Execise 1 Image formation

1.1 Pen and Paper

Homogeneous Coordinates

笛卡尔坐标和齐次坐标的区别是：

- 笛卡尔坐标是用一组数来表示一个点的位置，例如平面上的点用 (x, y) 表示，空间中的点用 (x, y, z) 表示。

- 齐次坐标是用一组数来表示一个点的比例，例如平面上的点用 (x, y, w) 表示，空间中的点用 (x, y, z, w) 表示，其中 w 不等于 0。齐次坐标可以通过除以 w 来转换为笛卡尔坐标，例如 (x, y, w) 对应于 (x/w, y/w)。

- 齐次坐标的优点是可以用矩阵来表示仿射变换和投影变换，而笛卡尔坐标则需要额外的平移向量。齐次坐标还可以表示无穷远的点，例如 (1, 0, 0) 表示 x 轴上的无穷远点。

齐次坐标是一种用于投影几何的坐标系统，它可以表示包括无穷远点在内的所有点。齐次坐标的特点是，如果一个点的齐次坐标乘以一个非零常数，那么得到的还是同一个点的齐次坐标。齐次坐标可以用矩阵来表示各种变换，而不需要额外的平移向量。

穿过原点 （0，0） 的直线的方程可以写为 $nx + my = 0$，其中 n 和 m 不都是 0。在参数形式中，这可以写成 $x = mt$，$y = -nt$。设 $Z = 1/t$，因此可以写出直线上点的坐标 （m/Z，$-n/Z$）。在齐次坐标中，这变为 （m，$-n$，Z）。在极限中，当 t 接近无穷大时，换句话说，当点远离原点时，Z 接近 0，点的齐次坐标变为 （m，$-n$，0）。因此，我们将 （m，$-n$，0） 定义为无穷远处点的齐次坐标，对应于直线 $nx + my = 0$ 的方向。由于欧几里得平面的任何直线都平行于穿过原点的直线，并且由于平行线在无穷远处具有相同的点，因此欧几里得平面每条线上的无限点都被赋予了齐次坐标。

cross product:

$$\begin{pmatrix} i & j & k \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$$

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

$\tilde{x} = \tilde{l}_1 \times \tilde{l}_2 = (b_1 c_2 - b_2 c_1, a_2 c_1 - a_1 c_2, a_1 b_2 - a_2 b_1)$

如果两条直线不平行，那么$\tilde{x}$的第三个分量不为零，所以可以将其归一化，归一化后就将其转为欧几里得平面

即$\tilde{x} = (\frac{b_1 c_2 - b_2 c_1}{a_1 b_2 - a_2 b_1}, \frac{a_2 c_1 - a_1 c_2}{a_1 b_2 - a_2 b_1}, 1)$

带入x,y入直线即可验证其是否是在直线上。

$$a_i x + b_i y + c_i = 0, i = 1, 2$$

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

$\tilde{l} = \tilde{x}_1 \times \tilde{x}_2 = (b_1 c_2 - b_2 c_1, a_2 c_1 - a_1 c_2, a_1 b_2 - a_2 b_1)$

只需要验证下式是否成立

$$(b_1 c_2 - b_2 c_1)\frac{a_i}{c_i} + (a_2 c_1 - a_1 c_2)\frac{b_i}{c_i} + a_1 b_2 - a_2 b_1 = 0, i = 1, 2$$

基本思路 齐次转为笛卡尔坐标系坐标，带入验证比如点在直线上，或是两点都在直线上则两点确定了一条直线。

c)You are given the following two lines:

$$l_1 = \{(x, y)^T \in \mathbb{R}^3 | x + y + 3 = 0\}$$

$$l_2 = \{(x, y)^T \in \mathbb{R}^3 | -x - 2y + 7 = 0\}$$

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?

直接求解点：(-13,10)

cross product:

(1,1,3),(-1,-2,7)

res = (7+6,-3-7,-2+1) = (13,-10,-1)

由于齐次转欧几里要让最后一个坐标为1，因此每一项除以-1，得到(-13,10,1)相同的结果

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

$$ax + by + c = 0$$
$$3x + 4y + c = 0$$

到原点距离为3距离公式：

$$\frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} = 3$$
$$|c| = 15$$

求解完毕，c = 15 或 -15

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

参考上一题，类似

normal vector :(2,5)

距离带入即可：$\frac{1}{5}$

Transformations

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

.

$$\begin{pmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

易知

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

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

$$(x_i, y_i) = ((\ x_1^i \atop x_2^i\ ), (\ y_1^i \atop y_2^i\ ))$$

Find the 2×3 translation matrix mapping x_i onto y_i which is optimal in the least square sense.

