

Work Review on SLAM

项目目标

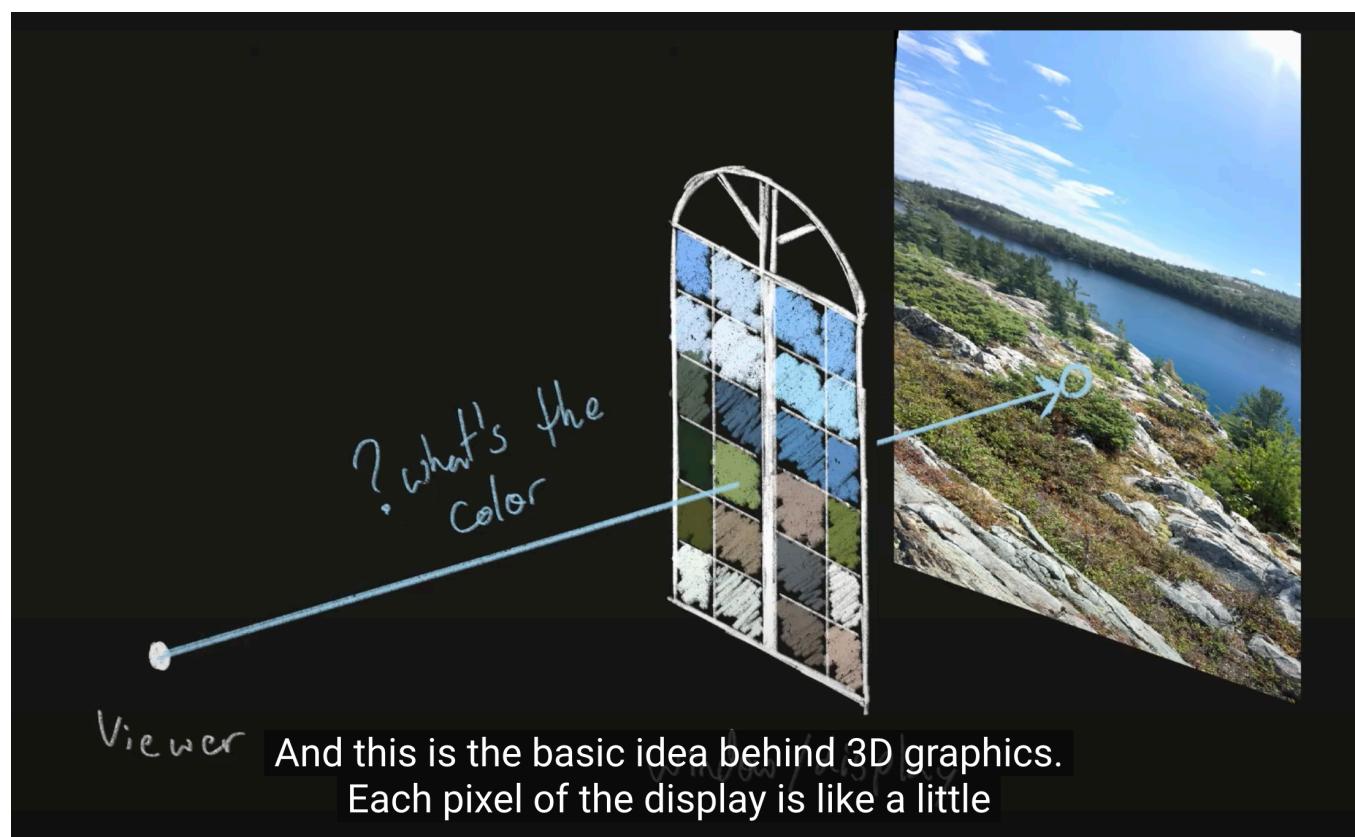
能实时地基于四目cv成像与三个imu数据，对手柄的空间位姿进行测算。

理论基础

单目摄像头能干什么

在只有单个摄像头的情况下，由于小孔成像的基本原理，我们无法准确感知物体的实际大小与相对摄像头的距离。

大多数情况下，我们之所以能根据照片想象出实际场景，实际上是因为我们具有一定的常识，以及大脑负责了对计算机来说最困难的部分：将平面图片逆向还原成三维空间。[\[1\]](#)

