

视觉 SLAM 十四讲

从理论到实践

高翔 张涛 等著

视觉 SLAM 十四讲

从理论到实践

高翔 张涛 刘毅 颜沁睿 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书系统介绍了视觉 SLAM（同时定位与地图构建）所需的基本知识与核心算法，既包括数学理论基础，如三维空间的刚体运动、非线性优化，又包括计算机视觉的算法实现，例如多视图几何、回环检测等。此外，我们还提供了大量的实例代码供读者学习研究，从而更深入地掌握这些内容。

本书可以作为对 SLAM 感兴趣的研究人员的入门自学材料，也可以作为 SLAM 相关的高校本科生或研究生课程教材使用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

视觉 SLAM 十四讲：从理论到实践 / 高翔等著. — 北京：电子工业出版社，2017.3

ISBN 978-7-121-31104-8

I. ①视…II. ①高…III. ①人工智能—视觉跟踪—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字（2017）第 053910 号

策划编辑：郑柳洁

责任编辑：白 涛

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：720×1000 1/16 印张：25 字数：560 千字

版 次：2017 年 3 月第 1 版

印 次：2017 年 3 月第 1 次印刷

定 价：75.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zltts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）51260888-819 faq@phei.com.cn。

作者简介

2008年就读清华大学自动化系，2012年免试进入清华大学自动化系攻读博士学位。研究课题为视觉SLAM，兴趣包括计算机视觉与机器学习。曾撰写过与SLAM相关的论文和技术博客。



· 高翔 ·

清华大学自动化系教授、党委书记、副系主任。1995年9月至1999年9月在清华大学自动化系检测技术与自动化装置专业学习，获博士学位。1999年10月至2002年9月在日本国立佐贺大学大学院工学系研究科系统控制专业学习，获博士学位。研究课题包括机器人、航空航天、计算机视觉等。



· 张涛 ·

华中科技大学图像与人工智能研究所 在读博士，本科毕业于武汉理工大学数学系。读博期间专注于图像处理、三维重建、视觉SLAM，以及传感器融合研究和应用，读博期间先后于深圳市大疆创新公司、英特尔中国研究院等单位实习。



· 刘毅 ·

比利时荷语鲁汶大学人工智能硕士，电子工程学士（GroupT）。电子科技大学信息显示与光电技术学士。长期致力于研究人工智能技术在机器人领域的应用，包括计算机视觉、机器学习和SLAM。现担任地平线机器人公司智能驾驶部算法工程师。



· 颜沁睿 ·

阅读须知

本书图片

为描述方便，书中介绍图片相关内容时直接使用了对应的色彩名称，对应到纸面上即灰色。此外，部分图片可能需要放大观察，纸面上无法呈现应有效果。为此，书中图片均在博文视点官方网站提供下载。

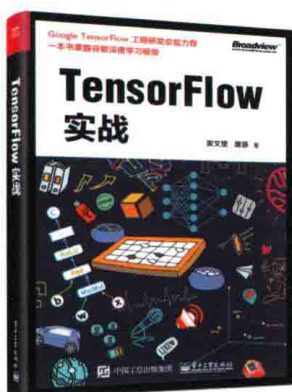
读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），您即可享受以下服务。

- **下载资源：**本书所提供的示例代码及资源文件均可在 [下载资源](#) 处下载。
- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **与作者交流：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与作者和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/31104>





《TensorFlow 实战》

黄文坚 唐源 著



《第一本无人驾驶技术书》（封面待定）

刘少山 唐洁 吴双 李力耘 著



三步加入“人工智能交流群”，实时获取资源共享

1. 扫码添加小编为微信好友。
2. 申请验证时输入“AI”。
3. 小编带你加入“人工智能交流群”。

目录

第 1 讲	预备知识	1
1.1	本书讲什么	1
1.2	如何使用本书	3
1.2.1	组织方式	3
1.2.2	代码	5
1.2.3	面向的读者	6
1.3	风格约定	6
1.4	致谢和声明	7
第 2 讲	初识 SLAM	9
2.1	引子：小萝卜的例子	11
2.2	经典视觉 SLAM 框架	17
2.2.1	视觉里程计	17
2.2.2	后端优化	19
2.2.3	回环检测	20
2.2.4	建图	21
2.3	SLAM 问题的数学表述	22
2.4	实践：编程基础	25
2.4.1	安装 Linux 操作系统	25
2.4.2	Hello SLAM	27
2.4.3	使用 cmake	28
2.4.4	使用库	30
2.4.5	使用 IDE	32
第 3 讲	三维空间刚体运动	37
3.1	旋转矩阵	39
3.1.1	点和向量，坐标系	39

