Camera calibration

- Necessary to recover 3D metric from image(s).
  - 3D reconstruction,
  - Object/camera localization, and
  - *etc.*
- Computes 3D (real world)–2D (camera image) relationship.

References

- Chapter 4 Zissermann - Estimation of 2D Projective Transformations
- Chapter 7 Zissermann - Computation of the Camera Matrix P

Slightly different formulas in these slides w.r.t. book; for more informations, please refer to Salvi et al 2002, Zhang 2005, Remondino and Fraser 2006

A point in camera geometry

A point is expressed with several coordinate system.

3D points in world coordinate

A point $\mathbf{X}_w = (X_w, Y_w, Z_w)^\mathsf{T}$ in a world coordinate.

3D points in camera coordinate

A point $\mathbf{X}_c = (X_c, Y_c, Z_c)^\mathsf{T}$ in a camera coordinate.

2D points in image coordinate

A point $\mathbf{x} = (x, y)^\mathsf{T}$ in an image plane.

Projection matrix

A 3×4 projection matrix $\mathbf{P}$ denotes relationship between $\mathbf{X}_w$ and $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{P}\mathbf{X}_w, \tag{1}$$

$$\rightarrow s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \tag{2}$$

Intrinsic and extrinsic parameters

A projection matrix can be decomposed into two components, intrinsic and extrinsic parameters, as

$$\mathbf{x} = \mathbf{P}\mathbf{X}_w = \mathbf{A}\left[\mathbf{R}|\mathbf{t}\right]\mathbf{X}_w, \tag{3}$$

$$\rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \tag{4}$$

where

- Intrinsic: 3×3 calibration matrix $\mathbf{A}$.
- Extrinsic: 3×3 Rotation matrix $\mathbf{R}$ and 3×1 translation vector $\mathbf{t}$.

Extrinsic parameters

Denotes transformation between $\mathbf{X}_w$ and $\mathbf{X}_c$ as

$$\mathbf{X}_c = [\mathbf{R}|\mathbf{t}]\,\mathbf{X}_w, \tag{5}$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \tag{6}$$

Intrinsic parameters

Project a 3D point $\mathbf{X}_c$ to image plane as

$$\mathbf{x} = \mathbf{A}\left[\mathbf{R}|\mathbf{t}\right]\mathbf{X}_w = \mathbf{A}\mathbf{X}_c, \tag{7}$$

$$\rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}, \tag{8}$$

where

- α_x and α_y are focal lengths in pixel unit.
- x_0 and y_0 are image center in pixel unit.
- s is skew parameter.

4 steps projecting a 3D world point to a 2D image point

A 3×4 projection matrix $\mathbf{P}$ denotes relationship between ${}^{W}\mathbf{X}_w$ and ${}^{I}\mathbf{x}$ as

$$
{}^{I}\mathbf{x} = \mathbf{P}\,{}^{W}\mathbf{X}_w, \tag{9}
$$

$$
\rightarrow s \begin{bmatrix} {}^{I}x_d \\ {}^{I}y_d \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} {}^{W}X_w \\ {}^{W}Y_w \\ {}^{W}Z_w \\ 1 \end{bmatrix}. \tag{10}
$$

1/4: A 3D world point to a 3D camera point

Change the world coordinate system to the camera one.

- From a 3D point ${}^{W}\mathbf{X}_w$ in metric system w.r.t. the world coordinate
- To a 3D point ${}^{C}\mathbf{X}_w$ in metric system w.r.t. the camera coordinate

$$
\begin{aligned}
{}^{C}\mathbf{X}_w &= [{}^{C}\mathbf{R}_w | {}^{C}\mathbf{T}_w] \, {}^{W}\mathbf{X}_w, && (11) \\
\rightarrow s \begin{bmatrix} {}^{C}X_w \\ {}^{C}Y_w \\ {}^{C}Z_w \\ 1 \end{bmatrix} &= \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_{14} \\ R_{21} & R_{22} & R_{23} & t_{24} \\ R_{31} & R_{32} & R_{33} & t_{34} \end{bmatrix} \begin{bmatrix} {}^{W}X_w \\ {}^{W}Y_w \\ {}^{W}Z_w \\ 1 \end{bmatrix}. && (12)
\end{aligned}
$$

2/4: A 3D camera point to a 2D camera point

Change the 3D camera coordinate system to the 2D camera one.

