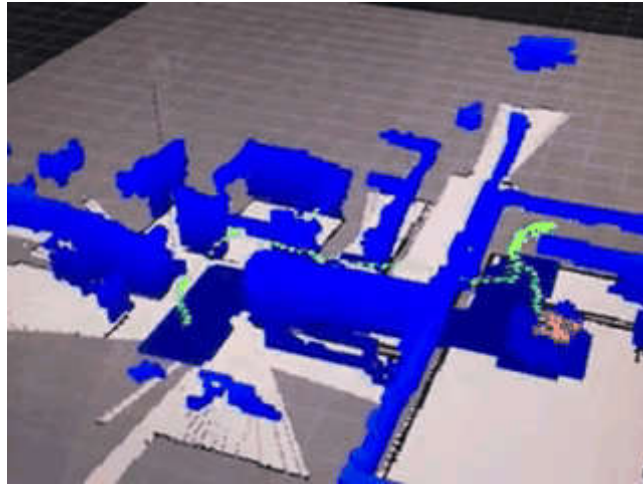


SLAM技术框架



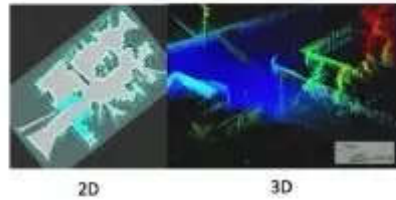
前一篇我们对SLAM有了个粗浅的认识，我们知道SLAM可以一边建图一边定位，还可以用建立好的地图进行导航。

有同学问现在手机上的百度/高德/腾讯等地图类App不就可以做到吗，为什么还需要SLAM？这里解释一下：目前地图类App在室外定位、导航方面确实做的很不错，而且衍生出很多基于地理位置的游戏、社交、生活类应用。这是因为地图类App背后使用的是GPS技术，据说美国军用GPS精度可以达到厘米级定位精度，而开放给大众使用的民用GPS也可以达到米级的定位精度。但是，**GPS只能在室外使用！而要想解决建筑物内、洞穴、海底等在GPS失效地域的定位、建图、姿态估计、路径规划，目前最有效的就是SLAM技术。**下面这个视频描述的是美国宾夕法尼亚大学的Vijay Kumar教授的团队在几年前的工作，展示了搭载SLAM技术的无人机是如何快速的对复杂建筑物内部建图的。

最早的SLAM雏形是在军事（核潜艇的海底定位）上的应用，主要传感器是军用雷达。SLAM技术发展到如今已经几十年，目前以激光雷达作为主传感器的SLAM技术比较稳定、可靠，仍然是主流的技术方案。但随着最近几年计算机视觉技术的快速发展，SLAM技术越来越多的应用于家用机器人、无人机、AR设备，基于视觉的Visual SLAM（简称VSLAM）逐渐开始崭露头角。