Hint: Define a cost function as

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

and find the optimal T^* which minimizes E:

$$T^* = \arg\min_T E(T)$$

You can find T^* by calculating the Jacobian J_E of E and setting it to 0^T:

$$J_E = [\frac{\partial E}{t_1}, \ldots, \frac{\partial E}{t_N}]! = 0^T$$

Can you give an intuitive explanation for the equation you derive for T^* ?

这个其实比较显然

> we need to take the derivative of E(T) with respect to the elements of *T* and set it to zero. This is because at the minimum of a function, its derivative is zero.

对应于一个坐标系结果转换到另一个坐标系结果，利用转换矩阵，最小化二乘法其实目的就是让各个转换矩阵的元素的偏导值为0即可，这个比较显然，因为在导数为0，那么就是

极值了(但是值得注意的是，这里的平移矩阵应该不是只有两个自由度的平移矩阵，而是拥有6个自由度的转换矩阵)

之前的理解是错误的，而应该是有两个自由度的平移矩阵(一直纠结这个translation matrix是不是标准形式)。如下：

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

$$E(T) = \sum_{i=1}^{N} ||Tx_i - y_i||_2^2$$

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

$$\frac{\partial E}{\partial t_1} = \sum_{i=1}^{N} 2(x_1^i + t_1 - y_1^i) = \sum_{i=1}^{N} 2(x_1^i - y_1^i) + 2Nt_1 = 0$$

求解得到

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

同理:

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

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 T^*.

上面纸质写法错误，换种写法，直接由b)得到结果:

t1 = 2, t2 = -4

Camera Projections

相关知识参考链接:一文带你搞懂相机内参外参(Intrinsics & Extrinsics) - 知乎 (zhihu.com)

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

The camera pose consists of a $90°$ rotation around the x axis and translation of (1, 0, 2) .

The focal lengths f_x, f_y are 100.

The principal point $(c_x, c_y)^T$ is $(25, 25)$.

先看旋转矩阵R，由于绕X转动，旋转矩阵R给定如下：

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & -sin(\theta) \\ 0 & sin(\theta) & cos(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

translation of (1, 0, 2) ,所以[R|t]如下：

$$[R|t] = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$$

$f_x = 100, f_y = 100, c_x = 25 \, and \, c_y = 25$

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

$$\tilde{P} = K[R|t]$$

再扩展一维[0,0,0,1]为齐次即为所求。

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

首先把 $\tilde{x}_s = (25, 50, 1, 0.25)^T$ 转换为标准的齐次形式通过除以第四维度，然后得到非齐次坐标是(100,200,4)非齐次坐标，然后通过 $\tilde{x}_s = \tilde{P}x_w \rightarrow x_w = \tilde{P}^{-1}\tilde{x}_s$ 求解，求出

上面的那个P的逆即可求得答案。

Let's perform our first projection of a geometric shape. We define C_0 as the cube centered at $c_c = (0, 0, 15)^T$ with equal side lengths s = 20.

i) Project the 8 corners of the cube to the image plane for the pinhole camera with focal lengths $f_x = f_y = 5$ and the principal point $(c_x, c_y)^T = (10, 10)^T$. Draw the projected points
and edges of the cube in a coordinate system on a paper.

给出那八个点吧

(±10,±10,25)^T 和 (±10,±10,5)^T

然后由 $f_x = f_y = 5$ and the principal point $(c_x, c_y)^T = (10, 10)^T$ 容易获得内参矩阵，即可求得。

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 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.

类似(1)，只不过8个顶点坐标向上移一下，然后内参矩阵改变一下也十分容易。

iii) Let's move it even further away along the z-axis while zooming in. Now, the distance of the
center of cube 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.

类似(1)

iv) Project the 8 corners of the first cube C_0 using an orthographic projection and add the principal
point $c_c = (10, 10)^T$ 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.

正交投影较为容易，就让8个角的坐标的z值取0就好了，然后画在一张图上，再加上 $c_c = (10, 10)^T$ 的principal point，即每个x y都加上10即可。

v) When is the perspective projection most similar to the orthographic projection?

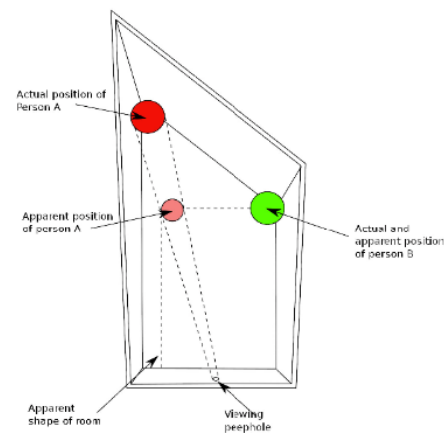当物体远离相机时，透视投影与正交投影最相似。 这是因为，当物体远离具有透视投影的相机时，它们在屏幕上的尺寸减小的速度比它们靠近时的速度要慢。 此效果类似于正交投影，其中对象保持恒定大小，无论距它的距离如何。

换句话说，物体离相机越远，透视畸变的影响就越不明显，透视投影就越类似于正交投影。

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.

