Camera Calibration

E. Hammer Johnson
johnsoeh@stolaf.edu

I. Introduction

When capturing images for scientific use, image accuracy is often desirable if not paramount. However, in the real world, cameras do not capture perfect perspective images, instead exhibiting both random and systematic error. In lieu of technological advancements facilitating the dissemination of super-accurate cameras, a natural solution to the problem of accuracy is camera calibration. If errors in the image capture process can be modeled, they can be corrected, and undistorted images acquired without the use of specialized, expensive, or nonexistent equipment.

A general algorithm for a widely practiced calibration process is adapted from [1] as follows:

1) Using a single camera, take multiple images of a planar target which has on it a set of feature points with known geometry.
2) Locate the feature points on the target in each image.
3) Providing a parameterized model for the camera, estimate parameters so as to minimize discrepancy between model and observed locations of feature points.

This paper focuses in particular upon the work of [1]; for a geometric interpretation of the following, the interested reader is referred to [2]. Section II introduces camera calibration in an ideal world. Section III addresses realworld noise in data and equipment. Section IV begins to explore causes and effects of error in the calibration process.

II. Ideal Camera

It is generally accepted that straight lines closely approximate light in most applications. An ideal camera—one producing an image of perfect perspective—can therefore be modeled by the intersection of some *image plane* and a pencil of lines (rays of light) through a single *focal point*. This is known as the pin-hole camera model for the singularity of its focal point. Where the captured scene is also a plane, which is here the case, the image capture process reduces to a projection of one plane onto another.

A. Pinhole Camera Model

Notation: Vectors and matrices appear boldface; all else is scalar. For all vectors $\mathbf{v}$, we notate $\sqrt{\mathbf{v}\cdot\mathbf{v}} = ||\mathbf{v}||$. For an introduction to homogeneous coordinates in the language of calibration, the interested reader is referred to [3].

Given an image of a calibration plane, let $\mathbf{R}$ be a rotation matrix and $\mathbf{t}$ a translation vector such that $\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$ is a 3×4 matrix for the change of basis from the realworld homogenous coordinate system (3D) to the camera homogenous coordinate system (2D). Given realworld coordinates and corresponding image coordinates of reference points,

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$$

respectively, we model the projection up to a scalar s with

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

where nonrigid

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

corrects the geometric effects of perspective, with α and β scaling factors to correct for nonsquare image pixels, γ to correct for skewed image axes, and (u_0, v_0) the so-called *principle point* of the image.

B. Calibration Computation

Without loss of generality fix the world coordinate system such that the calibration plane lies on $Z = 0$, and let the rotation matrix $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}$. Then for all points on the calibration plane, $Z = 0$ and so

$$\begin{aligned} s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \\ &= \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \end{aligned}$$

by multiplication.

1) Estimating Homographies: Given a set of corresponding calibration plane and image points

$$\mathbf{P}_i = \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$$

respectively, a variety of methods can be used to estimate the homography between them. Although Zhang presents a maximum likelihood estimation method [1], it reduces to a nonlinear least squares problem under the given assumptions of Gaussian noise applied to all $\mathbf{m}_i$ with mean $\mathbf{0}$ and covariance matrix a scalar multiple of $\mathbf{I}$. In particular, defining

$$\begin{bmatrix} \hat{p}_{i1} \\ \hat{p}_{i2} \\ \hat{p}_{i3} \end{bmatrix} = \mathbf{H}\mathbf{P}_i$$

and

$$\hat{\mathbf{p}}_i = (1)/(\hat{p}_{i3}) \begin{bmatrix} \hat{p}_{i1} \\ \hat{p}_{i2} \\ \hat{p}_{i3} \end{bmatrix}$$

we minimize

$$\sum_i ||\mathbf{p}_i - \hat{\mathbf{p}}_i||^2.$$

Any nonlinear minimization algorithm will suffice here; Zhang happens to use the Levenberg-Marquardt Algorithm [1].

