

无人机视觉 SLAM 环境感知发展研究

苗升, 刘小雄, 黄剑雄, 居玉婷, 章卫国

(西北工业大学 自动化学院, 西安 710072)

摘要: 无人机在进行搜索救援等高级任务的时候, 往往需要确定自己的位置和环境信息; 仿照于人类通过视觉感知环境信息, 视觉 SLAM 是计算机视觉领域里面通过视觉传感器感知环境的信息并快速跟踪自身的位置和建立环境地图的一种前沿技术; 文章首先阐述了 VSLAM 的重要组成部分: 前端处理 (特征点法和直接法)、数据关联、后端优化算法 (滤波方法和优化方法) 和建图; 然后总结了一些在无人机上成功应用的典型 VSLAM 算法, 以及在 VSLAM 发展的 30 多年的时间里涌现出许多出色的方案和研究机构; 接着论述了当前用于无人机 VSLAM 发展的几个重点问题, 多无人机协同的 C-SLAM、深度学习和语义分割在 SLAM 中的应用、以视觉惯导为代表的多传感器融合 SLAM; 最后, 对 VSLAM 方法进行总结, 给出了未来的发展方向, 希望对后续研究提供指导和帮助。

关键词: 视觉导航; 视觉 SLAM; 无人机; 非线性优化

Research on Development of UAV Visual SLAM Environment Perception

MIAO Sheng, LIU Xiaoxiong, HUANG Jianxiong, JU Yuting, ZHANG Weiguo

(College of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: When UAV perform advanced tasks such as search and rescue, they often need to determine their own location and environmental information. Modeled on humans perceiving environmental information through vision, visual SLAM (visual simultaneous localization and mapping, VSLAM) is a cutting-edge technology in the field of computer vision that uses visual sensors to perceive environmental information and quickly track its own location and build environmental maps. This article first explains the important components of VSLAM: front-end processing (feature point method and direct method), data association, back-end optimization algorithms (filtering methods and optimization methods) and mapping; then summarizes some of the successful applications on UAVs. Typical VSLAM algorithm, and many outstanding programs and research institutions that have emerged during the development of VSLAM for more than 30 years; Then it discusses several key issues currently used in the development of UAV VSLAM, the application of multi-UAV collaborative C-SLAM, deep learning and semantic segmentation in SLAM, and multi-sensor fusion SLAM represented by visual inertial navigation; Finally, the VSLAM method is summarized, and the future development direction is given, hoping to provide guidance and help for follow-up research.

Keywords: visual navigation; VSLAM (visual simultaneous localization and mapping); UAV; nonlinear optimization

0 引言

近年来, 随着无人机相关产业的大力发展, 其在非国防和民用等诸多领域扮演着越来越重要的作用。室内场景飞行与感知周围环境是旋翼无人机经常需要面临的情景, 如: 室内救援、室内搜索及环绕障碍等。在这些复杂的环境下旋翼无人机常常难以进行自我导航, 由于信号被遮挡或者存在缺失, 会导致传统的 GPS 导航定位方法失败。无人机具有轻量化的要求, 这便使得难以在上面防止过多过重的传感器, 而在没有先验的地图信息之下, 如何感知未知的周围环境和确定自己当前的位置和姿态便是我们首先需要解决的问题之一。准确的位姿的确定对于无人机在以后执行高级任务, 如: 路径规划、避障穿越、三维重建、

目标跟踪等方面具有十分重要的意义。

同时定位和建图 (SLAM, simultaneous localization and mapping)^[1-2]便是解决这个问题目前最受欢迎的方法, 它是由 Smith^[3]和 Cheeseman 在 1986 年在 IEEE 机器人技术和自动化会议首次提出的。SLAM 研究方法根据传感器的不一样, 可以分为基于激光雷达的 SLAM 和基于计算机视觉的 SLAM。随着视觉传感器的广泛使用和成本的降低, 基于视觉的 SLAM 逐渐引起人们的注意, 成为目前研究的一大热点, 依据实际使用的摄像头不同类型可以分为单目、双目、RGB-D 的 SLAM 等。

经典的 SLAM 框架包括: 前端视觉里程计、后端优化、建图、回环检测等 4 个主要部分。

收稿日期: 2020-11-04; 修回日期: 2020-12-28。

基金项目: 航空科学基金资助项目(201905053003); 国家自然科学基金项目(62073266)。

作者简介: 苗升(1997-), 男, 陕西西安人, 硕士研究生, 主要从事无人机视觉导航算法、飞行控制算法方向的研究。

刘小雄(1973-), 男, 陕西周至人, 博士, 副教授, 主要从事飞机飞行控制、导航与制导算法方向的研究。

章卫国(1956-), 男, 西安市人, 博士, 教授, 主要从事飞行控制技术方向的研究。

引用格式: 苗升, 刘小雄, 黄剑雄, 等. 无人机视觉 SLAM 环境感知发展研究[J]. 计算机测量与控制, 2021, 29(8): 1-6, 41.

