Chapter #

Camera Calibration

Finding the Intrinsic and Extrinsic Parameters

Zhigang Zhu
Department of Computer Science, The City College of New York, New York, NY 10031

1. INTRODUCTION

Outline of this chapter

- Calibration: Finding the intrinsic and extrinsic parameters
  - Problems and assumptions
  - Direct parameter estimation approach
  - Projection matrix approach
- Direct Parameter Estimation Approach
  - Basic equations (from the last chapter)
  - Homogeneous system
  - Estimating the image center using vanishing points
  - SVD (Singular Value Decomposition)
  - Focal length, aspect ratio, and extrinsic parameters
  - Discussion: why not do all the parameters together?
- Projection Matrix Approach
  - Estimating the projection matrix M
  - Computing the camera parameters from M
  - Discussion
- Comparison and Summary

1.1 Problems and Assumptions

In the previous chapter discussing camera models, our goal was to build the geometric relations between a 3D scene and its 2D images for both computer vision and computer graphics. The relations consist of two parts: 3D transformation that represents the viewpoints and viewing directions of the camera (real or virtual), and the perspective projection that maps 3D points onto 2D images. Recall that we can use the same sets of equations for both vision and graphics. However, we have also pointed out that vision is much more challenging than graphics: In graphics, 2D projections can easily be generated from 3D models, given the 3D models and the (virtual) camera parameters. In computer vision a real camera carries out the projections and the goal is to reconstruct the 3D models from these 2D images. These inverse problems are more difficult to solve, because (1) intrinsic and extrinsic camera parameters need to be extracted using the *calibration* procedure that we will discuss below; and (2) 3D information needs to be recovered from 2D images in which the third dimension is lost.

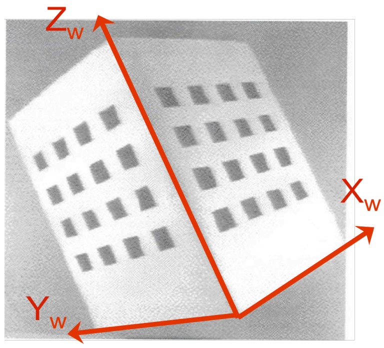

Figure 1. Calibration target pattern

In this chapter, we will discuss methods for solving the first problem. Informally, we can state the problem as the following:

Given one or more images of a calibration pattern, estimate the intrinsic parameters, the extrinsic parameters, or both.

We shall consider several aspects of the problem in order to achieve the desired accuracy of calibration.

(1) How shall we design and measure the calibration pattern? This determines the accuracy in 3D measurements. Figure 1 shows a typical calibration pattern, which has accurately measured chessboard patterns on the sides of a cube. The centers or corners of the black squares are called control points, which have accurate 3D coordinate measurements. The rule of thumb is to have a good distribution of the control points that are not coplanar and span across the camera's field of view, in order to assure stability of the

solution. Additionally, when constructing the calibration pattern, the tolerance of accuracy should be one or two orders of magnitude lower than the desired accuracy of calibration. For example, in order to achieve a desired accuracy of 0.1mm, we should use a 0.01 mm tolerance in construction.

(2) How can we extract the image correspondences? This determines the accuracy in 2D measurements. Techniques such as corner detection and line fitting can be used. Typically, sub-pixel accuracy (i.e., localization error below one pixel) for control point localization is needed in order to achieve the desired accuracy in calibration. This can often be done semi-automatically by handpicking the initial control points of the image and then refining them with sub-pixel localization techniques.

(3) How can we estimate the camera parameters by using the 3D-2D pair? This will be the focus of this chapter, assuming that both the 3D and 2D control points meet the requirements mentioned above.

Calibration is a tedious procedure, with its computational algorithms often sensitive to noise. So for some applications alternative approaches would be to obtain 3D information (structure, relations, etc.) from un-calibrated cameras.

1.2 Camera Model

In the previous chapter, we have discussed the perspective camera model: from world to camera (coordinate systems), and from camera (3D) to image frame (2D) (Figure 2).

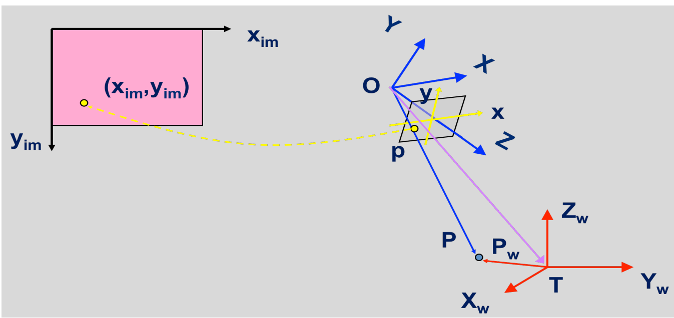

Figure 2. Intrinsic and extrinsic camera parameters

For the completeness of our discussion, we summarize it here. Overall, there are four points represented in four related coordinate systems:

(1) Point $p_{im}(x_{im}, y_{im})$ in the digital frame coordinate system, measured in pixels;
(2) Point p(x,y) in the physical image coordinate system, measured in mm;
(3) Point P(X,Y,Z) in the camera coordinate system, measured in mm;
(4) Point $P_w(X_w,Y_w,Z_w)$in the world coordinate system, measured in mm.

Note that the points P and P_w are actually two representations of the same 3D point, in the camera and the world coordinate systems, respectively. The camera parameters can be divided into two groups:

(1) Intrinsic parameters, which link the digital frame coordinates of an image point to its corresponding camera coordinates. They include the focal length of the camera, f, the image center, (o_x, o_y), the aspect ratio $\alpha=s_y/s_x$, and sometimes the radial distortion parameter(s).
(2) Extrinsic parameters, which define the location and orientation of the camera coordinate system with respect to the world coordinate system. We use a 3x3 rotation matrix $R = [r_{ij}]_{3x3}$ and a 3D translational vector $T = (T_x,T_y,T_z)^T$ to represent this set of parameters.

