

基于视觉的同时定位与地图构建的研究进展*

陈 常, 朱 华[†], 由韶泽

(中国矿业大学 机电工程学院, 江苏 徐州 221008)

摘 要: 基于视觉的同时定位与地图构建(VSLAM)是目前机器人定位方面的热门研究课题,在机器人自身的定位以及场景识别、任务执行、路径规划等方面发挥着重要的作用。针对VSLAM的应用领域和发展趋势进行总结和归纳,分析了VSLAM的基本原理;在此基础上,从间接法和直接法两个方面对VSLAM关键技术和最新的研究进展进行了阐述,对比分析不同方法的优点和实现难点。最后展望了VSLAM的未来发展趋势和研究方向。

关键词: 同时定位与地图构建; 特征匹配; 闭环检测; 机器人; 计算机视觉; 多视图几何

中图分类号: TP242.62 **文献标志码:** A **文章编号:** 1001-3695(2018)03-0641-07

doi:10.3969/j.issn.1001-3695.2018.03.001

Survey on vision-based simultaneous localization and mapping

Chen Chang, Zhu Hua[†], You Shaoze

(School of Mechatronic Engineering, China University of Mining & Technology, Xuzhou Jiangsu 221008, China)

Abstract: Vision-based simultaneous location and mapping(VSLAM) is a popular research topic in robot positioning area. It plays a significant role in robot positioning and scene recognition, task execution, path planning. This paper summarized the application areas and development trends of VSLAM, and analyzed the fundamental principle of VSLAM. On this basis, it surveyed the key technologies and latest research progress of VSLAM from indirect and direct methods, and discussed the comparative advantages and the implementation difficulties of different methods. Finally, it prospected the future development trend and research direction of VSLAM.

Key words: simultaneous localization and mapping; feature alignment; loop detection; robot; computer vision; multiple view geometry

0 引言

同时定位与地图构建(simultaneous localization and mapping, SLAM)最先应用于机器人领域^[1~3],其目标是在一个未知的环境根据传感器数据实时构建周围环境地图同时对机器人自身进行定位。基于视觉的同时定位与地图构建(vision-based simultaneous localization and mapping, VSLAM)对获得图像信息进行处理并创建三维地图,同时可以看成一种实时的运动估计结构(structure from motion, SFM)算法^[4]。

相比于传统的激光测距仪、高精度的惯导和GPS(全球定位系统),相机的成本低廉、功耗低,并且能够提供鲁棒和精准位置识别的环境信息。不论是在嘈杂的室内、高楼林立的街区,还是在暗流涌动的深海、地形复杂的河流峡谷,VSLAM都能使机器人进行定位和地图构建。

位置识别是VSLAM系统中用来闭环的关键模块,用来检测因遮挡、主动运动、系统重新初始化等导致跟踪丢失后的相机重定位和减小当传感器返回到构图区域的累积误差。因此以相机为主要传感器的VSLAM,在最近几年得到了快速的发展。随着算法的不断改进和处理器性能的不提升,VSLAM系统的精度也不断提升,获得的图像信息在更高级的功能上发挥着重要的作用,已经成为当今SLAM领域主要的研究对象。

VSLAM成功应用于无缆水下机器人(AUV)^[5,6]、无人驾驶汽车^[7,8]、无人机(UAV)^[9,10]、自动牵引车(AGV)^[11]、扫地机器人^[12]等机器人领域中,其用途包括传感器自身的定位,以及后续的场景理解、任务执行、路径规划。对于像救援、探测等特种机器人,它们的作业环境通常是未知的、变化的,同时也是复杂的、危险的,VSLAM对它们实时的自定位和周围环境地图的构建起着关键性的作用。随着增强现实(AR)和虚拟技术(VR)的应用进入市场,VSLAM已经成为越来越热门的研究领域。

根据图像处理的方法,VSLAM主要分为前端和后端。前端也称为视觉里程计(visual odometry, VO)^[13,14],它根据获取图像的信息,估计出相机运动,给后端优化提供较好的初始值;后端则解决地图的优化问题,根据前端提供的数值,得到全局的最优估计。后端图优化技术主要在G2O(general graph optimization)^[15]、Ceres^[16]等框架下实现,优化的时机和范围以及如何兼顾实时性是当前研究的热点问题。VSLAM根据视觉里程计是否提取特征分为直接法和间接法。间接法也称为特征点法。VSLAM长期以来由基于特征的间接法主导,如MonoSLAM^[17]、PTAM^[18]、ORB-SLAM^[19],但近年来随着DTAM^[20]、SVO^[21]、LSD-SLAM^[22]、DSO^[23]等算法的逐渐普及,直接法引起了广泛关注。

收稿日期: 2017-03-07; 修回日期: 2017-04-16 基金项目: 国家“863”计划资助项目(2012AA041504)

作者简介: 陈常(1992-),男,硕士研究生,主要研究方向为移动机器人、VSLAM、控制与路径规划;朱华(1960-),男(通信作者),教授,博导,博士,主要研究方向为救援机器人技术、机械动力学、噪声与振动控制、设备工况监测与故障诊断(zhuhua83591917@163.com);由韶泽(1993-),男,博士研究生,主要研究方向为移动机器人、跟踪与定位。

1 VSLAM 基本原理

经典的 VSLAM 系统一般包含前端视觉里程计、后端优化、闭环检测、构图四个主要部分。以传感器而论,VSLAM 研究主要分为单目、双目(或多目)、RGB-D 三大类,其余还有鱼眼、全景等特殊相机。VSLAM 技术是根据多视图几何原理^[24],通过相机拍摄的视频信息估计出相机在未知环境中的位姿,同时构建环境地图。使用单目相机的 VSLAM 原理如图 1 所示。其目标是同时恢复出相机每帧图像对应的相机运动参数 $C_1 \cdots C_m$,以及环境三维结构 $X_1 \cdots X_n$ 。VSLAM 将不同图像中对应相同场景点的图像点进行匹配(如将对应 X_1 的图像点 x_{11} 、 x_{21} 、 x_{31} 进行匹配)。通过优化求解目标函数:

$$\arg \min_{C_1 \cdots C_m, X_1 \cdots X_n} \sum_{i=1}^m \sum_{j=1}^n \|h(C_i - X_j) - \hat{x}_{ij}\|_{\Sigma_{ij}} \quad (1)$$