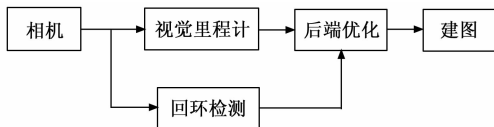


图 1 SLAM 的基本框架

SLAM 前端又称为视觉里程计 (VO, visual odometry), 主要分为基于图像特征变化的特征点法和基于图像灰度的直接法两类。前者基于特征点, 通过最小化两帧图像之间的重投影误差 (bundle adjustment), 后者是根据像素的亮度信息估计运动位姿, 不需要计算描述子和关键点, 在灰度假设不变的条件下, 通过最小化光度误差来求解机器人的运动位姿。

在后端优化的部分, 由于 SLAM 的运动方程和观测方程都是非线性函数, 因而我们主要是采取了非线性优化的方法来优化误差函数, 大致可以分为以只考虑 k 时刻的状态和 $k-1$ 时刻状态的具有一阶马尔可夫性的扩展卡尔曼滤波^[4-5]的方法和考虑 k 时刻与之前所有状态的非线性优化的方法。

早期的 SLAM 最开始的应用载体是一些地面移动机器人, 利用卡尔曼滤波或者粒子滤波去融合轮式里程计和传感器的信息。比较早的将 SLAM 技术用到无人机上是在 2003 年由 Aouf^[6] 等人, 他们将相机固定在飞机朝下, 控制输入信息传递到飞行器而不再是里程计。2006 年 Angeli 和 Filliat^[7] 等采用 2D 的纯视觉 SLAM 应用在飞艇和无人机上面。在视觉处理方面, 他们关联了 Harries 角点检测器和 SIFT 描述子, 同时为了提高地图的精度, 将径向的畸变系数也作为了卡尔曼滤波器的参数, 实验结果表明该 2D-SLAM 系统可以提供稳定的位姿估计。2014 年由 Forster 等人提出的著名的 SVO 算法^[8] 应运而生, 该 SLAM 算法专门以无人机的俯视相机为目标应用平台, 舍弃了对于特征点描述子的计算, 并且在关键点周围 4×4 的区域内进行块匹配, 因而速度极快且在无人机平台上实验效果显著。

无人机的 SLAM 相关工作已经有很多, 除了上述研究成果外, 吴显亮等^[9]、王希彬^[10] 和赵良玉^[11] 从不同角度对 SLAM 技术进行相关总结。基于上述研究成果, 本文从 SLAM 的基本结构展开, 讨论了 SLAM 的基本框架、相关算法以及 SLAM 领域最新的研究进展, 并对未来 SLAM 技术的发展进行了展望。

1 视觉 SLAM 的基本框架

1.1 前端提取

视觉的特征是由关键点 (Detectors) 和描述子 (Descriptor) 两部分组成的。被检测的对象一般有点特征, 直线特征以及边缘, 轮廓特征等。关键点常常是物体边界方向突然改变的点或两个或多个边缘段之间的交点。当图像局部或者全局存在亮度和光照变化的时候, 关键点仍然稳定, 在 SLAM 中应用较多, 故在 SLAM 的特征提取几乎提取的都是特征点 (Keypoint)。经典关键点检测器有: Har-

ris 角点检测器 (HARRIS, harris corner detector,)、SIFT 角点检测器 (SIFT, scale invariant transform)、SURF 角点检测器 (SURF, speed-up robust features) 等。现代关键点检测器有: FAST 角点检测 (FAST, features from accelerated segment test)、ORB 角点检测器^[12] (ORB, oriented FAST and rotated BRIEF) 等。目前效果比较好的特征点是 ORB 特征点, ORB 是在 FAST 关键点的检测速度快的基础之上添加了尺度和旋转的特征, 可以称为“Oriented FAST”——改进的 FAST 角点, 其快速性可以满足 SLAM 实时的要求。其尺度性是在构建的金字塔的每一层去检测角点, 而旋转不变性是通过灰度质心法去实现 (Intensity Centroid) 的。

