



北京化工大学

BEIJING UNIVERSITY OF CHEMICAL TECHNOLOGY

COMPUTING METHODS

---

## 自动化科学导论课程报告

---

李昊

### 目录

第 1 章 自动化大类各专业之间的相互关系	1
第 2 章 学习的总结，体会和感受	1
2.1 学习的总结	1
2.2 学习的体会	1
2.3 学习的感受	1
第 3 章 视觉 VSLAM 综述	1
3.1 AGV 的导航 (Navigation)	1
3.2 SLAM 定义	2
3.3 V-SLAM 的框架	2
3.4 基于单个视觉传感器的 V-SLAM 方法	3
3.4.1 特征法	3
3.4.2 直接法 (稠密、半稠密)	4

3.4.3 经典 VSLAM 方法总结 . . . . .	5
3.5 多传感器融合的 V-SLAM 方法 . . . . .	6
3.5.1 多传感器融合:VINS-Mono . . . . .	6
3.6 基于深度学习的 V-SLAM 方法 . . . . .	7
参考文献	7

创建日期: 2020 年 3 月 25 日  
更新日期: 2020 年 5 月 31 日

# 第 1 章 自动化大类各专业之间的相互关系

其实吧，说专业差距大，其实也没那么大。举图像识别做例子：大家都做图像识别，可能在研究生阶段区别比较大，但是对于本科来说差别真的几乎没有。如果一定要做区分的话

- 自动化：偏向系统集成，控制算法，感知算法
- 测控：就我们学校，MEMS 无疑
- 通信：各种雷达信号处理

# 第 2 章 学习的总结，体会和感受

## 2.1 学习的总结

自动化学科是一门多学科交叉的高技术学科。研究内容广，几乎无所不包。但是，控制、信息、系统是其核心。自动化学科知识体系中的核心知识是：控制与智能、对象和建模、系统和工程等三个知识领域，也是自动化学科与其他学科的最重大区别。

## 2.2 学习的体会

选自动化就对了！选自动化就对了！选自动化就对了！千万不要搞过程控制！千万不要搞过程控制！千万不要搞过程控制！更加坚定了自己选择自动化和进行移动机器人以及机器人感知相关方向研究的意志。

## 2.3 学习的感受

与其说选择专业，我觉得更应该选择研究方向后再决定专业，因为每个专业的研究方向都有很大的重叠部分。

# 第 3 章 视觉 VSLAM 综述

## 3.1 AGV 的导航 (Navigation)

导航技术是 AGV 技术的研究核心，没有导航功能的 AGV 系统是很难想象的。举个例子，如果 AGV 系统想要在工厂等工作场所运输货物，它必须要在不触碰到任何障碍物的情况下运动到不同的地点。随着信息化技术的提升，CPU 算力的提升和传感器的升级，AGV 的柔性路径导航吸引了越来越多的研究者进行研究，产业化应用越来越

越广。柔性路径导航技术应用较多的是视觉与激光导航、惯性 (IMU) 导航 [1]、基于 WIFI/UWB 等无线信号定位 [2] 的导航技术等。机器人实现其自主导航的前提是机器人自身位姿的确定，出于对定位精度的要求以及现实环境的复杂化，目前大多数导航系统都是由上述两种或者多种导航方式的结合组合导航技术 [3]。

## 3.2 SLAM 定义

同步定位与地图构建 (SLAM 或 Simultaneous localization and mapping) 是一种概念：希望机器人从未知环境的未知地点出发，在运动过程中通过重复观测到的地图特征（比如，墙角，柱子等）定位自身位置和姿态，再根据自身位置增量式的构建地图，从而达到同时定位和地图构建的目的。

在误差和噪音条件下，定位和地图构建技术上的复杂度不支持两者同时获得连续的解。即时定位与地图构建 (SLAM) 是这样一个概念：把两方面的进程都捆绑在一个循环之中，以此支持双方在各自进程中都求得连续解；不同进程中相互迭代的反馈对双方的连续解有改进作用。

