胡说八道——相机模型

小孔相机模型

小孔相机是最简单的相机模型。在这种模型下，相机是将空间中点$X = (x, y, z, 1)^T$变换为相片上点$p = (p_x, p_y, 1)$的装置（用齐次坐标），小孔相机模型认为$(x, y, z, 1)^T \longmapsto (p_x, p_y, 1)$是一种线性变换，也就是如下：

$$\begin{aligned} \lambda p_x &= h_{11}x + h_{12}y + h_{13}z + h_{14} \\ \lambda p_y &= h_{21}x + h_{22}y + h_{23}z + h_{24} \\ \lambda &= h_{31}x + h_{32}y + h_{33}z + h_{34} \end{aligned} \tag{1}$$

或者

$$\lambda p_{3\times1} = H_{3\times4}X_{4\times1} \tag{2}$$

即

$$\lambda \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3}$$

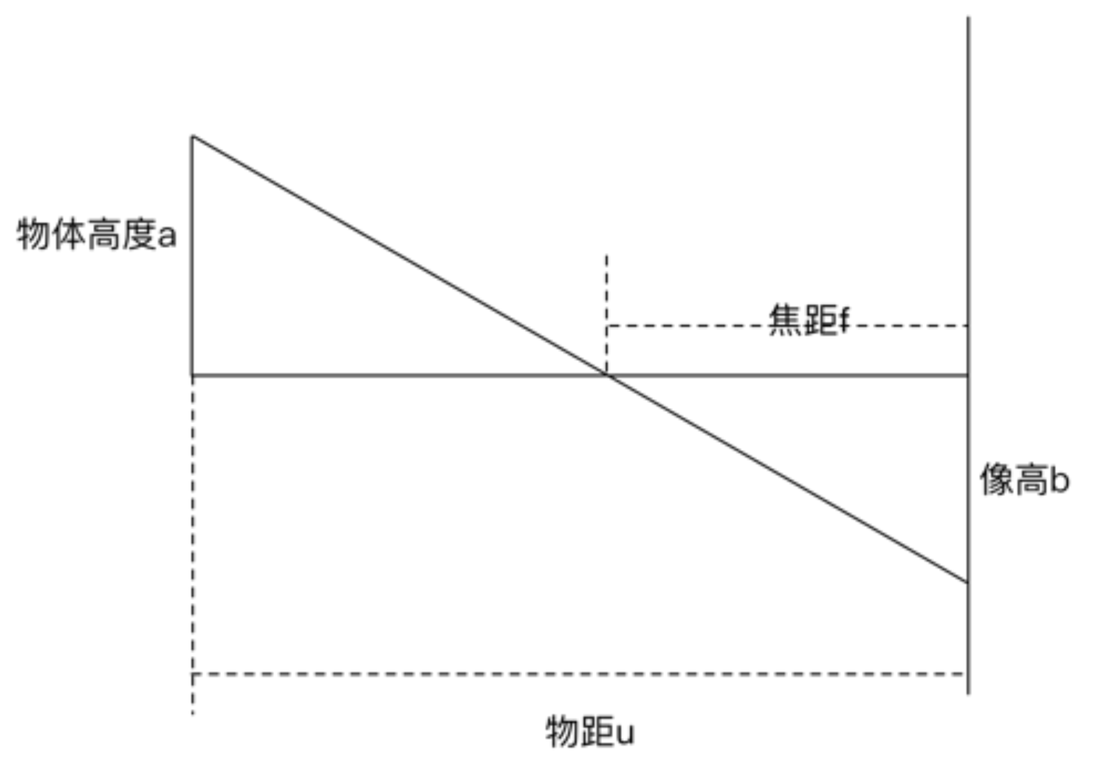
相机镜头是透镜成像，根据物理学公式$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$，相机显然不是线性的。现代相机镜头还由多个镜片构成，更加不是线性的。所以认为相机是一种线性变换，不过是一种简化。**搞计算机的毕竟不是搞物理的，高级的模型还是得物理学家来提炼，能折腾最简单小孔成像模型算不错的啦。**

