

双目立体视觉（Binocular Stereo Vision）是机器视觉的一种重要形式，它是基于视差原理并利用成像设备从不同的位置获取被测物体的两幅图像，通过计算图像对应点间的位置偏差，来获取物体三维几何信息的方法。



伪激光雷达-双目立体相机

深度学习和计算机视觉在自动驾驶系统中已经非常流行且被广泛应用。计算机视觉领域在过去的十年里得到了迅猛的发展，特别是在障碍物检测方面。障碍物检测算法，如YOLO或RetinaNet提供了二维边界框，用边界框给出了障碍物在图像中的位置。

目前，大多数的目标检测算法都是基于单目RGB摄像机的，不能返回每个障碍物的距离。为了能够返回每个障碍物的距离，工程师们将相机与激光雷达（LiDAR，光探测和测距）传感器进行融合，

后者使用激光来返回深度信息。将计算机视觉信息和激光雷达输出进行传感器的融合。这种方法的问题是使用激光雷达，就会导致价格昂贵。所以经常有人使用的一个双目摄像头进行替代，并使用几何信息来定义每个障碍物的距离，故可以将双目相机获取的数据称之为伪激光雷达。



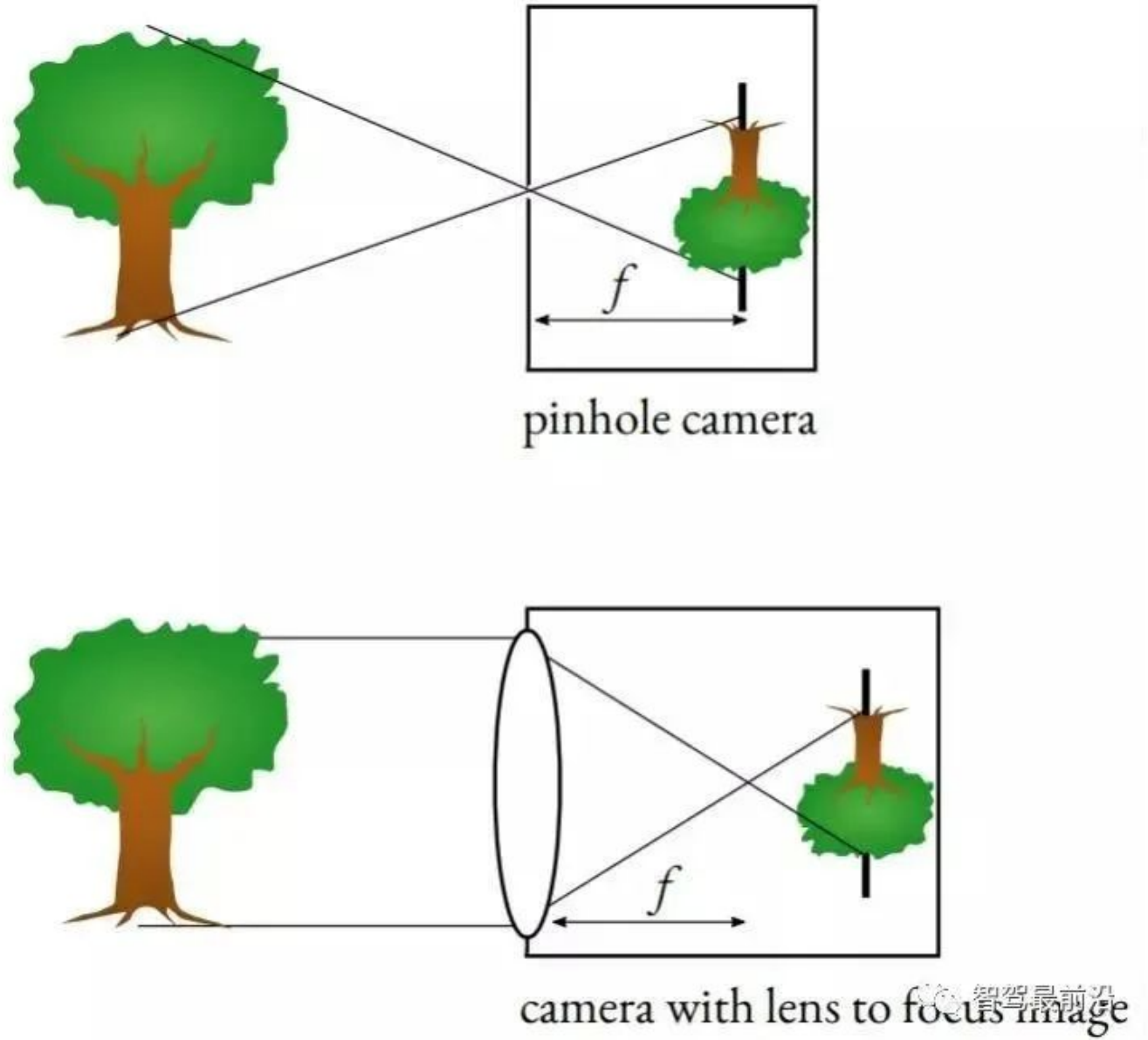
单目与双目的比较

双目视觉利用几何学来构建深度图，并将其与目标检测相结合以获得三维距离。那么如何利用立体视觉实现距离估计？以下是双目障碍物检测的5步伪代码：

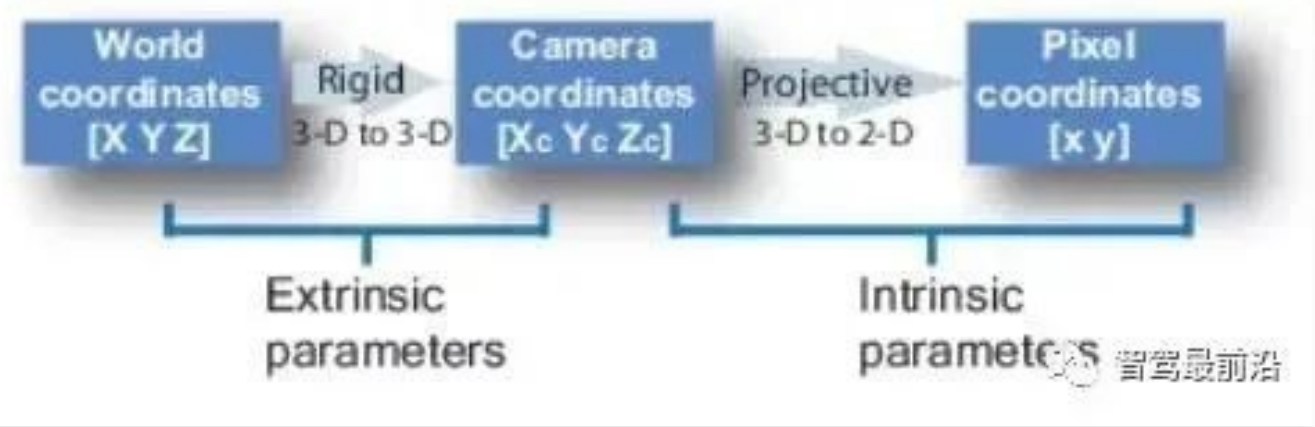
- 标定 2 个摄像头（内外参的标定）
- 创建极线约束
- 先构建视差图，然后构建深度图
- 然后将深度图与障碍物检测算法相结合
- 估计边界框内像素的深度。

相机内外参标定

每个摄像机都需要标定。相机的标定是指将三维世界中的 $[X, Y, Z]$ 坐标的三维点转换为具有 $[X, Y]$ 坐标的二维像素。这里简单的介绍一下针孔相机模型。顾名思义就是用一个针孔让少量光线穿过相机，从而得到清晰的图像。

