

分类号：TP242.6

学校代码：10363

密 级：公开

学 号：2210330137



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 多结构特征融合的室内视觉
SLAM 算法研究

论 文 作 者 龚光强

校 内 导 师 孙新柱 副教授

校 外 导 师 于晓东 高级工程师

专业学位类别 电子信息

研究方向（领域） 人工智能与系统集成

论文提交日期：2024 年 5 月 31 日

分类号: TP242.6

单位代码: 10363

密 级: 公 开

学 号: 2210330137

题 目 多结构特征融合的室内视觉 SLAM 算法研究

题 目 Research on Visual SLAM Algorithm for Indoor with Multi-structural Feature Fusion

学生姓名: 龚光强

校内教师: 孙新柱 (副教授)

校外教师: 于晓东 (高级工程师)

专业学位类别: 电子信息

研究方向 (领域): 人工智能与系统集成

论文答辩日期: 2024 年 5 月 16 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律结果由本人承担。

学位论文作者签名：虞礼强

日期：2024年 5 月 30 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：焦光强

指导教师签名：孙红书

日期：2024年 5 月 30 日

日期：2024年 5 月 30 日

多结构特征融合的室内视觉 SLAM 算法研究

摘 要

视觉同步定位与地图构建 (Simultaneous Localization and Mapping, SLAM) 是指移动机器人通过相机获取未知环境信息构建地图, 并使用地图实时查找自身位置的过程, 是机器人实现自主导航的前提, 已成为该领域研究的热点。目前大部分室内视觉 SLAM 系统仅利用图像中的点特征估计实时位姿, 忽视了室内结构化场景丰富的线特征和面特征。仅依靠点特征的纯视觉 SLAM 系统常常会因为特征退化导致性能下降, 严重时会使 SLAM 系统跟踪丢失。现代室内场景减少了不必要的装饰和繁杂的细节, 但存在明显的轮廓分界、空旷的房间和单调的墙壁。与点特征相比, 线特征和面特征包含更丰富的几何信息, 能提供更多的约束条件, 并且具有较强的空间连续性, 可以更直观地表示室内环境的空间结构, 有助于提高 SLAM 系统的稳定性和准确性。在室内结构化场景下, 针对视觉 SLAM 系统因特征退化引起定位失准的核心问题, 充分利用环境中的点、线和面等基础特征, 构建更稳定的结构特征, 重点解决纯视觉 SLAM 方法在室内结构化场景中的定位精度和鲁棒性下降等问题。

针对传统视觉里程计在室内纹理稀疏以及光照不均匀环境下存在特征退化和定位失准的问题, 提出基于点、线结构特征辅助的 RGB-D 视觉里程计算法。首先, 为克服因室内光照不均匀造成线特征提取稳定性差的问题, 通过引入水平与垂直约束的像素连接规则, 改进边

缘拟合线特征提取方法，使提取的线段具有更好的连续性和鲁棒性；然后利用线特征的相交关系构建更稳定的结点-线结构特征，减少旋转累计误差，并通过结点辅助匹配提高定位精度；最后，构造最小点线重投影误差函数，估计六自由度位姿。

针对基于点特征的视觉 SLAM 系统在室内结构化场景位姿估计精度降低的情况，提出基于曼哈顿假设的点、线和面多特征视觉 SLAM 方法。考虑到传统面特征提取算法检测精度差的问题，使用改进的 PlaneRecNet 网络提高平面分割的效果。针对曼哈顿世界假设的应用场景受限问题，利用平面特征的结构关系判断曼哈顿假设条件。在非曼哈顿场景使用基于点、线的结构特征辅助定位方法；对于曼哈顿场景，采用方向引导策略筛选主平面，比较和跟踪平面结构的正交性估计三自由度旋转矩阵，降低旋转累积误差。