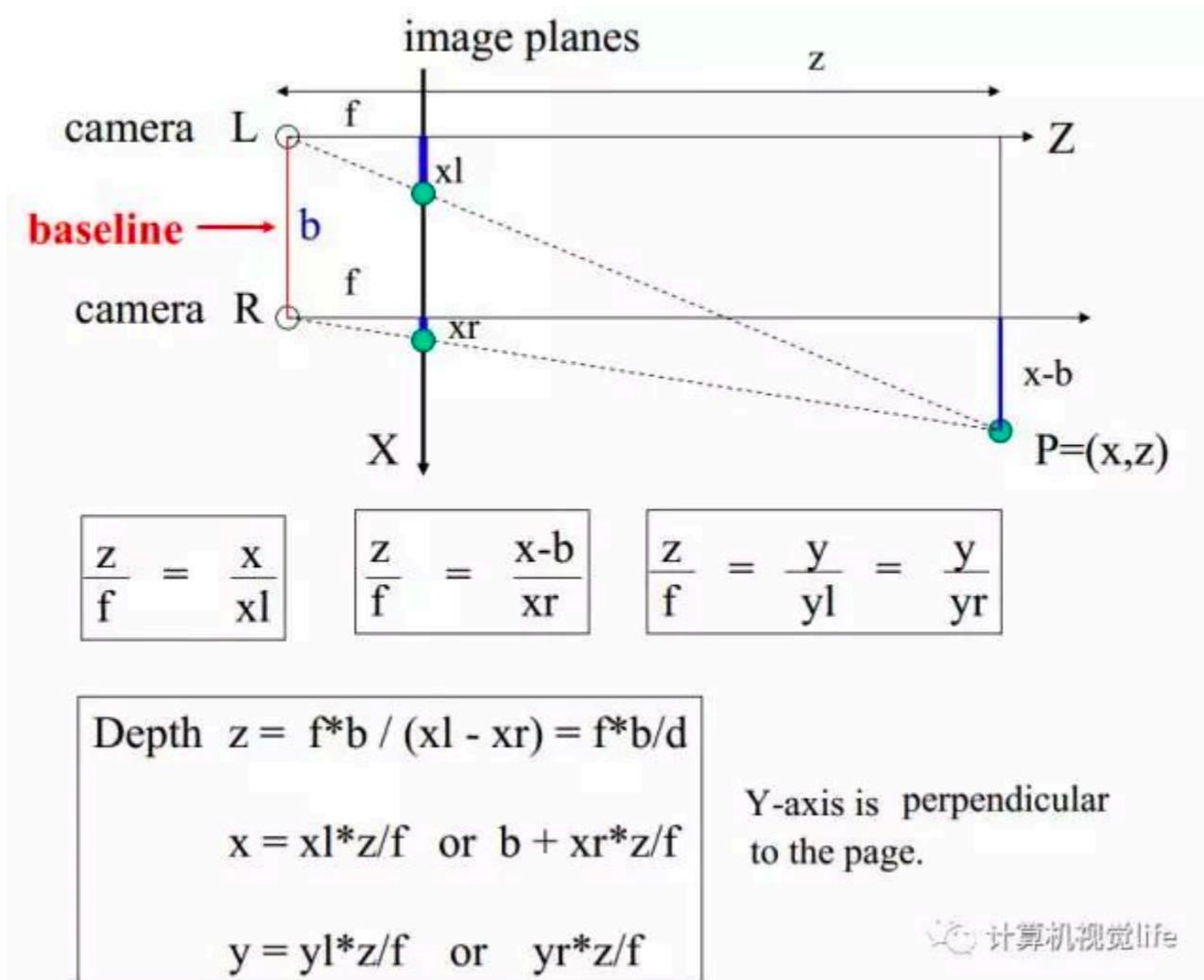


因此，单目摄像头缺少了深度信息，仅有单目成像，我们无法知道手柄的空间具体位姿。

双目摄像头能干什么

双目摄像头模拟了人的双眼，左眼+右眼的搭配能以不同的角度真实感知物体的距离，原理可以参考这篇文章：[【深度相机系列三】深度相机原理揭秘—双目立体视觉 - 知乎](#)

也就是假设两个相机位于同一平面（即光轴平行，即下图中L所在水平线与R所在水平线平行）时，P分别投影在两个绿点处，根据初中几何相似知识，我们在焦距（f）、基线（baseline）**xl**和**xr**（物体在图中的像素点位置）已知的情况下，就可以算出物体的实际空间距离。



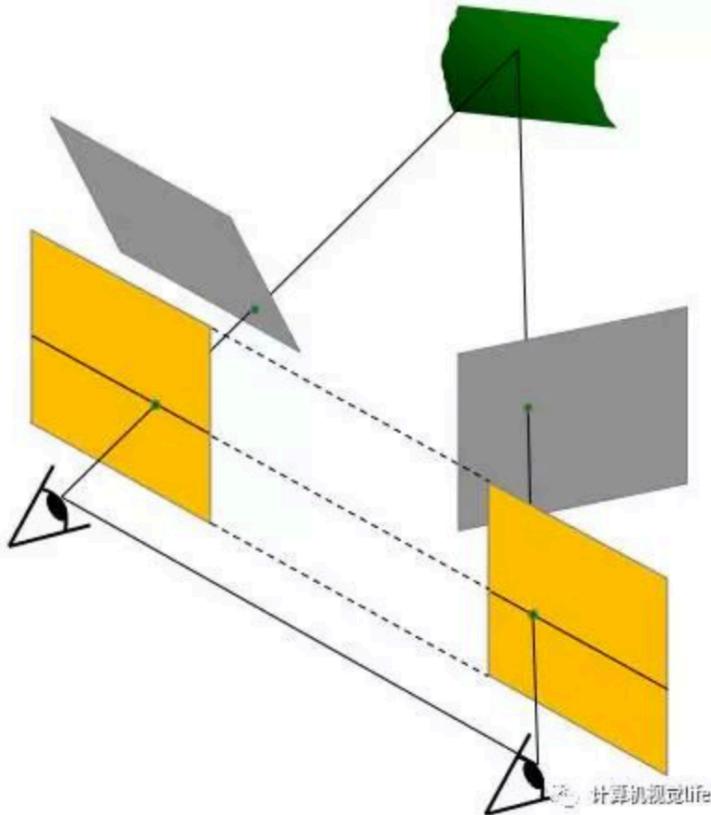
这其中，焦距、基线，都属于内参，可以在出厂时固定并通过与厂商联系获取，也可以自行使用业界的参数标定法（例如张正友标定法）。

但这个**xl**、**xr**却是我们需要通过算法去判定的，因为在两张图中，计算机不知道哪个点对哪个点，而一旦知道，就能测算出来了。这是问题一。

此外，两个相机的光轴并不往往在同一个平面（以我司眼镜上的四目摄像头为例，似乎都有一些

向外的倾角，以获取更大的视野），因此成像分析部分就会困难一些。而解决这个问题的办法就是基于单应矩阵^[2]实现图像矫正。

图像矫正是通过分别对两张图片用单应（homography）矩阵变换（可以通过标定获得）得到的，目的就是把两个不同方向的图像平面（下图中灰色平面）重新投影到同一个平面且光轴互相平行（下图中黄色平面），这样就可以用前面理想情况下的模型了，两个相机的极线也变成水平的了。



