Intrinsic and Extrinsic Parameters
Rigid Transformations, Homogeneous Coordinates, and Perspective Projection

Amir

Lecture Outline

- Camera models and coordinate systems
- Rigid transformations in 3D
- Homogeneous coordinates
- Intrinsic and extrinsic camera parameters
- Perspective projection model
- Examples and discussion questions

Why Camera Geometry?

- Relates 3D world points to 2D image measurements
- Foundation of computer vision tasks:
  - Camera calibration
  - 3D reconstruction
  - Visual odometry and SLAM
- Separates *geometry* from *appearance*

Coordinate Systems

- **World coordinate system**: fixed reference in the scene
- **Camera coordinate system**: origin at the camera center
- **Image coordinate system**: pixel coordinates on the image plane

A key goal: map a 3D point

$$\mathbf{X}_w = (X, Y, Z)^T$$

from world coordinates to an image point

$$\mathbf{x} = (u, v)^T.$$

Rigid Transformations in 3D

Definition: A rigid transformation preserves distances and angles.

It consists of:

- Rotation $\mathbf{R} \in SO(3)$
- Translation $\mathbf{t} \in \mathbb{R}^3$

Mapping from world to camera coordinates:

$$\mathbf{X}_c = \mathbf{R}\,\mathbf{X}_w + \mathbf{t}$$

Rotation Matrix Properties

A rotation matrix $\mathbf{R}$ satisfies:

- $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ (orthonormality)
- $\det(\mathbf{R}) = 1$

Interpretation:

- Rows (or columns) represent camera axes
- Encodes camera orientation in space

Homogeneous Coordinates

Motivation:

- Allow translation to be represented as matrix multiplication
- Enable projective transformations

A 3D point in homogeneous coordinates:

$$\tilde{\mathbf{X}} = (X, Y, Z, 1)^T$$

A 2D image point:

$$\tilde{\mathbf{x}} = (u, v, 1)^T$$

Rigid Transformation in Homogeneous Form

Using homogeneous coordinates:

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

This 4×4 matrix compactly represents rotation and translation.

Extrinsic Parameters

Definition: Extrinsic parameters describe the *pose* of the camera in the world.

They consist of:

- Rotation $\mathbf{R}$
- Translation $\mathbf{t}$

They answer the question:

Where is the camera, and how is it oriented?

Intrinsic Parameters

Definition: Intrinsic parameters describe the internal geometry of the camera.

They are encoded in the intrinsic matrix:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

- (f_x, f_y): focal lengths in pixel units
- (c_x, c_y): principal point
- s: skew (often zero)

Pinhole Camera Model

Assumption:

- All light rays pass through a single point (camera center)

Perspective projection from camera coordinates:

$$(x, y) = \left(\frac{X_c}{Z_c}, \frac{Y_c}{Z_c}\right)$$

This explains:

- Scale change with depth
- Vanishing points

Full Camera Projection Equation

Combining extrinsic and intrinsic parameters:

$$\tilde{\mathbf{x}} \sim \mathbf{K} \left[\mathbf{R} \mid \mathbf{t} \right] \tilde{\mathbf{X}}_w$$

Where:

- $\sim$ denotes equality up to scale
- $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ is the *camera projection matrix*

Example: Simple Projection

Assume:

- $\mathbf{R} = \mathbf{I}$, $\mathbf{t} = \mathbf{0}$
- $f_x = f_y = f$, $c_x = c_y = 0$

Then:

$$\mathbf{x} = \left(f\frac{X}{Z},\ f\frac{Y}{Z} \right)$$

Interpretation: Objects farther from the camera appear smaller.

Conceptual Questions

1. Why are intrinsic parameters independent of camera pose?
2. What happens to the projection when $Z \to 0$?
3. Why are homogeneous coordinates essential for perspective projection?
4. How would lens distortion violate the pinhole model assumptions?

References

- R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.
- R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011.
- O. Faugeras, *Three-Dimensional Computer Vision*, MIT Press, 1993.