Integrating the transformations we described in Chapter #, we can establish the transformation from the world coordinate system to the digital frame coordinate system, i.e., $(X_w,Y_w,Z_w)^T \rightarrow (x_{im}, y_{im})^T$, as

$$
\begin{aligned}
x' = x_{im} - o_x &= -f_x \frac{r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z} \\
y' = y_{im} - o_y &= -f_y \frac{r_{21}X_w + r_{22}Y_w + r_{23}Z_w + T_y}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z}
\end{aligned}
\quad (1)
$$

where we define the effective focal lengths in the x and y directions as

$$f_x = f/s_x,\ fy=f/s_y \quad (2)$$

Note that in equation (1) we neglect the radial distortion parameters, and define (x',y') in a coordinate system with the origin (o_x,o_y).

The intrinsic parameters should be listed by the manufacturers in the camera specifications. Unfortunately these parameters may be incomplete or inaccurate. The extrinsic parameters that describe the pose (position and orientation) of the camera looking at the scene are usually unknown. The goal of calibration is to find the two sets of parameters, given a number of corresponding points in 3D and 2D, { $(X_w,Y_w,Z_w)_i^T$, $(x_{im}, y_{im})_i^T$, i=1, 2, … N}.

2. DIRECT PARAMETER METHOD

2.1 Formulating a Homogeneous System

The two equations in (1) are actually not linear. In the so-called direct parameter method, the first step is to make them linear. To this end, let us have a careful look at the two sets of camera parameters, which are re-written here for easy references:

- Extrinsic parameters
  - $R=[r_{ij}]_{3x3}$: a 3x3 rotation matrix, which can be represented by three rotation angles
  - $T= (T_x,T_y,T_z)^T$: a 3D translation vector
- Intrinsic parameters
  - f_x, f_y: effective focal length in pixel, or we can use $\alpha = f_x/f_y$ and f_x to represent them.
  - (o_x, o_y): image center, which we assume known, therefore (x',y') is known.
  - k_1: radial distortion coefficient, which we neglect within the scope of the basic algorithm we will discuss.

We assume that the coordinates of the image center are known. The next question is, whether we can assume that the image center is the origin of the digital frame coordinate system. The answer is no, so we cannot eliminate (o_x,o_y) from the linearized equation. Therefore, we really have to know the image center. We will come back to them later if they are not known, but right now let us assume we know the center of the image. With this we notice that the two equations have the same denominator. By dividing x' by y' and then redistributing the terms we obtain

$$f_y(r_{21}X_w + r_{22}Y_w + r_{23}Z_w + T_y)/y' = f_x(r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x)/x'$$

$$x' f_y(r_{21}X_w + r_{22}Y_w + r_{23}Z_w + T_y) = y' f_x(r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x) \quad (3)$$

In Equation (3), we have (X_w,Y_w,Z_w) and its corresponding image point (x',y'), both of which are known or can be measured; the unknown parameters are T_x, T_y, f_x, f_y,, r_{1j} and r_{2j}. j=1,2,3.

Now we are one step away from constructing a system of linear equations. To accomplish this, we will use the aspect ratio: $\alpha = f_x/f_y$ y in equation (3), and use N pairs of control points: {(X_i, Y_i, Z_i) <-> (xi, yi), i=1,…,N }. Note that for ease of representation, we have already dropped the prime in both x_i and y_i and the subscript "w" in (X_i,Y_i, Z_i)

$$x_i X_i r_{21} + x_i Y_i r_{22} + x_i Z_i r_{23} + x_i T_y$$
$$- y_i X_i (\alpha r_{11}) - y_i Y_i (\alpha r_{12}) - y_i Z_i (\alpha r_{13}) - y_i (\alpha T_x) = 0$$

For each pair of points we have a linear equation of 8 unknowns $\mathbf{v} = (v_1,\dots, v_8)^T$

$$x_i X_i v_1 + x_i Y_i v_2 + x_i Z_i v_3 + x_i v_4 - y_i X_i v_5 - y_i Y_i v_6 - y_i Z_i v_7 - y_i v_8 = 0 \quad (4)$$

where

$$(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8) = (r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x) \quad (5)$$

With this, we have a homogeneous system of N linear equations

$$\mathbf{Av = 0} \quad (6)$$

where

$$\mathbf{A} = \begin{bmatrix} x_1X_1 & x_1Y_1 & x_1Z_1 & x_1 & -y_1X_1 & -y_1Y_1 & -y_1Z_1 & -y_1 \\ x_2X_2 & x_2Y_2 & x_2Z_2 & x_2 & -y_2X_2 & -y_2Y_2 & -y_2Z_2 & -y_2 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ x_NX_N & x_NY_N & x_NZ_N & x_N & -y_NX_N & -y_NY_N & -y_NZ_N & -y_N \end{bmatrix}$$

given N corresponding pairs {(Xi, Yi, Zi) <-> (xi, yi) }, i=1,2,…N. It is called the homogeneous system since the right-hand side of the equation is an all zero vector. In such a system, with 8 unknowns $\mathbf{v} = (v_1, \dots, v_8)^T$, we have only **7** independent parameters (see below for it's properties). Here are a few observations for this homogeneous system of linear equations.

- The system has a nontrivial solution (up to a scale)
  - If N >= 7 and N points are not coplanar, then rank (**A**) = 7
  - The solution can be determined from the SVD of A
- Obviously, the system also has a trivial solution **v**= **0**. However, it is not the correct solution of the problem since R cannot be all zeros
- If $\mathbf{v_0}$ is a solution of the homogeneous system, so is c $\mathbf{v_0}$ (where c is any real value including 0)

We will now give an overview of SVD and its applications, since it is the key to this as well as other problems in camera calibration and stereo vision.