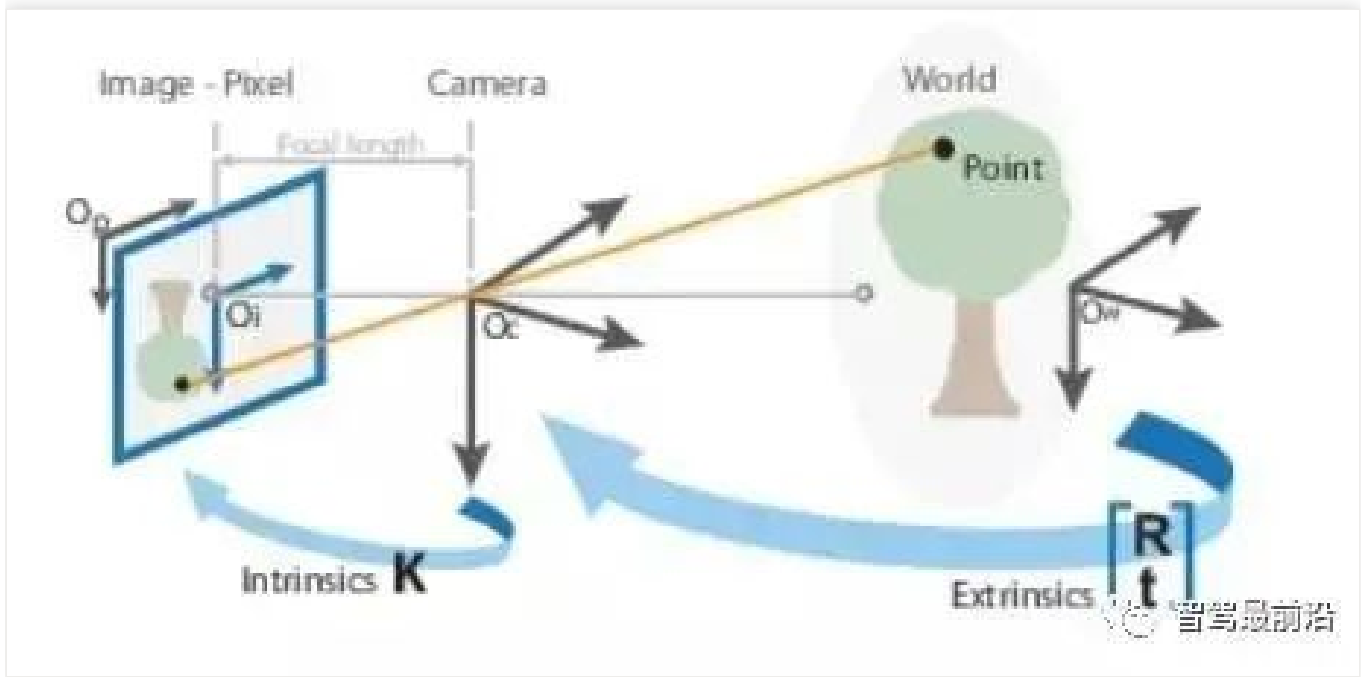


针孔相机模型可以设置焦距，使得图像更加的清晰。为了相机标定，我们需要通过摄像机坐标系计算世界坐标点到像素坐标的变换关系。

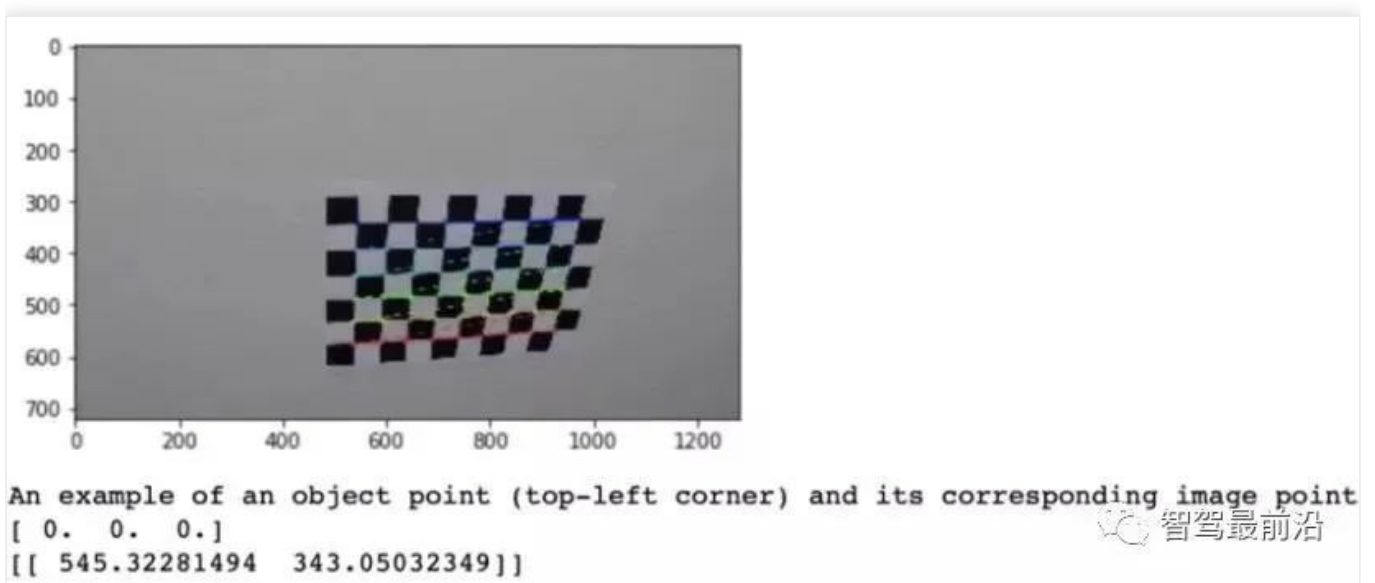


相机标定过程

- 从世界坐标系到相机坐标的转换称为外参标定，外部参数称为R（旋转矩阵）和T（平移矩阵）。
- 从摄像机坐标到像素坐标的转换称为内参标定，它获取的是相机的内部参数，如焦距、光心等...
- 内参我们常称之为K的矩阵。

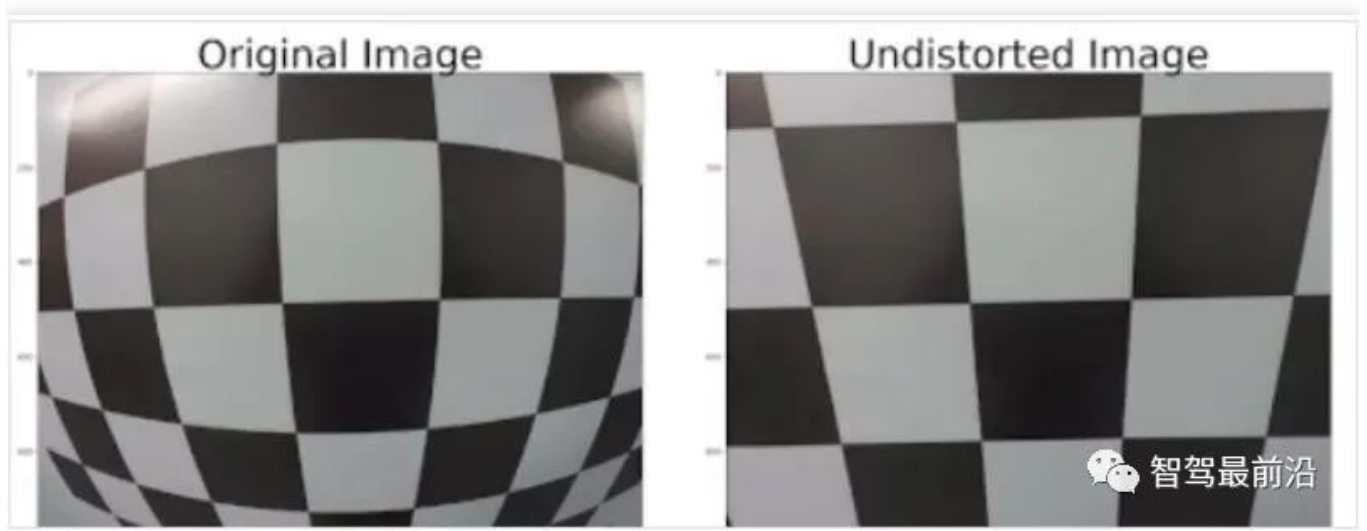


内参标定，通常使用棋盘和自动算法获得，如下图我们在采集标定板时，将告诉算法棋盘上的一个点（例如世界坐标系点 0, 0, 0）对应于图像中的一个像素为（545, 343）。



为此，相机标定必须用摄像机拍摄棋盘格的图像，在得到一些图像和对应的点之后，标定算法将通过最小化平方误差来确定摄像机的标定矩阵。得到标定参数后为了得到校正后的图像，需要进行畸

变校正。畸变可以是径向的，也可以是切向的。畸变校正有助于消除图像失真。



图像的畸变校正

以下是摄像机标定返回的矩阵形式

$$\begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

f 是焦距- (u_0, v_0) 是光学中心：这些是固有参数。

我认为每一个计算机视觉工程师都应该必须知道并掌握相机的标定，这是最基本且重要的要求。

在相机标定的过程中涉及到一些齐次坐标转换的问题，这里简单的介绍一下有两个公式可以得到从世界坐标系到像素坐标系的关系：

(1) 世界坐标系到相机坐标系的转换（外参标定公式）

$$o_{camera} = [R|t]O_{world}$$

(2) 相机坐标系到图像坐标系的转换（内参标定公式）

$$o_{image} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} o_{camera} = K o_{camera}$$

所以从三维空间坐标系到图像坐标系下的关系可以总结为

$$o_{image} = PO = K[R|t]O_{world}$$

但是我们知道这个公式是齐次坐标才可以这么写，也就是需要将 O_{world} 从 $[X \ Y \ Z]$ 修改为 $[X \ Y \ Z \ 1]$ ，加这个“1”后称为齐次坐标。

$$o_{image} = PO = \underbrace{K[R|t]}_{3 \times 4} O_{world} \rightarrow \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