剪纸？ 看简单介绍是一个视觉与实际不匹配的视觉错位设计。

(a) Photo of a real-world an Ames Room   (b) Schematic drawing of an Ames Room

Figure 1: **Ames Room.** The ames room creates an optical illusion for the viewer (see Fig. 1a). When looking through the peephole into the room, the room itself appears to be an ordinary cuboid despite its actual distorted shape (see Fig. 1b). As a result, objects in the distorted corner look smaller than they actually are. This is achieved by designing the room in such a way that crucial information for the geometric understanding of the real room are lost when looking through the peephole. It serves as an example of how different 3D scenes (here the perceived cuboid room and the real distorted room) can be projected to the same image when using a perspective projection.

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 = 102mm.

这个比较容易，写出薄透镜成像基本基本公式代入即可。

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

代入zs,f即可。

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.

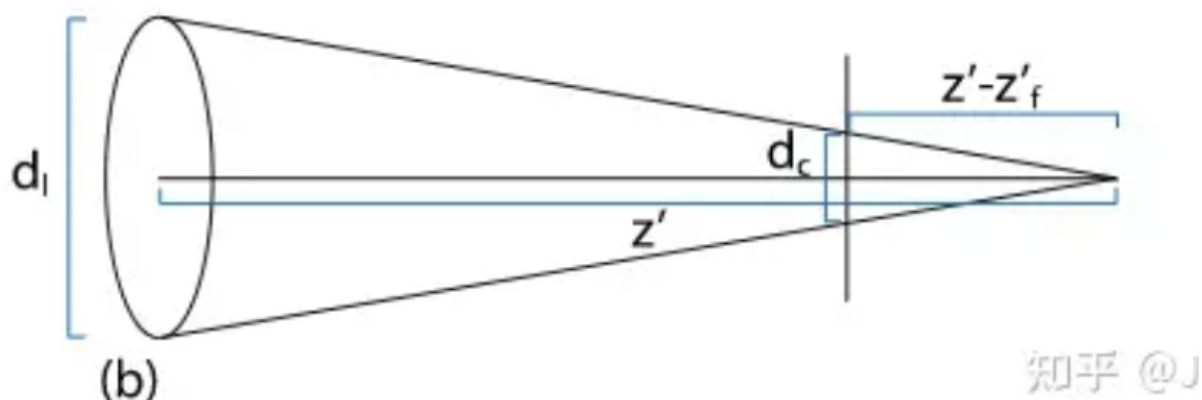对于the circle of confusion的理解

在摄影学中，焦平面是指相机镜头的成像平面，而焦深度是指在这个平面上，物体的前后焦点之间的距离。当一个物体在焦平面之前或之后时，它的图像就会变得模糊。这种模糊是由于光线不再汇聚到一个点上，而是散布到焦平面上的一个圆形区域内。这个圆形区域就被称为**混淆圆**（circle of confusion）。

当一个三维点投影到二维图像上时，它会被成像为一个二维点。然而，在实际情况下，由于光线的散射和透镜的缺陷等因素，这个二维点实际上是一个圆形区域。这个圆形区域就是混淆圆。如果这个圆形区域越小，那么图像就越清晰；如果这个圆形区域越大，那么图像就越模糊。

因此，当我们说一个物体在焦平面之前或之后时，实际上是指它的投影点落在了混淆圆之外。如果混淆圆的半径小于一定阈值，我们就认为这个物体是清晰的；否则我们就认为它是模糊的。

参考链接:6.2 投影相机模型 - 知乎 (zhihu.com)

类似下面这个图:

(b)

其中c是d_c ，之后几个变量分别对应即可。

得到几何关系:

$$\frac{c}{\Delta z_s} = \frac{d}{z_s}$$
$$N = \frac{f}{d}$$
$$c = \frac{f}{N z_s} \Delta z_s$$

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 Δz_s = 0.1mm as well as when Δz_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?

通过b)公式算c，然后比较c与每个像素(正方形)的边长即可。

64mm^2 400 × 400, 对应每个像素的面积 → 0.0004 mm^2 边长 0.02mm

然后分别带入 Δz_s = 0.1mm和 Δz_s = 0.03mm计算c 与0.02比较，需要比0.02小才能够保证不会模糊。