在图像中对于寻找到关键点之后, 我们需要一种方法将不同的图像关键点加以对比, 寻找到其中相似的关键点。描述子作为一种描述关键点周围的环境信息被人为设计出来。目前广泛被应用的描述子有基于梯度 HOG 描述子^[13] 和二进制 BRISK 描述子^[14]。HOG 的基本思想是通过物体在局部图像邻域中的强度梯度分布来描述物体的结构, 计算代价昂贵。实际工程中, 我们采用计算代价相对较少的二进制描述子, 其核心思想是仅仅依赖强度信息 (即图像本身), 并将关键点周围的信息编码为一串二进制数字。有了关键点周围的描述子, 我们便可以得到我们提取关键点周围的环境信息。

得到了带有描述子的关键点, 接下来便可以在不同的图像之中进行特征匹配。计算关键点描述子的汉明距离 (hamming distance), 来表征描述子的相似程度。当检测到特征点数量非常大时, 快速近似最近邻 (FLANN) 算法^[15] 能够满足 SLAM 的实时性需求。

在 SLAM 领域, 特征点法是比较流行的方法, 但提取特征点存在着耗时长、图像信息丢失严重等缺陷。针对在无人机的应用领域, 一种与特征法相对应的直接法也渐渐地受到了人们地重视。直接法基于灰度不变假设, 即认为同一个关键点在两幅图片之中, 其灰度值应该是不会变化的。在直接法中, 我们可以选择不计算关键点周围的描述子而得到其他图像中的配对点 (半直接法), 通过光流^[11] 来跟踪关键点的运动, 从而估计相机的运动; 或者既不计算关键点也不计算描述子 (直接法), 而是选择一些有明显梯度变化的像素 (半稠密直接法), 或者干脆选择所有的像素点 (稠密直接法) 参与计算, 之后根据最小化光度误差可以求解位姿。直接法对于场景的特征信息没有要求, 只需要场景中存在明暗的变化情况即可。

1.2 VO 里面的数据关联与 PnP 问题

不管是特征法, 还是直接法里面的光流追踪, 我们都可以得到在相邻图片中已经匹配好的特征点。如何由这相邻的两幅图片求解出相机的位姿变换是 SLAM 前端里程计的核心问题。根据匹配点的维数可以划分为: 2D-2D、2D-3D、3D-3D 三大类问题。在 SLAM 里面, 3D 点的深度信息可以由 RGB-D 相机给出, 或者在单目相机中, 可

以通过三角化去恢复 2D 点的深度信息。2D—2D 可以通过经典的八点法去求解本质矩阵 $E^{[16]}$, 恢复出旋转信息和平移信息。2D—3D 是在 SLAM 最常见的一种匹配方式, 它是在知道关键点的 3D 空间位置与其在图像上对应的 2D 点的像素坐标的前提下, 去求解此时的相机位姿, 这便是一个 PnP 问题。在 SLAM 中, 我们可以利用 3 对 2D—3D 的匹配点信息, 恢复出此时的相机位姿。大致的做法, 分为两类, 一种是线性化几何约束的解法, 包括 EPnP 等算法, 另一种是基于计算机视觉的非线性的解法, 将其转化为一个 BA (bundle adjustment) 问题。为了保障求解的精度, SLAM 主要采取的做法是构建最小二乘方法, 用 ξ 表示相机的位姿, P_i 表示第 i 个点对应的空间 3D 坐标, u_i 是第 i 个点对应的像素 2D 坐标:

$$\xi = \arg \min_{\xi} \sum_{i=1}^n \|u_i - f(P_i, \xi)\|_2 \quad (1)$$

通过不断调整相机的位姿 ξ 去使 3D 点重投影到图片上的 2D 点 $f(P_i)$ 与图片上该对应关键点所在的 2D 点 u_i 的误差函数的二范数最小, 来达到优化的目的。

1.3 后端优化

前端的视觉里程计仅仅可以给出短时间内的轨迹和地图, 随着时间的累积, 无人机的位姿不确定性会越来越大, 无法在实际中进行应用。因而, 后端的优化部分就是为了解决在较长时间, 数据量很大的情况下去求解最优的轨迹和地图。按照优化的思想大致可以分为基于滤波的方法和基于优化的方法。

