



视觉SLAM：从理论到实践



主讲人 高翔

清华大学 自动控制与工程 博士
慕尼黑工业大学计算机视觉组 博士后
Email: gao.xiang.thu@gmail.com



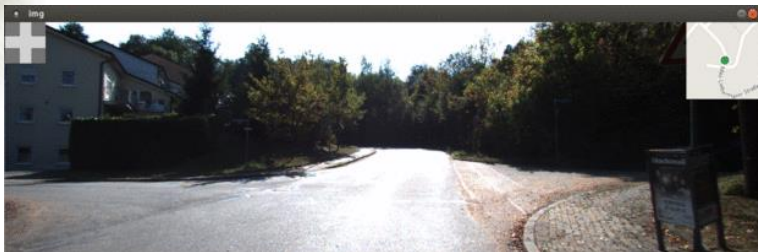
第一讲 概述与预备知识

1. 课程内容与预备知识
2. SLAM是什么
3. 视觉SLAM数学表述与框架
4. 实践：Linux下的C++编程基础

1. 课程内容与预备知识

课程内容与预备知识

- 计算机视觉



现实世界中的相机

视频序列

物体识别
物体跟踪
物体检测
语义分割

.....
SLAM

计算机视觉任务

课程内容与预备知识

- SLAM是什么?

室内/室外定位

稀疏-半稠密重建

稠密重建

课程内容与预备知识

- SLAM可以用在哪些地方？

手持设备定位

自动驾驶定位

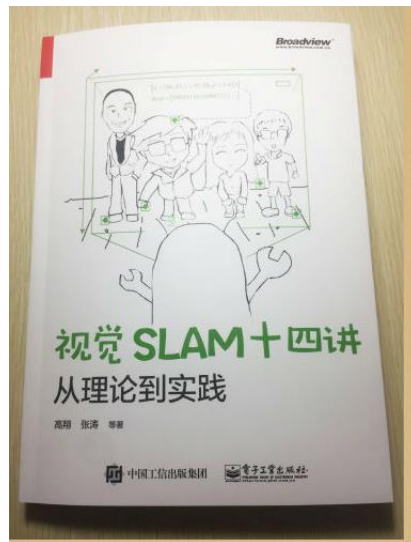
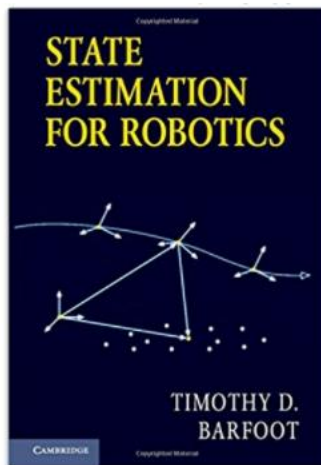
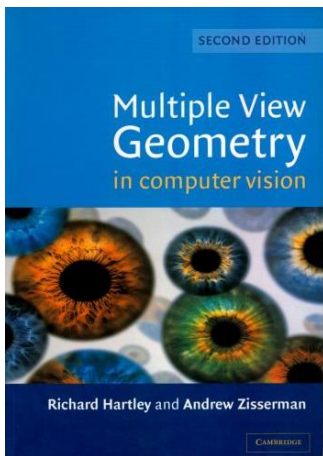
AR（增强现实）

课程内容与预备知识

- 课程内容
 - SLAM: Simultaneous Localization and Mapping
 - 同时定位与地图构建
 - SLAM的数学基础知识
 - SLAM相关的计算机视觉知识
 - 工程实践

课程内容与预备知识

- 课程内容
 - 教材：视觉SLAM十四讲：从理论到实践
 - 参考书：多视图几何 状态估计



课程内容与预备知识

- 课程特点
 - 基础、必要的理论知识
 - 大量的编程内容
 - 重视工程实践
 - 必要的习题
- 观念：只有亲自动手实现了算法，才能谈得上理解

课程内容与预备知识

- 内容提纲

1. 概述与预备知识（教材第1，2讲）
2. 三维空间的刚体运动（教材第3讲）
3. 李群与李代数（教材第4讲）
4. 相机模型与非线性优化（教材第5，6讲）
5. 特征点法视觉里程计（教材第7讲）
6. 直接法视觉里程计（教材第8讲）
7. 后端优化（教材第10，11讲）
8. 回环检测（教材第13讲）

- 周日讲课，一周习题时间

习题

- 每节课结束后会布置习题
- 习题花费时间约课内时间的3倍以上
- 需要至少达到8/10分才能通过
- 包含编程题、简答题
- 会提供详细的操作指导

例子：

2 熟悉 Linux (2 分, 约 2 小时)