3x3 3x4 3x1

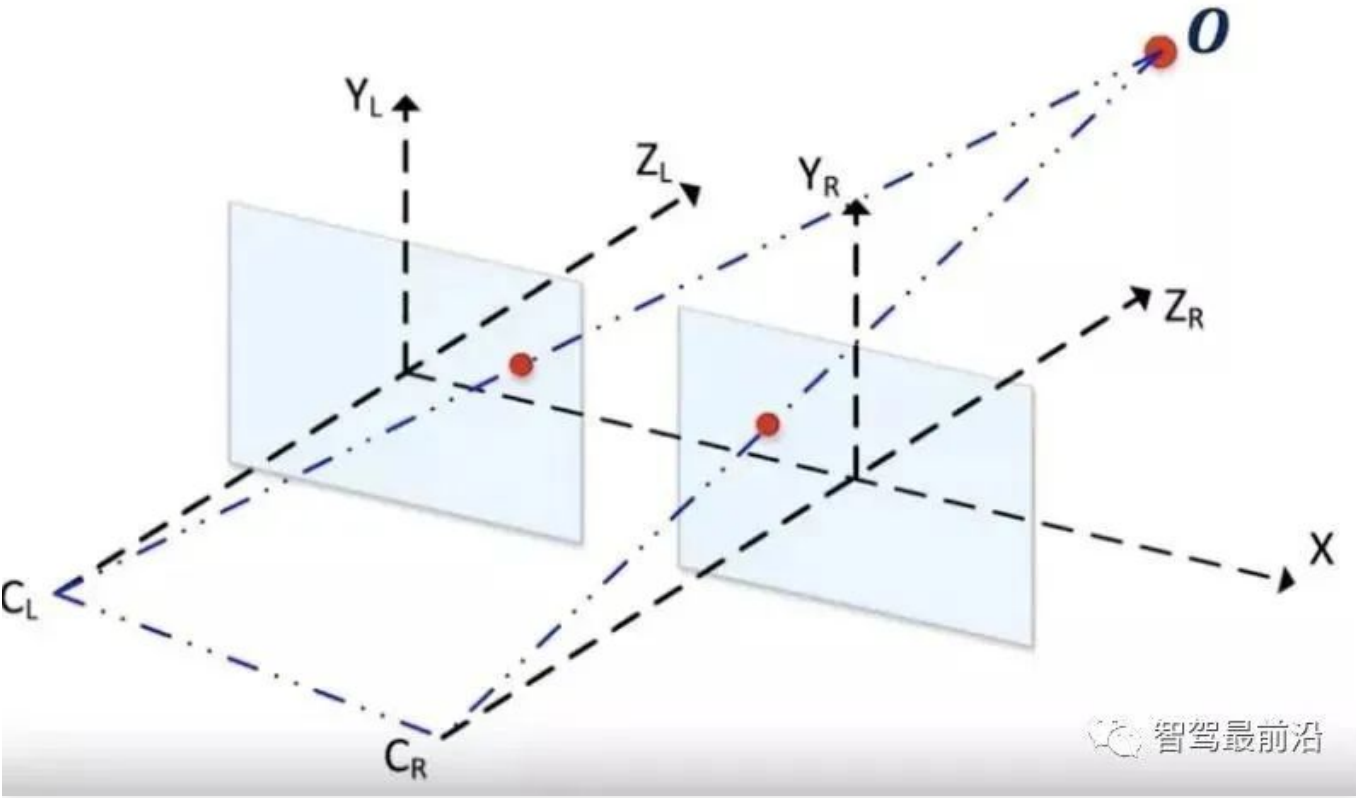
3x4

双目视觉的对极几何

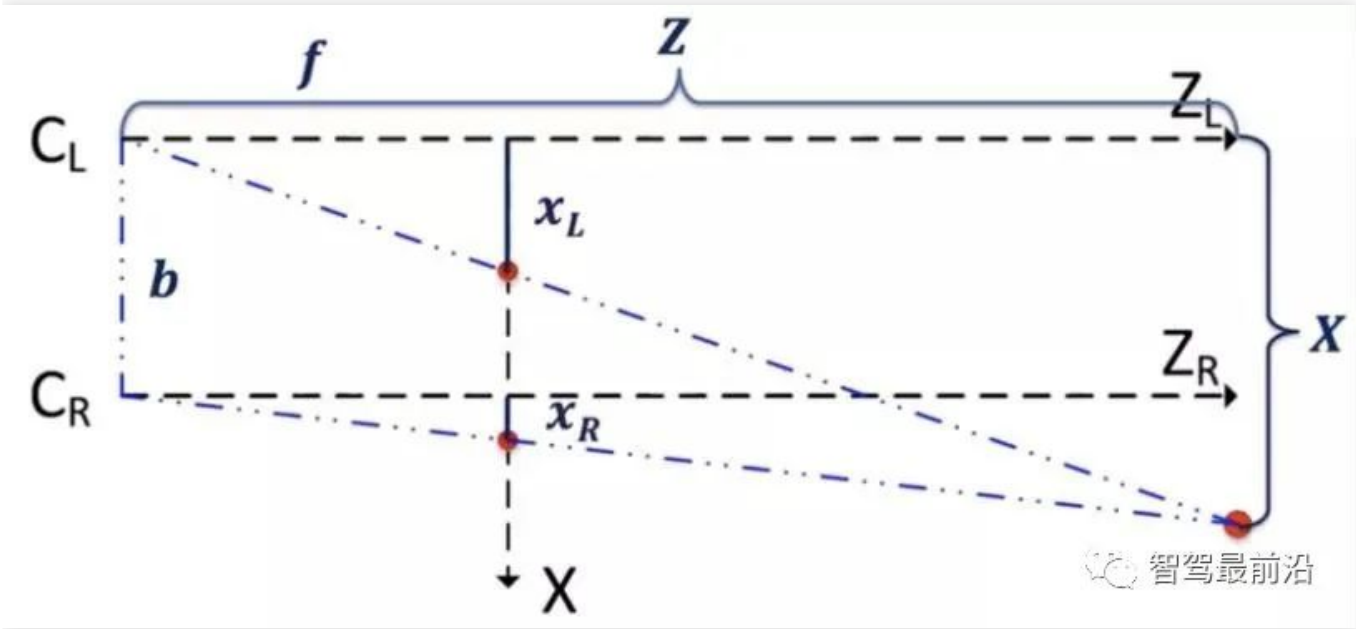
我们知道双目立体视觉是基于两幅图像来寻找深度的，人类的眼睛就像两个相机，因为两只眼睛从不同的角度观察图像，所以他们可以计算两个视角之间的差异，并建立距离估计。下图是一个双目立体相机的例子