同步定位与地图构建 (SLAM) 被定义为以下问题：在建立新地图模型或者改进已知地图的同时，在该地图模型上定位机器人。实际上，这两个核心问题如果分开解决，将毫无意义；必须同时求解。即通过控制数据  $u_{1:t}$  和观测数据  $z_{1:t}$  来求解位姿和地图的联合概率分布

$$p(x_{1:t}, m | z_{1:t}, u_{1:t-1}) \quad (3.1)$$

## 3.3 V-SLAM 的框架

视觉 SLAM 由前端（视觉里程计）、后端（位姿优化）、回环检测、建图 4 个部分组成。

- 视觉里程计 (Visual Odometry): 仅有视觉输入的姿态估计 [4]
- 后端优化 (Optimization): 后端接受不同时刻视觉里程计测量的相机位姿，以及闭环检测的信息，对它们进行优化，得到全局一致的轨迹和地图 [5]
- 闭环检测 (Loop Closing): 指机器人在地图构建过程中，通过视觉等传感器信息检测是否发生了轨迹闭环，即判断自身是否进入历史同一地点 [5]
- 建图 (Mapping): 根据估计的轨迹，建立与任务要求对应的地图 [5]

流程如下图所示：

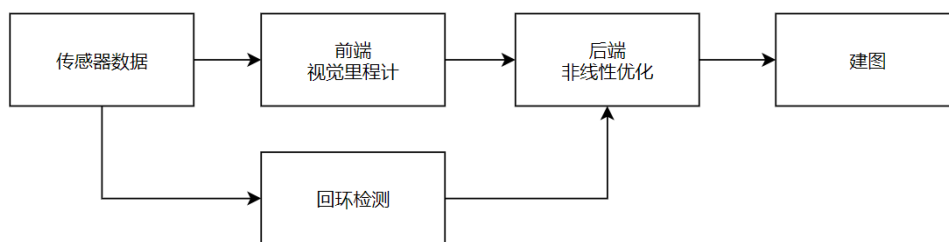


图 3.1: VSLAM 流程图 [5]

### 3.4 基于单个视觉传感器的 V-SLAM 方法

视觉 SLAM 算法可根据利用图像信息的不同分为特征法（稀疏法）和直接法（半稠密法、稠密法）。

#### 3.4.1 特征法

早期的单目视觉 SLAM 的实现是借助于滤波器而实现的 [6,7]。MonoSLAM [6] 是第一个实时的单目视觉 SLAM 系统。MonoSLAM 利用扩展卡尔曼滤波（extended Kalman filter, EKF）来实现同时定位和地图创建。但也正由于 EKF 的引入，导致 SLAM 算法会有计算复杂度的上升以及线性化带来的不确定问题。该方法的缺点是：场景窄、路标数有限、稀疏特征点易丢失等。



图 3.2: MonoSLAM 效果图

随后，基于关键帧的单目视觉 SLAM 发展起来。PTAM（parallel tracking and mapping [8] 是最具代表性的一个，其首次提出了关键帧（keyframe）的概念，在保持准确度的同时，不用精细地处理每一帧，显著提高了 SLAM 的实时性。同时，PTAM 首次实现了跟踪和建图的并行化，区分出前后端（跟踪需要实时响应图像数据，地图优化放在后端进行），为后续的 SLAM 方法奠定了基本模型。