- From a 3D point ${}^{C}\mathbf{X}_{w}$ in metric system w.r.t. the camera coordinate
- To a 2D point ${}^{C}\mathbf{X}_{u}$ in metric system w.r.t. the camera coordinate

$$
{}^{C}\mathbf{X}_{u} = [{}^{C}\mathbf{R}_{w} | {}^{C}\mathbf{T}_{w}]\, {}^{W}\mathbf{X}_{w}, \tag{13}
$$

$$
\rightarrow s \begin{bmatrix} {}^{C}X_{u} \\ {}^{C}Y_{u} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^{W}X_{w} \\ {}^{W}Y_{w} \\ {}^{W}Z_{w} \\ 1 \end{bmatrix}, \tag{14}
$$

$$
{}^{C}X_{u} = \frac{f}{{}^{W}Z_{w}}\, {}^{W}X_{w} \qquad {}^{C}Y_{u} = \frac{f}{{}^{W}Z_{w}}\, {}^{W}Y_{w},
$$

where f denotes focal length in metric system.

3/4: Lens distortion

Practical lens distort the previous 3D→2D projection.

$$ {}^{C}X_u = {}^{C}X_d + \delta_x \qquad {}^{C}Y_u = {}^{C}Y_d + \delta_y, \tag{15} $$

where δ_x and δ_y denote distortion parameter along with each axis.
In the case of no lens distortion,

$$ \delta_x = 0 \qquad \delta_y = 0 \tag{16} $$

- Radial distortion δ_{xr} and δ_{yr},
- Decentering distortoin δ_{xd} and δ_{yd},
- Thin prism distortion δ_{xp} and δ_{yp}.

4/4: A 2D camera point to a 2D image point

Change the 2D camera coordinate system to the 2D image one.

- From a 2D point ${}^{C}\mathbf{X}_d$ in metric system w.r.t. the camera coordinate
- To a 2D point ${}^{I}\mathbf{X}_d$ in pixel system w.r.t. the camera coordinate

$$s \begin{bmatrix} {}^{I}X_d \\ {}^{I}Y_d \\ 1 \end{bmatrix} = \begin{bmatrix} -k_u & 0 & u_0 \\ 0 & -k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^{C}X_d \\ {}^{C}Y_d \\ 1 \end{bmatrix}, \tag{24}$$

$${}^{I}X_d = -k_u {}^{C}X_d + u_0 \qquad {}^{I}Y_d = -k_v {}^{C}Y_d + v_0,$$

where

- parameters (k_u, k_v) transform from metric measures to pixel.
- (u_0, v_0) define the projection of the focal point in the plain.

Camera calibration: General idea

Task

Compute camera parameters:

- Packed parameters $\mathbf{P}$.
- Each components $\mathbf{A}$, $\mathbf{R}$, and $\mathbf{t}$.

Given

- Known 3D points $\{\mathbf{X}_i | i = 1, \ldots, N\}$.
- Observed 2D points $\{\mathbf{x}_i | i = 1, \ldots, N\}$.

Camera calibration: Projective matrix estimation

Setting $p_{34} = 1$, i-th image point $\mathbf{x}_i$ is written as

$$x_i = \frac{X_i p_{11} + Y_i p_{12} + Z_i p_{13} + p_{14}}{X_i p_{31} + Y_i p_{32} + Z_i p_{33} + 1} \tag{25}$$

$$y_i = \frac{X_i p_{21} + Y_i p_{22} + Z_i p_{23} + p_{24}}{X_i p_{31} + Y_i p_{32} + Z_i p_{33} + 1} \tag{26}$$

Solve as an optimization problem w.r.t. $\mathbf{P}$ such as

1. Linear method 1 solves as $\mathbf{A}\mathbf{x} = \mathbf{b}$.
2. Linear method 2 solves as $\mathbf{A}\mathbf{x} = \mathbf{0}$.
3. Non-linear method solves non-linearly.

