COMP90086 Assignment 1

YU LUO
Student ID: 1528667

August 22, 2025

1 Mapping between world and image coordinates

1.1 Calculating Real World Coordinates

Projection model. The relationship between a 3D world coordinate (X, Y, Z) and its 2D image plane projection (x, y) using a pinhole camera model is given by the perspective projection equations:

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}. \tag{1}$$

The camera center is at the origin $(0, 0, 0)$, the optical axis is aligned with the $+Z$ axis, and the image plane is located at a distance f (focal length) from the origin. The image has been flipped so that $+x \leftrightarrow +X$ and $+y \leftrightarrow +Y$. The inverse mapping, used to recover the world coordinates, is:

$$X = \frac{xZ}{f}, \quad Y = \frac{yZ}{f}. \tag{2}$$

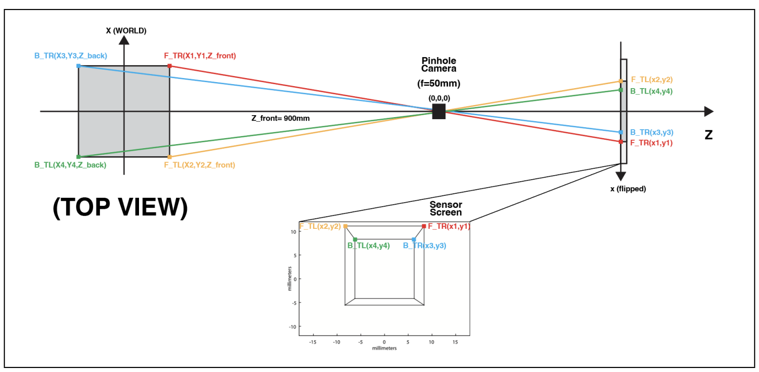

Figure 1: Top view of the cube and camera geometry.

Front face calculation. The closest face of the cube lies on the plane $Z_{\text{front}} = 90$ cm $= 900$ mm. The focal length is $f = 50$ mm. For the front face, the inverse mapping simplifies to:

$$X = \frac{x \cdot 900}{50} = 18x$$
$$Y = \frac{y \cdot 900}{50} = 18y$$

Applying these to the image coordinates of the front-top-right corner (Sees Figure 1), $F_{\text{TR}}(x, y) = (8.333333, 11.11111)$, we get:

$$X_1 = 18 \cdot 8.333333 = 149.999994 \approx 150 \text{ mm}$$
$$Y_1 = 18 \cdot 11.11111 = 199.99998 \approx 200 \text{ mm}$$
$$Z_1 = 900 \text{ mm}$$

Similarly, the front-bottom-left corner, $F_{\text{BL}}(x, y) = (-8.333333, -5.555556)$, has world coordinates:

$$X = 18 \cdot (-8.333333) = -149.999994 \approx -150 \text{ mm}$$
$$Y = 18 \cdot (-5.555556) = -100.000008 \approx -100 \text{ mm}$$
$$Z = 900 \text{ mm}$$

Back face calculation. The side length of the cube, s, can be determined from the front face's dimensions. For example, the difference between the X coordinates of the top-right and top-left corners is:

$$s = X_{\text{F\_TR}} - X_{\text{F\_TL}} = 150 - (-150) = 300 \text{ mm}$$

Since it is a cube, the depth is also 300 mm. The back face is located at:

$$Z_{\text{back}} = Z_{\text{front}} + s = 900 + 300 = 1200 \text{ mm}$$

For the back face, the inverse mapping becomes:

$$X = \frac{x \cdot 1200}{50} = 24x$$
$$Y = \frac{y \cdot 1200}{50} = 24y$$

Applying these to the image coordinates of the back-top-right corner (Sees Figure 1), $B_{\text{TR}}(x, y) = (6.25, 8.333333)$, we get:

$$X = 24 \cdot 6.25 = 150 \text{ mm}$$
$$Y = 24 \cdot 8.333333 = 199.999992 \approx 200 \text{ mm}$$
$$Z = 1200 \text{ mm}$$

The remaining corner coordinates can be calculated similarly.

Final result. The world coordinates for all eight cube vertices are summarised in the table below.

Table 1: Reconstructed cube vertices in world coordinates.

Corner	X (mm)	Y (mm)	Z (mm)
F_BL	-150	-100	900
F_BR	150	-100	900
F_TL	-150	200	900
F_TR	150	200	900
B_BL	-150	-100	1200
B_BR	150	-100	1200
B_TL	-150	200	1200
B_TR	150	200	1200

1.2 World $\rightarrow$ Image: Translation Check

I verify the World $\rightarrow$ Image conversion behaviours by translating the cube in world space: (a) baseline; (b) $+X$ shifts image right; (c) $+Y$ shifts image up; (d) $-Z$ (closer) enlarges the cube; (e) $+Z$ (farther) shrinks it; (f) combined $(-X, -Y, -Z)$ yields left/down shift with scaling up—consistent with perspective.

