Simple Calibration Without Metric Information Using an Isoceles Trapezoid

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 an isoceles trapezoid. Assuming a unit aspect ratio and zero skew, we introduce a novel and simple camera calibration approach. The key features of the proposed technique are its simplicity and the lack of need for 3D coordinate information about the calibrating object - i.e. the isosceles trapezoid. By ultilizing the symmetry of such trapezoid, we show that one can obtain both the internal and the external camera parameters. To demonstrate the effectiveness of the algorithm, we present the processing results on synthetic and real images, and compare our results to Zhang's flexible calibration method.

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 (up to a global scalar), and to determine the rigid motion of the camera with respect to the world coordinate frame. Classical techniques for camera calibration [3, 6, 9] require a so called calibration rig. The basic idea is to establish correspondences between a set of points in the world coordinate frame and their projections in the image plane, so that both the internal and the external geometry of the imaging system can be determined.

Recently, Zhang [11] has shown that it is possible to calibrate a camera using a planar point pattern shown at a few different orientations, and has given very accurate results. Zhang [12] has also presented a method for calibration using a 1D object with a fixed pivoting point, which filled the missing dimension in calibration. But these methods still require metric information about the calibrating object. Other techniques that do not require metric information [4, 5, 7] lead to solving more complex non-linear problems that are less tractable or ill-conditioned, and may not be always guarantied to converge to the global minimum. Motivated by making camera calibration an off-the-shelf tool available to a wider spectrum of users, we have developed an easy technique that uses a simple calibration object, but requires no metric information, and that reduces the problem to that of solving a set of simple second order equations.

*This work was partially supported by Sun Microsystems 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

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \mathbf{A}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3 \ \mathbf{t}] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

$$= \underbrace{\mathbf{A}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}]}_{\mathbf{H}} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2)$$

where $\sim$ indicates equality up to a scalar, $\mathbf{r}_i$ are the columns of the orthonormal rotation matrix R, $\mathbf{t}$ is the translation vector, $\mathbf{A}$ is the 3×3 upper triangular camera intrinsic matrix, and $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$ is the homography mapping the world plane to the image plane. Assuming a unit aspect ratio and zero camera skew, which is a reasonable assumption for modern CCD cameras [1, 2, 10], $\mathbf{A}$ will reduce to the form

$$\mathbf{A} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where f is the camera focal length, and $[u_0 \ \ v_0 \ \ 1]^T$ is the principal point $\mathbf{c}$.

Let $\{\mathbf{M}_i\}_{i=1...4}$ be four corners of an isosceles trapezoid in a world plane as shown in Figure 1, and without loss of generality assume that this plane is $Z = 0$. Using simple geometry one can readily show that the axis of symmetry of the trapezoid is then given by the line $\mathbf{SP}$, where $\mathbf{S}$ is the intersection of $\mathbf{M}_1\mathbf{M}_3$ and $\mathbf{M}_2\mathbf{M}_4$ projected to the point $\mathbf{s}$ in the image plane, and $\mathbf{P}$ is the intersection of $\mathbf{M}_1\mathbf{M}_4$ and $\mathbf{M}_2\mathbf{M}_3$ projected to the point $\mathbf{p}$ in the image plane. Since in the world plane the line $\mathbf{SP}$ is perpendicular to $\mathbf{M}_1\mathbf{M}_2$ and $\mathbf{M}_3\mathbf{M}_4$, we can choose the world coordinate frame as follows: origin at $\mathbf{O}$, i.e. the intersection of $\mathbf{SP}$ and $\mathbf{M}_1\mathbf{M}_2$,

x-axis along $\mathbf{M}_1\mathbf{M}_2$ with positive direction towards $\mathbf{M}_2$, y-axis along $\mathbf{SP}$ with positive direction towards $\mathbf{S}$, and z-axis given by the right-hand rule. The camera and the image coordinate systems are set as customary. We will derive the basic constraints on $\mathbf{H}$ given the configuration described above.

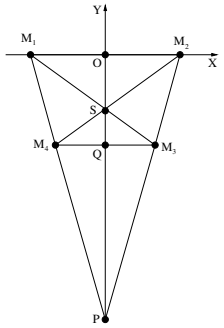


Figure 1. Four points forming an isosceles trapezoid.

