Homogeneous Coordinates

Motivation

Last time we defined coordinate systems and wrote equations for a *pinhole camera model*. Write the equations that map a 3D point (X, Y, Z) in 3D camera coordinates to a 2D pixel point (u, v). Label the units of each variable.

Problem: This operation is fundamentally nonlinear. Why?

Goal for today: Rewrite the projection as a linear operation using **homogeneous coordinates**.

Motivating Example

What are all the points in 3D that project to the same point (x', y')?

Definition

Homogeneous coordinates represent an N-dimensional point using $N + 1$ coordinates.

2D point: (x, y) becomes _____ in homogeneous coordinates

3D point: (X, Y, Z) becomes _____ in homogeneous coordinates

Homogenous Equivalence:

Converting back to regular coordinates:

Matrix Form of Projection

Write a matrix that performs the projection from 3D camera coordinates to 2D pixel coordinates using homogeneous coordinates:

The Intrinsic Matrix:

Write the matrix from above in its standard form: a 3x3 matrix called the **intrinsic matrix K**, and a 3x4 projection matrix:

Why is this called the "intrinsic" matrix?

World Coordinates

So far, all our 3D points have been in *camera-centered coordinates*. But what if we want to describe points in the world?

Multiple Coordinate Systems

Draw a scene with:

- A camera with its coordinate system
- A world coordinate system
- A 3D point **P**

Question: The same physical point **P** has different coordinates in each system. How are they related?

World-to-Camera Transformations

3D Camera coordinates and World coordinates are related by a **rigid body transformation**:

$$\mathbf{P}_c = \mathbf{R} \cdot \mathbf{P}_w + \mathbf{T}$$

This has two parts:

1. **Rotation: R** (a 3×3 matrix)

    - What does **R** represent?

    - What properties must **R** have?

2. **Translation: T** (a 3×1 vector)

    - What does **T** represent?

The rotation **R** and translation **T** together are called the **extrinsic parameters**.

Full Projection Pipeline

Now we can describe the complete process of projecting a world point onto the image:

Step 1: Transform from world to camera coordinates

Step 2: Project onto image using intrinsic matrix

Step 3: "Dehomogenize" to pixel coordinates

The Complete Projection Matrix

Why not combine all of these steps into a single matrix multiplication? Write it out below.

Summary: The Camera Model

Intrinsic Parameters (internal camera properties)

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

- f_x, f_y: focal length in pixels
- (c_x, c_y): principal point (pixel coordinates)

Extrinsic Parameters (camera pose in world)

- $\mathbf{R}$: 3×3 rotation matrix (world to camera)
- $\mathbf{T}$: 3×1 translation vector
- Together: $\mathbf{P}_c = \mathbf{R}\mathbf{P}_w + \mathbf{T}$

Complete Projection Equation

From world point to image pixel:

$$\tilde{\mathbf{p}} = \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\tilde{\mathbf{P}}_w$$

Or equivalently:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

where $\sim$ means "equal up to scale" (homogeneous coordinates).