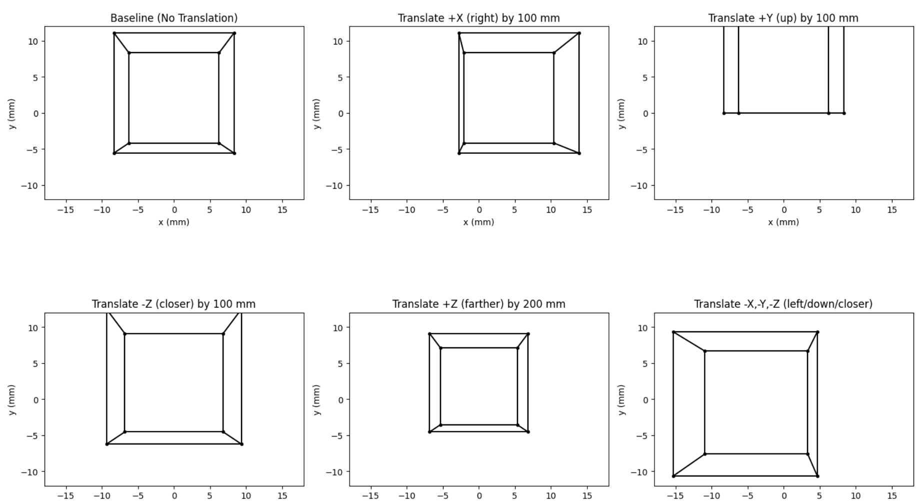

Figure 2: Projection after world translations.

2 Depth of Field Simulation

2.1 Implementation

Parameter choices. I used an odd blur kernel (51) to make the effect visually clear and to satisfy the Gaussian filter requirement (Sees Figure 3); the focus band is specified in millimetres to match the metric depth map. All the implementation details can be found in the code.

Figure 3: Side-by-side comparison for two scenes: original (left), depth map (centre, visualised), and simulated DoF (right). The same global blur is used in both scenes; only the focal plane and focus band are varied.

2.2 Discussion on DoF Simulation Flaws

The baseline DoF simulation produces results that are visually unconvincing due to its oversimplified blur model.

Abrupt Depth Transitions. Using a binary in-focus mask creates discrete boundaries between sharp and blurred regions (Sees Figure 4). This produces visible seams or halos around objects at depth discontinuities, which is unrealistic since real optics yield smooth, continuous transitions.

Uniform Blur. A single global Gaussian blur is applied to all out-of-focus pixels, ignoring the physics of real lenses where blur magnitude increases gradually with distance from the focal plane. As a result, near- and far-out-of-focus objects appear equally blurred, flattening the perceived depth (Sees Figure 5).

Dependence on Depth Map Quality. The method is highly sensitive to the accuracy of the depth map. Imperfections in the depth data directly introduce visible artifacts, including:

- **Halos and seams:** Boundary errors misclassify pixels near depth discontinuities, leading to jagged outlines and unnatural hard edges as shown in Figure 4.
- **Color bleeding:** Misalignment between the RGB image and the depth map may cause foreground colors to spill into blurred regions, or background colors to leak into sharp areas.
- **Background leakage:** Thin or fine structures (e.g., fences, tree branches) are especially vulnerable to noise; inaccuracies allow blurred background textures to "show through," making these objects appear fragmented or semi-transparent (See Figure 4-5).

Trade-off in Bandwidth. Increasing the focus band can reduce harsh seams but simultaneously weakens the DoF effect, leaving the image nearly uniformly sharp (Sees Figure 6). This highlights the fundamental limitation of threshold-based masking: it cannot reproduce the gradual, depth-dependent blur of real optical systems.

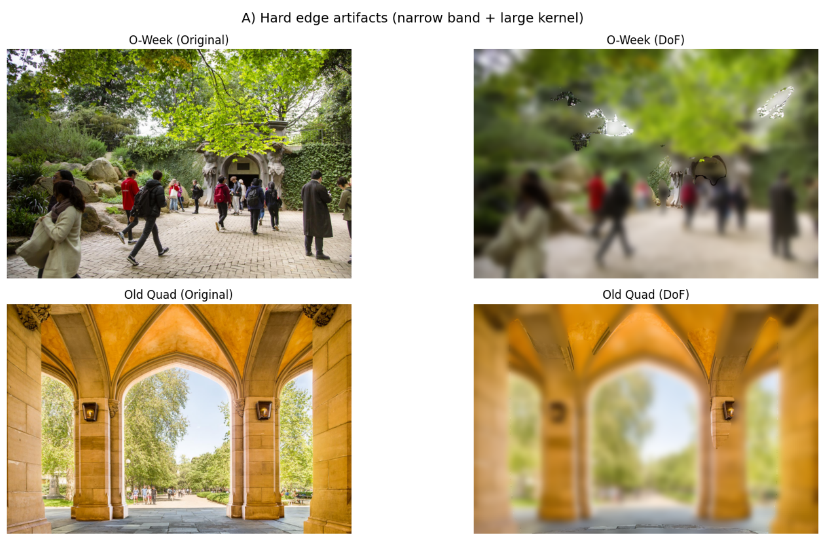

Figure 4: Scenario A: Narrow band + strong blur. Halos and sharp seams are visible.

Figure 5: Scenario B: Single global blur. All out-of-focus regions look equally blurred.

Figure 6: Scenario C: Wide focus band. Fewer seams, but flatter overall look.