那么我们如何根据双目立体相机如何估计深度？想象一下你有两个摄像头，一个左摄像头和一个右摄像头。这两个摄像头在同一Y轴和Z轴上对齐。那么唯一的区别是它们的X值。



根据上图我们的目标是估计O点（代表图像中的任何像素）的Z值，即距离。X是对齐轴，Y是高度值，Z是深度值，两个蓝色的平面图对应于每个摄像头的图像。假设我们从从俯视的角度来考虑这个问题。

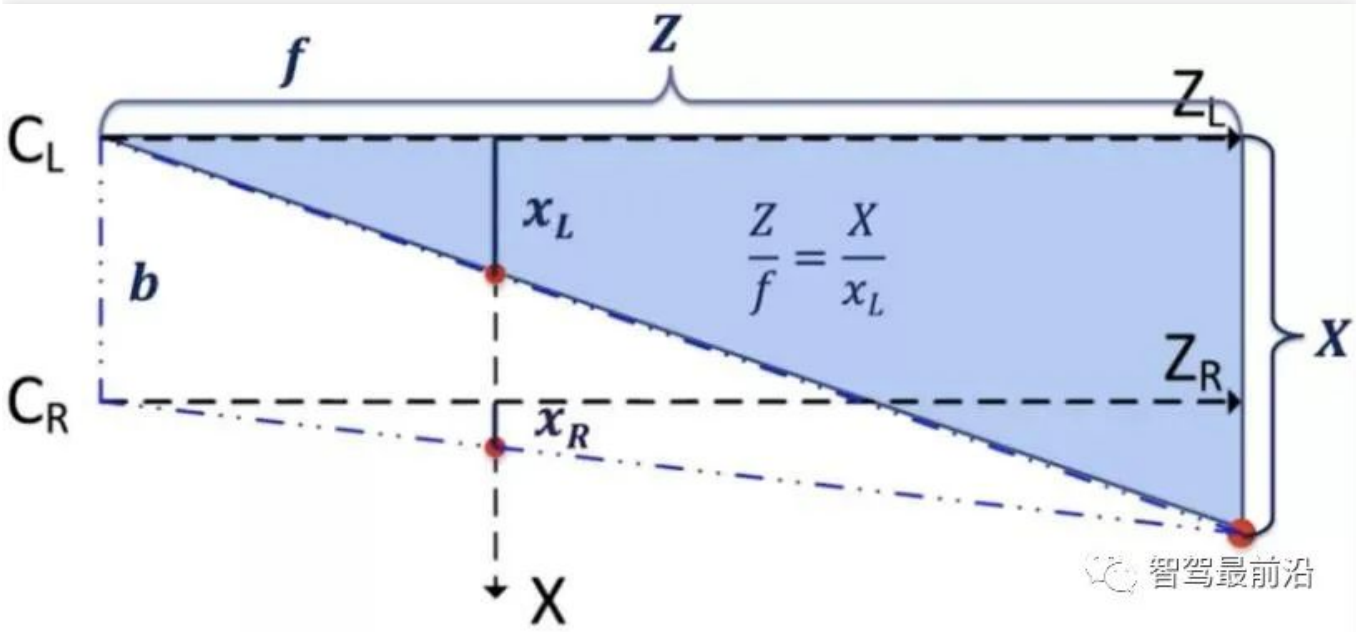


已知：

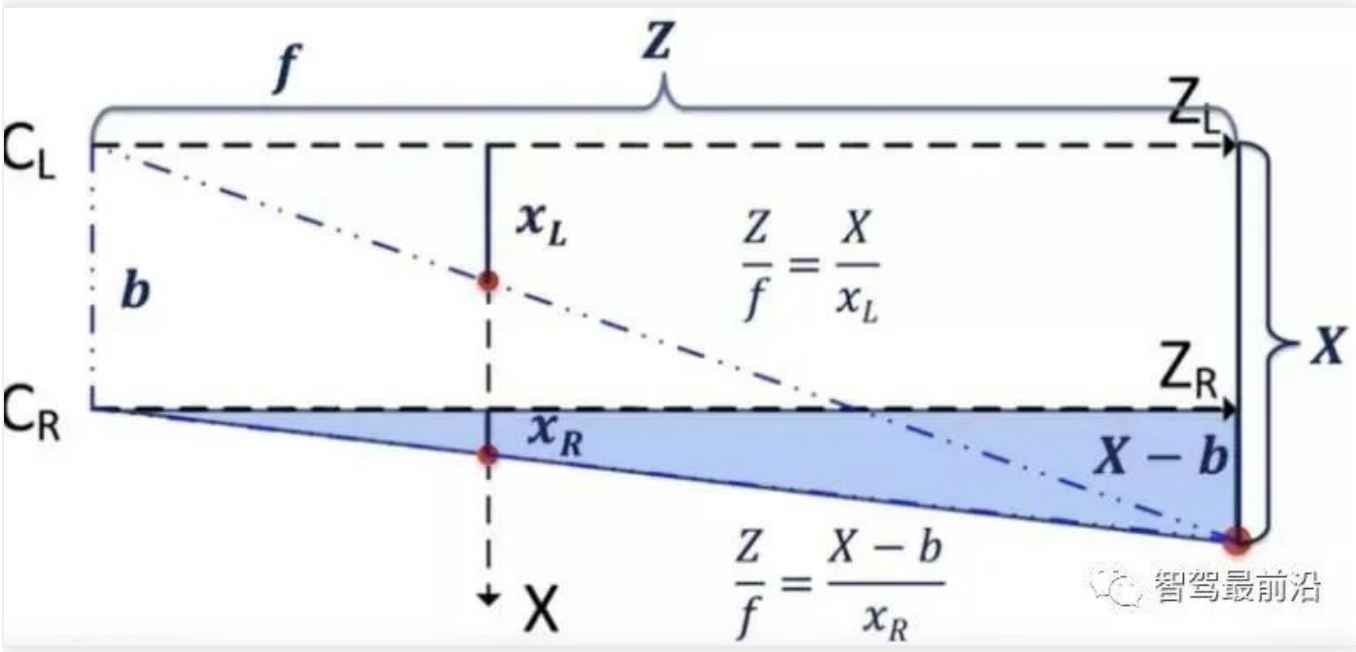
- (1) x_L 对应于左侧相机图像中的点。 x_R 是与左侧图像中该点的对应位置。