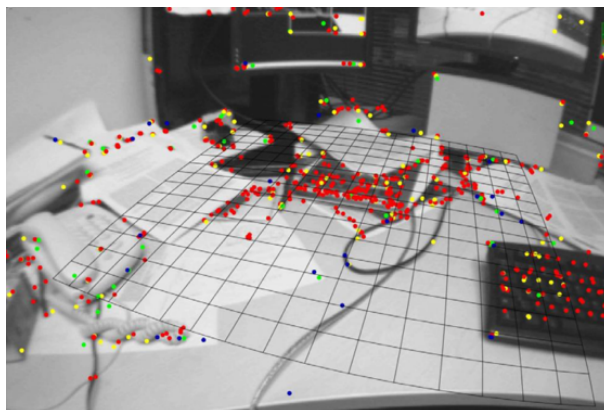


图 3.3: PTAM 效果图

ORB-SLAM2 算法 [9, 10] 是首个基于单目，双目和 RGB-D 相机的开源 SLAM 方案，在大规模的、小规模、室内室外的环境都可以运行。其利用 ORB 而非 SURF [11] 或者 SIFT [12] 特征进行运算，因此可以提高效率。同时，ORB-SLAM2 又创新性地使用三个线程完成 V-SLAM（将整个流程分成定位、创建地图、闭环三个部分），得益于 AMD 推动了多核心 CPU 的发展 [13–15]，ORB-SLAM2 的三线程可以更好地发挥 CPU 的性能从而提高 V-SLAM 的实时性。

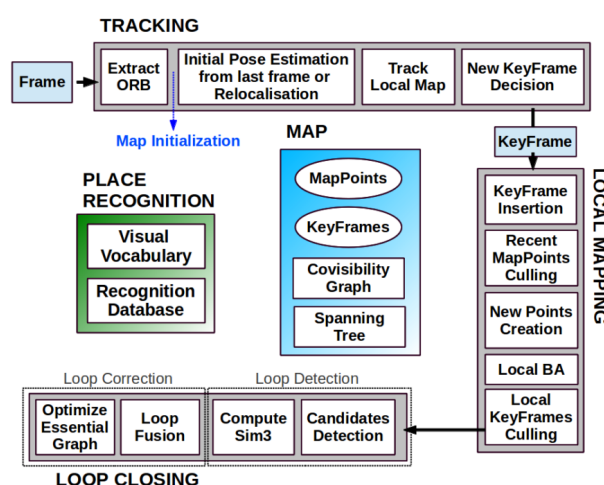


图 3.4: ORB-SLAM 主要分为三个线程进行，也就是如图所示的，分别是 Tracking、LocalMapping 和 LoopClosing [10]

### 3.4.2 直接法（稠密、半稠密）

SVO (semi-direct monocular visual odometry) 算法 [16] 主要面向无人机航拍场合，它将特征点法与直接法结合，跟踪关键点，不计算描述子，根据关键点周围的小图像块的信息估计相机的运动。主要分运动估计线程和地图构建线程两个线程。综上所述，由于 SVO 的目标是无人机平台，所以其没有闭环、没有重定位、也没有建图功能，但这

也使得它的实时性非常强（可以达到 100FPS）。

RGBD（RGB-depth map）能够同时获得图像信息和深度信息，这大大地简化了三维重建和 SLAM 的过程。随着微软公司 Kinect 深度相机的普及化，越来越多厂商（Intel，ZED，ASUS）也开始研制低成本、高精度的深度摄像头这导致深度摄像头的价格开始直线下降，于是越来越多的 SLAM 算法开始利用深度摄像头进行研究。

RTAB-MAP 开源库 [17] 在 2013 年建立，直至 2019 年已经发展成一个跨平台的独立 C++ 库和一个 ROS 包，这说明了 RTAB-MAP 本质上更偏向工程应用。其通过内存管理方法来实现回环检测，同时限制了地图的大小以保证建图的实时性。同时，通过图3.5我们可以看到 RTAB-MAP 开放了接口允许其他里程计数据（包括第三方视觉里程计）的接入，同时也可以选择输出：OctoMap（3D 栅格地图）[18]；稠密点云地图 [19]；2D 栅格地图 [20]，对第三方 VSLAM 的算法评判提供了量化评判基础。