SLAM分类

• 激光

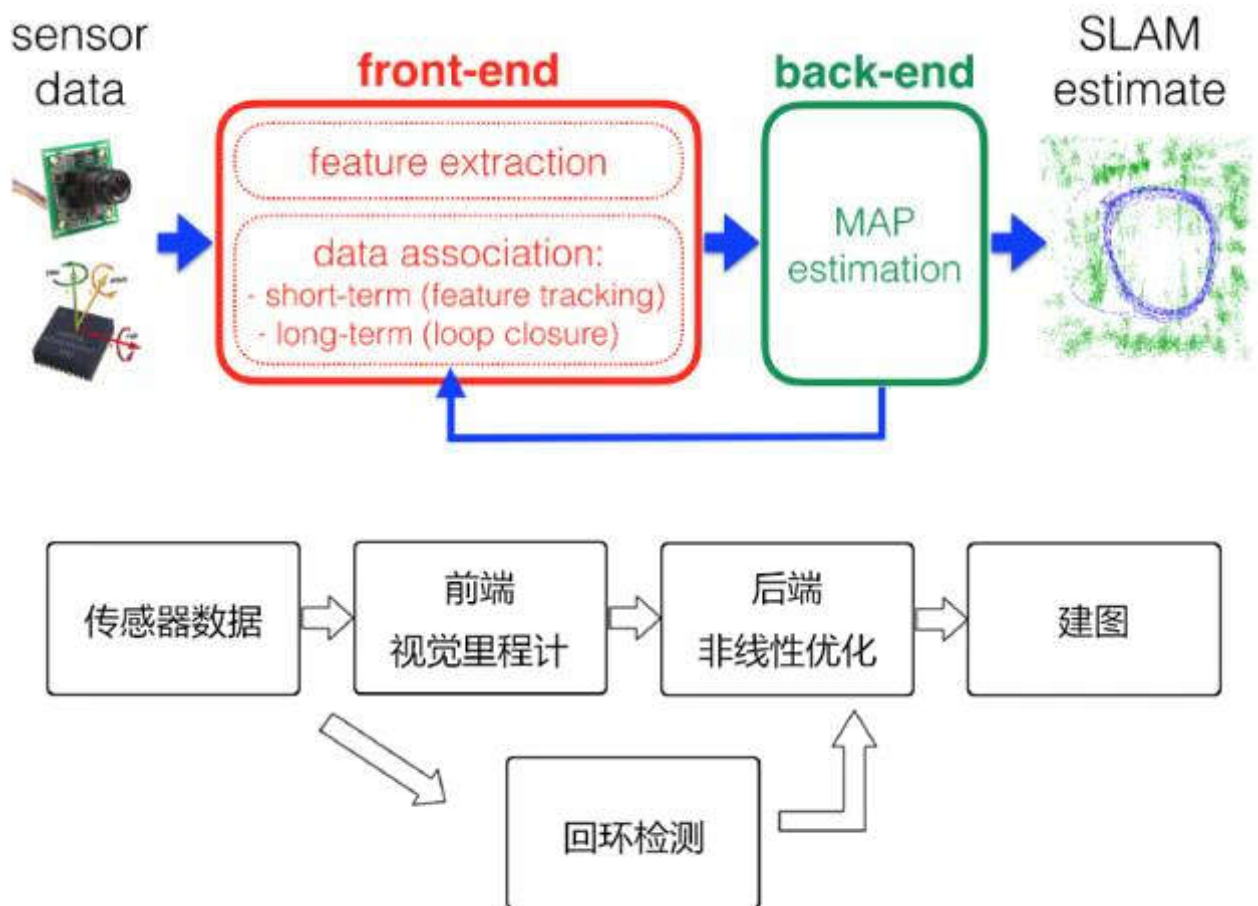


• 视觉



VSLAM 技术架构

本文主要介绍目前非常热门的VSLAM的技术框架，未来会有非常好的前景。VSLAM的技术框架如下，主要包括传感器数据预处理、前端、后端、回环检测、建图。



1 传感器数据

传感器数据预处理。这里的传感器包括摄像头、惯性测量单元（Inertial measurement unit, 简称IMU）等，涉及传感器选型、标定、多传感器数据同步等技术。