(2) b 是基线，是两个摄像头之间的距离。

针对左相机，如下图，我们可以得到一个公式： $Z = X \cdot f / x_L$ 。



针对右相机，如下图，我们可以得到另一个公式： $Z = (X - b) \cdot f / x_R$ 。



此时根据两个公式我们可以计算出正确的视差 $d = x_L - x_R$ 和一个物体的正确XYZ位置。

视差和深度图

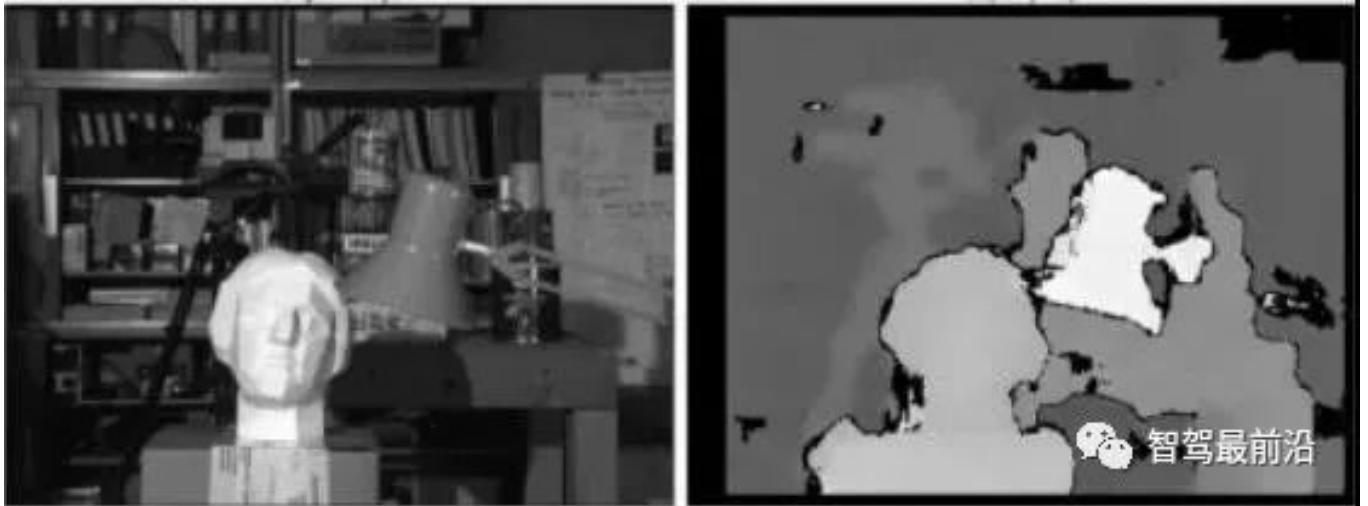
什么是视差？视差是指同一个三维点在两个不同的摄像机角度获得的图像中位置的差异。视差图是指一对立体图像之间明显的像素差异或运动。要体验这一点，试着闭上你的一只眼睛，然后快速地

闭上它，同时打开另一只眼睛。离你很近的物体看起来会跳一段很长的距离，而离你较远的物体移动很少，这种运动就是视差。

$$Z = \frac{fb}{x_L - x_R} = \frac{fb}{d},$$

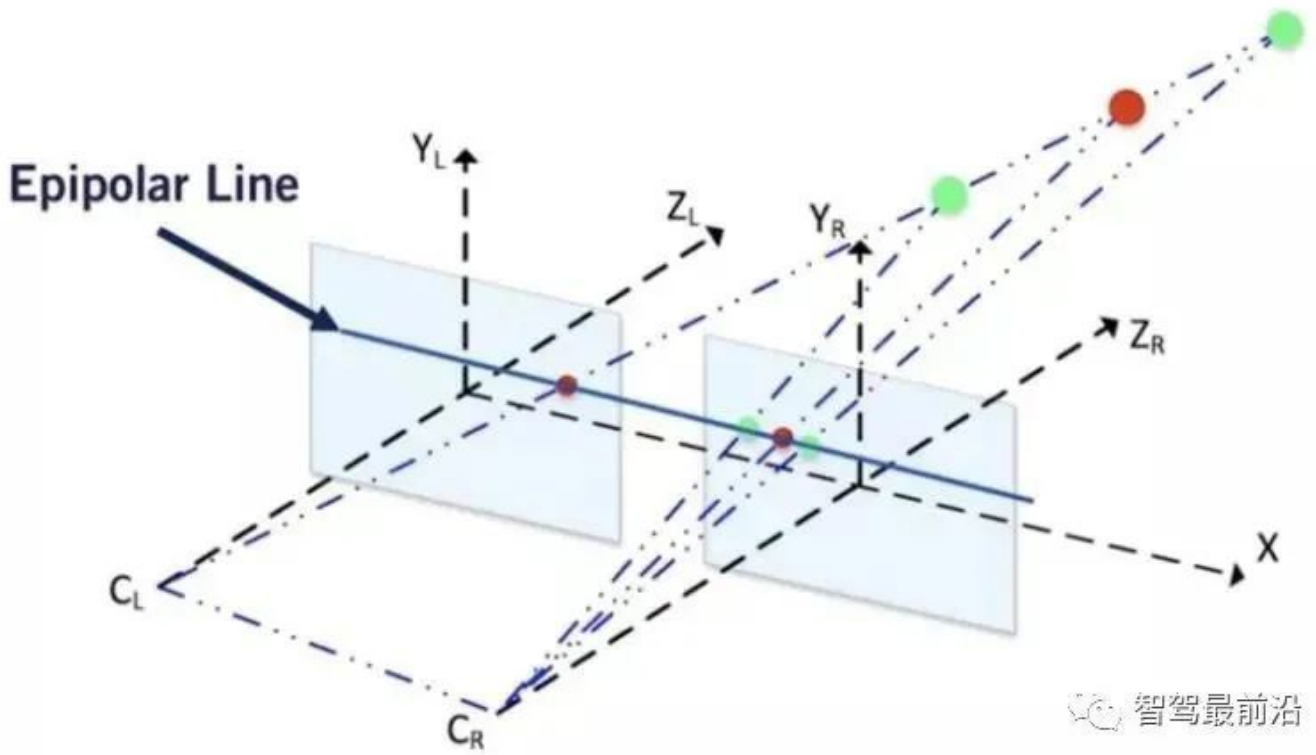
$$X = \frac{Zx_L}{f}, \quad Y = \frac{Zy_L}{f}$$