3.1.2	坐标系间的欧氏变换	40
3.1.3	变换矩阵与齐次坐标	42
3.2	实践: Eigen	44
3.3	旋转向量和欧拉角	48
3.3.1	旋转向量	48
3.3.2	欧拉角	50
3.4	四元数	51
3.4.1	四元数的定义	51
3.4.2	四元数的运算	53
3.4.3	用四元数表示旋转	55
3.4.4	四元数到旋转矩阵的转换	55
3.5	* 相似、仿射、射影变换	56
3.6	实践: Eigen 几何模块	57
3.7	可视化演示	60
第 4 讲	李群与李代数	62
4.1	李群与李代数基础	64
4.1.1	群	64
4.1.2	李代数的引出	65
4.1.3	李代数的定义	67
4.1.4	李代数 $\mathfrak{so}(3)$	67
4.1.5	李代数 $\mathfrak{se}(3)$	68
4.2	指数与对数映射	69
4.2.1	$SO(3)$ 上的指数映射	69
4.2.2	$SE(3)$ 上的指数映射	70
4.3	李代数求导与扰动模型	72
4.3.1	BCH 公式与近似形式	72
4.3.2	$SO(3)$ 李代数上的求导	73
4.3.3	李代数求导	74
4.3.4	扰动模型 (左乘)	75
4.3.5	$SE(3)$ 上的李代数求导	76
4.4	实践: Sophus	76
4.5	* 相似变换群与李代数	79
4.6	小结	81

第 5 讲	相机与图像	82
5.1	相机模型	84
5.1.1	针孔相机模型	84
5.1.2	畸变	87
5.1.3	双目相机模型	90
5.1.4	RGB-D 相机模型	92
5.2	图像	93
5.3	实践: 图像的存取与访问	95
5.3.1	安装 OpenCV	95
5.3.2	操作 OpenCV 图像	96
5.4	实践: 拼接点云	99
第 6 讲	非线性优化	104
6.1	状态估计问题	106
6.1.1	最大后验与最大似然	106
6.1.2	最小二乘的引出	108
6.2	非线性最小二乘	109
6.2.1	一阶和二阶梯度法	110
6.2.2	高斯牛顿法	111
6.2.3	列文伯格—马夸尔特方法	113
6.2.4	小结	114
6.3	实践: Ceres	115
6.3.1	Ceres 简介	116
6.3.2	安装 Ceres	116
6.3.3	使用 Ceres 拟合曲线	117
6.4	实践: g2o	121
6.4.1	图优化理论简介	121
6.4.2	g2o 的编译与安装	122
6.4.3	使用 g2o 拟合曲线	123
6.5	小结	128
第 7 讲	视觉里程计 1	130
7.1	特征点法	132
7.1.1	特征点	132
7.1.2	ORB 特征	134
7.1.3	特征匹配	137

7.2	实践: 特征提取和匹配	138
7.3	2D-2D: 对极几何	141
7.3.1	对极约束	141
7.3.2	本质矩阵	143
7.3.3	单应矩阵	146
7.4	实践: 对极约束求解相机运动	148
7.5	三角测量	153
7.6	实践: 三角测量	154
7.6.1	三角测量代码	154
7.6.2	讨论	156
7.7	3D-2D: PnP	157
7.7.1	直接线性变换	158
7.7.2	P3P	159
7.7.3	Bundle Adjustment	161
7.8	实践: 求解 PnP	165
7.8.1	使用 EPnP 求解位姿	165
7.8.2	使用 BA 优化	166
7.9	3D-3D: ICP	172
7.9.1	SVD 方法	173
7.9.2	非线性优化方法	175
7.10	实践: 求解 ICP	176
7.10.1	SVD 方法	176
7.10.2	非线性优化方法	178
7.11	小结	180
第 8 讲	视觉里程计 2	182
8.1	直接法的引出	184
8.2	光流 (Optical Flow)	185
8.3	实践: LK 光流	187
8.3.1	使用 TUM 公开数据集	187
8.3.2	使用 LK 光流	188
8.4	直接法 (Direct Method)	192
8.4.1	直接法的推导	192
8.4.2	直接法的讨论	195