1.3.1 基于滤波的后端

2007 年帝国理工学院的 A. J. Davison 提出了世界上第一个可以实时应用的 SLAM 系统——Mono-SLAM^[17], 这是第一个基于 EKF 方法的实时的单目 SLAM 系统, 虽然初步解决了实时的问题, 能够在线创建稀疏的地图, 但是因为存在较大的漂移量, 效果不是特别好。Montemerlo^[18-19] 则分别研究了基于粒子滤波器 (PF, particle filter) 的 FastSLAM 的方法。在 FastSLAM 包含了机器人定位和特征标志位置的两个过程。粒子滤波器中可能的运动路径用粒子表示, 1 个粒子对应着 1 种可能, 每条路径的好坏是由观测信息计算得到的粒子权重进行评价。在路标的估计当中, FastSLAM 则采用了传统的 EKF 进行估计。Mourikis 提出了一种融合 IMU 和视觉信息的滤波算法——MSCKF (multi-state constraint Kalman filter)^[20], 该算法不同于 EKF 将特征点加入状态变量, 而是将不同时刻的相机位置 p 和姿态 (四元数 q 表示) 加入状态变量进行优化更新。由于状态变量的维度大幅度的降低, 且加入了 IMU 信息, 因而 MSCKF 可以适用于更剧烈的运动, 且鲁棒性也有了显著性的提高。

1.3.2 基于优化的后端

2007 年, 同步追踪和建图算法^[21] (PTAM, parallel tracking and mapping) 这一非线性优化的 SLAM 首次成功应用。该算法由英国牛津大学的 G. Klein 和 D. Murry 提出

并开源。PTAM 的特点如下:

1) PTAM 首次引入了 SLAM 当中的前后端的概念, 提出了跟踪部分是需要实时的响应数据, 而对于地图优化的部分则没有必要进行实时的计算。

2) PTAM 引入了关键帧的机制, 减少了需要优化的图像数量, 确保了算法的实时性。

3) PTAM 是第一个采用 BA 非线性优化作为后端优化的方案, 在 PTAM 提出之后 SLAM 逐渐转向了非线性优化为主导的后端。在 PTAM 提出之后, 后端部分优化成为了主流的方法。

此后, 研究者们逐渐注意到 BA 问题系数矩阵具有一定的稀疏性 (即 Jacobian 矩阵中只有当前待优化变量的对应位置处为非零项, 其余位置均为零), 进而可基于 Schur 消元将无关变量边缘化, 实现 BA 中线程方程组的稀疏求解^[22]。稀疏化处理虽然可以缓解非线性优化的计算负担, 但计算规模依然会随着时间而累积。OKVIS^[23] 它是由苏黎世联邦理工大学 S. Leutenegger 等人提出的基于双目+惯导的视觉里程计, 是一种基于滑动窗口的非线性优化紧耦合算法。该算法采用滑动窗口方法约束优化的方法满足系统的实时性要求, 但它没有重定位和闭环检测等环节, 而且 CPU 的占有率较高。Vins-mono^[24] 是香港科技大学沈劭劭老师开源的单目视觉惯导 SLAM 方案, 其推荐的平台是微小旋翼无人机。该算法也是融合了惯导的单目视觉系统, 是目前很出名的无人机 SLAM 算法。该算法采用光流法进行特征点的追踪, IMU 以更快的速率进行预积分^[25] (IMU pre-integration), 把关键帧之间的 IMU 测量值积分成相对运动的约束, 两者的数据进行了视觉惯性联合初始化, 后端是基于滑动窗口的重投影误差和 IMU 残差非线性优化和基于词袋模型的回环检测。