2.2 SVD: Definition and Applications

2.2.1 Definition: Singular Value Decomposition (SVD)

Any mxn matrix A can be written as the product of three matrices

$$\mathbf{A} = \mathbf{UDV}^T \tag{7}$$

$$\begin{bmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ & & & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & & u_{1m} \\ u_{21} & u_{22} & & u_{2m} \\ & & & \\ u_{m1} & u_{m2} & & u_{mm} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & & 0 \\ 0 & \sigma_2 & & \\ 0 & & & \sigma_n \\ 0 & & & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & & v_{n1} \\ v_{12} & v_{22} & & v_{n2} \\ & & & \\ v_{1n} & v_{2n} & & v_{nn} \end{bmatrix}$$

where the following hold:

(1). D is a mxn diagonal matrix [dij] with dij =0 if i≠j; dii = σi (i=1,2,…,n). *Singular values* σi (i=1,…, n) are fully determined by A, and they are sorted $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

(2). Neither U nor V is unique, but the columns of each matrix are mutually orthogonal vectors. Note that we use V^T instead of V, simply because we want to show that the columns of both U and V are eigenvectors (below).

2.2.2 SVD: properties

The following properties will be very useful in a number of applications.

1. Singularity and Condition Number

We say an nxn matrix A is *nonsingular* iff all singular values are nonzero. For a nonsingular matrix, its *condition number*, i.e., degree of singularity of A is defined as

$$C = \sigma_1 / \sigma_n \tag{8}$$

We thus have the following important observation: A is *ill-conditioned* if 1/C is comparable to the arithmetic precision of your machine; in other words, we say A is almost singular. The condition number can be used to examine how robust the linear system is: the smaller the condition number is, the more robust is the linear system.

2. Rank of a square matrix A

Rank (A) = number of nonzero singular values of A

3. Inverse of a square Matrix

If A is nonsingular then the inverse of A is

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T \tag{9}$$

where the diagonal terms of diagonal matrix D^{-1} is simply ${d_{ij}}^{-1} = 1/d_{ij}$. We can easily prove that $A^{-1} A = I$.

If some of the singular values of A are zeros, i.e., A is singular but we still want to find its inverse, we can approximate it with the *pseudo-inverse* of A

$$\mathbf{A}^{+} = \mathbf{V}\mathbf{D}_0^{-1}\mathbf{U}^T \tag{10}$$

where D_0^{-1} is a diagonal matrix with its diagonal terms equal to $1/d_{ij}$ for all nonzero singular values and zeros otherwise.

4. Eigenvalues and Eigenvectors (questions)

The eigenvalues of both A^TA and AA^T are σ_i^2 ($\sigma_i > 0$), while the columns of U are the eigenvectors of AA^T (mxm)

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_i = \sigma_i^2\mathbf{u}_i \tag{11}$$

and the columns of V are the eigenvectors of A^TA (nxn)

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \sigma_i^2\mathbf{v}_i \tag{12}$$

2.2.3 SVD Application 1: Solving the Least Square Problem

The Least Square Method is used to solve a system of m equations for n unknowns **x** ($m \geq n$):

$$\mathbf{Ax} = \mathbf{b} \tag{13}$$

where **A** is an mxn matrix of the coefficients, and **b** ($\neq \mathbf{0}$) is the m-D vector of the data. The least square method is equivalent to multiplying the transpose of A to both sides and turns the original system into the following system:

$$\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b} \tag{14}$$

where the coefficient matrix becomes a square matrix (i.e., the numbers of equations and the unknowns are the same). Therefore the solution for x can be found by multiplying the pseudo-inverse of the coefficient matrix, as:

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^+ \mathbf{A}^T \mathbf{b} \tag{15}$$

The reason to use the pseudo-inverse is that A^TA might be singular. We know we can compute the pseudo-inverse of A^TA by using its SVD, and in practice, $(A^TA)^+$ is more likely to coincide with $(A^TA)^{-1}$ given $m > n$. But it is always a good idea to look at the condition number of A^TA to get a sense of how stable your linear system is. This result will be used later in this chapter.

2.2.4 SVD Application 2: Solving the Homogeneous System

A homogeneous system has m equations for n unknowns **x**($m \geq n-1$) in the form of

$$\mathbf{Ax} = \mathbf{0} \tag{15}$$

where rank (A) = n-1. In practice, this can be verified by looking at the SVD of A. A non-trivial solution (up to an arbitrary scale) by using SVD is simply proportional to the eigenvector corresponding to the only zero eigenvalue of A^TA (nxn matrix). The correctness can be proven using equation (12) when the singular value is zero. Note that all the other eigenvalues should be positive because rank (A)=n-1. In practice, we use the eigenvector (i.e. v_n) corresponding to the minimum eigenvalue of A^TA, i.e. σ_n^2. This result will be used in solving the 7 unknown equations for calibration (Eq. (6)).

2.2.5 SVD Application 3: Enforcing Orthogonality

We often have problems finding the numerical estimate of a matrix A whose entries are not independent. Errors introduced by noise alter the estimate to Â. Then we can enforce constraints by using SVD. Let us take orthogonal matrix A as an example. Our goal is to find the closest matrix to the estimate Â, which exactly satisfies the constraints for orthogonality. Upon obtaining the estimate Â, we perform the SVD of Â as

$$\hat{\mathbf{A}} = \mathbf{UDV}^T \tag{16}$$

We observe that D should be an identity matrix I (all the singular values are 1) if A is orthogonal. However, in reality, it might not be the case. What we can do is to replace D with I so that the final solution for A will be an orthogonal matrix:

$$\mathbf{A} = \mathbf{UIV}^T \tag{17}$$

This result will be used in the rotation matrix estimation.