3 Constraints on H

From equation (2) it follows that

$$\mathbf{H} = \begin{bmatrix} fr_{11} + u_0 r_{31} & fr_{12} + u_0 r_{32} & ft_x + u_0 t_z \\ fr_{21} + v_0 r_{31} & fr_{22} + v_0 r_{32} & ft_y + v_0 t_z \\ r_{31} & r_{32} & t_z \end{bmatrix} \tag{4}$$

where r_{ij} are the components of the rotation matrix $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$ given by

$$r_{11} = \cos(\theta_y)\cos(\theta_z) \tag{5}$$
$$r_{21} = \cos(\theta_y)\sin(\theta_z) \tag{6}$$
$$r_{31} = -\sin(\theta_y) \tag{7}$$
$$r_{12} = \sin(\theta_x)\sin(\theta_y)\cos(\theta_z) - \cos(\theta_x)\sin(\theta_z) \tag{8}$$
$$r_{22} = \sin(\theta_x)\sin(\theta_y)\sin(\theta_z) + \cos(\theta_x)\cos(\theta_z) \tag{9}$$
$$r_{32} = \cos(\theta_y)\sin(\theta_x) \tag{10}$$

and t_x, t_y, and t_z are the components of the translation vector.

Now, let us denote the homogeneous x and y vanishing points in the image coordinate system by $\mathbf{v}_x = [v_{xx} \;\; v_{xy} \;\; 1]^T$ and $\mathbf{v}_y = [v_{yx} \;\; v_{yy} \;\; 1]^T$, and the image point corresponding to the projection of $\mathbf{O}$ by $\mathbf{o} = [o_x \;\; o_y \;\; 1]^T$. As is well known [1, 2, 3], $\mathbf{h}_1 = r_{31}\mathbf{v}_x$, $\mathbf{h}_2 = r_{32}\mathbf{v}_y$, and $\mathbf{h}_3 = t_z\mathbf{o}$. On the other hand, for the configuration of the points proposed above, we have

$$\mathbf{v}_x \sim (\mathbf{m}_1 \times \mathbf{m}_2) \times (\mathbf{m}_3 \times \mathbf{m}_4) \tag{11}$$
$$\mathbf{v}_y^T(\mathbf{s} \times \mathbf{p}) = 0 \tag{12}$$
$$\mathbf{o} \sim (\mathbf{m}_1 \times \mathbf{m}_2) \times (\mathbf{s} \times \mathbf{p}) \tag{13}$$

These are the five basic constraints on our homography $\mathbf{H}$, given a single image of an isosceles trapezoid described above. In general, a 2D homography has eight degrees of freedom. 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 $\mathbf{v}_z$ [1]. Thus,

$$(\mathbf{v}_x - \mathbf{c})^T(\mathbf{v}_y - \mathbf{c}) + f^2 = 0 \tag{14}$$

By combining equations (11),(12), and (14), we can show that $\mathbf{v}_y$ is given by

$$\mathbf{v}_y \sim \begin{bmatrix} v_{xx} - u_0 \\ v_{xy} - v_0 \\ f^2 - u_0 v_{xx} + u_0^2 - v_0 v_{xy} + v_0^2 \end{bmatrix} \times (\mathbf{s} \times \mathbf{p}) \tag{15}$$

which after some algebraic manipulations indicates that $\mathbf{v}_y$ is of the form

$$\mathbf{v}_y = \begin{bmatrix} k_1 f^2 + k_2 & k_3 f^2 + k_4 & 1 \end{bmatrix}^T \tag{16}$$

where k_i depend on the principal point $\mathbf{c}$.

4 Solving Camera Calibration

Since cross-ratio is a projective invariant, a natural solution in the absence of metric information can be obtained by equating a cross-ratio across two or more images. In particular, since $\mathbf{v}_y$ is defined as a function of f and the principal point $\mathbf{c}$ in equation (16), a cross-ratio involving $\mathbf{v}_y$ along the line $\mathbf{s} \times \mathbf{p}$ would yield a relation between f and the principal point $\mathbf{c}$.