其中: $h_{ij} = h(C_i, X_j)$ 为三维点在图像中投影位置函数。对于相机,它在欧氏空间里运动,经常需要估计一个三维空间的变换矩阵 SE3 或 Sim3。求解这个矩阵是 VO 的核心问题,而求解的思路则分为基于特征的间接法和直接对像素进行处理的直接法。

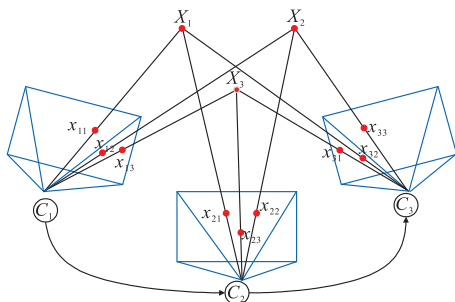


图1 多视图几何原理

2 间接法

2.1 构建稀疏地图的算法

间接法的基本思路是,先从获得的图像中提取特征点;然后在图像上进行特征匹配,得到匹配好的点;再根据这些点进行相机位姿的求解。相机位姿求解部分不再直接使用图像本身的信息。

MonoSLAM 是 Davison 等人^[17]提出的第一个成功用于单目相机的纯视觉 SLAM 系统,也是实时的 SFM 系统。MonoSLAM 利用概率模型,将场景点投影至图像的形状为一个概率椭圆,为每帧图像抽取 Shi-Tomasi 角点^[25],在投影椭圆中主动搜索特征点进行匹配,后端采用扩展卡尔曼滤波器(extended Kalman filter, EKF)进行优化。因为 VSLAM 问题通常都是非线性的, MonoSLAM 和早期的文献[26~28]都是使用滤波器进行优化,与使用 Levenberg-Marquardt^[19]、Gauss-Newton^[22]、Dog-Leg^[29]等迭代的非线性优化的算法相比更容易积累误差,不能保证全局优化。

捆集优化(bundle adjustment, BA)^[30]诞生于 1950 年的摄影学中,之后迅速地被计算机视觉应用,作为 SFM 等视觉重建算法的最后一步,利用 Levenberg-Marquardt 算法最小化观测的图像点坐标与预测的图像点坐标之间的误差。Mouragnon 等人^[31]第一次将 BA 用于视觉里程计当中,应用全局 BA 优化计算轨迹。然而自从 2007 年 Klein 等人^[18]将 BA 用于 PTAM 中以来,BA 得到了突破性的发展。

PTAM 是实时的 SFM 系统,也是实现基于 BA 的实时的单

目 SLAM 系统。它将跟踪和构图作为两个独立的任务在两个线程并行执行,跟踪结果不再以一定的概率依赖于构图过程。通过使用鲁棒性高的跟踪算法,跟踪与构图之间的数据关联不需要共享,因此解决了每帧更新地图时的计算负担,还使得跟踪线程可以对图像进行更彻底的处理,从而极大地改善系统性能。系统框架如图 2 所示。

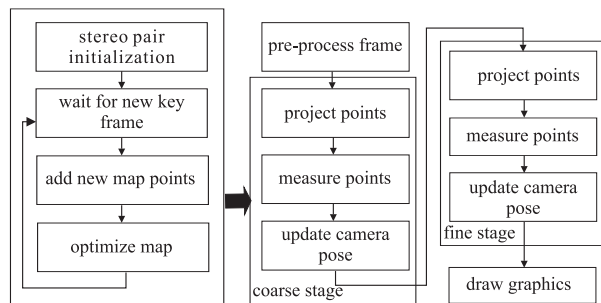


图2 PTAM系统框架

跟踪线程通过一个匀速运动模型预测当前帧的位置,以搜索地图中的三维点在当前帧图像中对应的 FAST 角点^[32],并根据匹配关系优化当前帧的位姿:

$$\arg \min_{C_t} \sum_{j=1}^n w_j \|h(C_t, X_j) - \hat{x}_j\|_{\Sigma_j} \quad (2)$$

其中: w_j 是 Tukey 函数^[33]对应的权值,用于缓解异常值对结果的影响。构图线程用极线搜索(epipolar search)初始化,沿极线匹配不同关键帧之间对应相同场景点的图像特征点,通过三角化法恢复这些场景点的三维位置,并对所有关键帧和三维点运行 BA,恢复精准的三维地图。

PTAM 对关键帧的选择、特征匹配、点云三角化、每帧相机位置估计、追踪失败后的重定位都非常有效;然而仍有一些局限性:缺少闭环控制和足够的遮挡处理、较差的视图不变特性、在形成地图过程中需要人工干预、不能处理纯旋转场景。除此之外, PTAM 设计的初衷是应用于 AR 中,局限于小范围当中,并未解决 SLAM 场景中的大尺度闭环问题。

PTAM 的开源对 VSLAM 的发展产生深远的影响,如今很多 VSLAM 系统都借鉴 PTAM 的算法框架。针对 PTAM 不能应用于大场景以及在高频纹理的场景中不够鲁棒,文献[34]展开对这些问题的研究并解决。DT-SLAM^[35]通过增加三角化特征,很好地解决了相机纯旋转的问题。浙江大学 CAD&CG 国家重点实验室章国锋课题组吸收了 PTAM 的优点,研发了应用于动态环境的 RDSLAM^[36]和处理快速移动和强旋转的 RK-SLAM^[37]。王可^[38]进一步改进了 PTAM 算法,提出了基于视觉反馈的变分深度估计模型,实现了环境的稠密三维建模。