2) Obtaining Constraints: By the preceding, let a 3×3 homography $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix}$ be estimated such that

$$k \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

for some scalar k. Whereas $\mathbf{H}$ has eight degrees of freedom as a homography, $\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$ has only six parameters, so we can procure two constraints on $\mathbf{A}$ by assuming that $\mathbf{H}$ fits the pin-hole model. In particular, we have

$$\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}$$

for some scalar λ. Rearranging, we get

$$\mathbf{A}^{-1} \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}$$

(note that $\mathbf{A}$ is invertible by design) whence

$$\mathbf{A}^{-1}\mathbf{h}_1 = \lambda \mathbf{r}_1 \quad \text{and} \quad \mathbf{A}^{-1}\mathbf{h}_2 = \lambda \mathbf{r}_2$$

as we multiply through.

Now, $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal since $\mathbf{R}$ is a rotation matrix. So $\mathbf{r}_1 \cdot \mathbf{r}_2 = 0$, and in particular

$$\begin{aligned} 0 &= \lambda \mathbf{r}_1 \cdot \lambda \mathbf{r}_2 \\ &= \mathbf{A}^{-1}\mathbf{h}_1 \cdot \mathbf{A}^{-1}\mathbf{h}_2 \\ &= \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2. \end{aligned}$$

Furthermore, $||\mathbf{r}_1|| = 1 = ||\mathbf{r}_2||$, and in particular

$$\begin{aligned} \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 &= ||\mathbf{A}^{-1}\mathbf{h}_1|| \\ &= \lambda ||\mathbf{r}_1|| = \lambda ||\mathbf{r}_2|| \\ &= ||\mathbf{A}^{-1}\mathbf{h}_2|| \\ &= \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2. \end{aligned}$$

Hence each such homography $\mathbf{H}$ gives

$$\begin{aligned} 0 &= \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \\ \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 &= \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2, \end{aligned}$$

two additional constraints on $\mathbf{A}$.

3) Final Solution: We can rearrange the aforementioned constraints on $\mathbf{A}$ to determine all parameters of a camera calibration. With $\mathbf{A}$ as before, let $\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1}$; then simple calculation gives $\mathbf{B} =$

$$\begin{bmatrix} \frac{1}{\alpha^2} & \frac{-\gamma}{\alpha^2\beta} & \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ \frac{-\gamma}{\alpha^2\beta} & \frac{\gamma^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix}.$$

Since $\mathbf{B}$ is symmetric, we can now write our constraints on $\mathbf{A}$ from (2) above in terms of the 6D vector

$$\mathbf{b} = \begin{bmatrix} \frac{1}{\alpha^2} \\ \frac{-\gamma}{\alpha^2\beta} \\ \frac{\gamma^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} \\ \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \\ \frac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix}.$$

Where

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{21} & h_{31} \\ h_{12} & h_{22} & h_{32} \\ h_{13} & h_{23} & h_{33} \end{bmatrix}$$

is a homography from (1) as before and

$$\mathbf{v}_{ij} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix},$$

our two constraints can be written as homogenous equations

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = \mathbf{0}.$$

This notation intuitively shows that for $n \geq 3$ distinct homographies, $\mathbf{b}$ (whence $\mathbf{B}$ whence $\mathbf{A}$) is in general defined up to a scale factor. For discussion of degenerate cases in which $n \geq 3$ is not sufficient the interested reader is referred to [1].

If each $\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$ transformation is of interest, determination follows directly from that of $\mathbf{A}$ using the preceding definitions in terms of $\mathbf{A}$ and $\mathbf{H}$.

III. Realworld Camera

In practice, one cannot with infinite precision construct cameras, locate image reference points, or even evaluate expressions. Accordingly, to overcome "random" noise in data, over-constraint is accepted as inevitable and error is minimized over as large a dataset as possible. Solutions to the systems of equations in Section II can all be minimized with respect to squared error using minimization algorithms such as Levenberg-Marquardt [4]. Similarly, to overcome the "systematic" physical imperfections of a camera, the pinhole camera model is augmented to correct image distortion.

A. Distortion Models

Camera lenses vary widely in number, type, and configuration of optical elements; the effect of distortion can be quite complex. Most distortion models—including those of widest use—consider only idealized categories of distortion.

1) Radial: In an ideal camera, varying the curvature of one lens surface produces *radial distortion*, a per-radius displacement of image points symmetric about some *principle point* from which radii are measured. Previous work indicates that distortion observed in most cameras can be reasonably approximated by simple radial distortion models alone [1]. The usefulness of more complex radial distortion models is of continued debate; cf. [1], [5], [6].

