Computer Vision

Exercise 1 – Image Formation

1 Pen and Paper

1.1 Homogeneous Coordinates

a) Proof that in homogeneous coordinates, the intersection point $\tilde{\mathbf{x}}$ of the two lines $\tilde{\mathbf{l}}_1$ and $\tilde{\mathbf{l}}_2$ is given by $\tilde{\mathbf{x}} = \tilde{\mathbf{l}}_1 \times \tilde{\mathbf{l}}_2$.

we can define $\tilde{x} := \tilde{I}_1 \times \tilde{I}_2$. and then show that $\tilde{x}$ lies in both lines.

For any point who locates on the line, it satisfies the equation.

$\{\bar{x} \mid \tilde{I}^T \bar{x} = 0\}$.

we can clearly show that

$$\tilde{I}_1^T \cdot \bar{x} = \tilde{I}_1^T \cdot \tilde{x} = \tilde{I}_1 \cdot \tilde{I}_1 \times \tilde{I}_2 = 0.$$

$$\tilde{I}_2 \cdot \bar{x} = \tilde{I}_2^T \cdot \tilde{x} = \tilde{I}_2 \cdot \tilde{I}_1 \times \tilde{I}_2 = 0.$$

As $\tilde{x}$ lies on both lines, it is the intersection point.

b) Similarly, proof that the line that joins two points $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ is given by $\tilde{\mathbf{l}} = \tilde{\mathbf{x}}_1 \times \tilde{\mathbf{x}}_2$.

Similarly, we can also define $\tilde{I} = \tilde{x}_1 \times \tilde{x}_2$.

For any point who locates on the line, it satisfies the equation.

$\{\bar{x} \mid \tilde{I}^T \bar{x} = 0\}$.

we can clearly show that

$$\tilde{I} \cdot \tilde{x}_1 = \tilde{x}_1 \times \tilde{x}_2 \cdot \tilde{x}_1 = 0.$$

$$\tilde{I} \cdot \tilde{x}_2 = \tilde{x}_1 \times \tilde{x}_2 \cdot \tilde{x}_2 = 0.$$

So the line $\tilde{I}$ joins both points.

c) You are given the following two lines:

$$\mathbf{l}_1 = \{(x, y)^\top \in \mathbb{R}^3 | x + y + 3 = 0\}$$
$$\mathbf{l}_2 = \{(x, y)^\top \in \mathbb{R}^3 | -x - 2y + 7 = 0\}$$

First, find the intersection point of the two lines by solving the system of linear equations. Next write the lines using homogeneous coordinates and calculate the intersection point using the cross product. Do you obtain the same intersection point?

$x = -y - 3$.

Inserting this into the second gives.

$y + 3 - 2y + 7 = 0$.

$y = 10$
$x = -13.$

the intersection point is $(-13, 10)$.

$\tilde{I}_1$ can be expressed using homogeneous coordinate $\tilde{I}_1 = (1, 1, 3)^T$

Similarly, $\tilde{I}_2 = (-1, -2, 7)^T$

$$\tilde{I}_1 \times \tilde{I}_2 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \times \begin{pmatrix} -1 \\ -2 \\ 7 \end{pmatrix} = \begin{pmatrix} 13 \\ -10 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ 1 \end{pmatrix} \Rightarrow x = \begin{pmatrix} 13 \\ -10 \end{pmatrix}$$

d) Write down the line whose normal vector is pointing into the direction $(3, 4)^\top$ and which has a distance of 3 from the origin.

The normalized normal vector: $n = \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \end{pmatrix}$

Hence, we obtain the line as

$$\tilde{I} = \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 3 \end{pmatrix}$$

e) What distance from the origin and what (normalized) normal vector does the homogeneous line $\tilde{\mathrm{l}} = (2, 5, \frac{\sqrt{29}}{5})^\top$ have?

$$\tilde{I} = (2, 5, \frac{\sqrt{29}}{5})^T = \sqrt{29} \left(\frac{2\sqrt{29}}{29}, \frac{5\sqrt{29}}{\sqrt{29}}, \frac{1}{5} \right)^T$$

Hence, we obtain

$$n = \left(\frac{2\sqrt{29}}{29}, \frac{5\sqrt{29}}{29} \right)^T$$

$$d = \frac{1}{5}$$

1.2 Transformations

a) Write down the 2×3 translation matrix which maps $(1, 2)^\top$ onto $(0, 3)^\top$.

$$T = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1-1 \\ 2+1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

b) Let's assume that you are given N 2D correspondence pairs

$$(\mathbf{x}_i, \mathbf{y}_i) = \left(\begin{pmatrix} x_1^i \\ x_2^i \end{pmatrix}, \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix} \right)$$

Find the 2×3 translation matrix mapping $\mathbf{x}_i$ onto $\mathbf{y}_i$ which is optimal in the least square sense.