8.5	实践: RGB-D 的直接法	196
8.5.1	稀疏直接法	196
8.5.2	定义直接法的边	197
8.5.3	使用直接法估计相机运动	199
8.5.4	半稠密直接法	200
8.5.5	直接法的讨论	202
8.5.6	直接法优缺点总结	203
第 9 讲	实践: 设计前端	205
9.1	搭建 VO 框架	206
9.1.1	确定程序框架	207
9.1.2	确定基本数据结构	208
9.1.3	Camera 类	210
9.1.4	Frame 类	212
9.1.5	MapPoint 类	213
9.1.6	Map 类	213
9.1.7	Config 类	214
9.2	基本的 VO: 特征提取和匹配	216
9.2.1	两两帧的视觉里程计	216
9.2.2	讨论	224
9.3	改进: 优化 PnP 的结果	224
9.4	改进: 局部地图	227
9.5	小结	233
第 10 讲	后端 1	235
10.1	概述	237
10.1.1	状态估计的概率解释	237
10.1.2	线性系统和 KF	239
10.1.3	非线性系统和 EKF	242
10.1.4	EKF 的讨论	243
10.2	BA 与图优化	245
10.2.1	投影模型和 BA 代价函数	245
10.2.2	BA 的求解	247
10.2.3	稀疏性和边缘化	248
10.2.4	鲁棒核函数	255
10.2.5	小结	256

10.3	实践: g2o.....	257
10.3.1	BA 数据集	257
10.3.2	g2o 求解 BA	258
10.3.3	求解.....	262
10.4	实践: Ceres.....	264
10.4.1	Ceres 求解 BA.....	265
10.4.2	求解.....	267
10.5	小结.....	269
第 11 讲	后端 2.....	270
11.1	位姿图 (Pose Graph)	271
11.1.1	Pose Graph 的意义	271
11.1.2	Pose Graph 的优化	272
11.2	实践: 位姿图优化.....	274
11.2.1	g2o 原生位姿图	274
11.2.2	李代数上的位姿图优化	278
11.2.3	小结.....	284
11.3	* 因子图优化初步.....	285
11.3.1	贝叶斯网络.....	285
11.3.2	因子图	286
11.3.3	增量特性	288
11.4	* 实践: gtsam	289
11.4.1	安装 gtsam 4.0	289
11.4.2	位姿图优化.....	290
第 12 讲	回环检测	297
12.1	回环检测概述	299
12.1.1	回环检测的意义	299
12.1.2	方法.....	300
12.1.3	准确率和召回率	301
12.2	词袋模型.....	303
12.3	字典.....	305
12.3.1	字典的结构.....	305
12.3.2	实践: 创建字典	306
12.4	相似度计算.....	309
12.4.1	理论部分	309

12.4.2	实践：相似度的计算	310
12.5	实验分析与评述	314
12.5.1	增加字典规模	314
12.5.2	相似性评分的处理	316
12.5.3	关键帧的处理	316
12.5.4	检测之后的验证	317
12.5.5	与机器学习的关系	317
第 13 讲	建图	319
13.1	概述	320
13.2	单目稠密重建	322
13.2.1	立体视觉	322
13.2.2	极线搜索与块匹配	323
13.2.3	高斯分布的深度滤波器	325
13.3	实践：单目稠密重建	328
13.4	实验分析与讨论	339
13.4.1	像素梯度的问题	339
13.4.2	逆深度	340
13.4.3	图像间的变换	341
13.4.4	并行化：效率的问题	342
13.4.5	其他的改进	343
13.5	RGB-D 稠密建图	343
13.5.1	实践：点云地图	344
13.5.2	八叉树地图	347
13.5.3	实践：八叉树地图	350
13.6	*TSDF 地图和 Fusion 系列	352
13.7	小结	356
第 14 讲	SLAM：现在与未来	357
14.1	当前的开源方案	358
14.1.1	MonoSLAM	358
14.1.2	PTAM	359
14.1.3	ORB-SLAM	361
14.1.4	LSD-SLAM	363
14.1.5	SVO	364
14.1.6	RTAB-MAP	366