2.3 Solving the Homogeneous System

With this preparation, we are now ready to solve our homogeneous system of N linear equations in (6), given N corresponding pairs $\{(X_i, Y_i,, Z_i) \leftrightarrow (x_i, y_i)\}$, i=1,2,…N, for the 8 unknowns $\mathbf{v} = (v_1,\ldots,v_8)^T$, 7 of them are independent parameters. We know that the system has a nontrivial solution (up to a scale) if N >= 7 and the N points are not coplanar, thus Rank (**A**) = 7. It can be determined from the SVD of A, as

$$\mathbf{A} = \mathbf{UDV}^T \tag{18}$$

where the rows of $\mathbf{V}^T$ are eigenvectors $\{\mathbf{e}_i\}$of $\mathbf{A^TA}$. Therefore, the solution is proportional to the 8^{th} row $\mathbf{e}_8$ corresponding to the only zero singular value σ_8=0, as

$$\bar{\mathbf{v}} = c\mathbf{e_8} \tag{19}$$

where c is a constant that can be of any value (negative, zero or positive) for just satisfying the homogeneous system, but we have to find out what it should be for estimating the real camera parameters. There is also another practical consideration, the errors in localizing image and world points may make the rank of A to be maximum (i.e., 8). In this case we simply select the eigenvector corresponding to the smallest eigenvalue.

2.4 Extracting Camera Parameters from the Solution

2.4.1 Scale Factor and Aspect Ratio

Due to the fact that the any multiple of v makes equation (6) true, we cannot determine the scale by just using equation (6). Therefore we turn to the physical meanings of the intrinsic and extrinsic parameters to solve this problem and eventually extract these camera parameters. Without loss of generality, let us write $\overline{\mathbf{v}}$ as $\mathbf{e_8}$, (i.e., by set c=1), we could instead define a scale factor γ and write the following equation, after equation (5):

$$\overline{\mathbf{v}} = \gamma\,(r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x) \tag{20}$$

We obtain a lot of useful constraints by employing the fact that that the rotation matrix R is an orthogonal matrix. Let us write matrix R as three row vectors:

$$\mathbf{R} = \left(r_{ij}\right)_{3\times 3} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1^T \\ \mathbf{R}_2^T \\ \mathbf{R}_3^T \end{bmatrix}$$

Then we have

$$\mathbf{R}_i^T \mathbf{R}_j = \begin{cases} 1 & \text{if } \mathrm{i = j} \\ 0 & \text{if } \mathrm{i \neq j} \end{cases} \tag{21}$$

which has two aspects: (1) the length (magnitude) of each vector R_i is 1; and (2) the dot product of two different vectors is zero. We will now use the first observation. From the second row of the rotation matrix, i.e., i=j=2, we have

$$r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$$

Then by comparing the first three components of the vectors in the left and right sides of equation (20) we obtain the scale up to a sign:

$$|\,\gamma\,| = \sqrt{\overline{v}_1^2 + \overline{v}_2^2 + \overline{v}_3^2} \tag{22}$$

We will still have to determine the sign. Similarly, from first row of R (i=j=1), we have

$$r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$$

Then by comparing the 5th and 7th components of the vectors on the left and right sides of equation (20) we obtain

$$\alpha\,|\,\gamma\,| = \sqrt{\bar{v}_5^2 + \bar{v}_6^2 + \bar{v}_7^2} \tag{23}$$

and the aspect ratio α can be directly calculated by dividing equation (23) by equation (22).

$$\alpha = \sqrt{\bar{v}_5^2 + \bar{v}_6^2 + \bar{v}_7^2} \,/\, \sqrt{\bar{v}_1^2 + \bar{v}_2^2 + \bar{v}_3^2} \tag{24}$$

2.4.2 Rotation R and Translation T

Equation (20) now yields the first two rows of R and the first two components of T, given that the values of both α and |γ| have been obtained.

$$\bar{\mathbf{v}} = s\,|\,\gamma\,|\,(r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x) \tag{25}$$

where s is a common sign (+ or -) to be determined. Hence, the first two rows of R and the first two components of T can be found up to a common sign s (+ or -), as

$$s\mathbf{R}_1^T, s\mathbf{R}_2^T, sT_x, sT_y \tag{25a}$$

The third row of the rotation matrix can easily be computed as the vector product of the first and second row:

$$\mathbf{R}_3^T = \mathbf{R}_1^T \times \mathbf{R}_2^T = s\mathbf{R}_1^T \times s\mathbf{R}_2^T \tag{25b}$$

Upon this point, there are still three remaining questions:

- How to find the sign s?
- Is R orthogonal?
- How to find T_z and f_x, f_y?

We will work on this in the following subsections.

2.4.3 Finding the sign s

Referring to Figure 1, we re-write related entities of the perspective camera model as

$$\begin{aligned} x &= -f_x \frac{X}{Z} = -f_x \frac{r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z} \\ y &= -f_y \frac{Y}{Z} = -f_y \frac{r_{21}X_w + r_{22}Y_w + r_{23}Z_w + T_y}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z} \end{aligned} \tag{26}$$

Here are a few facts that we can use for a pair of points (X_w, Y_w, Z_w) and (x,y). Let us focus on the first equation in (26) to figure out how to find the sign s:

- The effective focal length is always positive, i.e., $f_x > 0$
- the point (X_w, Y_w, Z_w) is in front of the camera, i.e., its Z coordinate in the camera coordinate system is always positive, i.e., Z > 0
- both x and (X_w, Y_w, Z_w) are known values

The solution is to check the sign of X, either positive or negative, which should be opposite to x. One way to do this is to assume the sign s is, for example, positive, then compute the value X as $r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x$. If it has the same sign as x, the assumption is correct, otherwise the sign is opposite. Upon the determination of the sign and the rotation matrix R, the first components of the translational vector T are fully determined.