图像校正示意图

此外，从最原始的办法出发，要想在另一张图中找到第一张图中与某个像素点匹配点，可以通过遍历全图实现，复杂度达到了 $O(n^2)$ ，此外，假设每个点都需要周遭K个点才能确定其匹配度的话，那复杂度就是 $O(kn^2)$ 甚至 $O(n^3)$ ，显然难以接受。

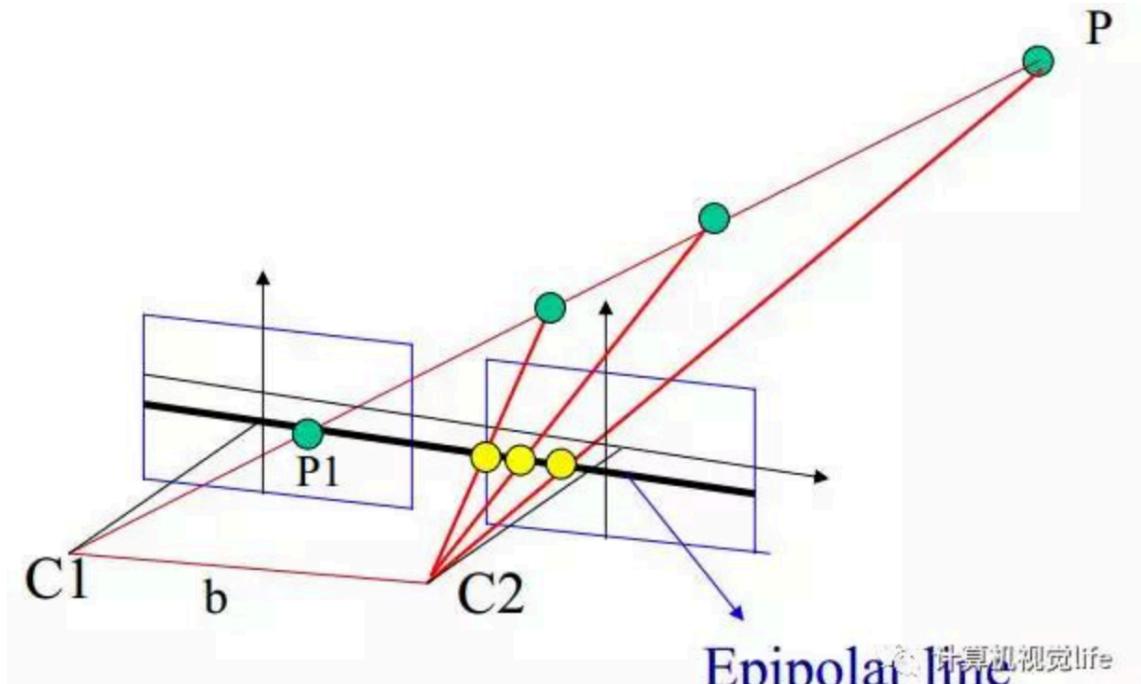
如果能有一种办法可以快速地缩小查找范围，就可以显著加快匹配速度。第一个答案就是极线约束。

那什么是极线呢？如下图所示。C₁, C₂是两个相机，P是空间中的一个点，P和两个相机中心点C₁、C₂形成了三维空间中的一个平面PC₁C₂，称为极平面（Epipolar plane）。极平面和两幅图像相交于两条直线，这两条直线称为极线（Epipolar line）。P在相机C₁中的成像点是P₁，在相机C₂中的成像点是P₂，但是P的位置事先是未知的。

我们的目标是：对于左图的P₁点，寻找它在右图中的对应点P₂，这样就能确定P点的空间位置，也就是我们想要的空间物体和相机的距离（深度）。

所谓极线约束（Epipolar Constraint）就是指当同一个空间点在两幅图像上分别成像时，已知左图投影点p₁，那么对应右图投影点p₂一定在相对于p₁的极线上，这样可以极大的缩小匹配范围。