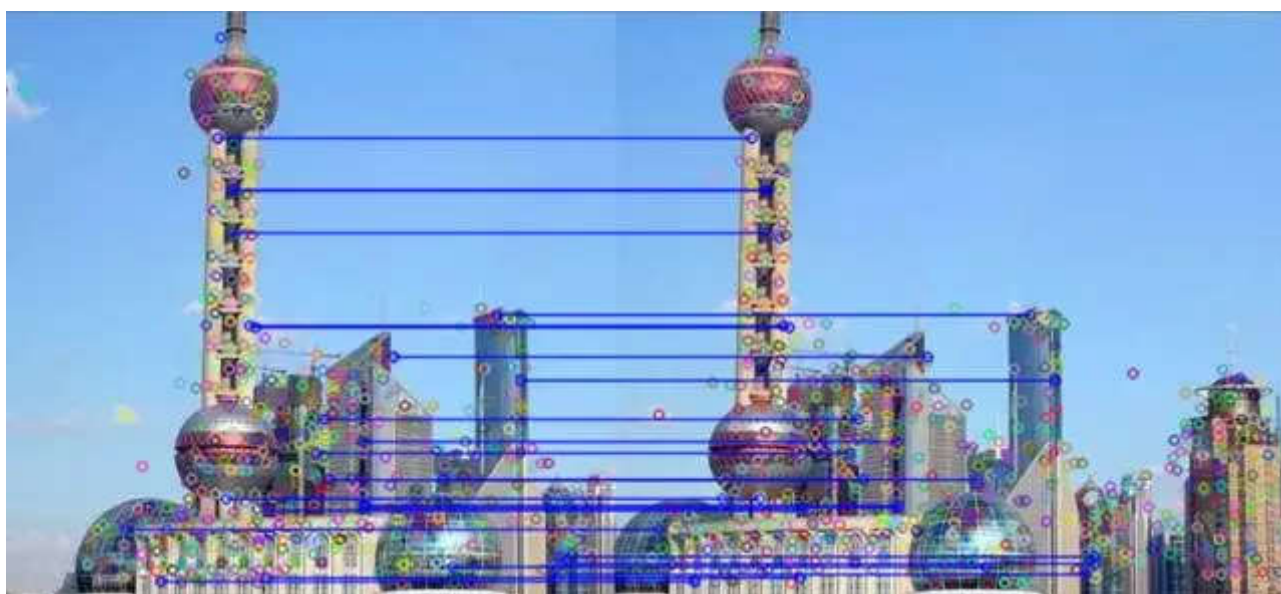


不同类型视觉传感器的对比

2 前端

又称为视觉里程计（visual odometry，简称VO）。主要是研究如何根据相邻帧图像定量估算帧间相机的运动。通过把相邻帧的运动轨迹串起来，就构成了相机载体（如机器人）的运动轨迹，解决了定位的问题。然后根据估算的每个时刻相机的位置，计算出各像素的空间点的位置，就得到了地图。

VSLAM中，前端主要涉及计算机视觉相关的算法。典型做法一般是：首先提取每帧图像特征点，对相邻帧进行特征点粗匹配，然后利用RANSAC（随机抽样一致）算法去除不合理的匹配对，然后得到位置和姿态信息。整个过程涉及到特征提取、特征匹配、对极几何、PnP、刚体运动、李代数等多视图几何知识。



相邻图像特征点匹配

前面说视觉里程计只计算相邻帧的运动，进行局部估计，这会不可避免的出现累积漂移，这是因为每次估计两个图像间的运动时都有一定的误差，经过相邻帧多次传递，前面的误差会逐渐累积，轨迹漂移（drift）的越来越厉害。



轨迹漂移现象

解决轨迹漂移的方法有两个：后端优化、回环检测。

3 后端

主要是对前端的结果进行优化，得到最优的位姿估计。主要有两种方法：

- 一种是基于滤波理论的优化，主要有 EKF, PF, RBPF, UKF 等方法，其中 EKF（扩展卡尔曼滤波）在早期是主流的方法。它的思路是将状态估计模型线性化，并用高斯分布近似其噪声，然后按照卡尔曼滤波进行预测来更新。但是实际上，这种对噪声的高斯分布大部分情况下是不成立的，此外，线性化过程中丢失了高阶项。
- 另一种就是非线性优化（图优化）。它的基本思想是将优化的变量作为图的节点，误差项作为图的边，在给定初值后，就可以迭代优化更新。由于图优化的稀疏性，可以在保证精度的同时，降低计算量。

后端优化涉及到的数学知识比较多，具有较高的难度。总的来说，从状态估计的角度来讲，SLAM 是一个非线性非高斯系统。因此传统的滤波理论已经逐渐被抛弃，而图优化已经成为主流方法。

4 回环检测

主要目的是让机器人能够认识自己曾经去过的地方，从而解决位置随时间漂移的问题。视觉回环检测一般通过判断图像之间的相似性完成，这和我们人类用眼睛来判断两个相同的地点是一样的道理。因为图像信息丰富，因此VSLAM在回环检测中具有很大的优势。