2.4.4 Rotation R: Orthogonality

Note that when we used equation (6) we calculated the first 2 rows of R without using their mutual orthogonal constraint. This means that the eight unknowns in the equation system are not independent. So, unless everything is fully accurate, there is no guarantee they will be orthogonal. Using a rotation matrix estimate that does not meet the physical constraints would be problematic. So the question is, does the following hold:

$$\hat{\mathbf{R}}^T \hat{\mathbf{R}} = \mathbf{I}\,? \tag{26}$$

If not, how do we enforce the orthogonality? One of the solutions is to compute the SVD of estimate R

$$\hat{\mathbf{R}} = \mathbf{UDV}^\mathrm{T} \tag{27}$$

Then to replace the diagonal matrix D with the 3x3 identity matrix, so the final result of the rotation matrix is

$$\mathbf{R} = \mathbf{UIV}^\mathrm{T} \tag{28}$$

2.4.5 Find Tz, Fx and Fy

We still need to find T_z and f_x (or f_y). The solution is to use one of the equations in (1),

$$x = -f_x \frac{r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + T_z}$$

to construct a system of N linear equations with two unknowns

$$\underbrace{x}_{\mathbf{a_{i1}}} T_z + \underbrace{(r_{11}X_w + r_{12}Y_w + r_{13}Z_w + T_x)}_{\mathbf{a_{i2}}} f_x = -x\underbrace{(r_{31}X_w + r_{32}Y_w + r_{33}Z_w)}_{\mathbf{bi}}$$

$$\mathbf{A}\begin{pmatrix} T_z \\ f_x \end{pmatrix} = \mathbf{b} \qquad (28)$$

given N pairs of points (i=1, …N). The solution can be easily found by using the least square method:

$$\begin{pmatrix} T_z \\ f_x \end{pmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \qquad (29)$$

Once again, we can use the SVD method to find the inverse or pseudo-inverse, so that we obtain Tx and fx. Then we can also find $f_y = f_x / \alpha$.

2.5 Estimating the Image Center

At the beginning of the direct parameter method, we have assumed that the center of the image (o_x,o_y) is known, so that we obtain a linear equation with 8 unknowns. Now we need to come back to the starting point to see if we can estimate the image center before we perform the calibration procedure we discussed above. We will introduce a technique using the typical calibration pattern of two faces of a cube seen from a camera's view. This will also lead us to the definition of vanishing points, a very important concept in perspective geometry for both computer vision and computer graphics.

2.5.1 Vanishing points

Due to perspective, all parallel lines in 3D space appear to meet at a point on the image - the *vanishing point*, which is the common intersection of all the image lines. Figure 3 shows the vanishing point VP1 in the image of a set of parallel lines on the calibration cube in space. There is a very important property of this vanishing point:

The vector from the center of projection to the vanishing point is parallel to the set of parallel lines.

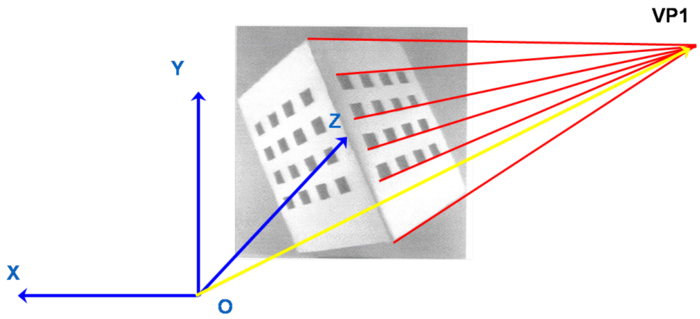

Figure 3. Vanishing point

This property will be used to find the image center without knowing much about the camera's intrinsic and extrinsic parameters.

2.5.2 Orthocenter Theorem

Given three mutually orthogonal sets of parallel lines in an image, a triangle T is formed on the physical image plane, defined by the three vanishing points of these three sets of parallel lines. Then the image center is the orthocenter of triangle T -- the common intersection of the three altitudes of the triangle.

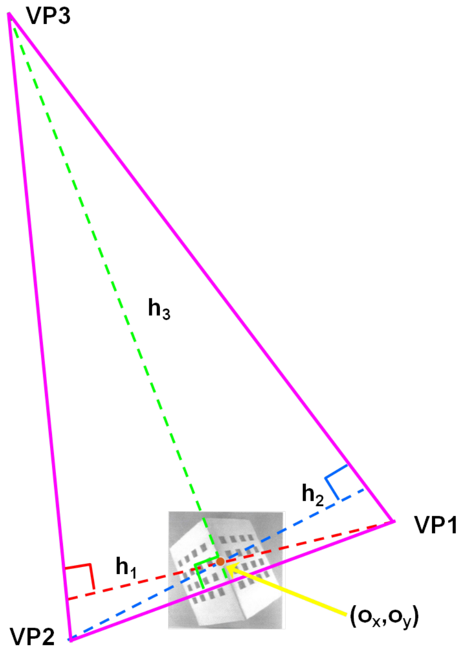

Figure 4. Orthocenter Theorem

Figure 4 shows the orthocenter theorem. First, three vanishing points VP1, VP2 and VP3 are found for the images of three sets of parallel lines, respectively. Then a triangle is formed with the three vanishing points as the vertices, and its three altitudes can be defined. An altitude is the perpendicular line segment from a vertex to its opposite base. Finally, the orthocenter is calculated as the intersection of these three altitudes, which is exactly the center of the image: (o_x,o_y).

In the following, we will give a proof to the orthocenter theorem. By doing this, we hope to obtain a better understanding of the perspective geometry, and meanwhile figure out what we need to know in an image in order to use the orthocenter theorem to find the image center.