根据极线约束的定义，我们可以在下图中直观的看到P₂一定在对极线上，所以我们只需要沿着极线搜索一定可以找到和P₁的对应点P₂。



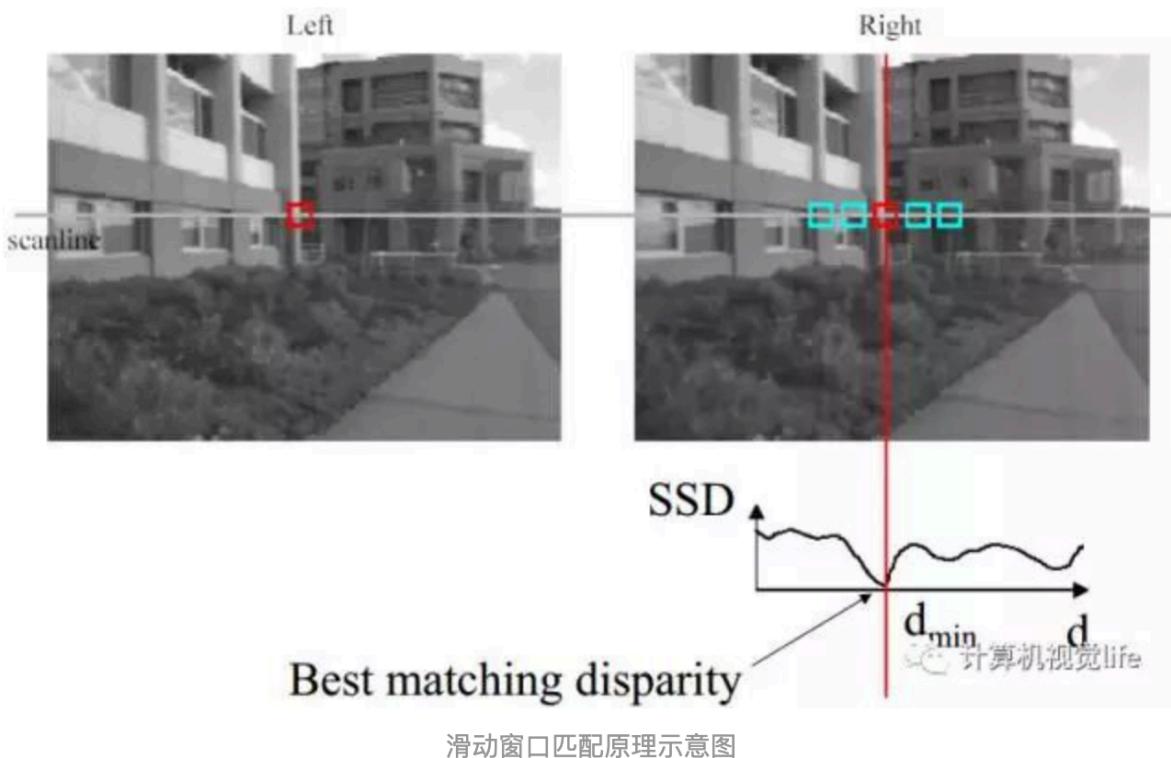
极线约束示意图

这样以来，从需要在一个平面找，变成了在一条直线上找，因此复杂度降到了 $O(n\sqrt{n})$ 。

当然，实际情况下，肯定会有误差，因此不必然一定在一条直线上，因此可以使用滑动窗口法适当放宽搜查范围，也就是 $O(kn\sqrt{n})$ 。

4) 基于滑动窗口的图像匹配

上述问题的解决方法：使用滑动窗口来进行匹配。如下图所示。对于左图中的一个像素点（左图中红色方框中心），在右图中从左到右用一个同尺寸滑动窗口内的像素和它计算相似程度，相似度的度量有很多种方法，比如 误差平方和法（Sum of Squared Differences，简称SSD），左右图中两个窗口越相似，SSD越小。下图中下方的SSD曲线显示了计算结果，SSD值最小的位置对应的像素点就是最佳的匹配结果。



