Jason Corso
Lecture Notes for Camera Calibration
Computer Vision, Fall 2002, JHU.

1 Motivation for the Problem

1.1 Computer Vision

Compute properties of the 3d world from one or more digital images.

It is likely that a good amount of information can be obtained through the geometric properties of the projections (images): shape, position, etc.

1.2 Perspective Projection

Recall that under full perspective projection, a large amount of information is lost during the forward projection (imaging) process. Think about each optical ray shooting out from the optical center of the camera; for each ray, the result of only the first collision is known.

The linear matrix equation of perspective projection is:

$$\underbrace{\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}}_{(3\times 1)} = \underbrace{M_{int}}_{(3\times 3)} \underbrace{M_{ext}}_{(3\times 4)} \underbrace{\begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}}_{(4\times 1)}, \tag{1}$$

where x_i, y_i, z_i represent the projective coordinates of a point p in the image space and $X_w, Y_w, Z_w, 1$ represent the projective coordinates of a point P in world coordinates. The actual pixel location of the projected point p denoted $(\mathbf{x}, \mathbf{y})$ is $(\frac{x_i}{z_i}, \frac{y_i}{z_i})$. In (1), M_{int} and M_{ext} are the projection matrices containing the parameters of the projection for the current camera system. M_{int} is termed the *intrinsic matrix* as it contains those (linear) parameters intrinsic to the camera itself: focal length (f), pixel scale (s_x, s_y), optical center (o_x, o_y), and pixel skew (γ). The intrinsic matrix is

$$M_{int} = \begin{pmatrix} fs_x & -fs_x \cot\gamma & o_x \\ 0 & \frac{fs_y}{\sin\gamma} & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

If we assume that our pixels are square, i.e. the skew is $\frac{\pi}{2}$, the matrix simplifies to

$$M_{int} \approx \begin{pmatrix} fs_x & 0 & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

Notice that we cannot distinguish between a change in focal length and a change in pixel scale. In this discussion, we are neglecting any non-linear parameters that may affect pixel distortion (radial, tangential).

M_{ext} is termed the *extrinsic matrix* as it contains those terms external to the camera itself that relate the camera coordinate frame to the world coordinate frame. There are six parameters which define a transformation between coordinates frames, three for rotation encoded into a 3x3 rotation matrix R and three for position, T. The extrinsic matrix is

$$M_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \tag{4}$$

Forward projection may also written in a more condensed form:

$$P_c = RP_w + T \tag{5}$$

$$p_i = M_{int} P_c \tag{6}$$

1.3 Problem Statement

Given this model of forward projection, we compute the camera parameters enabling us compute some metric properties of the 3d world from images: camera calibration.

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

2 Types of Calibration

2.1 Photogrammetric Calibration

Calibration is performed through imaging a calibration pattern whose geometry in 3d is known with high precision. The accuracy of the calibration is proportional to the accuracy of the feature detection employed while locating the calibration pattern in the images.

Pros: Calibration can be performed very efficiently.

Cons: An expensive set-up apparatus is required: 2 or 3 (perfectly) orthogonal planes.

For a detailed explanation of these methods, see Trucco and Verri [3].

2.1.1 Direct Parameter Calibration

With knowledge of the forward projection process, set-up a system of equations and solve for the calibration parameters.

2.1.2 Projection Matrix Estimation

Estimate the projection matrix linking the world and image coordinates.

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \underbrace{M}_{(3\times 4)} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}, \tag{7}$$

We can write the equations per-pixel-coordinate:

$$\mathbf{x} = \frac{x_i}{z_i} = \frac{m_{11}X_w + m_{12}Y_w + m_{13}Z_w + m_{14}}{m_{31}X_w + m_{32}Y_w + m_{33}Z_w + m_{34}} \tag{8}$$

$$\mathbf{y} = \frac{y_i}{z_i} = \frac{m_{21}X_w + m_{22}Y_w + m_{23}Z_w + m_{24}}{m_{31}X_w + m_{32}Y_w + m_{33}Z_w + m_{34}} \tag{9}$$

M is defined up to an arbitrary scale factor and has 11 independent entries. Thus, we need ≥ 6 points. M can be estimated through least squares techniques. An example of such a technique is in Trucco and Verri on page 133 [3].

2.2 Self-Calibration

Makes use of some important properties from Projective Geometry to calculate the intrinsic parameters solely from point correspondences between images of a static scene assuming the intrinsic parameters do not change for the set of images. The approach is flexible but not well-established yet; since there are many parameters to estimate, it is often difficult to accurately compute them. A good reference is [2].

2.3 Multiplane Approach

Refer to [4]. It is a hybrid method pulling theory from both photogrammetric and self-calibration approaches. It uses a calibration pattern that is different than those used in Photogrammetric approachs in that it is planar and simple to produce: you print out a pattern and attach is to a planar surface, a text book cover.

Multiple images of the pattern are taken, and a technique from projective geometry is used to compute the intrinsic and extrinsic parameters from the images. The technique is based on a well-known transformation called a homography which maps points on a plane to points on another plane with a linear transformation; i.e. it acts as a change of basis for the plane.

This technique has no expensive set-up, is much more practical and fairly easy to implement or use an existing implementation. It is more flexible than standard photogrammetric methods and more robust than self-calibration.

3 The Multiplane Approach In Detail

3.1 Set-Up Notation

Let a 2D point $m = [u \quad v]^T$ and a 3D point $M = [x \quad y \quad z]^T$. Let s be an arbitrary scale factor and $A = M_{int}$.