Figure 5. Orthocenter proof: a geometric approach (TO DO)

[Proof] Let us define the center of projection of the camera in 3D space as O (see above figure on the right). L_1, L_2, and L_3 are three mutually orthogonal sets of parallel lines, and V_1, V_2, and V_3 are the three vanishing points, forming a triangle $V_1V_2V_3$. From question 1 at the end of this chapter you should have proved that the vector OV_i from the center of projection O (viewpoint or pinhole) to a vanishing point V_i is parallel to the mutually orthogonal set of parallel lines L_i in 3D space (i=1,2,3). This means that OV_i is perpendicular to OV_j as long as $i \neq j$, therefore OV_1 is perpendicular to V_2V_3, and OV_2 is perpendicular to V_1V_3, and OV_3 is perpendicular to V_1V_2. V_ih_i is the altitude from V_i (i=1,2,3), therefore V_1h_1 is perpendicular to V_2V_3, and V_2h_2 is perpendicular to V_1V_3, and V_3h_3 is perpendicular to V_1V_2. Let us define the center of the image as o, therefore line Oo is perpendicular to the image plane. Also note that point o is the projection of point O in the image plane, which lies on all the three altitudes V_ih_i, so o is the intersection point of V_1h_1, V_2h_2, and V_3h_3. #

We do not need to know anything about the camera parameters in order to find the three vanishing points. For example, in the above proof, we don't use the focal length information. But since we need to use the orthogonal relations of the altitude, we need to assume that we know the aspect ratio of the image. In Figure 6, if the assumed aspect ratio was wrong, we would have obtained a scaled image in the y direction. The vanishing points would still be correct, but the image center will not be at (o_x, o_y), since h_i is not the altitude any more.

Secondly, we need to have three vanishing points generated from three sets of parallel lines that are mutually orthogonal to each other. It is not a problem to have three sets of mutually orthogonal parallel-line sets.

However this is no guarantee that a set of parallel lines will generate a vanishing point. The camera view angles matter. For example, if the camera is perpendicular to the plane which includes the set of parallel lines, then the images of these parallel lines would still be in parallel and would not have an intersection. Therefore, their vanishing point will be at infinity. In this case we cannot form a triangle to estimate the image center.

Of course the orthocenter theorem only works on images without lens distortions or with lens distortions removed.

Finally, we have two interesting questions: (1) Can we solve both the aspect ratio and the image center? (2) Can we find the focal length of a camera by using the three vanishing points? These questions will be left for a homework project.

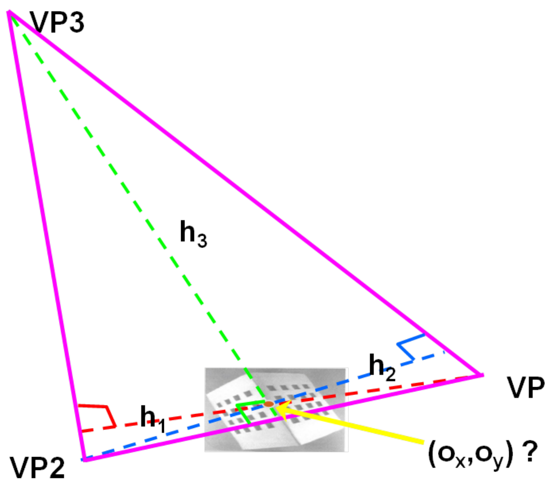

Figure 6. Orthocenter is not the image center if using an incorrect aspect ratio

2.6 Direct Parameter Calibration Summary

We would like to summarize the direction parameter method into an algorithm and then offer some discussions.

2.6.1 Direct Parameter Calibration Algorithm

We list the steps of the direct parameter method.

1. Estimate image center (and aspect ratio?)
2. Measure N 3D coordinates (X_i, Y_i, Z_i), i = 1,2, …, N

3. Locate their corresponding image points (x_i,y_i) - using edge / corner detection and/or Hough Transform
4. Build matrix **A** of a homogeneous system **Av = 0**
5. Compute SVD of **A** , solution **v**
6. Determine aspect ratio α and scale $|\gamma|$
7. Recover the first two rows of **R** and the first two components of **T** up to the same sign
8. Determine sign s of γ by checking the projection equation
9. Compute the 3^{rd} row of **R** by vector product, and enforce orthogonality constraint by SVD
10. Solve T_z and f_x using Least Square and SVD, then $f_y = f_x / \alpha$

Note that there are three important techniques in the direct calibration approach and three applications of SVD in the method. The three important techniques are: (1) the use of the divide and conquer technique -- particularly with a unique technique in finding the image center first in order to simplify the problem; (2) the creation of linear systems of fewer unknowns; and (3) the decomposition of intrinsic and extrinsic parameters using physical constraints of the camera (orthogonality of R, sign of depth, etc.). SVD is used three times in (1) solving the homogeneous system; (2) enforcing orthogonality constraints; and (3) solving the least square system.

2.6.2 Remaining Issues and Possible Solutions

In using the direct parameter method, we need to ask the following questions:

(1). Can we select an arbitrary image center for solving other parameters?

(2). How to find the image center (o_x,o_y)? Do we need to know the aspect ratio in order to use the orthocenter theorem?

(3). Can we include the radial distortion in calibration?

Note that the original assumptions of the direct parameter method are: (1) The camera is without lens distortions. (2) The aspect ratio is known when estimating image center. (3) The image center is known when estimating other parameters including the aspect ratio. We will not discuss the radial distortion here. Even without radial distortion, we have one problem: using the direct parameter method we cannot find the image center without knowing the aspect ratio, and the estimation of the aspect ratio needs the knowledge of the image center.

One solution to this problem is to first assume that the aspect ratio is approximately 1:1 (or 4:3 by using the manufacturer's information) to find image center (o_x, o_y) using the orthocenter theorem. Then, we can use the estimated center to find other parameters including the aspect ratio. After that, we can refine the image center using the newly estimated aspect ratio. We could perform a few iterations until the changes of the estimations of both the aspect ratio and the image centers are not significant.

