Image Formation, Camera Model and Calibration

Ying Wu

Electrical Engineering and Computer Science
Northwestern University, Evanston, IL 60208

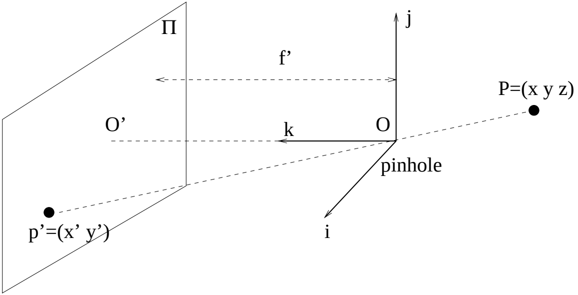
yingwu@northwestern.edu

http://www.eecs.northwestern.edu/~yingwu

Outline

Perspective Projection

- A pinhole camera



- Perspective projection

$$\begin{cases} x' = f'x/z \\ y' = f'y/z \end{cases} \tag{1}$$

- Thin lenses cameras has the same geometry.

Orthographic Projection

- the camera remains at a roughly constant distance from the scene
- the scene centers the optic axis
- orthographic projection:

$$\begin{cases} x' &= x \\ y' &= y \end{cases} \tag{2}$$

Weak Perspective Projection

- when the depth of the scene if “flat”
- a linear approximation to the perspective projection

$$\left\{ \begin{array}{lcl} x' & = & (f'/z_0)x \\ y' & = & (f'/z_0)y \end{array} \right. \tag{3}$$

- also called *scaled-orthographic projection*.

Outline

Homogeneous Coordinates

- A 3D point is $\mathbf{P} = [x \ y \ z]^T$ and a plane $ax + by + cz - d = 0$.
- Homogeneous coordinates unify points and lines.
- for points:

$$\mathbf{P} = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

- for planes:

$$\mathbf{\Pi} = \begin{pmatrix} a \\ b \\ c \\ -d \end{pmatrix},$$

  where the plane $\mathbf{\Pi}$ is defined up to a scale.
- We have

$$\mathbf{\Pi} \cdot \mathbf{P} = 0$$

Translation and Rotation

- "Craig notation". ${}^{F}\mathbf{P}$ means point P in frame F.
- Translation

$$
{}^{B}\mathbf{P} = {}^{A}\mathbf{P} + {}^{B}\mathbf{O}_A \tag{4}
$$

  where ${}^{B}\mathbf{O}_A$ is the coordinate of the origin $\mathbf{O}_A$ of frame A in the new coordinate system B.
- Rotation

$$
{}^{B}_{A}\mathbf{R} = ({}^{B}\mathbf{i}_A \ {}^{B}\mathbf{j}_A \ {}^{B}\mathbf{k}_A) = \begin{pmatrix} {}^{A}\mathbf{i}_B^T \\ {}^{A}\mathbf{j}_B^T \\ {}^{A}\mathbf{k}_B^T \end{pmatrix} \tag{5}
$$

  where ${}^{B}\mathbf{i}_A$ is the coordinate of the axis $\mathbf{i}_A$ of frame A in the new coordinate system B.
- Then we have,

$$
{}^{B}\mathbf{P} = {}^{B}_{A}\mathbf{R}\,{}^{A}\mathbf{P}
$$

Rigid Transformation

- $$ {}^{B}\mathbf{P} = {}^{B}_{A}\mathbf{R}\,{}^{A}\mathbf{P} + {}^{B}\mathbf{O}_{A} \tag{6} $$
- If we make two consecutive rigid transformation, i.e., from A→ B → C, then:

  $$ {}^{C}\mathbf{P} = {}^{C}_{B}\mathbf{R}({}^{B}_{A}\mathbf{R}\,{}^{A}\mathbf{P} + {}^{B}\mathbf{O}_{A}) + {}^{C}\mathbf{O}_{B} = {}^{C}_{B}\mathbf{R}\,{}^{B}_{A}\mathbf{R}\,{}^{A}\mathbf{P} + ({}^{C}_{B}\mathbf{R}\,{}^{B}\mathbf{O}_{A} + {}^{C}\mathbf{O}_{B}) $$
- It looks very awkward.
- Homogeneous coordinates make it concise.

  $$ \begin{bmatrix} {}^{B}\mathbf{P} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{B}_{A}\mathbf{R} & {}^{B}\mathbf{O}_{A} \\ \mathbf{0}^{T} & 1 \end{bmatrix} \begin{bmatrix} {}^{A}\mathbf{P} \\ 1 \end{bmatrix} $$

  and

  $$ \begin{bmatrix} {}^{C}\mathbf{P} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{C}_{B}\mathbf{R} & {}^{C}\mathbf{O}_{B} \\ \mathbf{0}^{T} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\mathbf{P} \\ 1 \end{bmatrix} $$
- $$ \begin{bmatrix} {}^{C}\mathbf{P} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{C}_{B}\mathbf{R} & {}^{C}\mathbf{O}_{B} \\ \mathbf{0}^{T} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}_{A}\mathbf{R} & {}^{B}\mathbf{O}_{A} \\ \mathbf{0}^{T} & 1 \end{bmatrix} \begin{bmatrix} {}^{A}\mathbf{P} \\ 1 \end{bmatrix} $$