2015 年 Mur-Artal 等人^[19]开源了当今基于稀疏特征最好的 VSLAM 算法——ORB-SLAM,极大地推动了 VSLAM 的发展。ORB-SLAM 在 PTAM 的基础上,根据文献[39]关于闭环检测的方法、文献[40]关于尺度感知闭环的方法和文献[41, 42]在大尺度环境下的局部相互可见地图的思想克服 PTAM 的局限性。该算法包含跟踪、局部构图、闭环检测三个并行线程。系统框架如图 3^[19]所示。跟踪线程采用恒速模型,并通过优化重投影误差来优化相机位姿;实时的相机重定位具有明显的旋转不变性,跟踪丢失可以恢复,增强地图的重复使用性;使用基于模型选择的初始化过程,允许创建平面或非平面的初始化地图。采用文献[36]的方法,去除冗余关键帧,增强程序可操作性;闭环检测线程采用 DBow2 算法^[39]加速闭环匹配帧

的筛选,通过全局 BA 优化关键帧和地图点,并使用 EPNP 算法^[43]完成重定位中的位姿估计。

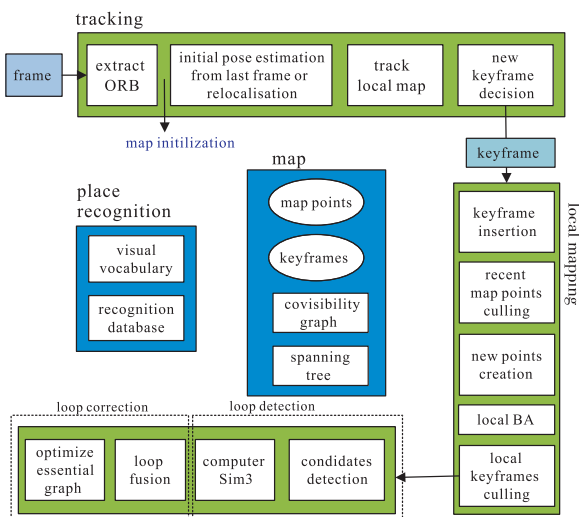


图3 ORB-SLAM系统框架

系统基于ORB特征^[44]而实现,并且通过G2O优化所有环节。相对于SIFT或SURF特征,ORB特征能够被快速提取,具有旋转不变性,并且能够利用金字塔构建出尺度不变性。使用统一的ORB特征使系统在特征提取与跟踪、关键帧选取、三维重建、闭环检测等步骤具有内生的一致性,使整个系统更加可靠。然而ORB-SLAM对特征缺失非常敏感,在弱纹理环境下缺乏鲁棒性,同时提取特征需要花费大量的时间。

单目相机是能够用来实施VSLAM最便宜和最小的传感器设备,然而仅通过一个相机不能够观测到深度,地图的尺度和估计的轨迹是未知的。因为它不能从第一帧进行三角测量,所以系统引导必须通过多视图或者滤波技术来产生一个初始的地图。最重要的是单目SLAM受尺度漂移影响,并且在探索中执行纯旋转使得定位容易失败。

通过使用立体相机或RGB-D相机,这些问题都能够得到很好的解决,并且能够成为可靠的VSLAM解决方案。2016年Mur-Artal等人^[45]在自己的ORB-SLAM算法的基础上增加了一个输入预处理线程(图4),提出ORB-SLAM2算法并开源,将应用拓展到单目、立体和RGB-D相机这三种主流视觉传感器。这也是对性能优越的SLAM系统扩展应用的思路之一。

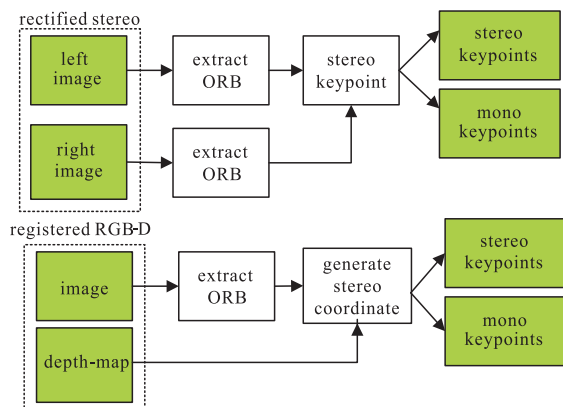


图4 输入预处理线程

针对ORB-SLAM2算法的固有缺点,Mur-Artal等人^[46]在系统中增加了IMU来实现零漂移定位。

2.2 构建稠密地图的算法

像PTAM、ORB-SLAM等系统构建的稀疏地图,并不能应

用于地图导航和人机交互等任务中。Mur-Artal等人^[47]在自己作品ORB-SLAM的基础之上,又进一步提出了基于概率的半稠密构图技术,处理单目SALM系统提供的关键帧,在不需要GPU加速的情况下实时地重建表面纹理和物体轮廓。

MonoSLAM、PTAM、ORB-SLAM和文献^[48]都是通过匹配特征点集,对环境和相机位姿进行估计,构建稀疏的地图。文献^[47]和上述都是间接的方法,不过它是基于概率的半稠密SALM系统。系统框架如图5^[47]所示。

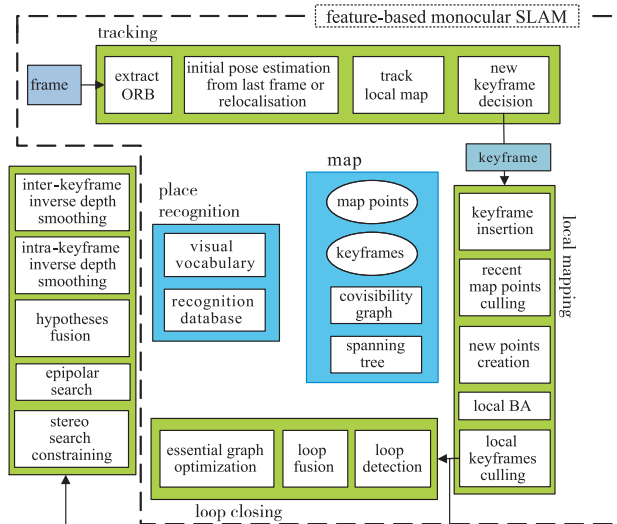


图5 ORB-SLAM2系统框架

文献^[47]在ORB-SLAM的基础上增加了一个半稠密构图线程来构建半稠密地图。与基于特征的SLAM^[49,50]、视觉里程计^[51,52]等构建稠密地图的算法相比,文献^[47]能够在不需要GPU加速的情况下实时地重建表面纹理和物体轮廓。通过局部BA和闭环后的位姿图优化,关键帧的位姿估计精度提高,从而实现高精度的半稠密重建。采用关键帧之间异常值检测的方法^[51]以及概率不确定的模型,提高了地图重建的精度。然而该算法主要限制是在构建半稠密地图时伴随着几个关键帧的延迟,以及并没有用于相机跟踪。