Another solution is to solve all the parameters at once, which leads to the projection matrix method that we are going to discuss below. Here we give a quick comparison:

The core of the direct parameter method: One point pair provides 1 linear equation of 8 unknowns (of which 7 are independent).

Projection matrix method: One point pair will provide 2 linear equations of 12 = (4+4+4) unknowns (of which 11 are independent)

3. PROJECTION MATRIX APPROACH

3.1 Estimation of Projection Matrix

We start with the projection matrix equation that we discussed in the chapter about camera models, with the relation of the world to frame transformation. For convenience, we drop subscripts "im" and "w" , so for N pairs of points: (xi,yi) ↔ (Xi,Yi,Zi), i=1,…,N, we have

$$\begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} = \mathbf{M} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \tag{30}$$

where $(x_i,y_i) = (u_i/w_i,\ v_i/w_i)$, and $M = [m_{ij}]$ is the 3x4 projection matrix, which includes both the four intrinsic parameters (f_x,f_y,o_x,o_y) and the six extrinsic parameters (α,β,γ, Tx, Ty, Tz), altogether 10 independent parameters. However, in order to derive linear equations, we use the nine elements of the rotation matrix $[r_{ij}]$3x3 instead of the three rotation angles α,β,γ:

$$\mathbf{M} = \begin{bmatrix} -f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ -f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + +o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix} \tag{31}$$

Hence it includes 16 =4+9+3 parameters. Inserting (31) into (30), we will have the expansion forms of two equations:

$$x_i = \frac{u_i}{w_i} = \frac{m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

$$y_i = \frac{u_i}{w_i} = \frac{m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}} \qquad (32)$$

Then, for N pairs of world-frame points, we have the following homogeneous linear equation system of the unknown vector **m**:

$$\mathbf{Am} = \mathbf{0} \qquad (33)$$

where

$$\mathbf{A} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1Y_1 & -y_1Y_1 & -y_1 \\ . & . & . & . & . & . & & & & & & \end{bmatrix}$$

$$\mathbf{m} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{21} & m_{22} & m_{23} & m_{24} & m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}^T$$

In (33), we have 2N equations and 11 independent variables. Therefore, if N >=6 , we will have 12 equations, and using SVD we can find the solution for **m** up to an unknown scale.

3.2 Computing camera parameters

We will take a similar approach here as we did in the direct parameter method. After we have an estimation of **m** up to an unknown scale, we can construct an estimated projection matrix for the estimation of the projection

$$\hat{\mathbf{M}} = \begin{bmatrix} \mathbf{q}_1 \; q_{41} \\ \mathbf{q}_2 \; q_{42} \\ \mathbf{q}_3 \; q_{43} \end{bmatrix} \qquad (34)$$

which is a scaled version of the real M, as

$$\hat{\mathbf{M}} = s\mathbf{M} \qquad (35)$$

where s is the scale between the two; $\mathbf{q_1}$, $\mathbf{q_2}$, and $\mathbf{q_3}$ are three row vectors. For easy comparison, we rewrite the M matrix here:

$$\mathbf{M} = \begin{bmatrix} -f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ -f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + +o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}$$

With this we summarize the steps of obtaining the intrinsic and the extrinsic parameters of the camera from the matrix $\hat{\mathbf{M}}$:

[Projection Matrix Calibration Algorithm]

1. Find the absolute value of the scale |s| by using the known values of $\mathbf{q_3}$ and the first three components of the last row of M, which is the unit row vector R_3^T
2. Determine T_z and the sign of s from the knowledge that $m_{34} = T_z > 0$ and the value of q_{43} is known.
3. Obtain R_3^T following steps 1 and 2.
4. Find (o_x, o_y) by the dot products of the two pairs of the row vectors, corresponding to $\mathbf{q_1} . \mathbf{q_3}$, and $\mathbf{q_2} . \mathbf{q_3}$, using the orthogonal constraints of **R,** and the known value of the scale s.
5. Determine f_x and f_y from the corresponding components in **M** to $\mathbf{q_1}$ and $\mathbf{q_2}$, using the known values of (o_x, o_y), and the knowledge of unit row vector R_1^T and R_2^T
6. All the rest: R_1^T, R_2^T, T_x, T_y, can be found easily
7. Enforce orthogonality to R using SVD.

4. COMPARISONS AND DISCUSSIONS

By comparing the direct parameter method and the projection matrix method, we can make the following observations:

There are two common properties:

(1) Both methods try to solve a linear system first, and then perform parameter decomposition second.
(2) Results should be exactly the same if everything is accurate.

There are two major differences:

(1) Number of variables in the homogeneous systems: In the projection matrix method, all the parameters are solved at once, and we have 2N Equations of 12 variables. The the direct parameter method uses a divide and conquer approach in three steps: image center using the orthocenter theorem, N equations of 8 variables and N equations of

2 variables. With fewer parameters in each step, the systems may be more robust than the projection matrix method.

(2) Assumptions. The projection matrix method is simpler and more general; sometimes a projection matrix is sufficient so there is no need for parameter decomposition. The direct parameter method assumes a known image center in the first two steps, and a known aspect ratio in estimating image center.