H是个什么鬼

$\lambda p = HX$中λ不重要（因为是齐次坐标），假设p和X是已知量，H又是什么？根据矩阵乘法的性质，我们知道H是一个3×4的矩阵，但是内容呢？显然H的值与相机有关。与相机的哪些有关？至少是这些因素：

1. 相机的位置
2. 相机的方向
3. 相片的大小
4. 相机镜头的焦距

换言之，只要知道了以上4个信息，H就可以推导出来。很难吗？其实一点不难，如果做一些条件限制，这个问题只是一个初等数学题（中学物理水平）。例如如下的题目应该是见过的：



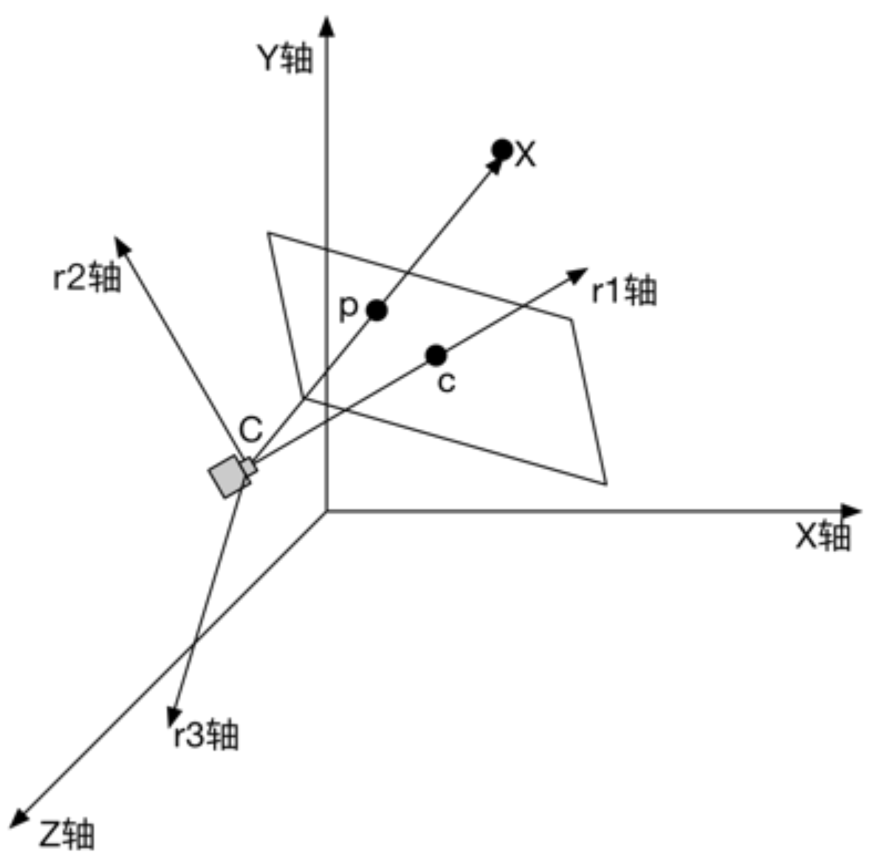
由图可知

$$b = \frac{af}{Z - f}$$

类似可以求得水平方向的像宽，进而求得物体坐标与相片坐标间的转换关系。因此H完全可以靠几何法导出（利用相似三角形的性质）。

H的代数法推导

数学家笛卡尔的一大贡献就是建立了解析几何。解析几何为什么重要？因为它在几何系统中引入了坐标系的概念，从而将代数方法带到几何领域。没有坐标系的几何，就是传统的欧几里德空间几何。在传统的几何学中，证明平行、全等、相似是不能用直尺和量角器的，只允许用公理、定理来推导。记得我读中学的时候，看到几何证明题就感到很困惑，明明用直尺量一下就可以解决的问题，为什么要搞那么复杂。后来才知道，原来古希腊**没有带刻度的直尺**。
解析几何引入坐标系，相当于加上了直尺。有了直尺，很多问题就不再依赖于定理推导，而是通过计算得出结论。谢天谢地，解析几何中判断两个图形面积相等，可以直接把面积算出来判断啦。解析几何的另外一大优势，就是可以利用高等代数的性质来求解几何问题。几何和代数关系一通，相当于和其他数学领域的关系也通了。**好处是几何问题变得简单，坏处是几何问题的解法让人看不懂（因为使用的不再是单纯的几何方法）**。
对于相机，可以建立如下的坐标系：



