

1 SLAM 中的几何与学习方法
2 Geometric Approaches and Neural Networks
3 for SLAM Systems



4

5

李言

Federico Tombari

6

updated 2024 年 7 月 16 日



8 **说明：**文章中的实验结果图片，除特殊说明以外，均来自我们自己的实
9 验。读者可以引用，引用时，请说明出处。

```
10       @misc{gl-SLAM,  
11       author = {Yanyan Li},  
12       title = {SLAM中的几何与学习方法},  
13       year = {2020},  
14       publisher = {GitHub},  
15       howpublished = {\url{https://github.com/yanyan-  
16       li/SLAM-BOOK}  
17       }  
18  
19
```

20 1 书稿概述 (Introduction)

21 1.1 执笔初衷 (Motivation)

22 SLAM (Simultaneous Localization and Mapping) 系统研究的重点是
23 传感器在未知环境中的定位与建图问题。在这个领域，定位部分主要研究传
24 感器的六自由度位姿估计问题，而建图的目标则是还原未知环境的三维信
25 息。SLAM 系统提供的位姿和环境信息可广泛应用于增强现实、虚拟现实、
26 机器人和无人驾驶系统，为机器与环境的智能互动提供支持。随着技术的不断
27 演进，传统视觉 SLAM 的概念已经扩展到包括深度相机、惯性测量单元
28 (IMU)、激光雷达、WIFI 等各种传感器。在 SLAM 领域的发展中，基础
29 环境重建任务也逐渐演变为包括语义重建、地图补全、渲染等多个方向的研究。
30 这些拓展使得 SLAM 系统不仅仅能够还原环境的几何结构，还能够理解
31 环境中的语义信息，从而更好地适应不同应用场景。总体而言，SLAM 技术
32 在不同领域的应用将继续为人工智能和自动化系统的发展提供关键支持。

33 在机器人感知、状态估计和视图几何领域，有许多优秀的中英文读
34 物。《State Estimation for Robotics》从概率和概率系统模型的角度入手，
35 详细阐述了三维空间运动的原理，并提供了运动状态的优化模型。该书通过
36 概率的视角为读者提供了深入的理解，为状态估计领域的学习和研究提供
37 了坚实的基础。《Factor Graphs for Robot Perception》聚焦于因子图优化
38 理论，通过将机器人的感知与定位问题转变成最大后验概率模型，提供了一
39 种与传统 BA (Bundle Adjustment) 优化模型有所不同的新视角。书中
40 介绍的消元排序和增量平滑等技术为深入探索定位与重建技术创造了新的
41 研究方向。此外，书中提到的因子图优化库 GTSAM 为读者提供了实际操
42 作的工具，帮助他们通过编写代码更好地理解书中的理论。《Multiple View
43 Geometry in Computer Vision》是传统视觉几何领域的经典书籍，内容涵盖
44 相当广泛。从相机模型到三角测量，从视觉约束的灭点、灭线到位姿计算的
45 基础矩阵、单应矩阵，该书提供了详细而全面的知识体系。当读者对某一概
46 念不够清晰时，可以在这本书中找到答案。在处理 SLAM 问题中的李群与
47 李代数问题时，《Stochastic Models, Information Theory and Lie Groups》
48 更全面地解释了矩阵李群的基本理论，以及流形范围内的李群与李代数的
49 映射关系。这本书为研究者提供了有关 SLAM 问题中数学基础的深入理解，
50 有助于解决复杂的机器人感知和定位任务。《视觉 SLAM 十四讲》是一本备
51 受中文读者欢迎的精彩而丰富的书籍。作者高翔博士在该书中展现了扎实