由于立体视觉，我们可以估计任何物体的深度，假设我们得到了正确的矩阵参数，则可以计算深度图或视差图：



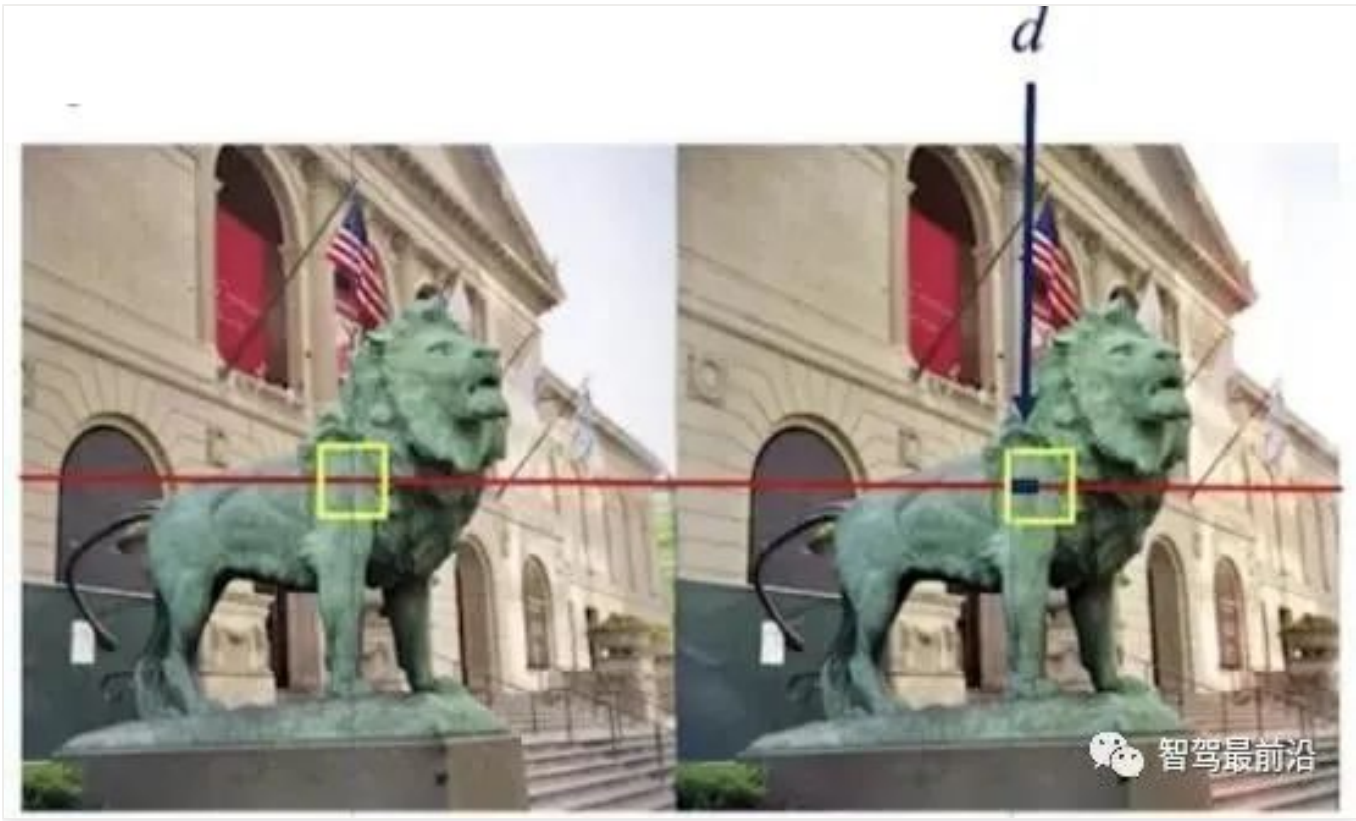
视差图

为了计算视差，我们必须从左边的图像中找到每个像素，并将其与右边图像中的每个像素进行匹配。这就是所谓的双目相机的立体匹配的问题。为了解决像素匹配的问题，引入对极几何约束，只需在对极线上搜索它，就不需要二维搜索，对应点一定是位于这条线上，搜索范围缩小到一维。