以及，最终，我们还可以把它理解成一个最优化问题：

5) 、基于能量优化的图像匹配

目前比较主流的方法都是基于能量优化的方法来实现匹配的。能量优化通常会先定义一个能量函数。比如对于两张图中像素点的匹配问题来说，我们定义的能量函数如下图公式1。我们的目的是：

(1) 、在左图中所有的像素点和右图中对应的像素点越近似越好，反映在图像里就是灰度值越接近越好，也就是下图公式2的描述。

(2) 、在同一张图片里，两个相邻的像素点视差（深度值）也应该相近。也就是下图公式3的描述。

$$\begin{aligned} \textcircled{1} \quad & Energy = matchCost + smoothnessCost \\ \textcircled{2} \quad & matchCost = \sum_{x,y} \|I(x,y) - J(x+d_{xy},y)\| \\ \textcircled{3} \quad & smoothnessCost = \sum_{neighbor \ pixels \ p,q} |d_p - d_q| \end{aligned}$$

能量函数

上述公式1代表的能量函数就是著名的马尔科夫随机场(Markov Random Field)模型。通过对能量函数最小化，我们最后得到了一个最佳的匹配结果。有了左右图的每个像素的匹配结果，根据前面的深度计算公式就可以得到每个像素点的深度值，最终得到一幅深度图。

关于马尔科夫随机场，这篇文章写的很通俗易懂：[马尔科夫随机场\(MRF\)在深度学习图像处理中的应用-图像分割、纹理迁移 - 知乎](#)

其他关于双目视觉的参考：

- 一篇文章认识《双目立体视觉》 - [腾讯云开发者社区-腾讯云](#)

单目+imu

这篇文章不错：[计算机视觉方向简介\(三\) | 单目微运动生成深度图 - 知乎](#)

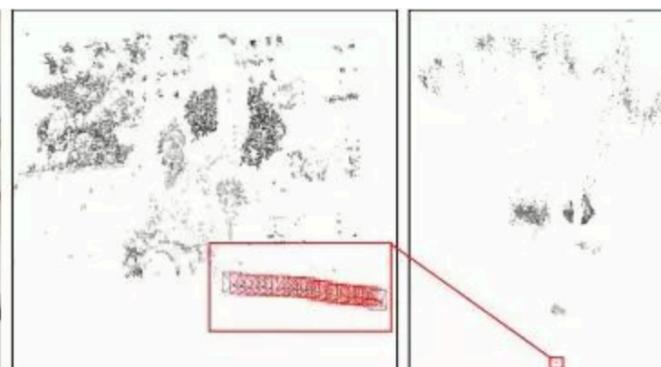
但是很多时候由于硬件的限制，我们不能通过深度相机获得深度图。只能利用单目相机通过相关算法来间接计算深度图。比较有名的方法就是运动恢复结构（Structure from Motion），也就是说，我们需要移动（通常需要较大的移动）单目相机从而获得不同视角的多张图片才能间接得到深度图。

从微小运动中获取深度图（depth from small motion）是其中一个比较巧妙的、用单目相机间接获得深度图的方法。该方法利用非常微小的运动来计算深度图，这个“非常微小的运动”的目的是在用户察觉不到的时间（比如手机用户寻找最佳拍摄位置时的微小移动，或者用户拿着相机按快门前的预览时间，或者类似live photo等）内得到深度图。如果该方法可以获得较高质量的深度图，就可以一定程度（静态场景下）上替代基于RGB双目立体视觉的深度相机（如手机双摄，手机双摄介绍见《为什么会出现双摄像头手机？》系列文章）的功能。

下面来介绍该技术的一个应用。如下图所示，(a) 是输入的一个微运动视频中所有帧的平均叠加图，可以看到运动真的是非常微小的。(c) 是算法计算的深度图，从放大的细节来看，边缘还是很锐利的，(d) 是利用得到的深度图进行重对焦的效果。我们看到对焦主体比较清晰，而位于主体前后景深的物体都已经虚化。