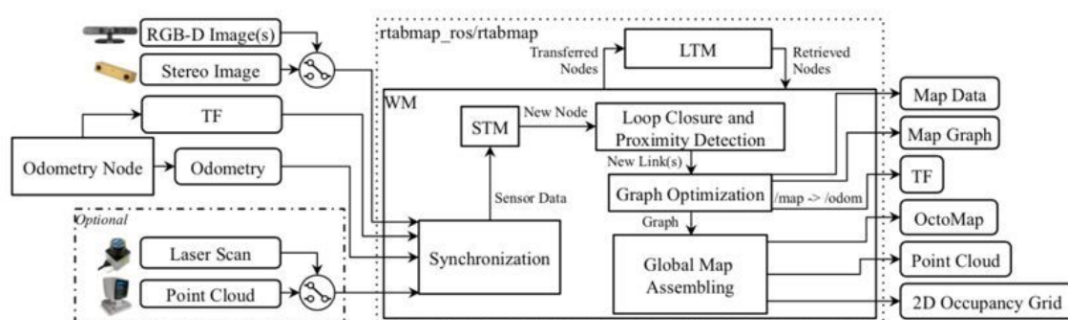


图 3.5: RTAB-Map 框架

### 3.4.3 经典 VSLAM 方法总结

由于篇幅所限，我们无法一一列举所有 VSLAM 经典方法。于是我们对大部分的知名 VSLAM 算法概要汇总整理至表3.1。

相关文献	方法	所用传感器	算法名称	发表年份
[6]	稀疏法 (特征点)	单目 SLAM	MonoSLAM	2007
[8]	稀疏法 (特征点)	单目 SLAM	PTAM	2007
[21]	半稠密法	单目 SLAM	LSD-SLAM	2014
[9]	稀疏法 (特征点)	单目 SLAM	ORB-SLAM	2015
[16]	半稠密法	单目 SLAM	SVO	2014
[10]	稀疏法 (特征点)	双目 SLAM	ORB-SLAM2	2017
[22]	稀疏法 (特征点)	双目 SLAM	S-PTAM	2015
[23]	稠密法	RGBD-SLAM	RGBD-SLAMv2	2014
[17]	稠密法	RGBD-SLAM	RTAB-SLAM	2019

表 3.1: 传统 SLAM 算法的总结

### 3.5 多传感器融合的 V-SLAM 方法

传统的 V-SLAM 方法的重大挑战是要同时实现高帧率和低延迟跟踪，这需要极大的计算成本（也意味着计算和耗电量的不足）。同时，如果相机帧率过高，那么曝光时间就越短，入射光线越少。于是，一个更好的工程方法被提出来：利用高速 IMU 来取代 V-SLAM 中 VIO 流程中的“ $I$ ”，下表很好地总结了 VSLAM 和 IMU 如何相互补充。

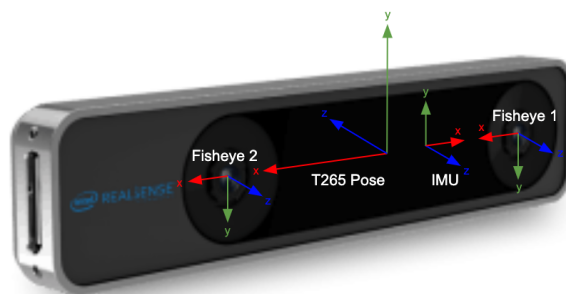


图 3.6: 集成了 IMU 模块的深度摄像头

比较项目	视觉里程计	IMU 里程计
更新频率	低	高
长时间稳定性	高	低
场景依赖度	高	低
重定位能力	高	低
CPU 功耗	高	低

表 3.2: 视觉里程计和 IMU 的特性对比

#### 3.5.1 多传感器融合:VINS-Mono

该算法 [24] 是一种鲁棒且通用的单目视觉惯性状态估计器。采用一种基于紧耦合、非线性优化的方法，通过融合预积分后的 IMU 测量值和特征观测值，获得高精度的视觉惯性里程计。结合紧耦合方法，回环检测模块能够以最小的计算代价实现重定位。处理视觉和惯性测量的最简单的方法是松耦合的传感器融合 [25,26]，其中 IMU 被视为一个独立的模块，用于辅助运动的视觉结构 (sfm) 获得的纯视觉位姿估计。融合通常由扩展卡尔曼滤波 (EKF) 完成，其中 IMU 用于状态传播，而视觉位姿用于更新。