2) Tangential: In another ideal camera, varying the alignment of two optical elements produces *tangential distortion*, less commonly called *decentering distortion*. Compared to radial distortion, tangential distortion describes much less of the distortion observed in most cameras today. For this reason it is generally the case that distortion models emphasize radial over tangential distortion, in some cases ignoring tangential distortion altogether [6]. Brief descriptions of the most popular distortion models follow.

Brown-Conrady: The Brown-Conrady model essentially fits power series to observed distortion [5]. In practice, the involved power series are truncated to finite polynomials, so the Brown model is also called the *polynomial model*, and less commonly the *radial model* in reference to the indeterminate of its polynomials. Writing undistorted image coordinates (x, y) in terms of center of distortion $(0, 0)$, and with $r^2 = x^2 + y^2$, distorted image coordinates (x', y') are

$$\begin{aligned} x' &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots) \\ &\quad +(P_1(r^2 + 2x^2) + 2P_2 xy)(1 + P_3 r^2 + \cdots) \\ y' &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots) \\ &\quad +(P_2(r^2 + 2y^2) + 2P_1 xy)(1 + P_3 r^2 + \cdots) \end{aligned}$$

where k_i and P_i are the ith radial and tangential distortion parameters, respectively. Even-numbered powers of r ensure symmetry about $(0, 0)$ for radial distortion, whereas tangential distortion allows asymmetry with the roles of P_1 and P_2 reversed when computing x' and y'.

Although Brown-Conrady is a very popular distortion model, it remains a matter of debate as to whether higher degree polynomials can increase its accuracy [1], [5], [6].

Division: The division model is a lightweight revision of the Brown model, disregarding tangential distortion and often appearing with only one radial distortion parameter [7]. With variables as before,

$$\begin{aligned} x' &= (x)/(1 + k_1 r^2 + \cdots) \\ y' &= (y)/(1 + k_1 r^2 + \cdots). \end{aligned}$$

Despite its relative simplicity, the division model regularly attains visually acceptable distortion correction with only one distortion parameter [7].

Polynomial: In contrast to the Brown-Conrady model with polynomials in r, the polynomial model utilizes polynomials in x and y; each coefficient is then a distinct distortion parameter [6]. For example,

$$\begin{aligned} x' &= a_1 x^3 + a_2 x^2 y + a_3 x y^2 + a_4 y^3 \\ &\quad + a_5 x^2 + a_6 xy + a_7 y^2 + a_8 x + a_9 y + a_{10} \\ y' &= b_1 x^3 + b_2 x^2 y + b_3 x y^2 + b_4 y^3 \\ &\quad + b_5 x^2 + b_6 xy + b_7 y^2 + b_8 x + b_9 y + b_{10} \end{aligned}$$

where a_i and b_i are all distortion parameters. The order three polynomial model shown above is particularly referred to as the *bicubic model* [8].

Gioi, et al. in [6] compare those models listed here and achieve least error with high-order polynomial models.

Rational Function: The rational function model, also abbreviated the *rational model*, extends the polynomial model to allow rational polynomials:

$$\begin{aligned} x' &= \frac{a_1 x^3 + a_2 x^2 y + \cdots + a_9 y + a_{10}}{c_1 x^3 + c_2 x^2 y + \cdots + c_9 y + c_{10}} \\ y' &= \frac{b_1 x^3 + b_2 x^2 y + \cdots + b_9 y + b_{10}}{c_1 x^3 + c_2 x^2 y + \cdots + c_9 y + c_{10}} \end{aligned}$$

with a_i, b_i, and c_i all distortion parameters [8].

B. Least Squares Minimization

In most cases, calibration algorithms outsource minimization work to pre-existing minimization algorithms. However, despite the crucial role of minimization in the calibration process, calibration results from multiple minimization algorithms are rarely given in any one publication. Literature search yields little insight into the choice of particular minimization algorithm or the properties of calibration error functions encountered in practice. For instance, although the Levenberg-Marquardt algorithm [4] is one of the most widely used such minimization algorithms, de Villiers, et al. find it yields error highest of four minimization algorithms compared in [5].

IV. Error Analysis

There is great potential for error in the calibration process. We explore some sources of error below.