非线性优化因为数据量的巨大, 以至于学者们认为该方法在本世纪初都无法实现机器人的实时 SLAM, 所以减少计算量就显得尤为重要。无人机的视觉导航, 若我们仅仅关注无人机的定位问题, 那么在后端的优化部分进行路标点的优化便会耗费大量的计算资源。有论文指出, 在 BA 问题中放弃对于路标点的优化, 而仅仅关心相机位姿之间的关联, 构建位姿图 (PG, pose graph) 便可以大幅度降低优化的规模^[26]。另一种基于概率图模型的因子图优化 (FGO, factor graph optimization) 算法也在降低非线性优化计算量也引起了学者的重视。因子图是一种求解最大后验概率的概率图模型^[27], 这种方法在 2008 年首次应用在 SLAM 中位姿图的求解, 该团队研究和开发了因子图优化算法 iSAM2^[28], 并发表了因子图算法库 gtsam^[29]。因子图优化的最大优势在于增量特性: 在新的变量因子加入图模型前, 先筛选出与其关联的因子; 随后, 只对受新变量因子影响的关联因子进行更新, 而保持其余因子不变, 如图 4 所示。按照这样的处理思想, 我们不必再每次新增加节点的时候, 对于整个图都进行优化, 节省了大量的计算成本。

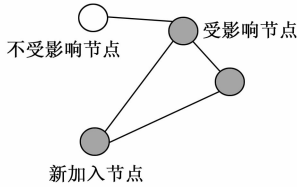


图 3 因子图增量特性

往往，完整的 VSLAM 还包括全局优化部分。后端优化能够有效地保证估计精度，而对于长时间运行的情况下，系统累积的漂移误差便会成为问题，因而后端属于局部优化。闭环检测的模块主要是对于整个模块的位姿图执行优化来消除漂移误差，如图 1 所示，这样便可以在检测到回环的时候校正位姿，消除 SLAM 的累计误差，使我们得到全局一致的轨迹和地图。闭环检测一般可以按以下 4 步进行：

- 1) 检测回环发生；
- 2) 计算回环和当前帧的运动，理论上 $T=I$ （单位阵）；
- 3) 验证闭环成立；
- 4) 闭环。

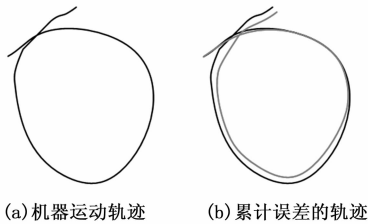


图 4 轨迹图

全局优化中的回环检测很像是一个分类的问题，因而机器学习在这一部分大有作用。回环检测的类型数量很大，但是每一类的样本却很少。因而，回环检测相当于是一个对“图像相似性”的学习，即对场景的识别和理解。回环检测本身很像是一个非监督的机器学习，我们即可以利用词袋去检测回环发生，也可以用深度学习的方式去检测回环^[34]。

1.4 建图

建图（mapping）作为 SLAM 的主要目标之一，对于无人机的导航定位、避障重建及人机交互等具有重要的意义。常见的地图可以分为：拓扑地图、栅格地图、路标地图、点云地图和认知地图。

拓扑地图使用关键帧作为地图的节点，关键帧之间共同的数据关联作为地图的边，该地图可以用作路径规划。栅格地图是将三维空间划分为独立的，均匀的立方体，每个立方体代表真实环境中的一个三维空间区域。立方体的值表示障碍物的概率。栅格地图可以用于导航，路径规划等应用，并且易于统合其他传感器的信息，但是随着时间的累计需要的存储空间也在增加。路标地图也称为稀疏点云地图，其可以满足 SLAM 的定位需求，ORB-SLAM2^[30] 所创建的便是这种稀疏地图。点云地图是一种稠密的地图，

实际上是环境中所有物体的 3D 点的集合。在稠密地图中，除了特征点的其他 3D 点可以通过极线搜索和块匹配技术恢复出它的深度信息。点云地图需要较大的存储空间去存储所有的点云信息，可以用于无人机的导航，定位避障和三维重建等任务。认知地图将语义的概念和地图中的元素联系起来，使移动机器人可以在语义地图中更加智能的完成任务，执行更加复杂的行动。