文献^[47,53,54]的算法都属于间接稠密的方法,不过它们是通过匹配光流场,重建稠密的环境模型。它们支持通过一种凸形的最小化分割,将完整的动态场景分割成许多独立的运动部分。这种场景分割的方法是构建稠密地图的新思路。

3 直接法

3.1 构建稠密地图的算法

间接法通常都需要在图像中提取并匹配特征点,根据它们在相机中的投影位置,通过优化重投影误差来优化相机运动,对相机拍摄的图像中存在的光度误差(photometric error)和几何失真具有高鲁棒性。但是计算和匹配特征需要大量的计算量,除此之外间接法对环境特征的丰富程度和图像质量十分敏感。相比之下,直接法不依赖于特征点的提取和匹配,而是从所有图像像素点(包括边缘和平滑强度变化的像素点)中采样,生成更完整的环境模型,通过最小化光度误差来估计相机运动,这提高了其在弱纹理特征环境中的鲁棒性。

文献^[55]是最早将直接法应用于实时的SLAM系统,但是在小数据集上只有很少的选定平面区域,并没有得到广泛的应用。DTAM是Newcombe等人^[20]于2011年提出的单目VS-

LAM 系统,是一种直接稠密的方法。该算法通过最小化全局空间规范能量函数来计算关键帧构建稠密深度图,而相机的位姿则使用深度地图通过直接图像匹配来计算得到。基于直接跟踪的 DTAM 对特征缺失、图像模糊有很好的鲁棒性,但由于 DTAM 为每个像素都恢复深度图,并且采用全局优化,所以计算量非常大,需要在 GPU 并行计算的情况下才能实现。相对于高加速的运动轨迹,PTAM 容易跟踪丢失并采取重定位的方法,而 DTAM 对于快速运动具有极度弹性进行更平滑的运动估计。DTAM 假设光度恒定,虽然能够处理局部照明变化,但对全局照明处理不够鲁棒。

2014 年 Engel 等人^[22]提出 LSD-SLAM 算法。LSD-SLAM 与 DTAM 一样都是用光度误差和几何先验信息来构建稠密或半稠密的地图。LSD-SLAM 是一种在 Sim3 上进行直接匹配的算法,以点云为代表实时地重构半稠密高精度的 3D 环境地图。

LSD-SLAM 的两个主要贡献是:a)构建了一个大尺度直接

单目 SLAM 的框架,提出了一种用来直接估计关键帧之间相似变换、尺度感知的图像匹配算法;b)在跟踪线程中结合深度估计的不确定性。该系统能够在 CPU 上实时运行,并且能够作半稠密 VO 在智能手机上实现 AR 的功能^[56]。

LSD-SLAM 系统主要分为跟踪、深度地图估计和地图优化三部分,如图 6^[22]所示。跟踪部分持续地跟踪相机新获取的图像,使用前一帧的位姿作为初始值,最小化光度误差,估计当前帧物体的位姿;深度地图估计部分将跟踪的帧精炼或取代当前关键帧,通过对像素、小基线立体的比较进行滤波和空间规则化来优化深度。如果相机移动距离现有地图太远,则将已存在邻近关键帧上的点投影到当前帧来初始化一个新的关键帧;一旦一个关键帧被作为跟踪参考,深度图将不会进一步重新定义,它将由地图优化部分合并到全局地图中。为了检测回环闭合和尺度漂移,使用尺度感知、直接图像匹配的方法来估计和相邻关键帧的相似变换。