1) Reference Points: The parameters of our calibration model are determined solely by ideal-distorted pairs of image coordinates; whereas the ideal coordinates of reference points are assumed *a priori*, the coordinates of corresponding points in the distorted image must be located. Zhang simulates calibration with the pinhole and radial distortion models and applies both random noise and systematic non-planarity to points on the calibration plane: systematic bias of reference points caused absolute error of more than 800% that caused by Gaussian noise [1]. Thus in both calibration target construction and reference point location, accuracy is paramount and precision of secondary concern.

2) Modeling: As observed in [6], there exists no literature relating the pinhole camera model to physical knowledge of cameras. It is as of yet unknown how accurately the pinhole model can be expected to approximate a given realworld camera when augemented with a distortion function. Nevertheless, overall modeling error as presented in literature is generally associated with the distortion function alone, and the potential accuracy of the pinhole-distortion combination is not questioned [6]. This practice likely reflects the ready availability of many distortion models.

3) Minimization: Although Levenberg-Marquardt and other minimization algorithms may be guaranteed to converge, it is in general not the case that this occurs at the desired minimum and with infinite precision, to say nothing of whether or not the algebraic distance being minimized is meaningful in the first place.

De Villiers, et al. find that Levenberg-Marquardt performed worst of four minimization algorithms used for calibration [5]. In particular, the Leapfrog Algorithm [9] consistently gave lowest calibration error, while the Fletcher-Reeves algorithm [10] exhibited faster convergence and less sensitivity to input, with error only slightly more than Leapfrog [5]. Leapfrog simulates a physical particle under the action of acceleration according to the gradient of the function to be minimized; Fletcher-Reeves considers past search directions in addition to gradient information to determine next direction [5]. Since minimization algorithms require an initial estimate of the solution, it is difficult to say whether results from [5] indicate that Leapfrog and Fletcher-Reeves are better-suited to the given minimization tasks in general or simply outperform Levenberg-Marquardt given certain input.

References

[1] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000. [Online]. Available: http://dx.doi.org/10.1109/34.888718

[2] O. Faugeras, “Stratification of 3-d vision: projective, affine, and metric representations,” *Journal of the Optical Society of America A*, vol. 12, pp. 46 548–4, 1995.

[3] P. Hillman, “White paper: Camera calibration and stereo vision,” *Square Eyes Software*, October 2005. [Online]. Available: www.peterhillman.org.uk/downloads/whitepapers/calibration.pdf

[4] M. I. A. Lourakis, “A brief description of the levenberg-marquardt algorithm implemened by levmar,” *Matrix*, vol. 3, p. 2, 2005. [Online]. Available: http://www.ics.forth.gr/~lourakis/

[5] J. P. De Villiers, F. W. Leuschner, and R. Geldenhuys, “Centi-pixel accurate real-time inverse distortion correction,” *Proceedings of SPIE*, vol. 7266, no. 1, pp. 726 611–726 611–8, 2008. [Online]. Available: http://link.aip.org/link/PSISDG/v7266/i1/p726611/s1&Agg=doi

[6] R. Grompone von Gioi, P. Monasse, J.-M. Morel, and Z. Tang, “Self-consistency and universality of camera lens distortion models,” *Prepublication*, May 2011. [Online]. Available: http://www.cmla.ens-cachan.fr/fileadmin/Documentation/Prepublications/2011/CMLA2011-01.pdf

[7] A. Wang, T. Qiu, and L. Shao, “A simple method of radial distortion correction with centre of distortion estimation,” *J. Math. Imaging Vis.*, vol. 35, no. 3, pp. 165–172, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1007/s10851-009-0162-1

[8] D. Claus and A. W. Fitzgibbon, “A rational function lens distortion model for general cameras,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, june 2005, pp. 213–219. [Online]. Available: www.robots.ox.ac.uk/~dclaus/publications/claus05rf_model.html

[9] J. Snyman, “An improved version of the original leap-frog dynamic method for unconstrained minimization: Lfop1(b),” *Applied Mathematical Modelling*, vol. 7, no. 3, pp. 216 – 218, 1983. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0307904X83900112

[10] R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients,” *The Computer Journal*, vol. 7, no. 2, pp. 149–154, 1964. [Online]. Available: http://comjnl.oupjournals.org/cgi/doi/10.1093/comjnl/7.2.149