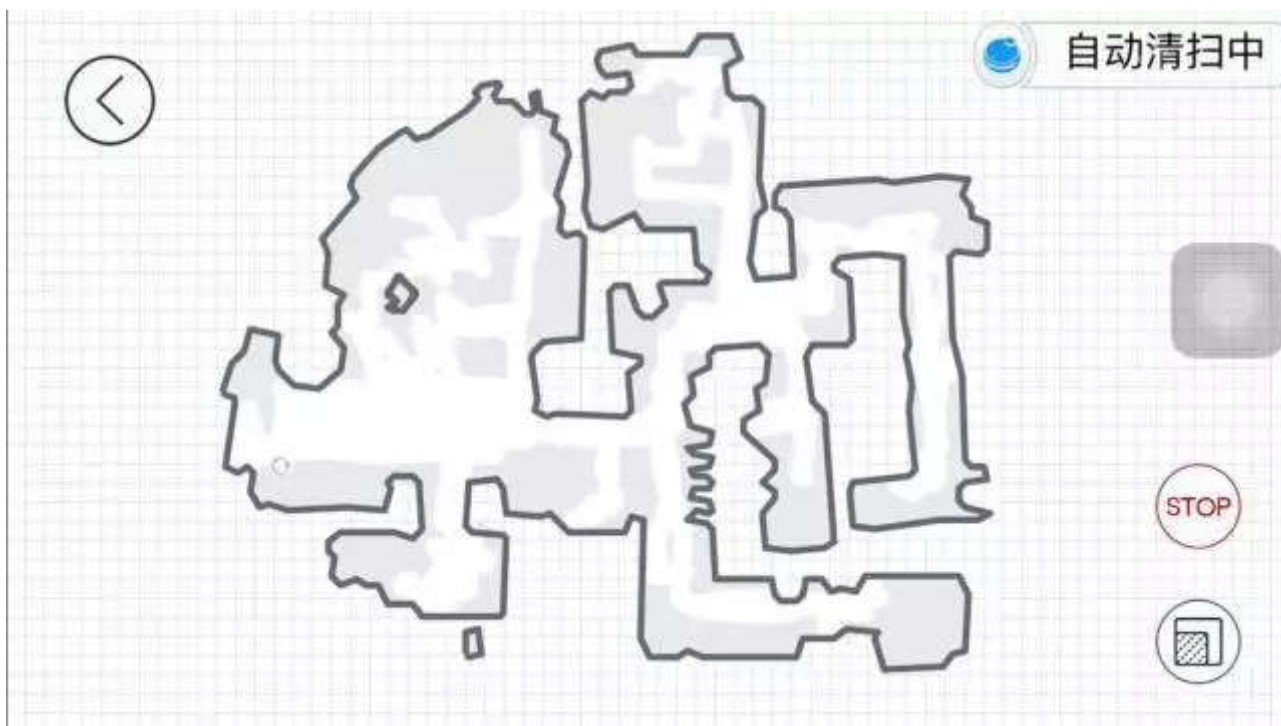
回环检测效果

当回环检测成功后，就会建立现在的图像和过去曾经见过图像的对应关系，后端优化算法可以根据这些信息来重新调整轨迹和地图，从而最大限度地消除累积误差。

5 建立地图

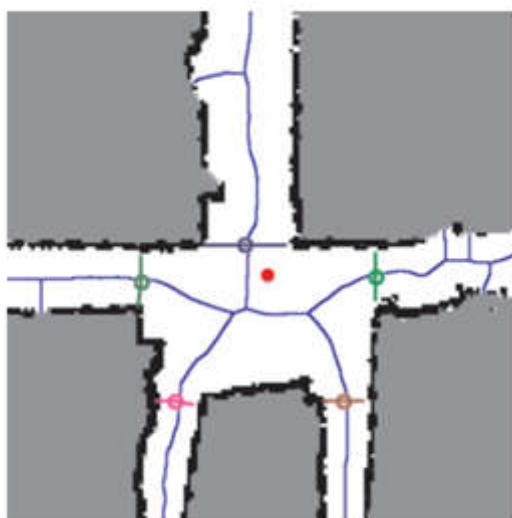
SLAM根据不同的传感器类型和应用需求建立不同的地图。常见的有2D栅格地图、2D拓扑地图、3D点云地图等。

比如前面提到过的扫地机器人，它只需要知道房屋内部的简单二维地图就可以了，不需要知道房屋到底有多高；它只需要知道哪里可以通过，哪里是障碍物，而不需要知道这个障碍物到底是什么，长什么样子；因此目前大部分具有SLAM功能的扫地机器人几乎都是采用廉价的消费级激光雷达方案，很少采用视觉SLAM方案（VSLAM也不够稳定）。