的理论知识和深厚的代码功力，以深入浅出的方式向中文读者介绍 SLAM 技术。这本书为读者提供了一个清晰而全面的视角，使他们能够更好地理解和应用 SLAM 技术。同时，对于中文读者，有幸《State Estimation for Robotics》和《Factor Graphs for Robot Perception》都有中文译本，并由电子工业出版社出版。这使得这些优秀的国际著作更容易为中文读者所接触，为他们学习 SLAM 技术提供了极大的帮助。与这些书籍相同的是，本书稿基于作者对 SLAM 的深刻理解，通过重新梳理行业内的书籍和论文，用更简单的语言向读者介绍基础理论和 SLAM 系统开发方法。这种简明扼要的方式有助于降低学习门槛，使更多的读者能够轻松理解和应用 SLAM 技术。这种努力有望为中文读者在机器人感知和定位领域的学术和实践研究提供宝贵的参考和指导。

学界与工业界先行者为推动 SLAM 发展作出了巨大贡献，当今随着设备算力的快速提升、深度学习领域的瞩目发展，诸多与 SLAM 任务相关的技术方向都处于快速发展阶段。因此，有必要帮助读者从海量的论文中梳理出来最新的理论和研究方向。基于此初衷，我们开始了本书的写作与编著工作，本书稿和社区中成书的不同点可以归纳为：

1. 传统书籍从 2D、3D 点的角度来介绍 SLAM 系统中前后端的细节，而本书介绍更多特征（直线、平面、灭点等）在 SLAM 系统中的全生命周期使用方法。 点特征是环境中最常见的信息，因此基于图像中 2D 特征点进行位姿和环境重建使得 SLAM 系统有更好的通用性。(1) 在特殊的，单一的场景中，如室内场景，已经成为很多应用的主流场景。由于室内或人造中存在很多更丰富的结构信息，利用这些结构信息提高位姿估计精度和环境重建效率具有非常很高的价值。(2) 其次，从技术的角度上说，将直线、平面、灭点等结构信息加入到系统的初值估计和状态优化中并非是点处理过程的重复，而是牵涉到不同的图像处理方法，参数化方法，和优化理论，因此对我们更好的理解 SLAM 的运行原理同样有着积极意义。

2. 传统书籍聚焦于纯视觉几何方案的 SLAM 方法，本书介绍深度学习与视觉几何的结合方法以及实现细节。 近年来，深度神经网络已经在物体识别，实例语义分割，深度估计，特征提取，场景补全，渲染等领域取得了令人瞩目的进展，而深度神经网络在这些高维任务 (high-level tasks) 中的能力，正好是传统 SLAM 方法所需要的。本书稿不仅关注深入探究视觉几何在 SLAM 系统中的应用，并提出深度学习方法与传统 SLAM 的融合的策略。

84 **3. 为帮助读者掌握构建 SLAM 系统的构建能力，并丰富本书中所提供的**
85 **理论，我们提供三个自研 SLAM 软件系统，SLAM 模块仿真软件、基于**
86 **结构约束的 RGB-D SLAM 系统和基于 Lidar-RGB 的追踪、理解与重**
87 **建系统。** 为了让读者更容易理解书稿中涉及到的理论，并帮助初学读者更
88 快的开发自己的 SLAM 系统，本书稿使用三个 SLAM 系统来讲述 SLAM
89 系统开发中需要注意的细节。(1)SLAM 仿真系统弱化了特征提取和特征匹
90 配的步骤，直接从 txt 文件中读取点和直线的观测信息，通过绑定 3D 路标
91 点线和 2D 图像的投影建立全局一致的同名点，线信息。在点线观测的基础
92 上，仿真系统介绍如何对点、线、平行线进行参数化，已经不同参数化形式
93 对目标方程收敛特性的影响。

94 1.2 主要内容

95 本书稿分为三个部分，第一个部分介绍 SLAM 问题的定义，第二部分
96 将 SLAM 系统中的基础理论分成多视图几何、深度网络基础、相机模型、初
97 始状态估计和有优化理论，第三部分侧重于 SLA 系统的工程实践，通过代
98 码将系统中的重要概念描述的更加清楚，并为初学者提供快速开发的基础
99 框架和模块。