Camera calibration: Linear method 1

Proposed by [Hall et al., 1982] [3].
Eq. (25) and Eq. (26) is rewritten as

$$X_i p_{11} + Y_i p_{12} + Z_i p_{13} + p_{14} - x_i X_i p_{31} - x_i Y_i p_{32} - x_i Z_i p_{33} = x_i \tag{27}$$
$$X_i p_{21} + Y_i p_{22} + Z_i p_{23} + p_{24} - y_i X_i p_{31} - y_i Y_i p_{32} - y_i Z_i p_{33} = y_i \tag{28}$$

Given N corresponding points $\{\mathbf{X}_i\}$ and $\{\mathbf{x}_i\}$, generate following equation:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_N X_N & -x_N Y_N & -x_N Z_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -y_N X_N & -y_N Y_N & -y_N Z_N \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{32} \\ p_{33} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_N \\ y_N \end{bmatrix} \tag{29}$$

$$\rightarrow \mathbf{Ap} = \mathbf{b}$$

where $\mathbf{A} \in \mathbb{R}^{2N \times 11}$, $\mathbf{p} \in \mathbb{R}^{11}$, and $\mathbf{b} \in \mathbb{R}^{2N}$.

[3] E. L. Hall, J. B. K. Tio, C. A. McPherson, and F. A. Sadjadi. Measuring curved surfaces for robot vision. *Computer*, 15

Camera calibration: Linear method 1 cont.

Considering an energy function $E_1 = \|\mathbf{Ap} - \mathbf{b}\|^2$, projection matrix is obtained by minimizing E_1 as

$$\hat{\mathbf{p}} = \arg\min_{p} E_1 = \arg\min_{p} (\mathbf{Ap} - \mathbf{b})^{\mathsf{T}} (\mathbf{Ap} - \mathbf{b}) \tag{30}$$

Differentiating E_1 w.r.t. $\mathbf{p}$,

$$\begin{aligned} &\frac{\partial E_1}{\partial \mathbf{p}} = 0 \\ &\rightarrow \mathbf{A}^{\mathsf{T}}(\mathbf{A}\hat{\mathbf{p}} - \mathbf{b}) = 0 \\ &\rightarrow \mathbf{A}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{p}} = \mathbf{A}^{\mathsf{T}}\mathbf{b} \\ &\rightarrow \hat{\mathbf{p}} = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}}\mathbf{b} \end{aligned} \tag{31}$$

$\mathbf{p}$ can be estimated if $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ is invertible.

Camera calibration: Linear method 1 cont.

This method heavily relies on whether the matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$ is invertible or not. Alternatively, we solve the problem by solving $\mathbf{Ax} = \mathbf{0}$ as Linear method 2 does.

Camera calibration: Linear method 2

Eq. (25) and Eq. (26) is rewritten as

$$X_i p_{11} + Y_i p_{12} + Z_i p_{13} + p_{14} - x_i X_i p_{31} - x_i Y_i p_{32} - x_i Z_i p_{33} - x_i p_{34} = 0 \tag{32}$$

$$X_i p_{21} + Y_i p_{22} + Z_i p_{23} + p_{24} - y_i X_i p_{31} - y_i Y_i p_{32} - y_i Z_i p_{33} - y_i p_{34} = 0 \tag{33}$$

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 & -y_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_N X_N & -x_N Y_N & -x_N Z_N & -x_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -y_N X_N & -y_N Y_N & -y_N Z_N & -y_N \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{33} \\ p_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{34}$$

$\rightarrow \mathbf{Ap} = \mathbf{0}$

where $\mathbf{A} \in \mathbb{R}^{2N \times 12}$ is points matrix, $\mathbf{p} \in \mathbb{R}^{12}$ is unknown projection matrix parameters vector, and $\mathbf{b} \in \mathbb{R}^{2N}$ is 2D points vector

Camera calibration: Linear method 2 cont.

To obtain the non-trivial solution of homogeneous system $\mathbf{Ap} = \mathbf{0}$, apply constrained optimization.
Considering an energy function $E_2 = \|\mathbf{Ap}\|^2$ subject to the constraint $\|\mathbf{p}\|^2 - 1 = 0$, prevents $\mathbf{p}$ from becoming a zero vector.
With a Lagrange multiplier $\lambda > 0$, we obtain the following energy function

$$\begin{aligned} E_2(\mathbf{p}, \lambda) &= \|\mathbf{Ap}\|^2 - \lambda(\|\mathbf{p}\|^2 - 1) \\ &= (\mathbf{Ap})^\mathbf{T}(\mathbf{Ap}) - \lambda(\mathbf{p}^\mathbf{T}\mathbf{p} - 1). \end{aligned} \tag{35}$$

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} E_2(\mathbf{p}, \lambda) = \arg\min_{\mathbf{p}} (\mathbf{Ap})^\mathbf{T}(\mathbf{Ap}) - \lambda(\mathbf{p}^\mathbf{T}\mathbf{p} - 1) \tag{36}$$

Camera calibration: Linear method 2 cont.