As shown in Figure 1, the four points $\mathbf{q}$, $\mathbf{v}_y$, $\mathbf{p}$, and $\mathbf{o}$, are colinear, and their cross-ratio is preserved under perspective projection. Therefore, we can write the following equality between two given images

$$\{\mathbf{q}, \mathbf{v}_y; \mathbf{p}, \mathbf{o}\}^1 = \{\mathbf{q}, \mathbf{v}_y; \mathbf{p}, \mathbf{o}\}^2 \tag{17}$$

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

From (16) it follows that (17) 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. Once the correct solutions for f^2 and $\mathbf{c}$ are found by minimizing this error, we can compute $\mathbf{v}_y$ using (16), from which we can find the three rotation angles as follows

$$\theta_z^j = \tan^{-1}\left(\frac{v_{xy}^j - v_0^j}{v_{xx}^j - u_0^j}\right) \tag{18}$$

$$\theta_y^j = -\tan^{-1}\left(\frac{f\sin(\theta_z^j)}{v_{xy}^j - v_0^j}\right) \tag{19}$$

$$\theta_x^j = \tan^{-1}\left(\frac{f\cos(\theta_z^j)}{(v_{yy}^j - v_0^j)\cos(\theta_y^j) - f\sin(\theta_z^j)\sin(\theta_y^j)}\right) \tag{20}$$

where j is the image 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.

Proposition 1. Let the true world-to-image homography be $\bar{\mathbf{H}} = [\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3]$, and the estimated world-to-image homography be $\hat{\mathbf{H}} = [\mathbf{h}_1\ \mathbf{h}_2\ \frac{1}{t_z}\mathbf{h}_3]$.

1. If $\bar{\mathbf{H}}^{-1} = [\mathbf{g}_1^T\ \mathbf{g}_2^T\ \mathbf{g}_3^T]^T$, then $\hat{\mathbf{H}}^{-1} = [\mathbf{g}_1^T\ \mathbf{g}_2^T\ t_z\mathbf{g}_3^T]^T$.

2. Given an image point $\mathbf{m}$, its corresponding true inhomogeneous world point $\bar{\mathbf{M}}$ and its estimated inhomogeneous world point $\hat{\mathbf{M}}$, we have: $\hat{\mathbf{M}} = \frac{1}{t_z}\bar{\mathbf{M}}$.

From proposition 1, we can 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

$$[\hat{u}_0\ \hat{v}_0] = \arg\min_{u_0,v_0\in\Omega} \sum_i d(x_i, \mathbf{Q}^{-1}x'_i))^2 + d(x'_i, \mathbf{Q}x_i)^2 \tag{21}$$

where Ω is the 2D searching space of u_0 and v_0, and $\mathbf{Q} = \hat{\mathbf{H}}'\hat{\mathbf{H}}^{-1}$ is the inter-image homography. 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 centered on 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 (21) and (17).

5 Experimental Results

The proposed approach has been tested on an extensive set of simulated and real data. Due to lack of space only some are presented below.

image	θ_x	θ_y	θ_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 1. Ground truth of simulated data

5.1 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 herein, we generated seven planes randomly with parameters listed in Table 1. The 3D search space Ω' is $25 \times 25 \times 41$ with 0.5 pixel interval for both u_0 and v_0 and 1 pixel interval for focal length.

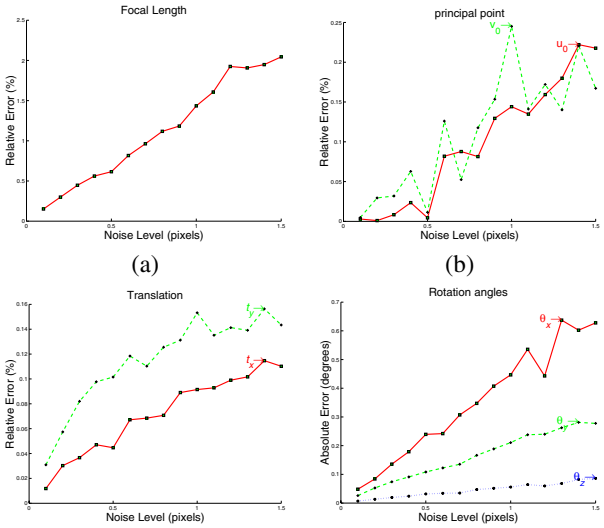


Figure 2. Performance vs noise (in pixels) using $u_0 - v_0 - f$ search space, averaged over 100 independent trials: (a), (b) and (c) relative errors of f, principal point and translation, (d) absolute error of rotation angles.

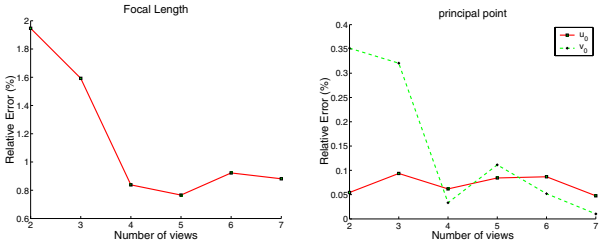
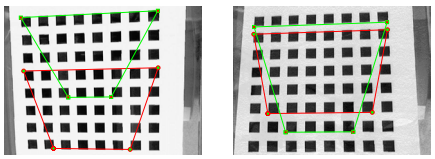


Figure 3. Performance vs number of images using $u_0 - v_0$ search space, averaged over 100 independent trials: (a) relative error of f, (b) absolute errors of principal point.