(a) An averaged image of small motion clip



(b) Front and top view of recovered 3D scene and camera poses



(c) Depth map



(d) Synthetic refocusing

计算机视觉life

如何确定两张图内的像素点之间的匹配对应关系

理解与使用**ORB-SLAM**

物体检测技术（基于yolo）

- [【经典论文解读】YOLOv4 目标检测_一颗小树x的博客-CSDN博客_yolov4](#)
- [\(21\) YOLOv3 - YouTube](#)

如何测算imu位姿

- [基于 IMU 的位姿解算_一颗小树x的博客-CSDN博客_imu位姿解算](#)

ORB-SLAM

source: [ORB-SLAM Project Webpage](#)

两篇论文分别发表于2015、2016年：

Related Publications

[1] [Raúl Mur-Artal](#), and [Juan D. Tardós](#).

ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras

ArXiv preprint arXiv:1610.06475, 2016.

[\[pdf\]](#)

[2] [Raúl Mur-Artal](#), [J. M. M. Montiel](#) and [Juan D. Tardós](#).

ORB-SLAM: A Versatile and Accurate Monocular SLAM System.

IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, October 2015.

2015 IEEE Transactions on Robotics Best Paper Award.

DOI: [10.1109/TRO.2015.2463671](https://doi.org/10.1109/TRO.2015.2463671)

[\[pdf\]](#)

[3] [Raúl Mur-Artal](#) and [Juan D. Tardós](#).

Probabilistic Semi-Dense Mapping from Highly Accurate Feature-Based Monocular SLAM.

Robotics: Science and Systems. Rome, Italy, July 2015.

[\[pdf\]](#) [\[poster\]](#)

[4] [Raúl Mur-Artal](#) and [Juan D. Tardós](#).

ORB-SLAM: Tracking and Mapping Recognizable Features.

Robotics: Science and Systems (RSS) Workshop on Multi View Geometry in RObotics (MVIGRO), Berkeley, USA, July 2014. *Oral presentation*

[\[pdf\]](#)

[5] [Raúl Mur-Artal](#) and [Juan D. Tardós](#).

Fast Relocalisation and Loop Closing in Keyframe-Based SLAM.

IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, June 2014.

[\[pdf\]](#)

[6] [Dorian Gálvez-López](#) and [Juan D. Tardós](#).

Bags of Binary Words for Fast Place Recognition in Image Sequences.

IEEE Transactions on Robotics, vol. 28, no. 5, pp. 1188-1197, October 2012.

DOI: [10.1109/TRO.2012.2197158](https://doi.org/10.1109/TRO.2012.2197158)

[\[pdf\]](#)

项目条件

四目摄像头

在眼镜上配有左上（LT）、右上（RT）、左下（LD）、右下（RD）四个摄像头。

其内参为（数据来源：金天4月1日答复）：

```
# intrinsic
Camera.fx: 274.1932382064952
Camera.fy: 275.36181187641506
Camera.cx: 316.23880553912664
Camera.cy: 236.82068245645195

# fish eye
Camera.k1: 0.012230065289018368
Camera.k2: -0.020113887598814294
Camera.k3: 0.020268290416176487
Camera.k4: -0.00615665479360849

# baseline (unit: m)
baseline: 0.065
```

路径分析

会议记录

2022-07-08, Fri

可见光问题

逆向 fpc 灯带

灯环要重新做

1. 灯圈设计之前，用惠普的灯圈去做测试
2. 通线调通后，再看看追踪能不能做

1. (21) The Math behind (most) 3D games - Perspective Projection - YouTube ↵
2. 关于单应矩阵，可以看这篇，我觉得写得非常生动形象：从零开始一起学习SLAM | 神奇的单应矩阵 - 知乎 ↵