图 5 稀疏点云地图

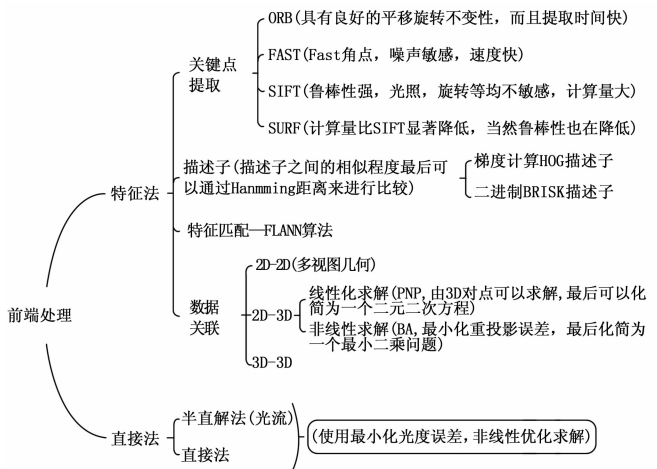


图 6 前端处理概述图

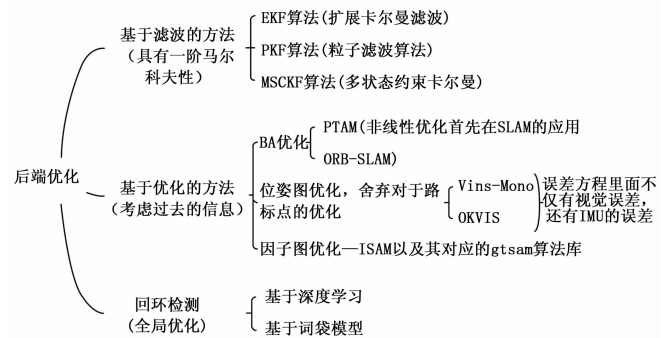


图 7 后端优化概述图

2 无人机应用的开源 SLAM 框架的对比和世界著名实验室

已面世的 SLAM 方案虽多达上百种，世界上研究 SLAM 的研究者也越来越多。表 1 给出了目前主流和具有代表性的 SLAM 框架，表 2 给出了世界上著名的研究机构和其方向。

表 1 SLAM 开源框架的基本特点归纳

开源方案	平台	前端	后端处理	回环检测
ORB-SLAM	MAV	特征点法	BA&PGO	DBow2
SVO	MAV	半直接法	BA&PGO	无
DSO	MAV	直接法	BA&PGO	无
VINS	MAV	半直接法	BA&PGO	DBow2
MSCKF	MAV	特征法	EKF	无
VI-ORB	MAV	特征法	BA&PGO	DBow2

表 2 SLAM 世界著名研究机构

实验室	地区	研究方向	主页
麻省理工学院 SPARK 实验室	美国	语义 SLAM, VIO 流形预积分	http://web.mit.edu/sparklab/
香港科技大学空中机器人实验室	香港	机器人状态估计, 建图与运动规划, 视觉惯性里程计研究	http://uav.ust.hk/
萨拉戈萨大学机器人感知实验室	西班牙	视觉 SLAM, 非刚性 SLAM 和增强现实	http://robots.unizar.es/slamlab/
苏黎世大学机器人与感知课题组	瑞士	移动机器人、无人机环境感知与导航, VIS-LAM	http://www.cvg.ethz.ch/index.php
慕尼黑工业大学计算机视觉组	德国	三维重建、机器人视觉、视觉 SLAM	https://vision.in.tum.de/research/SLAM

3 用于无人机的 SLAM 当前的研究重点

3.1 多无人机协同的 SLAM

相比于单无人机的 SLAM 建图与定位, 多无人机协同建图定位渐渐也成为了研究者关注的焦点问题。多无人机协同 SLAM (C-SLAM, cooperative visual simultaneous localization and mapping) 可以提高定位精度并不断与队友共享信息, 加速地图的创建速度, 扩大建图的范围。传统的 CSLAM 是各个无人机将信息传递给中央的卡尔曼滤波器, 但是这样的 SLAM 结构将会对噪声信息非常的敏感, 影响定位和建图的精度。Nemra 提出一种非线性 H_{∞} 的滤波方法^[31], 不对噪声信号统计并且将线性化的结果降至最低。通过作者给出的仿真实验表明, C-SLAM 的 xyz3 轴位置计算精度要比一架无人机的精度高。Trujillo^[32]为每一个 UAV 提供了单目相机, 测量每一个无人机之间的相对距离来测量更新 EKF 以提高系统的精度, 并且多无人机协作的系统可观性优于单个无人机系统的可观性, 减少了地标的确定性。

3.2 人工智能技术促进 SLAM

3.2.1 深度学习技术

随着深度学习的不断发展, 目前 SLAM 存在的很多缺陷是可以深度通过学习去解决的。目前已经发表两者结合的文章包括, 深度学习的算法可以用于帧间的估计, SLAM