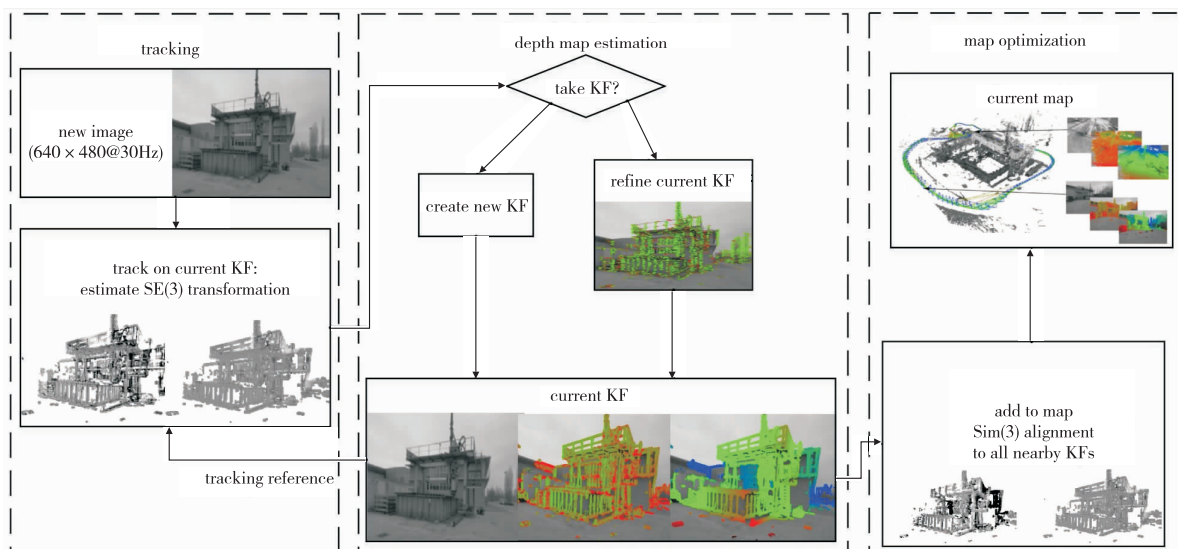


图6 LSD-SLAM系统框架

LSD-SLAM 采用一种基于外观构图的算法^[57]检测大尺度回环,通过交叉跟踪检验避免插入错误的跟踪回环,提高地图闭环精度。LSD-SLAM 最早应用于单目相机。文献[58,59]相继提出了 LSD-SLAM 在全向相机和立体相机中的使用方法,文献[60]又将 LSD-SLAM 成功应用于行驶汽车来构建街道的场景。

Forster 等人^[21]于 2014 年提出一种半直接的视觉里程计算法——SVO。与文献[61,62]相同 SVO 也是一种混合的算法。它融合间接法跟踪多特征点、平行跟踪和构图、关键帧选择的优点,在直接法的框架中进行运算,具有直接法的准确性和速度。虽然 SVO 也提取特征,但与 PTAM、ORB-SLAM 等间接法不同,特征只在构图线程有新的关键帧时才进行提取。与 PTAM 相同,SVO 也具有运动估计和构图两个线程。其流程如图 7^[21]所示。

在运动估计线程,SVO 采用直接法计算相机运动和特征匹配的初始猜想,优化基于特征的非线性重投影误差,使用 Lucas-Kanade^[55]算法来解决定位问题,通过优化相机位姿将重定位残差转换成 BA 问题,用迭代非线性最小二乘法解决。在构图线程,结合概率建图的方法增加异常值检测的鲁棒性,采用文献[63]的方法对深度估计的特征进行建模,更新贝叶斯框架中的分布,当滤波器的不确定性变得足够小时,估计的 3D 点将会查入地图中并用于运动估计。

从精度、速度和鲁棒性三个方面对 SVO 与文献[34]的方

法相比,实验表明 SVO 在这三个方面具有良好的性能,除此之外 SVO 甚至能在嵌入式平台上实现高帧率。SVO 应用于无人机中,使用逆深度坐标处理大的场景。但是算法仅跟踪相机的局部运动,不能构建一致性的全局地图。除此之外,SVO 只是单纯的视觉里程计,并没有闭环检测功能。

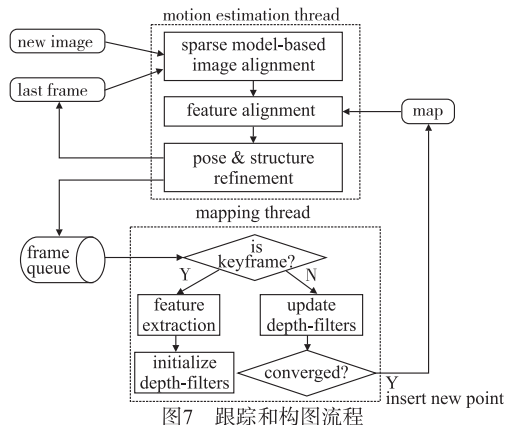


图7 跟踪和构图流程

3.2 构建稀疏地图的算法

长期以来直接法(如 DTAM、SVO 和 LSD-SLAM)不仅使用几何先验信息,还需要最小化光度误差来重建稠密的环境模型。然而自 DSO 提出之后,完全改变了直接法的现状。DSO 不考虑几何先验信息,能够直接优化光度误差。Jin 等人在文

献[61]提出类似的方法,但该方法基于 EKF 框架不能进行全局优化,而 DSO 基于非线性优化框架。

DSO 中除了完善直接法位姿估计的误差模型外,还加入了仿射亮度变换、光度标定、深度优化;另一个核心就是像 Okvis^[64] 一样使用滑动窗口算法(dynamic window approach, DWA)^[65]来优化位姿。为了满足实时性,DSO 并没有像其他直接法那样对图像进行光滑处理,而是对整个图像均匀采样。3D 重建和轨迹跟踪如图 8^[23]所示。

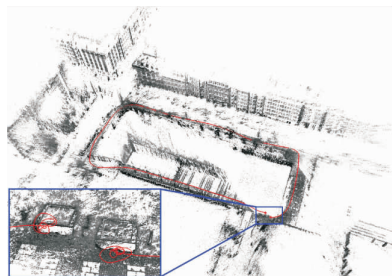


图8 3D重建和轨迹跟踪

DSO 采用直接优化光度误差的算法,考虑了光度标定模型,其优化范围不是所有帧,而是由最近帧及其前几帧形成的滑动窗口,并且保持这个窗口有七个关键帧。与 PTAM、ORB-SLAM 等间接法不同,DSO 优化包括相机内参、相机外参和逆深度值等所有模型参数,但算法效率与间接法中常用的 BA 算法相近。同时,该系统采用与其他直接法相同的地图点表示方法,即三维点被表示为相应坐标下的逆深度值,因此只有一个自由度。因为 DSO 的优化框架中保持固定 2 000 个激活点均匀分布在关键帧当中,所以在纹理密集的场景中深度图变得稀疏,在弱纹理场景中生成类似 LSD-SLAM 和文献[47]中的半稠密图。

DSO 采用滑动窗口优化算法,边缘化旧的图像帧和不在视角范围内的地图点。与 LSD-SLAM 和 SVO 假定光度恒定忽略曝光时间变化不同,DSO 充分利用了相机光度标定,包括透镜衰减、伽马校正和已知的曝光时间,这一步提高了视觉里程计的精度和鲁棒性。除了应用于单目相机,直接法同样已经用于 RGB-D 和立体相机上^[66,67]。