计算机领域的绝大多数科研人员都在 Linux 下工作，不掌握 Linux 会使你在研究道路上寸步难行。Linux 系统的基本知识亦是学习本课程的先决条件。如果你还未接触过 Linux，请阅读本次习题中提供的材料（见 books/目录下）。我建议阅读《鸟哥的 Linux 私房菜》第 1、2 章了解 Linux 历史，第 5-8 章了解基础操作。如果你在用 Ubuntu，也可以参考 Ubuntu 维基上自带的指南：<http://wiki.ubuntu.org.cn/Ubuntu>

不要把 Linux 想得太困难。现代的 Linux 系统多数具有方便的图形界面，非常容易上手。最好的学习方式可能是马上安装一个 Linux 然后熟悉它的操作界面，多数时候和 Windows/mac 差别不大。我们在本书中使用 Ubuntu 16.04。读者也可以按个人口味选择任意适合你的发行版，不过最好使用 Ubuntu 系列，这样我和你的操作方式会比较相似。

等你熟悉 Linux 后，请回答以下问题（如果你已经熟悉，就跳过上面的阅读内容，直接回答即可）：

1. 如何在 Ubuntu 中安装软件（命令行界面）？它们通常被安装在什么地方？
2. linux 的环境变量是什么？我如何定义新的环境变量？
3. linux 根目录下面的目录结构是什么样的？至少说出 3 个目录的用途。
4. 假设我要给 a.sh 加上可执行权限，该输入什么命令？
5. 假设我要将 a.sh 文件的所有者改成 xiang:xiang，该输入什么命令？

课程内容与预备知识

- 预备知识与课程使用的环境
 - 数学：高等数学、线性代数（矩阵论）、概率论
 - 编程：C++、Linux，了解语法和基本命令即可
 - 英语：文献、文档阅读能力
 - 环境：Ubuntu 16.04
 - 不提供windows环境下的方案！
 - 阅读教材第一章习题作为自测

2.SLAM是什么

SLAM是什么

- 自主运动的两大基本问题
 - 我在什么地方？——定位
 - 周围环境是什么样子？——建图
 - 定位与建图=内外兼修 定位侧重对自身的了解，建图侧重对外在的了解
- 相互关联
 - 准确的定位需要精确的地图
 - 精确的地图来自准确的定位



SLAM是什么

- 传感器

- 机器人感知外界环境的手段
- 种类：内质的/外质的
 - 内质：感受机器人本体信息
 - 加速度计、陀螺仪、编码器
 - 外质：测量外界的信息
 - 相机、激光、导轨、磁条



(a)



(b)



(c)



(d)



(e)



(f)

- 环境限制了传感器的形式

- GPS：需要能接收到卫星信号的环境
- Marker、导轨：需要环境允许安装
- 相比之下，激光、相机等携带式传感器更加自由

SLAM是什么

- 相机
 - 以一定速率采集图像、形成视频
- 分类
 - 单目相机 Monocular
 - 双目相机（立体相机） Stereo
 - 深度相机 RGB-D
 - 其他 全景、Event Camera
- 相机的特点
 - 以二维投影形式记录了三维世界的信息
 - 该过程丢掉了一个维度：距离



图 2-3 形形色色的相机：单目，双目和深度相机。



SLAM是什么

- 各类相机的区别
 - 单目：无深度 深度需要其他手段估计
 - 双目：通过视差计算深度
 - RGB-D：通过物理方法测量深度

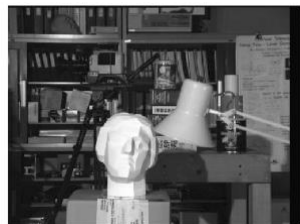
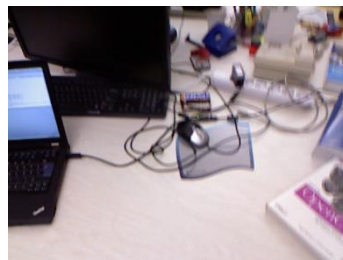
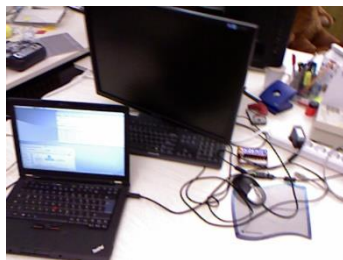


图 2-4 单目视觉中的尴尬：不知道深度时，手掌上的人是真人还是模型？



当相机运动时，可通过视差确定深度

深度即第三维信息，对SLAM来说至关重要

SLAM是什么

- 视觉SLAM框架
 - 前端：Visual Odometry
 - 后端：Optimization
 - 回环：Loop Closing
 - 建图：Mapping