Performance Versus Noise Level: In this experiment, we use the first five images in Table 1. 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} image in degrees. We measured the relative error of f, principal point, and translation with respect to the focal length, as argued by [8, 12] that the relative difference with respect to the focal length rather than the absolute error is a geometrically meaningful error measure, 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 around 2.0% , 0.55% better than that that by using 2D search space. The maximum relative error of princpal points is around 0.25% which is 2.5 pixels. Excellent performance is achieved for all extrinsic parameters, i.e. relative errors less that 0.16% for the translations, absolute error less than 0.65 degree for θ_x, less than 0.3 degree for θ_y and less than 0.09 degree for θ_z.

Performance Versus Number of Images: We also examined the performance with respect to the number of images. The orientation and position of the model planes are same as shown in Table 1. We vary the number of images 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 images, the relative error of f drops sharply to an average of 0.8519%, and the relative errors of u_0 and v_0 drops sharply to an average of 0.715% and 0.5275% separately. The more points or images we have, the more accurate camera calibration will be in practice because data redundancy can combat the noise in image data.



(a) (b)

Figure 4. Four trapezoids projected to two images.

quadruple	(1234)	(1235)	(1245)	(1345)	(2345)	mean	dev
α [11]	831.81	832.09	837.53	829.69	833.14	832.85	2.90
β [11]	831.82	832.10	837.53	829.91	833.11	832.90	2.84
f **(Ours)**	**833.55**	**834.29**	**835.96**	**832.33**	**834.28**	**834.08**	**1.32**
u_0 [11]	304.53	304.32	304.57	303.95	303.53	304.18	0.44
u_0 **(Ours)**	**303.50**	**304.25**	**303.25**	**304.00**	**304.50**	**303.90**	**0.52**
v_0 [11]	206.79	206.23	207.30	207.16	206.33	206.76	0.48
v_0 **(Ours)**	**217.25**	**213.25**	**212.25**	**205.25**	**208.50**	**211.30**	**4.60**

Table 2. Results for real data compared to Zhang's.

5.2 Real Data

In this subsection, we compared our algorithm to Zhang's flexible calibration method [11]. We use four trapezoids (i.e. 16 corners) as shown in Figure 4. The radial distortion was not considered in this paper and was removed according to [11]'s experimental results. For every trapezoid, we apply our algorithm using $u_0 - v_0$ search space independently, and the results are averaged ones. In order to evaluate our results, we also used an approach similar to [11] based on estimating the uncertainty of the results using the standard deviation of the estimated internal parameters f(compared to α and β in Zhangs algorithm), u_0 and v_0. We evaluated the variation of calibration results among all quadruples of images. Results are compared to those by Zhang in Table 2.

6 Conclusion

We propose a calibration technique that makes camera calibration an easy off-the-shelf tool available to a wide spectrum of users. Similar to classical photometric techniques our method requires a calibration object, i.e. an isosceles trapezoid. However, unlike these classical techniques our approach does not require metric information about the calibration object. The latter property makes our approach more similar to auto-calibration techniques. But with the advantage that unlike auto-calibration techniques our method does not require bootstraping the calibration from an intermediate projective reconstruction. Based on these observations, one can view our method as residing halfway between classical photogrammetric calibration methods and more recent auto-calibration techniques.

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *Int. J. Computer Vision*, 4(2):127–140, 1990.

[2] Criminisi, Reid, and A. Zisserman. Single view metrology. *Int. J. Computer Vision*, 2001.

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

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

[5] R. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. European Conference on Computer Vision*, pages 471–478, 1994.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *Int. J. Computer Vision*, 8(2):123–152, 1992.

[8] B. Triggs. Autocalibration from planar scenes. In *Proc. European Conference on Computer Vision*, pages 89–105, 1998.

[9] R. 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 and Automation*, 3(4):323–344, 1987.

[10] K.-Y. Wong, R. Mendona, and R. Cipolla. Camera calibration from surfaces of revolution. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 2003.

[11] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 22(11):1330–1334, 2000.

[12] Z. Zhang. Camera calibration with one-dimensional objects. pages 161–174, 2002.