viewpoints (i.e. the number of image pairs). The orientation and position of the model planes are same as shown in Table 2. We varied the number of available viewpoints from 2 to 7. Results are shown in Figure 3. For these set of experimentations the noise level was kept at 0.5 pixels and the results were again averaged over 100 independent trials. For four or more viewpoints, the relative error of f drops sharply to an average of 0.8519%, and the relative errors of u_0 and v_0 drop sharply to an average of 0.715% and 0.5275%, respectively. The more viewpoints we have, the more accurate camera calibration will be in practice, since data redundancy compensates for the noise in the data.

image set	f	rel. err.	u_0	rel. err.	v_0	rel. err.
(1 2 3 4 5 6 7)	2443.22	-0.61%	1137.00	0.12%	838.67	-0.02%
(1 2 3 4 5 6 8)	2604.29	5.94%	1143.00	0.36%	844.00	0.20%
(1 2 3 4 5 7 8)	2510.88	2.14%	1125.33	-0.36%	838.00	-0.05%
(1 2 3 4 6 7 8)	2347.97	-4.48%	1137.83	0.15%	838.00	-0.05%
(1 2 3 5 6 7 8)	2472.31	0.57%	1131.00	-0.13%	838.00	-0.05%
(1 2 4 5 6 7 8)	2411.91	-1.88%	1137.00	0.12%	838.00	-0.05%
(1 3 4 5 6 7 8)	2456.22	-0.08%	1131.00	-0.13%	840.67	0.06%
(2 3 4 5 6 7 8)	2418.83	-1.60%	1131.00	-0.13%	838.00	-0.05%

Table 3. Intrinsic parameters for real data



(a) (b)

Fig. 3. Performance vs number of viewpoints using $u_0 - u_0$ search space, averaged over 100 independent trials: (a) relative error for f, (b) absolute errors in principal point.

Fig. 4. Three collinear points along a TV antenna.

4.2. Real Data

For the real data, we used a consumer market Canon PowerShot S45 camera and a cheap tripod. The 1D object is the antenna of a home TV, which has one fixed point. We took 16 images with 8 different viewpoints. One pair of images observed from the same viewpoint is shown in Figure 4. Three points along one antenna highlighted in red are chosen to generate three pairs of similar isosceles triangles. For each pair of similar triangle, we applied our algorithm using $u_0 - v_0$ search space independently, and the results were averaged. The calculated intrinsic parameters using these images are listed in Table 3. Since ground truth is not available, in order to validate our parameters we use the sample mean as the ground truth and also show the relative difference with respect to the mean value of f as discussed in section 4.1. The largest relative distance, in our case, is less than 6%. The error could be attributed to several sources. Besides noise, non-linear distortion, and imprecision of the extracted image points, one source is the casual experimental setup, which is deliberately targeted for non-expert users. Despite all these factors, our experimentations indicate that the proposed algorithm provides very good results.

5. CONCLUSION

We propose a calibration technique that makes camera calibration an easy off-the-shelf tool available to non-expert users. However, unlike classical techniques our approach does not require metric information about the calibration object. Also the 1D object can be chosen among many readily available objects such as the edge of the cover of a textbook, a TV or radio antenna, etc.

6. REFERENCES

[1] O.D. Faugeras, *Computer Vision: a Geometric Viewpoint*, MIT Press, 1993.

[2] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, 2000.

[3] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics & Automation*, vol. 3, pp. 323–344, 1987.

[4] P.F. Sturm and S.J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Proc. of CVPR*, 1999, pp. 432–437.

[5] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Analysis & Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[6] Z. Zhang, "Camera calibration with one-dimensional object," in *Proc. European Conference on Computer Vision*, 2002, pp. 161–174.

[7] P. Baker and Y. Aloimonos, "Calibration of a multicamera network," in *Proc. of IEEE Workshop on Omnidirectional Vision and Camera Networks*, 2003.

[8] O. Faugeras, T. Luong, and S. Maybank, "Camera self-calibration: theory and experiments," in *Proc. of European Conference on Computer Vision*, 1992, pp. 321–334.

[9] Richard I. Hartley, "Self-calibration from multiple views with a rotating camera," in *Proc. European Conference on Computer Vision*, 1994, pp. 471–478.

[10] S.J. Maybank and O.D. Faugeras, "A theory of self-calibration of a moving camera," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 123–152, 1992.

[11] X. Cao and H. Foroosh, "Simple calibration using an isoceles trapezoid," in *Proc. of the 17th International Conference on Pattern Recognition*, 2004.

[12] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer Vision*, vol. 4, no. 2, pp. 127–140, 1990.

[13] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, 2001.

[14] B. Triggs, "Autocalibration from planar scenes," in *Proc. European Conference on Computer Vision*, 1998, pp. 89–105.