100 **第一部分** 从概率论的角度，构建状态、场景估计与观测信息的极大似然估
101 计模型。通过假设观测信息服从高斯分布的条件下，我们可以获得熟悉的非
102 线性最小二乘的形式。社区在每一个部分的重大突破，就构成了 SLAM 的
103 发展脉络。

104 **第二部分** 对于不同的传感器，SLAM 系统在位姿估计和重建中应用的方
105 法也大相径庭。对于依赖相机的（包括 RGB-D）系统，位姿估计可以分为
106 $2D-2D$ ， $2D-3D$ 和 $3D-3D$ 几种方式；对于激光雷达，有 $scan-scan$
107 和 $scan-map$ 的位姿估计策略；而对于 IMU 和 GPS，则由预积分和差分
108 法等方法进行位姿估计。而重建部分则包括稀疏，半稠密和稠密几种方式。

109 近年来，深度学习在更高层次的视觉任务中取得瞩目的成绩，如：物体
110 识别，语义分割等，这些课题曾是传统视觉无法或很难解决的任务。正是由
111 于深度学习拓展了我们对高层次视觉任务的想象，越来越多的 SLAM 开始
112 在他们的框架中融合学习的方法，来改进位姿估计的准确程度和环境重建
113 的效果。但是深度学习是一个非常宽广的领域，和 SLAM 相关的课题只是

114 它的一个分支。本书将会挑选和聚焦和 SLAM 相关的深度学习话题，我们
115 希望能通过这本书来介绍 SLAM 系统中使用的几何和深度学习的方法，帮
116 助读者掌握最新的进展。从整体上理解基于几何的 SLAM 系统的技术细节，
117 同时还能了解深度学习相关课题对传统几何 SLAM 的影响。

118 本书首先分析单帧图像入手：单帧图像的几何特征提取，单帧图像的深
119 度学习任务分析。通过相机模型和刚体变换的规则，我们把图像帧联系起
120 来，构建帧与帧之间的关系。然后按照 SLAM 中的每一个重要环节来详细
121 介绍相关的知识。

122 **第三部分** SLAM 领域对工程实践能力有较高的要求，为满足应用的实时
123 性要求，绝大多数 SLAM 系统都是通过 C++ 程序来实现的，其中部分系
124 统为了进一步提高运行效率，也会考虑使用指令集进行加速。因此，我们的
125 第三部分将会在 SLAM 项目的工程实践中，帮助读者加深对前面章节介绍
126 的理解，并能将这些理论快速的应用在今后的系统开发任务中。

127 1.3 面向读者

128 几何与学习就像是 SLAM 系统的一双翅膀，视觉几何、优化理论搭建
129 了鲁棒的、高精度的定位与重建系统，而深度学习相关的理论和应用帮助
130 SLAM 突破传统图像学的处理困境，让 SLAM 系统走向更加高维的任务，
131 在几何与深度学习的加持下，SLAM 有望展现出更加智能的特点。

132 本书稿聚焦在以相机为核心或相机参与的 SLAM，一方面是视觉 SLAM
133 领域具有更加成熟的研究成果，另一方面，视觉信息同样也是深度学习方法
134 大展身手的场景。本书在介绍基础理论的同时，希望给读者带来以下几个方
135 面的思考：

- 136 • 如何在位姿估计、三维重建和优化模块中使用点、线、面特征来提高
137 模块表现；
- 138 • 深度学习在诸多视觉任务中取得了令人瞩目的表现，哪些学习信息可
139 以被融合到 SLAM 系统中，对于几何与学习融合的项目，我们该如何
140 进行去伪存真；
- 141 • 未来的 SLAM 将走向何处。

142 我们希望这本书稿能够给为研究 SLAM 的学生和相关从业者系统介绍

- ¹⁴³ 几何特征与学习信息在 SLAM 系统中的提取和使用过程。更简单的说，如
¹⁴⁴ 果诸位对《SLAM 视觉十四讲》感兴趣，那这本书稿就是为你而写的。