Hint: Define a cost function as

$$E(\mathbf{T}) = \sum_{i=1}^{N} ||\mathbf{T}\bar{\mathbf{x}}_i - \mathbf{y}_i||_2^2$$

and find the optimal $\mathbf{T}^\star$ which minimizes E:

$$\mathbf{T}^\star = \underset{\mathbf{T}}{\text{argmin}} \ E(\mathbf{T})$$

You can find $\mathbf{T}^\star$ by calculating the Jacobian $\mathbf{J}_E$ of E and setting it to $\mathbf{0}^\top$:

$$\mathbf{J}_E = \left[\frac{\partial E}{\partial t_1}, \dots, \frac{\partial E}{\partial t_N} \right] \stackrel{!}{=} \mathbf{0}^\top$$

Can you give an intuitive explanation for the equation you derive for $\mathbf{T}^*$?

First, $T = \begin{bmatrix} 1 & 0 & t_1 \\ 0 & 1 & t_2 \end{bmatrix}$

Next, $E(T) = \sum_{i=1}^{N} || T\bar{x}_i - y_i ||_2^2$

$= \sum_{i=1}^{N} (x_1^i + t_1 - y_1^i)^2 + (x_2^i + t_2 - y_2^i)^2$

Then. $\frac{E(T)}{\partial t_1} = \sum_{i=1}^{N} 2(x_1^i + t_1 - y_1^i)$

$= \sum_{i=1}^{N} 2(x_1^i - y_1^i) + 2N t_1$

$\stackrel{!}{=} 0$

Rearraging gives $t_1 = \frac{1}{N} \sum_{i=1}^{N} (y_1^i - x_1^i)$

Similarly, $t_2 = \frac{1}{N} \sum_{i=1}^{N} (y_2^i - x_2^i)$

c) You are given the following three correspondence pairs:

$$\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -5 \end{pmatrix} \right)$$
$$\left(\begin{pmatrix} 5 \\ 7 \end{pmatrix}, \begin{pmatrix} 7 \\ 6 \end{pmatrix} \right)$$
$$\left(\begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ -4 \end{pmatrix} \right)$$

Using your derived equation, calculate the optimal 2×3 translation matrix $\mathbf{T}^\star$.

$t_1^* = \frac{(3-0) + (7-5) + (5-4)}{3} = 2$

$t_2^* = \frac{(-5-1) + (6-7) + (-4-1)}{3} = -4$

$T^* = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \end{bmatrix}$

1.3 Camera Projections

a) Calculate the full rank 4×4 projection matrix $\tilde{\mathbf{P}}$ for the following scenario:

- The camera pose consists of a 90° rotation around the x axis and translation of $(1, 0, 2)^\top$.
- The focal lengths f_x, f_y are 100.
- The principal point $(c_x, c_y)^\top$ is $(25, 25)$.

$$K = \begin{bmatrix} 100 & 0 & 25 \\ 0 & 100 & 25 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P = K\,[R \;\; t] = \begin{bmatrix} 100 & 0 & 25 & 0 \\ 0 & 100 & 25 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 100 & 25 & 0 & 150 \\ 0 & 25 & -100 & 50 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

b) For the previously defined projection, find the world point in inhomogeneous coordinates $\mathbf{x}_w$ which corresponds to the projected homogeneous point in screen space $\tilde{\mathbf{x}}_s = (25, 50, 1, 0.25)^\top$.

$$P^{-1} = \begin{bmatrix} \frac{1}{100} & 0 & -\frac{1}{4} & -1 \\ 0 & 0 & 1 & -2 \\ 0 & -\frac{1}{100} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{X}_w = P^{-1}\tilde{X}_s = \begin{pmatrix} -0.25 \\ 0.5 \\ -0.25 \\ 0.25 \end{pmatrix} \sim \begin{pmatrix} -1 \\ 2 \\ -1 \\ 1 \end{pmatrix}$$

In inhomogeneous coordinate, we write

$$x_w = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

c) Let's perform our first projection of a geometric shape. We define $\mathcal{C}_0$ as the cube centered at $\mathbf{c}_c = (0, 0, 15)^\top$ with equal side lengths $s = 20$.