第 1 讲

预备知识

1.1 本书讲什么

这是一本介绍视觉 SLAM 的书，也很可能是第一本以视觉 SLAM 为主题的中文书。那么，SLAM 是什么？

SLAM 是 Simultaneous Localization and Mapping 的缩写，中文译作“同时定位与地图构建”^[1]。它是指搭载特定传感器的主体，在没有环境先验信息的情况下，于运动过程中建立环境的模型，同时估计自己的运动^[2]。如果这里的传感器主要为相机，那就称为“视觉 SLAM”。

本书的主题就是视觉 SLAM。这里我们刻意把许多个定义放到一句话中，是希望读者有一个较明确的概念。首先，SLAM 的目的是解决“定位”与“地图构建”这两个问题。也就是说，一边要估计传感器自身的位置，一边要建立周围环境的模型。那么怎么解决呢？这需要用到传感器的信息。传感器能以一定形式观察外部的世界，不过不同传感器观察的方式是不同的。而之所以要花一本书的内容去讨论这个问题，是因为它很难——特别是我们希望实时地、在没有先验知识的情况下进行 SLAM。当用相机作为传感器时，我们要做的就是根据一张张连续运动的图像（它们形成了一段视频），从中推断相机的运动，以及周围环境的情况。

这似乎是个很直观的问题。我们自己走进陌生的环境时不就是这么做的吗？

在计算机视觉（Computer Vision）创立之初，人们就想象着有朝一日计算机将和人一样，通过眼睛去观察世界，理解周遭的物体，探索未知的领域——这是一个美妙而又浪漫的梦想，吸引了无数的科研人员日夜为之奋斗^[3]。我们曾经以为这件事情并不困难，然而进展却远不如预想的那么顺利。我们眼中的花草树木、虫鱼鸟兽，在计算机中却是那样的不同：它们只是一个个由数字排列而成的矩阵（Matrix）。让计算机理解图

像的内容，就像让我们自己理解这些数字一样困难。我们既不了解自己如何理解图像，也不知道计算机该如何理解、探索这个世界。于是我们困惑了很久，直到几十年后的今天，才发现了一点点成功的迹象：通过人工智能（Artificial Intelligence）和机器学习（Machine Learning）技术，计算机渐渐能够辨别出物体、人脸、声音、文字——尽管它所用的方式（概率学建模）与我们是如此不同。另一方面，在 SLAM 发展了将近 30 年之后，我们的相机才渐渐开始能够认识到自身的位置，发觉自己在运动——虽然方式还是和我们人类有巨大的差异。不过，至少研究者们已经成功地搭建出种种实时 SLAM 系统，有的能够快速跟踪自身位置，有的甚至能够进行实时的三维重建。

这件事情确实很困难，但我们已经有了很大的进展。更令人兴奋的是，近年来随着科技的发展，涌现出了一大批与 SLAM 相关的应用点。在许多地方，我们都希望知道自身的位置：室内的扫地机和移动机器人需要定位，野外的自动驾驶汽车需要定位，空中的无人机需要定位，虚拟现实和增强现实的设备也需要定位。SLAM 是那样重要。没有它，扫地机就无法在房间自主地移动，只能盲目地游荡；家用机器人就无法按照指令准确到达某个房间；虚拟现实也将永远固定在座椅之上——所有这些新奇的事物都无法出现在现实生活中，那将多么令人遗憾。

今天的研究者和应用开发人员，逐渐意识到了 SLAM 技术的重要性。在国际上，SLAM 已经有近三十年的研究历史，也一直是机器人和计算机视觉的研究热点。21 世纪以来，以视觉传感器为中心的视觉 SLAM 技术，在理论和实践上都经历了明显的转变与突破，正逐步从实验室研究迈向市场应用。同时，我们又遗憾地发现，至少在国内，与 SLAM 相关的论文、书籍仍然非常匮乏，让许多对 SLAM 技术感兴趣的初学者无从一窥门径。虽然 SLAM 的理论框架基本趋于稳定，但其编程实现仍然较为复杂，有着较高的技术门槛。刚步入 SLAM 领域的研究者，不得不花很长的时间，学习大量的知识，往往要走过许多弯路才得以接近 SLAM 技术的核心。