Let $\tilde{x}$ be an augmented vector in homogeneous coordinates: $\tilde{m} = [u \quad v \quad 1]^T$ and $\tilde{M} = [x \quad y \quad z \quad 1]^T$.

Projection is written as

$$s\tilde{m} = A[R \quad T]\tilde{M} \tag{10}$$

3.2 Planar Homography

The foundation of this approach rests in the First Fundamental Theorem of Projective Geometry: There exists a unique homography that performs a change of basis between two projective spaces of the same dimension. A proof can be found in [1].

What does this mean for us? Given a plane in the 3d world, there is a unique mapping (transformation matrix) that maps this plane to the imaged plane. The mapping is up to scale.

Let us derive it. Let r_i be the i^{th} column of the matrix R. We assume the model plane is at the coordinate $Z = 0$. s is an arbitrary scale factor.

$$\begin{aligned}
s[u \quad v \quad 1]^T &= A[r_1 \quad r_2 \quad r_3 \quad t][X \quad Y \quad Z \quad 1]^T \\
s[u \quad v \quad 1]^T &= A[r_1 \quad r_2 \quad r_3 \quad t][X \quad Y \quad 0 \quad 1]^T \\
s[u \quad v \quad 1]^T &= A[r_1 \quad r_2 \quad t][X \quad Y \quad 1]^T \\
s[u \quad v \quad 1]^T &= H[X \quad Y \quad 1]^T
\end{aligned}$$

Here $H = A[r_1 \quad r_2 \quad t]$. Thus, a point on the model plane is mapped to the plane in the image by

$$s\tilde{m} = H\tilde{M} \tag{11}$$

Homographies can be estimated from at least four points (they have 8 degrees of freedom) in a fashion similar to estimating the projection matrix in the second algorithm from photogrammetric calibration. See Appendix A of [4].

3.3 Constraints On The Intrinsic Parameters

From (11), we have

$$[h_1 \quad h_2 \quad h_3] = sA[r_1 \quad r_2 \quad t]. \tag{12}$$

Because R is a orthogonal matrix, we can use some of its properties to derive some basic constraints on the intrinsic parameters.

$$r_i^T r_j = 0; \tag{13}$$

$$r_i^T r_i = r_j^T r_j \tag{14}$$

Writing the (12) in terms of its columns, we get

$$\begin{aligned} h_1 &= sAr_1 \\ h_2 &= sAr_2 \\ h_3 &= sAt \end{aligned} \tag{15}$$

Then, we derive the two basic constraints.

$$\begin{aligned} h_1 &= sAr_1 \\ \frac{1}{s}A^{-1}h_1 &= r_1 \\ \frac{1}{s}A^{-1}h_2 &= r_2 \\ \\ r_1^T r_2 &= 0 \\ h_1^T A^{-T}A^{-1}h_2 &= 0 \end{aligned} \tag{16}$$

$$\begin{aligned} r_1^T r_1 &= r_2^T r_2 \\ h_1^T A^{-T}A^{-1}h_1 &= h_2^T A^{-T}A^{-1}h_2 \end{aligned} \tag{17}$$

There are two constraints because there are 6 degrees of freedom for the extrinsic parameters; thus, for each known homography we can only obtain two constraints on the 5 intrinsic parameters. We need ≥ 3 homography to fully determine the intrinsics unless we set the pixel skew to 0 after which we would need ≥ 2 homographies to determine the intrinsics.

3.4 Solving the Camera Calibration Problem

So, from this, we can compute a closed-form solution to the problem.

Let $B = A^{-T}A^{-1} = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2\beta} & \frac{v_0\gamma-u_0\beta}{\alpha^2\beta} \\ -\frac{\gamma}{\alpha^2\beta} & \frac{\gamma^2}{\alpha^2\beta^2}+\frac{1}{\beta^2} & -\frac{\gamma(v_0\gamma-u_0\beta))}{\alpha^2\beta^2}-\frac{v_0}{\beta^2} \\ \frac{v_0\gamma-u_0\beta}{\alpha^2\beta} & -\frac{\gamma(v_0\gamma-u_0\beta))}{\alpha^2\beta^2}-\frac{v_0}{\beta^2} & \frac{(v_0\gamma-u_0\beta)^2}{\alpha^2\beta^2}+\frac{v_0^2}{\beta^2}+1 \end{bmatrix}$ where $A = M_{int} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$

Since, B is symmetric, we can define a vector based on its 6 parameters $b = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T$. Then, we can use our two constraints on the intrinsics to build a system of homogeneous equations. $h_i^T B h_j = v_{ij}^T$

$$\begin{bmatrix} v_{ij}^T \\ (v_{11}-v_{22})^T \end{bmatrix} b = 0; \tag{18}$$

Thus, with n images, we can stack n such equations into $Vb = 0$ and solve for b. Given b, we can solve for the intrinsics by the equations in Appendix B. The extrinsics are also readily computed.

Now, we have an algebraic solution. We need to refine our solution. We can use maximum likelihood inference to do so.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} ||m_{ij} - \hat{m}(A, R_k, T_k, M_j)||^2, \tag{19}$$

where $\hat{m}(A, R_k, T_k, M_j)$ is the projection of point M_j in image k.

References

[1] O. Faugeras and Q. Luong. *The Geometry Of Multiple Images*. The MIT Press, 2001.

[2] Q. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.

[3] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

[4] Z. Zhang. A flexible new technique for camera calibration. Technical report, Microsoft Research, 1998.