间接法的 VSLAM 由于特征描述一般会有光照不变性,对图像的亮度值并不敏感。而在直接法中,由于相机位姿估计以图像的亮度值为出发点,亮度值的准确度会影响算法的精度和稳定性。因此,DSO 引入了光度标定的概念,利用精细的相机成像模型,标定成像过程中的光度参数,并用这些参数校正图像亮度值。

直接法假设同一个空间点在各视角下测到的灰度不变:

$$J(x; y) = I(x + u(x, y) + v(x, y)) \quad (3)$$

其中: x, y 是像素坐标; u, v 表示同一场景 I 和 J 两张图片中像素 $(x; y)$ 的函数。当场景亮度发生变化时,直接法容易受到影响。除此之外现有的直接 VSLAM 算法主要是单纯的视觉里程计如 SVO、DSO,与 ORB-SLAM 等间接法相比,在闭环检测和重定位方面显示出明显的不足。

4 间接和直接 VASLAM 对比

如图 9(a)所示,间接法首先对测量数据进行预处理来产生中间层,通过稀疏的特征点提取和匹配来实现,也可以采用稠密规则的光流,或者提取直线或曲线特征来实现。然后计算出地图点坐标或光流向量等几何量。因此间接法优化的是几何误差。

$$T_{k-1, k} = \arg \min_T \sum_i \|u'_i - \pi(p_i)\|_2^2 \quad (4)$$

其中: u_i 为 I_{k-1} 中任意像素点;它投影到空间点的坐标为 p_i ; u'_i 是 p_i 投影到 I_k 上的坐标。之后利用中间层的数值来估计周围环境的三维模型和相机运动。

直接法跳过预处理步骤直接使用实际传感器测量值,如在特定时间内从某个方向接收的光,如图 9(b)所示。

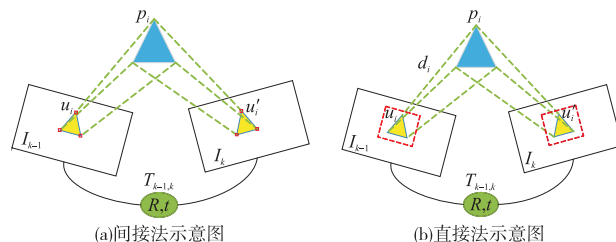


图9 间接法和直接法本质区别

在被动视觉的情况下,由于相机提供光度测量,所以直接法优化的是光度误差。

$$T_{k-1, k} = \arg \min_T \sum_i \|I_k(u'_i) - I_{k-1}(u_i)\|_\sigma^2 \quad (5)$$

$$u'_i = \pi(T \times (\pi^{-1}(u_i) \times d)) \quad (6)$$

然而对于 RGB-D 相机或激光扫描器等直接测量几何量的传感器,直接法优化的是几何误差。

直接法直接对图像中的像素进行处理,而间接法只处理包含特征点的信息,因此直接法的信息利用率更高,构建地图的完整性更好;然而整体处理速度较慢,一般需要 GPU 加速。间接法由于提取特征和构建描述符,在跟踪丢失情况下更容易重定位,所以重定位精度比直接法较高。在弱纹理环境下直接法的鲁棒性明显高于间接法。在动态场景下,间接法的鲁棒性高于直接法。像 DTAM、SVO、DSO 都没有回环检测线程,回环检测能力整体低于 ORB-SLAM 代表的直接法。间接法和直接法性能对比如表 1 所示。

表1 间接法和直接法 VSLAM 对比

比较项	间接法	直接法	比较项	间接法	直接法
信息利用率	√	√√√	弱纹理环境鲁棒性	√	√√√
处理速度	√√√	√√	构建地图完整性	√√	√√√
回环检测能力	√√√	√√	动态场景鲁棒性	√√√	√√
重定位能力	√√√	√√			

5 SLAM 发展趋势

5.1 主动 SLAM

到目前为止,所研究的 SLAM 问题仍然是机器人领域被动出现的估计问题,例如机器人执行 SLAM 是基于给定的传感器数据,没有自由主动地收集数据。控制机器人运动问题的主要目的,就是控制机器人在运动中最小化机器人地图表示和定位的不确定性。针对这一目的,主动 SLAM^[68] 逐渐引起人们的关注。主动 SLAM 算法有如下作用:能够让机器人识别可能的位置探索并使用,如在地图当前估计中的有利位置;使机器人计算访问过的每个有利点,并选择最高效的行动方案;在机器人执行选择好的行动方案,决定是否有必要继续或终结任务。文献[69,70]对主动 SLAM 进行探索研究,结合主动 SLAM 能够让机器人变得更加智能和实用。

5.2 语义 SLAM

目前 VSLAM 方案中所采用的图像特征的语义级别太低,造成特征的可区别性太弱,还有采用当前方法构建出来的点云地图对不同的物体并未进行区分。因此,人们已经认识到纯粹

几何地图的局限,产生了很多工作构建环境语义地图来提高机器人的自动性和鲁棒性,处理更复杂的任务(驾驶避开泥泞的道路),从路径规划到任务规划,处理更高级的人机交互。虽然SLAM中的语义问题研究仍然处于早期阶段,将语义与SLAM进行结合将会使机器人更有效地定位,用环境中类别、关系和属性的语义概念更加持续地建图。

5.3 SLAM 中新的传感器

二维激光扫描仪使SLAM系统变得非常具有鲁棒性,然而只能构建二维的地图。由于无人驾驶汽车和无人机等研究的推动,三维激光扫描仪成为主流的研究对象。同时在一些领域,为了获得更高的灵敏度,需要更快的传感器和更低延迟的处理时间,普通的相机无法达到这种效果,然而事件触发的相机如动态视觉相机(DVS)或异步图像传感器(ATIS),可以完美取代当前的图像传感器,正在改变基于图像的感知领域。文献[71]已经将事件相机应用于特征检测和跟踪当中。结合这些新的传感器,可以将SLAM系统变得效率更高、鲁棒性更好、精度更高。