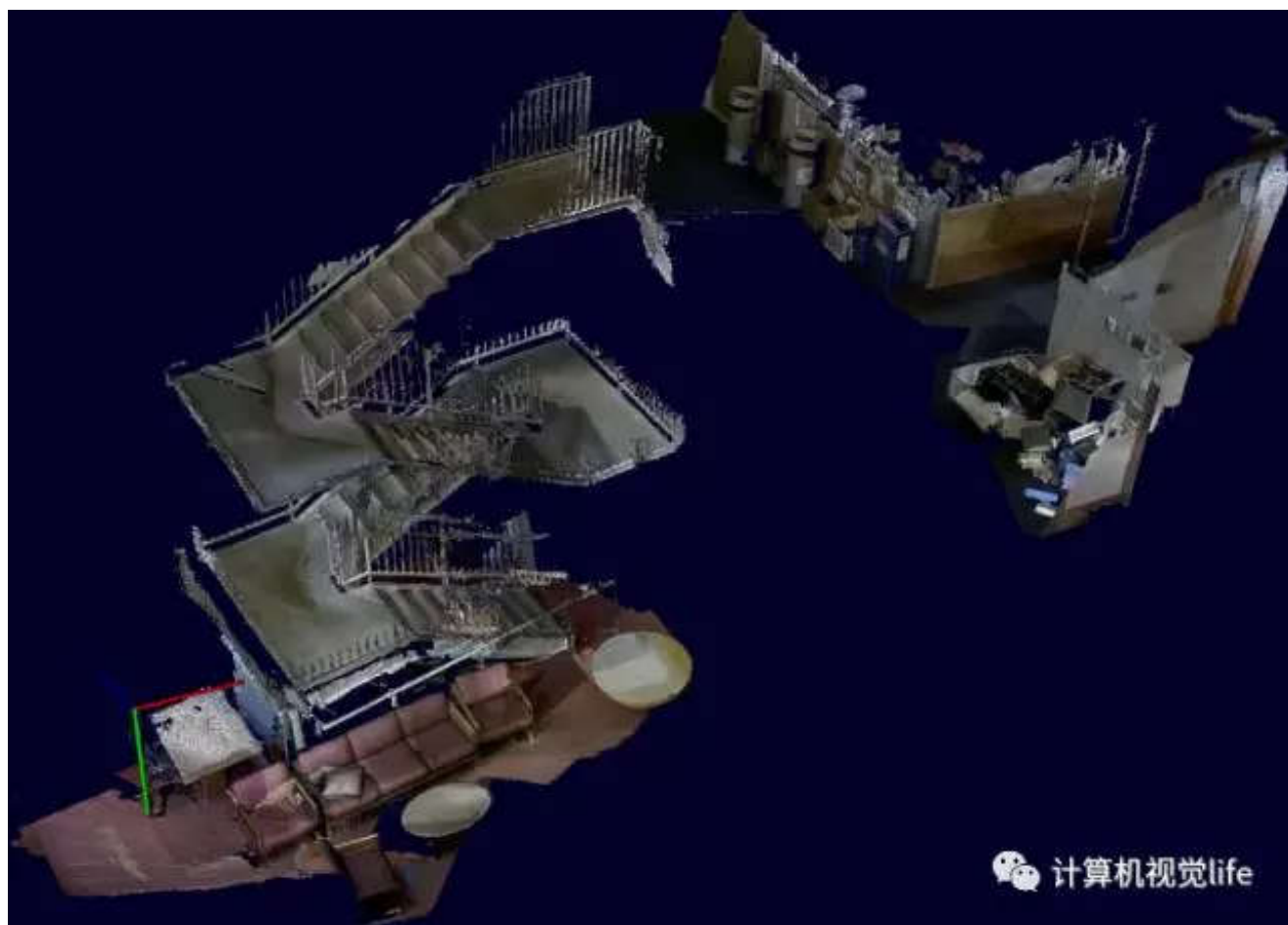
扫地机器人建立的2D地图

2D拓扑地图更强调地图元素之间的连通关系，而对精确的位置要求不高，去掉了大量地图的细节，是一种非常紧凑的地图表达方式。如下所示：



2D拓扑地图

3D点云地图在VSLAM中用的比较多，主要用于真实场景的视觉重建，重建的地图非常直观漂亮。但是点云地图通常规模很大，比如一张VGA分辨率（640 x 480）的点云图像，就会产生30万个空间点，这会占据非常大的存储空间，而且存在很多冗余信息。



3D点云地图

总结

前面介绍了VSLAM的典型技术框架。我们可以看到，将SLAM算法拆解后，用到的技术多是传统的计算机视觉算法，尤其是多视角几何相关知识。与当前大热的深度学习“黑箱模型”不同，SLAM的各个环节基本都是白箱，能够解释得非常清楚。但SLAM算法并不是上述各种算法的简单叠加，而是一个需要相互折中、密切配合的复杂系统工程。

敬请关注后续更多有关SLAM的技术。

注：本文使用了清华大学高翔博士的部分公开资料。

相关文章

[SLAM初识](#)