Differentiating E_2 w.r.t. $\mathbf{p}$

$$\begin{aligned}\frac{\partial E_2}{\partial \mathbf{p}} &= 0 \\ &\rightarrow \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{p}} - \lambda\hat{\mathbf{p}} = 0 \\ &\rightarrow \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{p}} = \lambda\hat{\mathbf{p}}\end{aligned} \tag{37}$$

Differentiating E_2 w.r.t. λ

$$\begin{aligned}\frac{\partial E_2}{\partial \lambda} &= 0 \\ &\rightarrow \hat{\mathbf{p}}^\mathsf{T}\hat{\mathbf{p}} - 1 = 0 \\ &\rightarrow \|\hat{\mathbf{p}}\|^2 = 1\end{aligned} \tag{38}$$

Camera calibration: Linear method 2 cont.

Pre-multiplying both sides of Eq. (37) by $\hat{\mathbf{p}}^\mathsf{T}$ gives

$$\begin{aligned} \hat{\mathbf{p}}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{p}} &= \lambda\hat{\mathbf{p}}^\mathsf{T}\hat{\mathbf{p}} \\ \rightarrow (\mathbf{A}\hat{\mathbf{p}})^\mathsf{T}(\mathbf{A}\hat{\mathbf{p}}) &= \lambda 1 \\ \rightarrow \|\mathbf{A}\hat{\mathbf{p}}\|^2 &= \lambda \end{aligned} \tag{39}$$

Eq. (39) is the same expression that $E_2 = \|\mathbf{A}\mathbf{p}\|^2$. This means that minimizing $\|\mathbf{A}\mathbf{p}\|^2$ is to minimize λ.

Camera calibration: Linear method 2 cont.

Differencing the energy function E_2 tells that

- Since Eq. (37) forms like $\mathbf{Ax} = \lambda\mathbf{x}$, $\hat{\mathbf{p}}$ should be an eigenvector of the matrix $\mathbf{A}^\mathbf{T}\mathbf{A}$ whose corresponding eigenvalue is λ.
- Eq. (38) minimizes λ as much as possible (ideally 0)

Thus, $\hat{\mathbf{p}}$ should be the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathbf{A}^\mathbf{T}\mathbf{A}$.

Camera calibration: Projective matrix decomposition

Now, we have

- An estimate of projective matrix $\mathbf{P}$.
- A set of corresponding points $\{\mathbf{X}_i\}$ and $\{\mathbf{x}_i\}$.

Next task is to decompose $\mathbf{P}$ into $\mathbf{A}$, $\mathbf{R}$, and $\mathbf{t}$.

Basically, we use constraint on matrix form

Decomposition of the Camera Matrix

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{array}\right]$$

P = A[R|t] or, equivalently,
P = K[R|t] = K[R|-Rc] = [M|-Mc]

Decomposition of the Camera Matrix

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{array}\right]$$

P = A[R|t] or, equivalently,

P = K[R|t] = K[R|-Rc] = [M|-Mc]

Find the camera center **C**

Find intrinsic **K** and rotation **R**

Decomposition of the Camera Matrix

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{array}\right]$$

P = A[R|t] or, equivalently,
P = K[R|t] = K[R|-Rc] = [M|-Mc]

Find the camera center **C**

$$\mathbf{Pc} = \mathbf{0}$$

Find intrinsic **K** and rotation **R**

Decomposition of the Camera Matrix

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{array}\right]$$

P = A[R|t] or, equivalently,
P = K[R|t] = K[R|-Rc] = [M|-Mc]

Find the camera center **C**

$$\mathbf{Pc} = \mathbf{0}$$

SVD of P!

c *is the Eigenvector corresponding to smallest Eigenvalue*

Find intrinsic **K** and rotation **R**

Decomposition of the Camera Matrix

$$\mathbf{P} = \left[\begin{array}{ccc|c} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{array}\right]$$

P = A[R|t] or, equivalently,
P = K[R|t] = K[R|-Rc] = [M|-Mc]

Find the camera center **C**

$$\mathbf{Pc} = \mathbf{0}$$

SVD of P!

***c** is the Eigenvector corresponding to smallest Eigenvalue*

Page 163 in Multiple Views Geometry

Find intrinsic **K** and rotation **R**

RQ DECOMPOSITION!

A4.1.1 in *Multiple view Geometry*