Summary

- *Projective transformation*

$$T = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix}$$

- *Affine transformation*

$$T = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

- *Euclidean transformation*, if $\mathbf{A} = \mathbf{R}$, i.e., a rotation matrix ($\mathbf{R}^T\mathbf{R} = \mathbf{I}$),

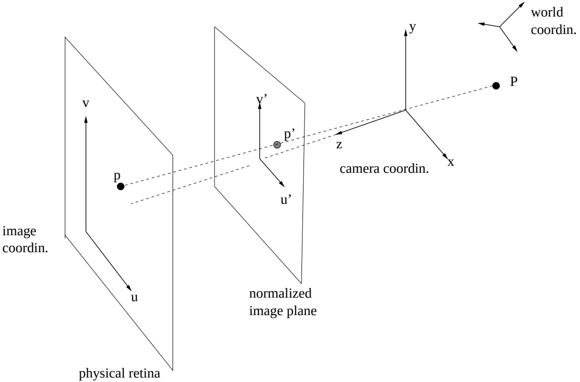
$$T = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

- Q: what are preserved under these transformations?

Outline

Geometric Image Formation

- Relation between 3D coordinates and image coordinates?
- For a 3D point $\mathbf{p}^w = [x^w, y^w, z^w]^T$ in the world coordinate system (W), it is mapped to a camera coordinate system (C), then to the physical image plane, and then the image coordinates $[u, v]^T$.



Transformation Concatenation

- Normalized image plane: located at the focal length $f = 1$.
- The pinhole (c) is mapped to the origin of the image plane ($\hat{c}$), and $\mathbf{p}$ is mapped to $\hat{\mathbf{p}} = [\hat{u}, \hat{v}]^T$.

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{u} \\ \hat{v} \\ 1 \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{p}^c \\ 1 \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} x_c \\ y^c \\ z^c \\ 1 \end{bmatrix}$$

- And we also have

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} kf & 0 & u_0 \\ 0 & lf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} kf & 0 & u_0 \\ 0 & lf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \\ 1 \end{bmatrix}$$

Intrinsic and Extrinsic Parameters

- Let $\alpha = kf$ and $\beta = lf$.
- *intrinsic parameters*: α,β,u_0 and v_0 *intrinsic parameters*.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \\ 1 \end{bmatrix} = \frac{1}{z^c} \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}$$

- *extrinsic parameters*: **R** and **t**, i.e., the camera pose.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z^c} \mathbf{M_1 M_2 p}^w = \frac{1}{z^c} \mathbf{M p}^w \tag{7}$$

- We call **M** the *projection matrix*.

Outline

The Setting of the Problem

- We are given:
  - a calibration rig, i.e., a reference object, to provide the world coordinate system
  - an image of the reference object.
- The problem is to solve:
  - the projection matrix,
  - the intrinsic and extrinsic parameters.
- Mathematically,
  - given: $[x_i^w, y_i^w, z_i^w]^T, i = 1, \ldots, n$, and $[u_i, v_i]^T, i = 1, \ldots, n$
  - to solve: $\mathbf{M}_1$ and $\mathbf{M}_2$, s.t.,

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{z_i^c}\mathbf{M}_1\mathbf{M}_2 \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix} = \frac{1}{z_i^c}\mathbf{M} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix}, \quad \forall i$$

Computing the Projection Matrix

- $$z_i^c \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \\ 1 \end{bmatrix}$$

- We can write

$$\begin{aligned} z_i^c u_i &= m_{11} x_i^w + m_{12} y_i^w + m_{13} z_i^w + m_{14} \\ z_i^c v_i &= m_{21} x_i^w + m_{22} y_i^w + m_{23} z_i^w + m_{24} \\ z_i^c &= m_{31} x_i^w + m_{32} y_i^w + m_{33} z_i^w + m_{34} \end{aligned}$$

- Then

$$\begin{aligned} x_i^w m_{11} + y_i^w m_{12} + z_i^w m_{13} + m_{14} - u_i x_i^w m_{31} - u_i y_i^w m_{32} - u_i z_i^w m_{33} &= u_i m_{34} \\ x_i^w m_{21} + y_i^w m_{22} + z_i^w m_{23} + m_{24} - v_i x_i^w m_{31} - v_i y_i^w m_{32} - v_i z_i^w m_{33} &= v_i m_{34} \end{aligned}$$

Least Squares Solution to the Projection Matrix

- Then,

$$\begin{bmatrix} x_1^w & y_1^w & z_1^w & 1 & 0 & 0 & 0 & 0 & -u_1x_1^w & -u_1y_1^w & -u_1z_1^w \\ 0 & 0 & 0 & 0 & x_1^w & y_1^w & z_1^w & 1 & -v_1x_1^w & -v_1y_1^w & -v_1z_1^w \\ \vdots & & & \vdots & & & \vdots & & & \vdots & \vdots \\ x_n^w & y_n^w & z_n^w & 1 & 0 & 0 & 0 & 0 & -u_nx_n^w & -u_ny_n^w & -u_nz_n^w \\ 0 & 0 & 0 & 0 & x_n^w & y_n^w & z_n^w & 1 & -v_nx_n^w & -v_ny_n^w & -v_nz_n^w \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ \vdots \\ m_{32} \\ m_{33} \end{bmatrix} = \begin{bmatrix} u_1 m_{34} \\ v_1 m_{34} \\ u_2 m_{34} \\ v_2 m_{34} \\ \vdots \\ u_n m_{34} \\ v_n m_{34} \end{bmatrix}$$