本书全面系统地介绍了以视觉传感器为主体的视觉 SLAM 技术，我们希望它能（部分地）填补这方面资料的空白。我们会详细地介绍 SLAM 的理论背景、系统架构，以及各个模块的主流做法。同时，**极其重视实践**：本书介绍的所有重要算法，都将给出可以运行的实际代码，以求加深读者的理解。之所以这么做，主要是考虑到 SLAM 毕竟是一项和实践紧密相关的技术。再漂亮的数学理论，如果不能转化为可以运行的代码，那就仍是可望而不可即的空中楼阁，没有实际意义。我们相信，实践出真知，实践出真爱。只有实际地演算过各种算法之后，你才能真正地认识 SLAM，真正地喜欢上科研。

自 1986 年提出以来^[4]，SLAM 一直是机器人领域的热点问题。关于它的文献数以千计，想要对 SLAM 发展史上的所有算法及变种做一个完整的说明，是十分困难而且没有必要的。本书中会介绍 SLAM 所牵涉的背景知识，例如射影几何、计算机视觉、状态估计理论、李群李代数等，并在这些背景知识之上，给出 SLAM 这棵大树的主干，而略去一部分形状奇特、纹理复杂的枝叶。我们认为这种做法是有效的。如果读者能够掌握主干的精髓，那么自然会有能力去探索那些边缘的、细节的、错综复杂的前沿知识。

所以，我们的目的是，让 SLAM 的初学者通过阅读本书快速地成长为能够探索这个领域边缘的研究者。另一方面，即便你已经是 SLAM 领域的研究人员，本书也可能有一些你还觉得陌生的地方，可以让你产生新的见解。

目前，与 SLAM 相关的书籍主要有《概率机器人》(*Probabilistic robotics*)^[5]、《计算机视觉中的多视图几何》(*Multiple View Geometry in Computer Vision*)^[3]、《机器人学中的状态估计》(*State Estimation for Robotics: A Matrix-Lie-Group Approach*)^[6] 等。它们内容丰富、论述全面、推导严谨，是 SLAM 研究者中脍炙人口的经典教材。然而就目前来看，还存在两个重要的问题：其一，这些图书的目的在于介绍基础理论，SLAM 只是其应用之一。因此，它们并不能算是专门讲解 SLAM 的书籍。其二，它们的内容偏重于数学理论，基本不涉及编程实现，导致读者经常出现“书能看懂却不会编程”的情况。而我们认为，只有读者亲自实现了算法，调试了各个参数，才能谈得上真正理解了问题本身。

我们会提及 SLAM 的历史、理论、算法、现状，并把完整的 SLAM 系统分成几个模块：视觉里程计、后端优化、建图，以及回环检测。我们将陪着读者一点点实现这些模块中的核心部分，探讨它们在什么情况下有效，什么情况下会出问题，并指导大家在自己的机器上运行这些代码。你会接触到一些必要的数学理论和许多编程知识，会用到 Eigen、OpenCV、PCL、g2o、Ceres 等库^①，掌握它们在 Linux 操作系统中的使用方法。

从写作风格上，我们不想把本书写成枯燥的理论书籍。技术类图书应该是严谨可靠的，但严谨并不意味着刻板。一本优秀的技术书应该是生动有趣而易于理解的。如果你觉得“这个作者怎么这么不正经”，敬请原谅，因为我并不是一个非常严肃的人^②。无论如何，有一件事是可以肯定的：只要你对这门新技术感兴趣，在学习本书的过程中肯定会有所收获！您会掌握与 SLAM 相关的理论知识，你的编程能力也将有明显的进步。在很多时候，您会有一种“我在陪你一起做科研”的感觉，这正是我所希望的。但愿您能在此过程中发现研究的乐趣，喜欢这种“通过一番努力，看到事情顺利运行”的成就感。

好了，话不多说，祝你旅行愉快！

1.2 如何使用本书

1.2.1 组织方式

本书名为“视觉 SLAM 十四讲”。顾名思义，我们会像在学校里讲课那样，以“讲”作为本书的基本单元。每一讲都对应一个固定的主题，其中会穿插“理论部分”和“实

^① 如果你完全没有听说过它们，那么应该感到兴奋，这说明你会从本书中收获很多知识。

^② 你会经常在脚注中发现一些神奇的东西。