之所以能够引入对极约束，这是因为两个相机是沿同一轴对齐的。以下是极线搜索的工作原理：

- 取左图中这一行上的每个像素
- 在同一极线上比较左图像像素和右图像中的每个像素
- 选择 cost 最低的像素
- 计算视差 d

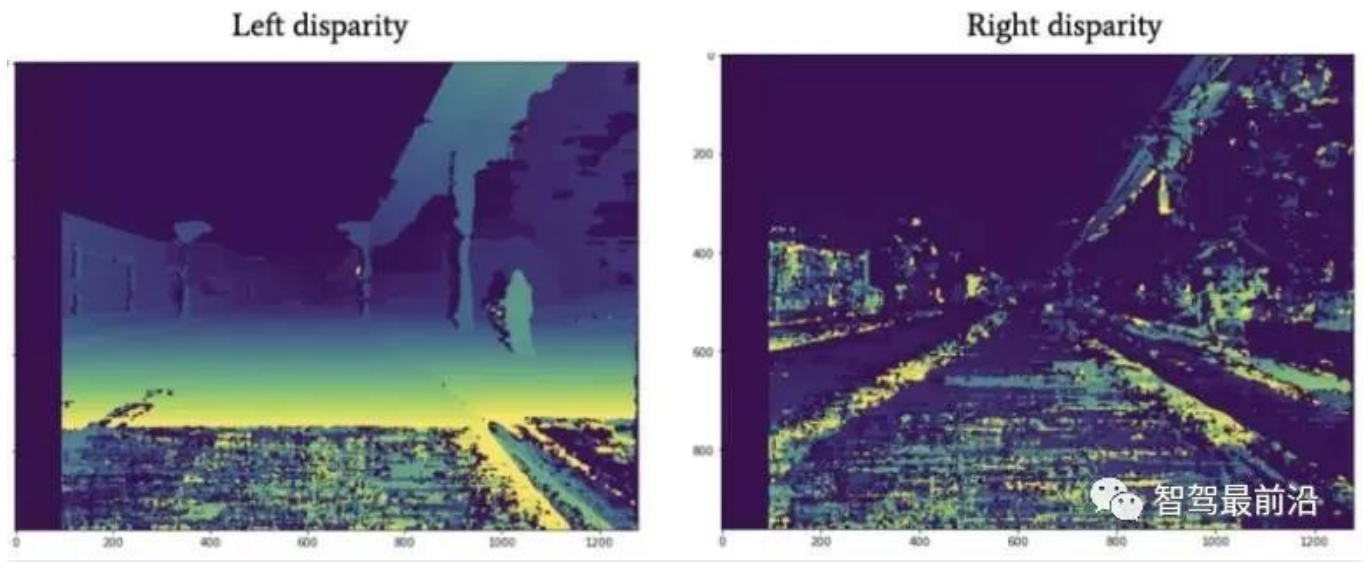


构建伪激光雷达效果

现在，是时候把这些应用到一个真实的场景中，看看我们如何使用双目立体视觉来估计物体的深度。假设我们有以下两张实际场景下的图片，并且我们我们已经获取了双目相机的外参矩阵。



此时我们计算视差图的步骤。将投影矩阵分解为摄像机内参矩阵 K ，以及外参 R, t 。使用我们在前面介绍的两个步骤估计深度，将得到左右图像的视差图。



从视差再到深度图的计算，由于我们有两个视差图，它基本上告诉我们两个图像之间像素的偏移。对于每台相机，我们都有一个投影矩阵 P_{left} 和 P_{right}

为了估计深度，我们需要利用投影矩阵 P 来估计 K ， R 和 t 。

一个名为`cv2.decomposeProjectMatrix()`的OpenCV函数可以这样做，将从 P 得到 K 、 R 和 t ；

现在是时候生成深度图了。深度图将使用另一幅图像和视差图来告诉我们该图像中每个像素的距离。

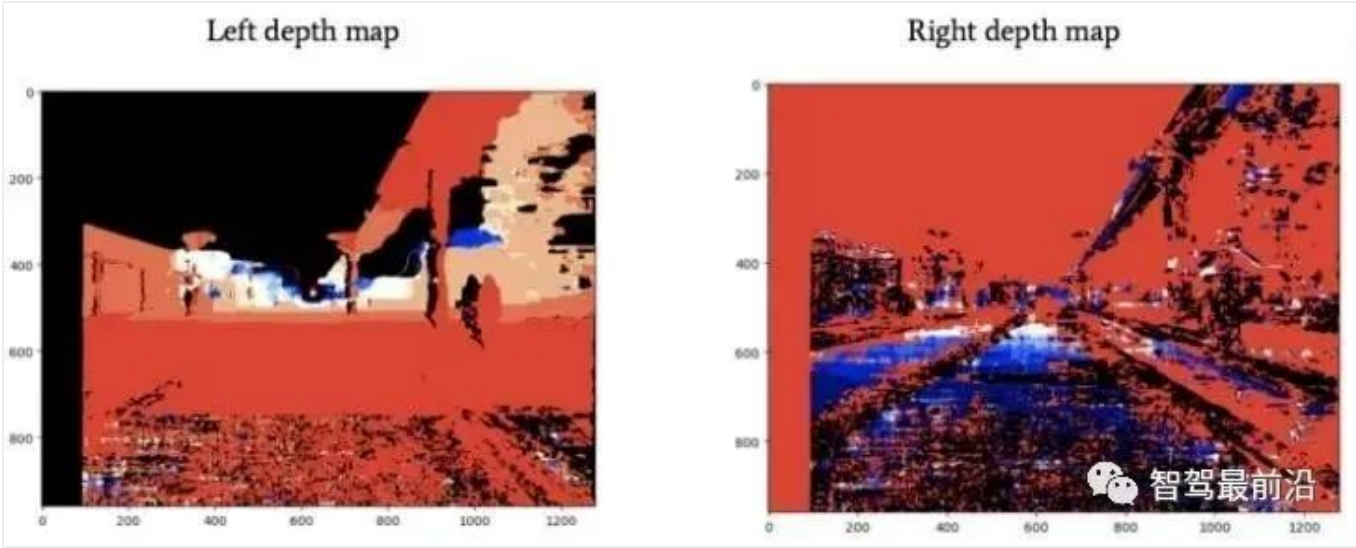
过程如下：

- 获取焦距 f 从 K 矩阵
- 计算基线 b 使用平移向量中的相应值 t

使用之前的公式和计算的视差图 d 计算图像的深度图：

$$Z = \frac{fb}{x_L - x_R} = \frac{fb}{d}$$

我们对每个像素进行计算，得到如下



估计障碍物的深度

现在有一个每个摄像头的深度图！假设我们将其与障碍物检测算法（如YOLO）相结合。对于每个障碍物，该算法将返回一个包含4个数字的边界框： $[x1; y1; x2; y2]$ 。这些数字表示框的左上点和右下点的坐标。假设我们在左边的图像上运行这个算法，并且使用左边的深度图。

如下图在这个边界框中，我们可以取最近的点。知道这一点后，可以通过深度图知道了图像中每个点的距离，而且边界框中的第一个点也就是相机到障碍物的距离。



双目测距实例

所以利用立体视觉，我们不仅知道图像中的障碍物，还知道障碍物与我们的距离！这个障碍物离我们28.927米远！

与使用激光雷达相比，它保持相对便宜的价格，并且仍然提供出色的性能。我们称之为“伪激光雷达”，因为它可以取代激光雷达的功能：探测障碍物，分类，以及三维视觉定位等领域。