- we can let $m_{34} = 1$ (Why?)
- We have:

$$\mathbf{Km} = \mathbf{U} \tag{8}$$

- The least squares solution:

$$\mathbf{m} = \mathbf{K}^{\dagger}\mathbf{U} = (\mathbf{K}^{\mathbf{T}}\mathbf{K})^{-1}\mathbf{K}^{T}\mathbf{U} \tag{9}$$

  where $\mathbf{K}^{\dagger}$ is the pseudoinverse of $\mathbf{K}$.
- Here, $\mathbf{m}$ and $m_{34} = 1$ constitute the projection matrix $\mathbf{M}$.

Computing Intrinsic and Extrinsic Parameters

- Note: $\mathbf{M}$ obtained is a scaled version of the true $\mathbf{M}$ (why?)
- Apparently, we have:

$$m_{34}\mathbf{M} = m_{34}\begin{bmatrix} \mathbf{m}_1^T & m_{14} \\ \mathbf{m}_2^T & m_{24} \\ \mathbf{m}_3^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^T & t_x \\ \mathbf{r}_2^T & t_y \\ \mathbf{r}_3^T & t_z \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{r}_1^T + u_0\mathbf{r}_3^T & \alpha t_x + u_0 t_x \\ \beta\mathbf{r}_2^T + v_0\mathbf{r}_3^T & \beta t_y + v_0 t_x \\ \mathbf{r}_3^T & t_x \end{bmatrix}$$

where $\mathbf{m}_i^T = [m_{i1}, m_{i2}, m_{i3}]$, $\mathbf{M} = \{m_{ij}\}$ is computed before.

- $\mathbf{r}_i^T = [r_{i1}, r_{i2}, r_{i3}]$, $\mathbf{R} = \{r_{ij}\}$ is the rotation matrix.
- To see it clearly, we have:

$$\begin{bmatrix} \alpha\mathbf{r}_1^T + u_0\mathbf{r}_3^T & \alpha t_x + u_0 t_x \\ \beta\mathbf{r}_2^T + v_0\mathbf{r}_3^T & \beta t_y + v_0 t_x \\ \mathbf{r}_3^T & t_x \end{bmatrix} = \begin{bmatrix} m_{34}\mathbf{m}_1^T & m_{34}m_{14} \\ m_{34}\mathbf{m}_2^T & m_{34}m_{24} \\ m_{34}\mathbf{m}_3^T & m_{34} \end{bmatrix}$$

Factorization

- easy to see $m_{34}\mathbf{m}_3 = \mathbf{r}_3$.
- Then, we have (why?)

$$m_{34} = \frac{1}{|\mathbf{m}_3|}$$

- Then, it is easy to figure out all the other parameters:

$$\begin{aligned}
\mathbf{r}_3 &= m_{34}\mathbf{m}_3 \\
u_0 &= (\alpha\mathbf{r}_1^T + u_0\mathbf{r}_3^T)\mathbf{r}^3 = m_{34}^2\mathbf{m}_1^T\mathbf{m}_3 \\
v_0 &= (\beta\mathbf{r}_2^T + v_0\mathbf{r}_3^T)\mathbf{r}^3 = m_{34}^2\mathbf{m}_2^T\mathbf{m}_3 \\
\alpha &= m_{34}^2|\mathbf{m}_1 \times \mathbf{m}_3| \\
\beta &= m_{34}^2|\mathbf{m}_2 \times \mathbf{m}_3|
\end{aligned}$$

Factorization (cont.)

- After that, it is also easy to get:

$$
\begin{aligned}
\mathbf{r}_1 &= \frac{m_{34}}{\alpha}(\mathbf{m}_1 - u_0\mathbf{m}_3) \\
\mathbf{r}_2 &= \frac{m_{34}}{\beta}(\mathbf{m}_2 - v_0\mathbf{m}_3) \\
t_z &= m_{34} \\
t_x &= \frac{m_{34}}{\alpha}(m_{14} - u_0) \\
t_y &= \frac{m_{34}}{\beta}(m_{24} - v_0)
\end{aligned}
$$

Questions to Discuss

- Enhance the accuracy?
    - The projection matrix **M** has 10 independent variables. (why?)
    - Even $m_{34} = 1$, we still have 11 parameters to determine.
    - However, these 11 parameters are not independent!
    - Can we use this dependency to enhance the accuracy?
- The method described above assumes that we know the 2D image coordinates. How can we get these 2D image points?
- If the detection of these 2D image points contains noise, how does the noise affect the accuracy of the result?
- If we are not using point-corespondences, can we use lines?