的闭环检测环节; 在建图方面, 结合深度学习可以将图像和语义进行关联, 构建环境的语义知识库, 供无人机进行高级智能任务的决策等。DeepVO^[33]中, 位姿估计去结合有监督的学习, 能够通过深度递归神经网络学习 (RNN) 隐式地学习图像间的动力学关系及内在联系。在特征提取方面, 相邻两帧图像在通道上进行组合, 堆叠成六通道的图像, 然后通过多层 CNN 网络对图像进行提取, 将提取的特征输入到 RNN 网络中, 再输出位姿。DeepVO 舍弃了前端的跟踪, 后端优化等一系列的过程, 提出了端到端的 SLAM, 节省了计算的时间。清华大学高翔^[34]提出基于无监督学习的 SDA (stacked denoising auto-encoder) 自动编码器去进行回环检测, 仿真结果显示该方法比传统基于词袋模型的回环检测有着更强的鲁棒性。

3.2.2 语义 SLAM

语义建图时当下比较热门的研究方向, 高智能的无人机在未来执行复杂任务的时候, 对于自己所处的环境信息要有一定的理解。语义地图, 这一既包含空间的路标信息又包含环境语义信息的地图渐渐引起研究者的注意^[35]。Davison^[36]提出使用卷积神经网络 (CNNs) 结合 SLAM 系统 (ElasticFusion)。经验证, 该算法在 25Hz 的 RGB-D 视频长时间之内可以接收 2D 图像并且返回每个像素的分类概率分布, 然后递增的贝叶斯更新模型将跟踪每个曲面的分类概率分布, 然后使用 SLAM 系统提供的数据关联来基于 CNN 的预测更新这些概率, 最终进行语义建图。对比单帧 CNN 语义的准确率, 该算法融合了额外的观测点来对于单帧的检测率进行了改进和提高, 性能从 43.6% 提升到 48.3%。东南大学^[37]在以 LSD-SLAM 为基础与 CNNs 进行了融合, 对于关键帧进行了语义分割, 从二维图进行语义提取然后映射成三维的语义信息。该方案对于周围的环境没有要求, 且计算量也显著降低。清华大学在 ORB-SLAM 的基础之上提出 DS-SLAM^[38], 加入了语义分割的部分, 语义分割采用 SegNet, 通过光流法估算图片的特征点的位置, 通过极线的距离去判断该物体是否是运动的物体, 并对动态的物体进行分割。

3.3 多传感器融合的 SLAM 用于无人机

无人机在单一传感器的条件下导航容易受噪声的影响, 效果可能不太理想。VLSAM 主要依靠环境的纹理信息, IMU 会随着时间的累计误差逐渐增大, GPS 的使用则局限于室外开阔的条件。若单一使用传感器则无人机的导航实用性不高, 实际应用中经常是将二者甚至是三者结合起来使用, 导航的精度会得到显著的提高。

目前的视觉惯性里程计 (VIO) 是在无人机是应用最多的导航设备之一, 在前面的部分已经略有涉及, IMU 提高的角速度和加速度在短期内的精度很高, 但是长期使用可能会发散。SLAM 则对于纯旋转或者无纹理的环境会跟丢, 但是两者的结合优势互补, 效果显著。ROVIO^[39]是 Bloesch 提出的一种将视觉和 IMU 进行融合的视觉惯性里程计, 该

算法将光度误差放入到迭代卡尔曼滤波进行更新。目前 VIO 的主要融合框架分为松耦合和紧耦合。松耦合是指 IMU 和 SLAM 给出各自的信息,通过一些融合算法,比如 EKF, UKF 等进行两者的信息融合。相比于松耦合,紧耦合是指 IMU 状态变量和相机的状态放入一个误差方程进行优化。通常情况下,紧耦合具有更好的定位效果。

4 结束语

本文从视觉 SLAM 的基本框架开始谈起,第一部分介绍了前端提取、数据关联、后端优化、建图等 4 个部分的基础知识。第二部分介绍了国内外的著名实验室和经典的 SLAM 方案。第三部分对用于 UAV 平台的 SLAM 研究重点问题进行了系统的介绍,包括:多无人机协同的 SLAM,深度学习应用于 SLAM 和语义建图,以及多传感器融合的 SLAM 在无人机上面的应用。

近年来,多旋翼无人机在民用领域和非国防的领域发挥着越来越重要的作用,相机等传感器成本价格的降低也在促进这以视觉为首的 SLAM 建图定位技术的发展。在未来,无人机对于传感器有着小型化和轻量化的需求,此外,动态场景下的环境感知、目标跟踪、语义建图和复杂环境下的避障和救援等都是未来的研究者需要优化和解决的问题。研究者还将致力于提高系统的鲁棒性和计算能力才能满足未来市场的需求,而这同样对于无人机搭载的硬件提出了更先进的要求。