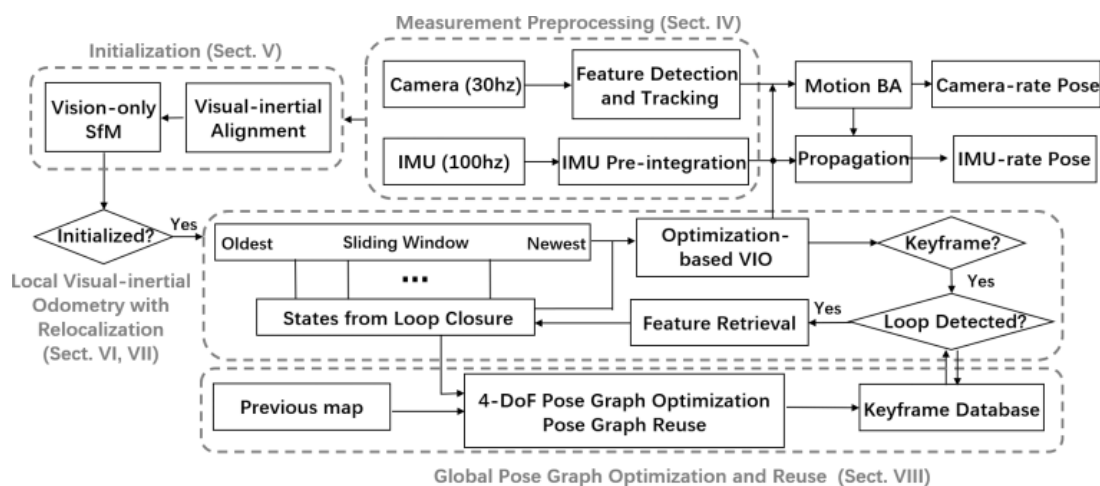


图 3.7: VINS-Mono 总体架构图

### 3.6 基于深度学习的 V-SLAM 方法

对于经典 V-SLAM，人工设计的稀疏图像特征当前有很多局限性，一方面如何设计稀疏图像特征最优地表示图像信息依然是计算机视觉领域未解决的重要问题 [27]，另外一方面系数图像特征在应对环境变化（光照变化，动态目标运动，相机参数改变，稀疏纹理等）方面依然有许多挑战。而基于深度学习的 VSLAM 方法拥有更高的准确率和一致性。下图给出了对比关系：

比较项目	传统 SLAM 算法	基于深度视觉的 SLAM 算法
模型参数调整难易程度	数据少，调参周期短	数据多，调参周期长
魔性物理含义	直观意义明确	缺少直观意义
模型泛化能力	信息利用不充分，泛化能力弱	信息利用充分，泛化能力强
适应能力	迁移能力弱	迁移能力强
设计流程	特征设计与分类训练分离	同步完成特征设计和分类器训练

表 3.3: 传统 SLAM 算法与基于深度学习的 SLAM 算法对比 [28]

## 参考文献

- [1] O. Wongwirat and C. Chaiyarat, "A position tracking experiment of mobile robot with inertial measurement unit (imu)," in *ICCAS 2010*, pp. 304–308, 2010.
- [2] S. Krishnan, P. Sharma, Z. Guoping, and O. H. Woon, "A uwb based localization system for indoor robot navigation," in *2007 IEEE International Conference on Ultra-Wideband*, pp. 77–82, 2007.