- **i)** Project the 8 corners of the cube $\mathcal{C}_0$ to the image plane for the pinhole camera with focal lengths $f_x = f_y = 5$ and the principal point $(c_x, c_y)^\top = (10, 10)^\top$. Draw the projected points and edges of the cube in a coordinate system on a paper.
- **ii)** Let's move the cube further away from the pinhole camera along the z-axis such that the distance of the center of the new cube $\mathcal{C}_1$ is now 20. Further, let's zoom in with our camera such that the focal lengths are $f_x = f_y = 10$ and the principal point remains the same. Draw the projected points and edges of the cube in a coordinate system on a paper.
- **iii)** Let's move it even further away along the z-axis while zooming in. Now, the distance of the center of cube $\mathcal{C}_2$ is 100 and the focal lengths are $f_x = f_y = 90$. Draw the projected points and edges of the cube in a coordinate system on a paper.
- **iv)** Project the 8 corners of the first cube $\mathcal{C}_0$ using an orthographic projection and add the principal point $\mathbf{c}_c = (10, 10)^\top$ onto the obtained pixel coordinates to be in the same coordinate system as before. Draw the projected points and edges of the cube in a coordinate system on a paper.
- **v)** When is the perspective projection most similar to the orthographic projection?

i). $K_1 = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 5 & 10 \\ 0 & 0 & 1 \end{bmatrix}$

$\tilde{x}_s = K_1 \tilde{x}_c.$

the cornor of the cube can be read off,

$c_1 = \begin{pmatrix} -10 \\ -10 \\ 5 \end{pmatrix}, c_2 = \begin{pmatrix} 10 \\ -10 \\ 5 \end{pmatrix}, c_3 = \begin{pmatrix} -10 \\ 10 \\ 5 \end{pmatrix}, c_4 = \begin{pmatrix} 10 \\ 10 \\ 5 \end{pmatrix}$

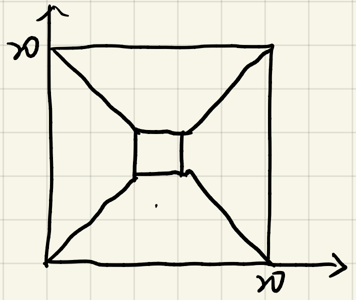
$c_5 = \begin{pmatrix} -10 \\ -10 \\ 25 \end{pmatrix}, c_6 = \begin{pmatrix} 10 \\ -10 \\ 25 \end{pmatrix}, c_7 = \begin{pmatrix} -10 \\ 10 \\ 25 \end{pmatrix}, c_8 = \begin{pmatrix} 10 \\ 10 \\ 25 \end{pmatrix}$

Then. the projected cornor can be shown as,

$P_1 = \begin{pmatrix} 0 \\ 0 \\ 5 \end{pmatrix} \sim \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, P_2 = \begin{pmatrix} 100 \\ 0 \\ 5 \end{pmatrix} \sim \begin{pmatrix} 20 \\ 0 \\ 1 \end{pmatrix}, P_3 = \begin{pmatrix} 0 \\ 100 \\ 5 \end{pmatrix} \sim \begin{pmatrix} 0 \\ 20 \\ 1 \end{pmatrix}$

$P_4 = \begin{pmatrix} 100 \\ 100 \\ 5 \end{pmatrix} \sim \begin{pmatrix} 20 \\ 20 \\ 1 \end{pmatrix}, P_5 = \begin{pmatrix} 200 \\ 200 \\ 25 \end{pmatrix} \sim \begin{pmatrix} 8 \\ 8 \\ 1 \end{pmatrix}, P_6 = \begin{pmatrix} 300 \\ 200 \\ 25 \end{pmatrix} \sim \begin{pmatrix} 12 \\ 8 \\ 1 \end{pmatrix}$

$P_7 = \begin{pmatrix} 200 \\ 300 \\ 25 \end{pmatrix} \sim \begin{pmatrix} 8 \\ 12 \\ 1 \end{pmatrix}, P_8 = \begin{pmatrix} 300 \\ 300 \\ 25 \end{pmatrix} \sim \begin{pmatrix} 12 \\ 12 \\ 1 \end{pmatrix}$



ii) $K_2 = \begin{bmatrix} 10 & 0 & 10 \\ 0 & 10 & 10 \\ 0 & 0 & 1 \end{bmatrix}$

$c_1 = \begin{pmatrix} -10 \\ -10 \\ 10 \end{pmatrix}, c_2 = \begin{pmatrix} 10 \\ -10 \\ 10 \end{pmatrix}, c_3 = \begin{pmatrix} -10 \\ 10 \\ 10 \end{pmatrix}, c_4 = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$

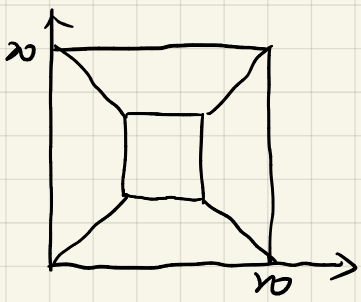
$$C_5=\begin{pmatrix}-10\\-10\\30\end{pmatrix},\ C_6=\begin{pmatrix}10\\-10\\30\end{pmatrix},\ C_7=\begin{pmatrix}-10\\10\\30\end{pmatrix},\ C_8=\begin{pmatrix}10\\10\\30\end{pmatrix}$$