参考文献:

- [1] 刘浩敏,章国锋,鲍虎军. 基于单目视觉的同时定位与地图构建方法综述 [J]. 计算机辅助设计与图形学学报, 2016, 28 (6): 855-868.
- [2] 于金霞,王璐,蔡自兴. 未知环境中移动机器人自定位技术 [M]. 北京:电子工业出版社, 2011.
- [3] SMITH R C, CHEESEMAN P. On the representation and estimation of spatial uncertainty [J]. The international journal of Robotics Research, 1986, 5 (4): 56-68.
- [4] HESCH J A, KOTTAS D G, BOWMAN S L, et al. Consistency analysis and improvement of vision-aided inertial navigation [J]. IEEE Transactions on Robotics, 2016, 30 (1): 158-176.
- [5] 冯国虎,吴文启,曹聚亮,等. 单目视觉里程计/惯性组合导航算法 [J]. 中国惯性技术学报, 2011, 19 (3): 302-306.
- [6] AOUF N, OLLERO A, SASIADEK J Z. Special issue on: airborne simultaneous localisation and map building (A-SLAM) [J]. Journal of Intelligent and Robotic Systems, 2009, 55 (4/5): 247-248.
- [7] ANGELI A, FILLIAT D, DONCIEUX S, et al. 2D simultaneous localization and mapping for micro aerial vehicles [C] // European Micro Aerial Vehicles (EMAV 2006), IEEE, 2006.
- [8] FORSTER C, PIZZOLI M, DAVIDE S. SVO: Fast semi-direct monocular visual odometry [C] // IEEE International Conference on Robotics & Automation, IEEE, 2014.
- [9] 吴显亮,石宗英,钟宜生. 无人机视觉导航研究综述 [J]. 系

统仿真学报, 2010, 22 (1): 62-65.

- [10] 王希彬,赵国荣,潘爽. 无人机视觉同时定位与地图创建技术综述 [A]. 第 30 届中国控制会议 [C]. 烟台, 2011: 3913-3918.
- [11] 赵良玉,朱叶青,金瑞. 多旋翼无人机单目 V-SLAM 研究综述 [J]. 航空兵器, 2020, 27 (2): 1-14.
- [12] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF [C] // International Conference on Computer Vision, IEEE, 2012.
- [13] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C] // IEEE Computer Society Conference on Computer Vision & Pattern Recognition, IEEE, 2005.
- [14] LEUTENEGGER S, CHLI M, SIEGWART R Y. BRISK: binary robust invariant scalable keypoints [C] // International Conference on Computer Vision, IEEE, 2011.
- [15] MUJA M, LOWE D G. Fast approximate nearest neighbors with automatic algorithm configuration [C] // Proceedings of the 4th International Conference on Computer Vision Theory and Applications, Lisbon, Portugal, 2009: 331-340.
- [16] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision [M]. 2nd ed. Cambridge University Press, 2003.
- [17] DAVISON A J, REID I D, MOLTON N D, et al. Monoslam: Real-time single camera slam [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007 (6): 1052-1067.
- [18] MONTEMERLOM, THRUN S, KOLLER D, et al. Fast-SLAM: a factored solution to the simultaneous localization and mapping problem [C] // Proceedings of the 18th International Conference on Artificial Intelligence, Pittsburgh, PA, USA, 2004.
- [19] MONTEMERLOM, THRUN S. Simultaneous localization and mapping with unknown data association using fast SLAM [C] // Proceedings of 2003 IEEE International Conference on Robotics and Automation, Taipei, China, 2003: 1985-1991.
- [20] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint Kalman filter for vision-aided inertial navigation [C] // IEEE International Conference on Robotics & Automation, IEEE, 2007.
- [21] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C] // Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara: IEEE Computer Society, 2007: 1-10.
- [22] DEUTSCH I, LIU M, SIEGWART R. A framework for multi-robot pose graph slam [C] // 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), Angkor Wat: IEEE, 2016: 567-572.
- [23] LEUTENEGGER S, LYNNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization [J]. The International Journal of Robotics Research, 2015, 34 (3): 314-334.

(下转第 41 页)