- [3] S. Ma, Y. Zhang, Y. Xu, B. Wang, J. Cheng, and Q. Zhao, “Indoor robot navigation by coupling imu, uwb, and encode,” in *2016 10th International Conference on Software, Knowledge, Information Management Applications (SKIMA)*, pp. 429–432, 2016.
- [4] D. Nister, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, pp. I–I, 2004.
- [5] 高翔, 视觉 SLAM 十四讲. 05 2020.
- [6] Russo, Ludovico, Rosa, Stefano, Bona, Basilio, Matteucci, and Matteo, “A ros implementation of the mono-slam algorithm,” vol. 4, pp. 339–351, 01 2014.
- [7] A. J. Davison, “Slam with a single camera,” *SLAM/CML Workshop at ICRA 2002*, 2002.
- [8] Klein, Georg, and D. Murray, “Parallel tracking and mapping for small ar workspaces.,” *Mixed and Augmented Reality*, 2007.
- [9] Mur-Artal, Raúl, Montiel, J. M. M., Tardós, and J. D., “ORB-SLAM: a versatile and accurate monocular SLAM system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [10] Mur-Artal, Raúl, Tardós, and J. D., “ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras,” *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [11] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool], “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.
- [12] Lowe and D.G., “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, 2004.
- [13] Wikipedia, “Multi-core processor — Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/w/index.php?title=Multi-core%20processor&oldid=958877109>, 2020. [Online; accessed 26-May-2020].
- [14] K. Hasan, J. Antonio, and S. Radhakrishnan, “A new composite cpu/memory model for predicting efficiency of multi-core processing,” 02 2014.

- [15] B. Schauer, “Multicore processors - a necessity,” *ProQuest Discovery Guides*, 01 2008.
- [16] C. Forster, M. Pizzoli, and D. Scaramuzza, “Svo: Fast semi-direct monocular visual odometry,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 15–22, 2014.
- [17] M. Labbé and F. Michaud, “Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation,” *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.
- [18] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, “OctoMap: An efficient probabilistic 3D mapping framework based on octrees,” *Autonomous Robots*, 2013. Software available at <http://octomap.github.com>.
- [19] Wikipedia, “Plagiarism — Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/w/index.php?title=Plagiarism&oldid=957645059>, 2020. [Online; accessed 27-May-2020].
- [20] D. V. Lu, D. Hershberger, and W. D. Smart, “Layered costmaps for context-sensitive navigation,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 709–715, 2014.
- [21] J. Engel, T. Schöps, and D. Cremers, “Lsd-slam: Large-scale direct monocular slam,” in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 834–849, Springer International Publishing, 2014.
- [22] T. Pire, T. Fischer, J. Civera, P. De Cristóforis, and J. Jacobo berlles, “Stereo Parallel Tracking and Mapping for robot localization,” in *Proc. of the International Conference on Intelligent Robots and Systems (IROS)*, pp. 1373–1378, September 2015.
- [23] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, “3-d mapping with an rgb-d camera,” *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, 2014.
- [24] T. Qin, P. Li, and S. Shen, “Vins-mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [25] S. Tsao and S. Jan, “Observability analysis and performance evaluation of ekf-based visual-inertial odometry with online intrinsic camera parameter calibration,” *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2695–2703, 2019.

- [26] J. Jackson, K. Brink, B. Forsgren, D. Wheeler, and T. McLain, “Direct relative edge optimization, a robust alternative for pose graph optimization,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1932–1939, 2019.
- [27] Fuentes-Pacheco, J. Ruiz-Ascencio, and J. . Rendón-Mancha, “Visual simultaneous localization and mapping: a survey,” *Artificial Intelligence Review*, 2015.
- [28] Z. Yang, L. Guoliang, T. Guohui, L. Yong, W. Ziren, Z. Wei, and L. Junwei, “A survey of visual slam based on deep learning,” *ROBOT*, vol. 39, no. 6, p. 889, 2017.