如上图所示，相机位于空间中$C = (C_x, C_y, C_z)$处，相机的方向由3个方向轴r_1、r_2、r_3确定，空间点$X = (x, y, z)^T$投影在相片上$p = (p_x, p_y)^T$处，$c = (c_x, c_y)^T$是相片的中心。此处省略若干推导过程，有兴趣的可以参考其他书籍（推荐Multiple View Geometry in Computer Vision 2nd第153页的推导）。最终可以得到如下的公式：

$$\lambda p = \lambda \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = K_{3\times3} \begin{bmatrix} R_{3\times3} & | & t_{3\times1} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = K_{3\times3}P_{3\times4}X \tag{4}$$

其中$R_{3\times3}$为旋转矩阵，$t_{3\times1}$平移矩阵，容易证明：

$$R_{3\times3} = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \tag{5}$$

$$t_{3\times1} = -R_{3\times3}C \tag{6}$$

$$K_{3\times3} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$P_{3\times4}$称为相机的外参（Extrinsic Parameters），表示点X从XYZ坐标系向相机坐标系的变换。其中r_1、r_2、r_3是表示相机方向的3×1**列向量**，并且**两两之间正交**。$t_{3\times1}$是在相机坐标系下像平面相对于C的位移量。$P_{3\times4}X$相当于先旋转R再移动t。

$K_{3\times3}$称为相机的内参（Intrinsic Parameters），f_x和f_y是相机x和y方向的焦距（通常可以认为$f_x = f_y$），c_x和c_y是相片的中心位置，s一般设置为0。$K_{3\times3}$故此一般可以简写为：

$$K_{3\times3} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$K_{3\times3}P_{3\times4}X$相当于将旋转移动后的坐标（相对于相机坐标系）变换到**相片坐标系**。

到此我们知道用于变换的$H_{3\times4} = K_{3\times3}P_{3\times4}$。

相机标定（Calibration）

计算$K_{3\times3}$的过程被称为相机标定（Calibration）。相机标定的做法很多，最靠谱的其实是光学标定法，也就是由相机厂商用光学方法给出K值。不过这种方法不怎么可行，因为相机厂商除非闲的没事不会干这种事。那么也就自己来做啦！可惜那么昂贵的光学仪器不是每个人都搞得定的，于是就有了廉价的方法——用算法来标定。OpenCV采用**张正友**的算法，比较简单好用。还有一种常见的方法是直接用图像文件的元数据（Exif）来计算，焦距转换一下单位（默认是英寸，需要根据dpi信息转换为像素），c_x和c_y就直接用相片宽度和高度除以2。在精度要求不高的情况下，这两种方法的效果差距不大（当然早期的图片可能不带Exif信息）。总之，这件事情现在不难。

相机位姿估计（Camera Pose Estimation）

$P_{3\times4}$包含相机的位置和姿态，所以称为位姿，有些书也称为Camera Motion。如果在拍摄前就已经设置好了相机的位姿，那么$P_{3\times4}$就很容易计算。比如把相机固定在三脚架上，三脚架上面可以直接读出相机的角度和位置信息。麻烦一些的情况是事先不知道物理位置。这个时候只能靠算法来估计。最常见的方法是**运动到结构（Structure From Motion）**，这个主题太大，本文就不讲啦。另一种方法是已知图像上某些物体的位置，逆推相机的位姿。

总结

相机模型是计算机视觉中比较重要的基础知识。小孔相机模型的数学原理不复杂，仅仅是中学水平，但是由于用了比较高级的矩阵表达，有些让人不太习惯。今天相机模型仍然有研究价值，但是小孔相机模型已经是非常初级的。多数情况下，只需要记住公式：

$$\lambda p = K_{3\times3}P_{3\times4}X \tag{4}$$

然后知道K和P如何求解就可以了。至于这个公式的推导，不知道也不影响啦。