4.1 Guidelines for Calibration

Here we summarize some guidelines for a good practice of camera calibration:

- Pick up a well-known technique (or a few)
- Design and construct calibration patterns (with known 3D)
- Make sure you know what parameters you want to find for your camera
- Run algorithms on ideal simulated data
    - You can either use the data of the real calibration pattern or computer generated data
    - Define a virtual camera with known intrinsic and extrinsic parameters
    - Generate 2D points from the 3D data using the virtual camera
    - Run algorithms on the 2D-3D data set
- Add noise to the simulated data to test the robustness
- Run algorithms on the real data (images of calibration target)
- If successful, you are all set!
- Otherwise:
    - Check your distribution of control points
    - Check the accuracy in 3D and 2D localizations
    - Check the robustness of your algorithms again
    - Develop your own algorithms

5. QUESTIONS AND PROJECTS

5.1 Questions

[Question 1]. Prove the Orthocenter Theorem by geometric arguments: Let T be the triangle on the image plane defined by the three vanishing points of three mutually orthogonal sets of parallel lines in space. Then the image center is the orthocenter of the triangle T (i.e., the common intersection of the three altitudes.

(1) Basic proof: use the result of Question 2 of the camera model chapter, assuming the aspect ratio of the camera is 1.

(2) If you do not know the focal length of the camera, can you still find the image center (together with the focal length) using the Orthocenter Theorem? Show why or why not.

(3) If you do not know the aspect ratio and the focal length of the camera, can you still find the image center using the Orthocenter Theorem? Show why or why not.

[Question 2]. Derive equations for the parameter estimation in the projection matrix algorithm listed in Section 3.2.

5.2 Projects

[Calibration Programming Exercises]: Implement the direct parameter calibration method in order to (1) learn how to use SVD to solve systems of linear equations; (2) understand the physical constraints of the camera parameters; and (3) understand important issues related to calibration such as calibration pattern design, point localization accuracy, and robustness of the algorithms. Since calibrating a real camera involves a lot of work in calibration pattern design, image processing, and error controls as well as solving the equations, we will mainly use simulated data to understand the algorithms. As a by-product we will also learn how to generate 2D images from 3D models using a "virtual" pinhole camera.

a. Calibration pattern "design". Generate data of a "virtual" 3D cube similar to the one shown in Figure 1. For example, you can hypothesize a $1\times1\times1$ m^3 cube and pick up coordinates of 3D points on one corner of each black square in your world coordinate system. Make sure that your data is sufficient for the following calibration procedures. In order to show the correctness of your data, draw your cube (with the control points marked) using Matlab (or whatever tools you are selecting). I have provided a piece of starting code in Matlab for you to use.
b. "Virtual" camera and images. Design a "virtual" camera with known intrinsic parameters including focal length f, image center (o_x, o_y) and pixel size (s_x, s_y). As an example, you can assume that the focal length is f = 16 mm, the image frame size is 512*512 (pixels) with (o_x,o_y) = (256, 256), and the size of the image sensor inside your camera is 8.8 mm *6.6 mm (so the pixel size is (s_x,s_y) = (8.8/512, 6.6/512)). Capture an image of your "virtual" calibration cube with

your virtual camera in a given pose (R and T). For example, you can take the picture of the cube 4 meters away and with a tilt angle of 30 degree. Use three rotation angles alpha, beta, gamma to generate the rotation matrix R. You may need to try different poses in order to have a suitable image of your calibration target.

c. Direction calibration method: Estimate the intrinsic (f_x, f_y, aspect ratio α, image center (o_x,o_y)) and extrinsic (R, T and further alpha, beta, gamma) parameters. Use SVD to solve the homogeneous linear system and the least square problem, and to enforce the orthogonality constraint on the estimate of R.

   i. Use the accurately simulated data (both 3D world coordinates and 2D image coordinates) to the algorithms, and compare the results with the "ground truth" data (which are given in step (a) and step (b)). Remember you are practicing a camera calibration, so you should pretend you know nothing about the camera parameters (i.e. you cannot use the ground truth data in your calibration process). However, in the direct calibration method, you could use the knowledge of the image center (in the homogeneous system to find extrinsic parameters) and the aspect ratio (in the Orthocenter theorem method to find the image center).

   ii. Study whether the unknown aspect ratio matters in estimating the image center, and how the initial estimation of image center affects the estimation of the remaining parameters. Give a solution to solve the problems if any.

   iii. Accuracy Issues. Add in some random noises to the simulated data and run the calibration algorithms again. See how the "design tolerance" of the calibration target and the localization errors of 2D image points affect the calibration accuracy. For example, you can add 0.1 mm random error to 3D points and 0.5 pixel random error to 2D points. Also analyze how sensitive the Orthocenter method is to the extrinsic parameters in imaging the three sets of the orthogonal parallel lines.

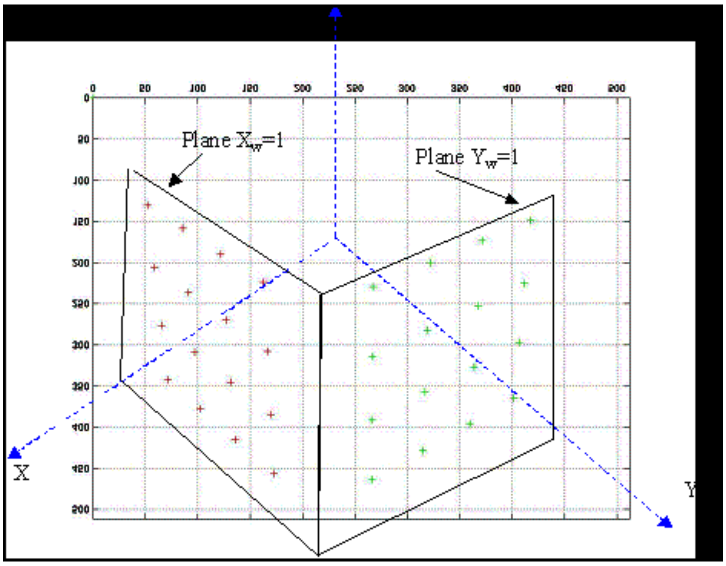

Figure 7. A 2D image of the “3D cube” with control 16+16 points

REFERENCES

Digital Image Processing
Computer Graphics
Photogrammetry