Then,

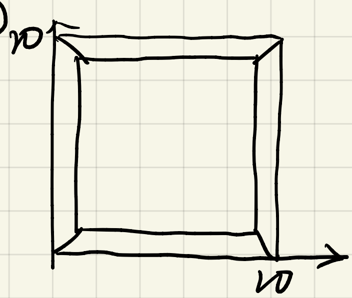
$$P_1=\begin{pmatrix}0\\0\\10\end{pmatrix}\sim\begin{pmatrix}0\\0\\1\end{pmatrix},\ P_2=\begin{pmatrix}200\\0\\10\end{pmatrix}\sim\begin{pmatrix}20\\0\\1\end{pmatrix},\ P_3=\begin{pmatrix}0\\200\\10\end{pmatrix}\sim\begin{pmatrix}0\\20\\1\end{pmatrix}$$

$$P_4=\begin{pmatrix}200\\200\\10\end{pmatrix}\sim\begin{pmatrix}20\\20\\1\end{pmatrix},\ P_5=\begin{pmatrix}200\\200\\30\end{pmatrix}\sim\begin{pmatrix}\frac{20}{3}\\\frac{20}{3}\\1\end{pmatrix},\ P_6=\begin{pmatrix}400\\200\\30\end{pmatrix}\sim\begin{pmatrix}\frac{40}{3}\\\frac{20}{3}\\1\end{pmatrix}$$

$$P_7=\begin{pmatrix}200\\400\\30\end{pmatrix}\sim\begin{pmatrix}\frac{20}{3}\\\frac{40}{3}\\1\end{pmatrix}.\quad P_8=\begin{pmatrix}400\\400\\30\end{pmatrix}\sim\begin{pmatrix}\frac{40}{3}\\\frac{40}{3}\\1\end{pmatrix}$$

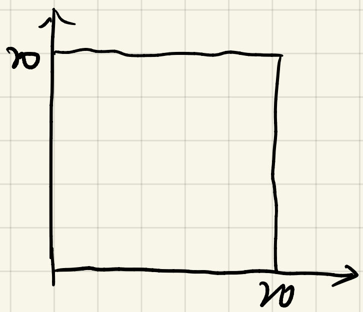


iii)



iv) for orthographic projection,

$$x_s=\begin{bmatrix}1&0&0\\0&1&0\end{bmatrix}x_c+\begin{pmatrix}10\\10\end{pmatrix}$$



v) For larger distances and focal lengths, the projection looks more similar to the orthographic projection and the perspective effect vanishes. In the limit, the perspective projection is equal to the orthographic projection.

d) Let's build our own Ames Room (see Fig. 1)! For this, print the pdf document `ames_room.pdf` and follow the instructions given on the document.

1.4 Photometric Image Formation

a) Write down the thin lens formula and calculate the focus distance z_c in meters for focal length $f = 100mm$ and the distance to the image plane $z_s = 104mm$.

$$\frac{1}{z_s} + \frac{1}{z_c} = \frac{1}{f}$$

$$z_c = \frac{1}{\frac{1}{f} - \frac{1}{z_s}}$$

$$= \frac{1}{\frac{1}{100} - \frac{1}{104}} \text{ mm} = 2600 \text{ mm} = 2.6 \text{ m}$$

b) Write the diameter of the circle of confusion c as a function of the focal length f, the image plane distance z_s as well as the distance Δz_s and the f-number N.

$$N = \frac{f}{d}. \quad \Rightarrow \quad d = \frac{f}{N}.$$

$$\frac{d}{z_s} = \frac{c}{\Delta z_s}.$$

Hence, we obtain $c = \frac{d}{z_s} \Delta z_s = \frac{f}{N z_s} \Delta z_s$

c) Using your derived formula, calculate the diameter of the circle of confusion for the setting $f = 35mm$, N = 1.4, $z_s = 40mm$ when $\Delta z_s = 0.1mm$ as well as when $\Delta z_s = 0.03mm$. Assuming the camera uses a sensor of size $64mm^2$ and a pixel resolution of 400×400 with squared pixels, are the calculated projections sharp or not?

When $\Delta z_s = 0.1mm$, $c_1 = 0.0625$ mm

when $\Delta z_s = 0.03mm$, $c_2 = 0.01875$ mm

The projection is sharp if the diameter is smaller than the smallest pixel side length.

One pixed has the surface area

$$a_p = \frac{64 mm^2}{400 \times 400} = 0.0004 mm^2$$

As the pixes are squared, we can obtain the side length

$$s_p = \sqrt{a_p} = 0.02 mm^2$$

Hence, only the second one is sharp.