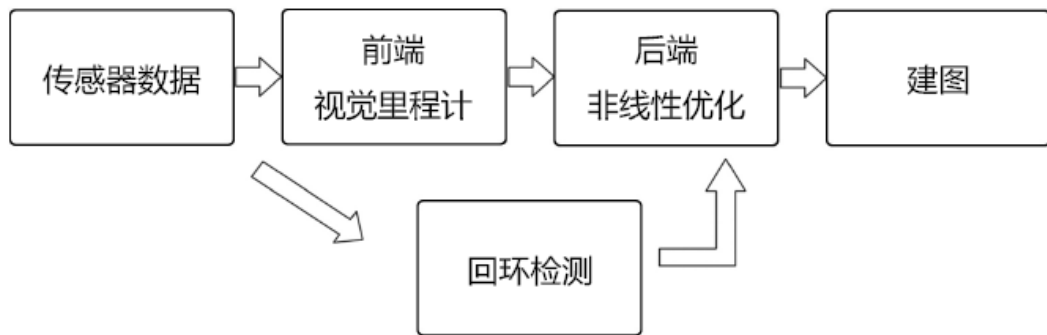


图 2-7 整体视觉 SLAM 流程图。

SLAM是什么

- 视觉里程计

- 估计邻近时刻的相机运动
- 最简化：两个图像的相对运动

- 方法

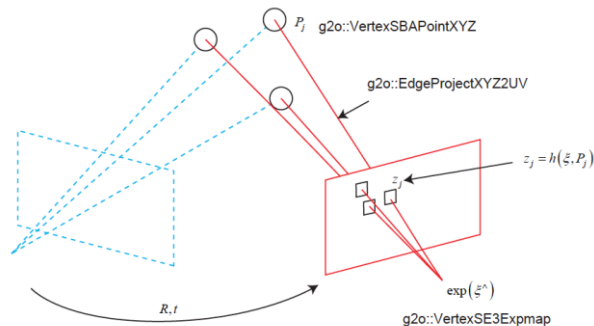
- 特征点法
- 直接法

- 后端

- 从带有噪声的数据中估计最优轨迹与地图
- 最大后验概率估计
- 滤波器
- 图优化



图 2-8 相机拍摄到的图片与人眼反应的运动方向。



SLAM是什么

回环检测

- 检测相机是否到达过之前位置
- 判断与之前位置的差异
- 计算图像间相似性
- 词袋模型

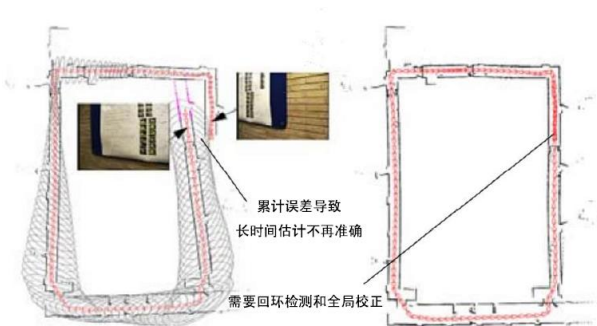


图 2-9 累计误差与回环检测的校正结果 [10]。

建图

- 导航、规划、通讯、交互、可视化
- 度量地图、拓扑地图
- 稀疏地图、稠密地图

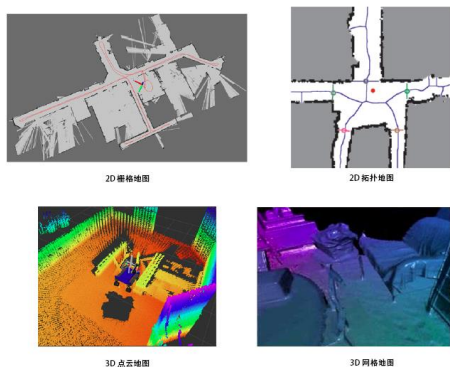


图 2-10 形形色色的地图：2D 栅格地图、拓扑地图以及 3D 点云地图和网格地图 [12]。

3.视觉SLAM的数学描述

SLAM是什么

- SLAM问题的数学描述

- SLAM问题的数学描述

$$\begin{array}{l} \text{运动方程} \\ \text{观测方程} \end{array} \quad \left\{ \begin{array}{l} \mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{w}_k) \\ \mathbf{z}_{k,j} = h(\mathbf{y}_j, \mathbf{x}_k, \mathbf{v}_{k,j}) \end{array} \right. .$$

问题：

- 位置是三维的，如何表述？——第2、3次课
- 观测模型如何表述？——第4次课
- 已知 \mathbf{u}, \mathbf{z} 时，如何推断 \mathbf{x}, \mathbf{y} ？——第5次课以后

4.实践：Linux下的C++基础



深蓝学院
shenlanxueyuan.com

感谢各位聆听!

Thanks for Listening