6 结束语

近年来,随着机器人、虚拟现实和增强现实等领域的发展,VSLAM得到了快速发展。长期以来间接法是VSLAM的主导,随着不断发展以及不断完善,直接法已经能与间接法相提并论。为了能够在实际应用中得到更好的应用,必须要有能够处理各种复杂环境的更加鲁棒的SLAM系统。随着研究的深入,VSLAM的这两种方法将会相互促进共同发展,或在一个系统中综合利用这两种方法取得更好的性能。

参考文献:

- [1] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: part I[J]. *IEEE Robotics & Automation Magazine*, 2006, 13(2):99-110.
- [2] Bailey T, Durrant-Whyte H. Simultaneous localization and mapping (SLAM): part II[J]. *IEEE Robotics & Automation Magazine*, 2006, 13(3):108-117.
- [3] 何俊学, 李战明. 基于视觉的同时定位与地图构建方法综述[J]. *计算机应用研究*, 2010, 27(8):2839-2844.
- [4] Pollefeys M, Van Gool L, Vergauwen M, et al. Visual modeling with a hand-held camera[J]. *International Journal of Computer Vision*, 2004, 59(3):207-232.
- [5] Jung J, Choi S, Choi H T, et al. Localization of AUVs using depth information of underwater structures from a monocular camera[C]//Proc of the 13th International Conference on Ubiquitous Robots and Ambient Intelligence. 2016: 444-446.
- [6] 王宏健, 王晶, 边信黔, 等. 基于组合EKF的自主水下航行器SLAM[J]. *机器人*, 2012, 34(1):56-64.
- [7] 姜岩, 赵熙俊, 龚建伟, 等. 简单城市环境下地面无人驾驶系统的设计研究[J]. *机械工程学报*, 2012, 48(20):103-112.
- [8] Lategahn H, Geiger A, Kitt B. Visual SLAM for autonomous ground vehicles[C]//Proc of IEEE International Conference on Robotics and Automation. 2011: 1732-1737.
- [9] 马跃龙, 赵勇, 曹雪峰. 一种基于单目SLAM的无人机序列图像拼接方法[J]. *系统仿真学报*, 2016, 28(10):2349-2355.
- [10] Engel J, Sturm J, Cremers D. Camera-based navigation of a low-cost quadcopter[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2012:2815-2821.
- [11] 毛曙源. 室内移动机器人自定位方法研究[D]. 杭州:浙江大学, 2016.
- [12] Vaussard F, Fink J, Bauwens V, et al. Lessons learned from robotic vacuum cleaners entering the home ecosystem[J]. *Robotics and Autonomous Systems*, 2014, 62(3):376-391.
- [13] Scaramuzza D, Fraundorfer F. Visual odometry[J]. *IEEE Robotics & Automation Magazine*, 2011, 18(4):80-92.
- [14] Fraundorfer F, Scaramuzza D. Visual odometry: part II: matching, robustness, optimization and applications[J]. *IEEE Robotics & Automation Magazine*, 2012, 19(2):78-90.
- [15] Kummerle R, Grisetti G, Strasdat H, et al. G2O: a general framework for graph optimization[C]//Proc of IEEE International Conference on Robotics and Automation. 2011:3607-3613.
- [16] Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM[C]//Proc of IEEE International Conference on Robotics and Automation. 2016:1271-1278.
- [17] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM[J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2007, 29(6):1052-1067.
- [18] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//Proc of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. [S. l.]: IEEE Computer Society, 2007:1-10.
- [19] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. *IEEE Trans on Robotics*, 2015, 31(5):1147-1163.
- [20] Newcombe R A, Lovegrove S J, Davison A J. DTAM: dense tracking and mapping in real-time[C]//Proc of IEEE International Conference on Computer Vision. 2011:2320-2327.
- [21] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[C]//Proc of IEEE International Conference on Robotics and Automation. 2014:15-22.
- [22] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[C]//Proc of the 13th European Conference on Computer Vision. Berlin: Springer, 2014: 834-849.
- [23] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2007, PP(99):1.
- [24] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2004.
- [25] Shi Jianbo, Tomasi C. Good features to track[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1994:593-600.
- [26] Davison A J. Real-time simultaneous localisation and mapping with a single camera[C]//Proc of IEEE International Conference on Computer Vision. 2003:1403-1410.
- [27] Meltzer J, Gupta R, Yang M H, et al. Simultaneous localization and mapping using multiple view feature descriptors[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2004:1550-1555.
- [28] Jeong W Y, Lee K M. CV-SLAM: a new ceiling vision-based SLAM technique[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2005:3195-3200.
- [29] Rosen D M, Kaess M, Leonard J J. An incremental trust-region method for robust online sparse least-squares estimation[C]//Proc of IEEE International Conference on Robotics and Automation. 2012: 1262-1269.
- [30] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM: why filter? [C]//Proc of IEEE International Conference on Robotics and Automation. 2010:2657-2664.
- [31] Mouragnon E, Lhuillier M, Dhome M, et al. Real-time localization and 3D reconstruction[C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC:

- IEEE Computer Society, 2006;363-370.
- [32] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]//Proc of European Conference on Computer Vision. [S. l.]:Springer-Verlag, 2006;430-443.
- [33] Huber P J. Robust statistics[J]. *Journal of the American Statistical Association*, 2011,78(381):1248-1251.
- [34] Weiss S, Achtelik M W, Lynen S, *et al.* Monocular vision for long-term micro aerial vehicle state estimation: a compendium[J]. *Journal of Field Robotics*, 2013,30(5):803-831.
- [35] Herrera C D, Kim K, Kannala J, *et al.* DT-SLAM: deferred triangulation for robust SLAM[C]//Proc of IEEE 3DV. 2014;609-616.
- [36] Tan Wei, Liu Haomin, Dong Zilong, *et al.* Robust monocular SLAM in dynamic environments[C]//Proc of IEEE International Symposium on Mixed and Augmented Reality. 2013;209-218.
- [37] Liu Haomin, Zhang Guofeng, Bao Hujun. Robust keyframe-based monocular SLAM for augmented reality[C]//Proc of IEEE International Symposium on Mixed and Augmented Reality. [S. l.]:IEEE Computer Society, 2016;1-10.
- [38] 王可. 基于变分模型的移动机器人三维环境建模方法研究[D]. 北京:北京工业大学,2016.
- [39] Galvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. *IEEE Trans on Robotics*, 2012,28(5):1188-1197.
- [40] Strasdat H, Montiel J M M, Davison A J. Scale drift-aware large scale monocular SLAM[C]//Proc of Robotics: Science and Systems VI. 2010.
- [41] Strasdat H, Davison A J, Montiel J M M, *et al.* Double window optimisation for constant time visual SLAM[C]//Proc of IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2011;2352-2359.
- [42] Mei C, Sibley G, Newman P. Closing loops without places[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2010;3738-3744.
- [43] Lepetit V, Moreno-Noguer F, Fua P. EPnP: an accurate $O(n)$ solution to the PnP problem[J]. *International Journal of Computer Vision*, 2009,81(2):155-166.
- [44] Rublee E, Rabaud V, Konolige K, *et al.* ORB: an efficient alternative to SIFT or SURF[C]//Proc of International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2011;2564-2571.
- [45] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras[J]. *IEEE Trans on Robotics*, 2017,33(5):1255-1262.
- [46] Mur-Artal R, Tardós J D. Visual-inertial monocular SLAM with map reuse[J]. *IEEE Robotics and Automation Letters*, 2017,2(2):796-803.
- [47] Mur-Artal R, Tardos J D. Probabilistic semi-dense mapping from highly accurate feature-based monocular SLAM[C]//Proc of Robotics: Science and Systems. 2015.
- [48] Bourmaud G, Mégret R. Robust large scale monocular visual SLAM[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015;1638-1647.
- [49] Newcombe R A, Davison A J. Live dense reconstruction with a single moving camera[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010;1498-1505.
- [50] Stühmer J, Gumhold S, Cremers D. Real-time dense geometry from a handheld camera[C]//Proc of Dagm Conference on Pattern Recognition. [S. l.]:Springer-Verlag, 2010;11-20.
- [51] Pizzoli M, Forster C, Scaramuzza D. REMODE: probabilistic, monocular dense reconstruction in real-time[C]//Proc of IEEE International Conference on Robotics and Automation. 2014;2609-2616.
- [52] Engel J, Sturm J, Cremers D. Semi-dense visual odometry for a monocular camera[C]//Proc of IEEE International Conference on Computer Vision. 2013;1449-1456.
- [53] Valgaerts L, Bruhn A, Mainberger M, *et al.* Dense versus sparse approaches for estimating the fundamental matrix[J]. *International Journal of Computer Vision*, 2012,96(2):212-234.
- [54] Ranftl R, Vineet V, Chen Qifeng, *et al.* Dense monocular depth estimation in complex dynamic scenes[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- [55] Silveira G, Malis E, Rives P. An efficient direct approach to visual SLAM[J]. *IEEE Trans on Robotics*, 2008,24(5):969-979.
- [56] Schops T, Enge J, Cremers D. Semi-dense visual odometry for AR on a smartphone[C]//Proc of IEEE International Symposium on Mixed and Augmented Reality-Media, Arts, Social Science, Humanities and Design. 2014;145-150.
- [57] Glover A, Maddern W, Warren M, *et al.* OpenFABMAP: an open source toolbox for appearance-based loop closure detection[C]//Proc of IEEE International Conference on Robotics and Automation. 2012;4730-4735.
- [58] Caruso D, Engel J, Cremers D. Large-scale direct SLAM for omnidirectional cameras[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2015;141-148.
- [59] Engel J, Stückler J, Cremers D. Large-scale direct SLAM with stereo cameras[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2015;1935-1942.
- [60] Usenko V, Engel J, Stückler J, *et al.* Reconstructing street-scenes in real-time from a driving car[C]//Proc of International Conference on 3D Vision. 2015;607-614.
- [61] Jin Hailin, Favaro P, Soatto S. A semi-direct approach to structure from motion[J]. *The Visual Computer*, 2003,19(6):377-394.
- [62] Bu Shuhui, Zhao Yong, Wan Gang, *et al.* Semi-direct tracking and mapping with RGB-D camera for MAV[J]. *Multimedia Tools & Applications*, 2017,76(3):1-25.
- [63] Baker S, Matthews I. Lucas-Kanade 20 years on: a unifying framework[J]. *International Journal of Computer Vision*, 2004,56(3):221-255.
- [64] Leutenegger S, Furgale P, Rabaud V, *et al.* Keyframe-based visual-inertial SLAM using nonlinear optimization[C]//Proc of Robotics: Science and Systems. 2013;789-795.
- [65] Leutenegger S, Lynen S, Bosse M, *et al.* Keyframe-based visual-inertial odometry using nonlinear optimization[J]. *International Journal of Robotics Research*, 2015,34(3):314-334.
- [66] Comport A I, Malis E, Rives P. Accurate quadrifocal tracking for robust 3D visual odometry[C]//Proc of IEEE International Conference on Robotics and Automation. 2007;40-45.
- [67] Kerl C, Sturm J, Cremers D. Dense visual SLAM for RGB-D cameras[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2013;2100-2106.
- [68] Leung C, Huang Shoudong, Dissanayake G. Active SLAM using model predictive control and attractor based exploration[C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2006;5026-5031.
- [69] Indelman V, Carlone L, Dellaert F. Planning in the continuous domain: a generalized belief space approach for autonomous navigation in unknown environments[J]. *International Journal of Robotics Research*, 2015,34(7):1021-1029.
- [70] Jur V D B, Patil S, Alterovitz R. Motion planning under uncertainty using iterative local optimization in belief space[J]. *International Journal of Robotics Research*, 2012,31(11):1263-1278.
- [71] Clady X, Ieng S H, Benosman R. Asynchronous event-based corner detection and matching[J]. *Neural Networks*, 2015,66(C):91-106.