为验证本文算法的性能，在公开的 TUM RGB-D 数据集、ICL-NUIM 数据集和室内真实场景分别对具有结构特征辅助的 RGB-D 视觉里程计算法和基于曼哈顿假设的多特征视觉 SLAM 方法进行了实验。与 ORB-SLAM3、MSC-VO、Planar-SLAM 和 Manhattan-SLAM 进行了对比分析，验证本文算法在室内结构化场景下的定位准确性。在视觉移动硬件平台上的测试，验证本文算法具有较强的环境适应能力，为视觉 SLAM 在室内场景的应用研究提供参考依据。

关键词：同步定位与地图构建，结构化场景，结构特征，曼哈顿世界假设，平面分割

RESEARCH ON VISUAL SLAM ALGORITHM FOR INDOOR WITH MULTI-STRUCTURAL FEATURE FUSION

ABSTRACT

Visual Simultaneous Localization and Mapping (SLAM) is a process in which a mobile robot constructs a map by acquiring unknown environment information through a camera and uses the map to find its own position in real time, which is a prerequisite for robots to achieve autonomous navigation and has become a hot spot of research in this field. Most of the current indoor visual SLAM systems only use point features in images to estimate real-time position, ignoring the rich line and surface features of indoor structured scenes. Pure visual SLAM systems relying only on point features often suffer from performance degradation due to feature degradation, and in severe cases, the SLAM system will lose tracking. Modern indoor scenes have reduced unnecessary decorations and busy details, but there are obvious contour demarcations, empty rooms and monotonous walls. Compared with point features, line features and surface features contain richer geometric information, provide more constraints, and have stronger spatial continuity, which can represent the spatial structure of indoor environments more intuitively and help to improve the stability and accuracy of the SLAM system. In indoor structured scenes, to address the core problem of localization misalignment caused by feature

degradation in visual SLAM systems, we make full use of the basic features such as points, lines and surfaces in the environment to construct more stable structural features, focusing on solving the problems such as degradation of localization accuracy and robustness of purely visual SLAM methods in indoor structured scenes.

Aiming at the problems of feature degradation and localization misalignment of traditional visual odometry in indoor environments with sparse texture and uneven illumination, the RGB-D visual odometry calculation method based on the assistance of point and line structural features is proposed. First, to overcome the problem of poor stability of line feature extraction due to uneven indoor illumination, the edge-fitting line feature extraction method is improved by introducing the pixel connection rule with horizontal and vertical constraints, so as to make the extracted line segments have better continuity and robustness; then, the intersection relationship of the line features is utilized to construct a more stable node-line structural feature, to reduce the rotational cumulative error and to improve the positioning accuracy; finally, the minimum point-line reprojection error function is constructed to estimate the six-degree-of-freedom position.

Aiming at the reduced accuracy of point-feature-based visual SLAM systems for indoor structured scene position estimation, a multi-feature visual SLAM method based on the Manhattan assumption for points, lines

and faces is proposed. Considering the poor detection accuracy of traditional face feature extraction algorithms, an improved PlaneRecNet network is used to improve the effectiveness of plane segmentation. To address the problem of limited application scenarios for the Manhattan world assumption, the structural relationships of planar features are used to determine the Manhattan assumption conditions. In the non-Manhattan scenarios, the structural features based on points and lines are used to assist the localization method; for the Manhattan scenarios, a direction-guided strategy is used to screen the main plane, and the orthogonality of the plane structures is compared and tracked to estimate the three-degree-of-freedom rotation matrix and reduce the rotational cumulative error.

To validate the performance of the algorithms in this paper, the RGB-D visual mileage computation method with structural feature assistance and the multi-feature visual SLAM method based on the Manhattan assumption are experimented on the publicly available TUM RGB-D dataset, the ICL-NUIM dataset, and indoor real-world scenes, respectively. Comparative analyses with ORB-SLAM3, MSC-VO, Planar-SLAM and Manhattan-SLAM are conducted to validate the localization accuracy of this paper's algorithms in indoor structured scenes. Tests on the visual mobile hardware platform verify that the algorithm in this paper has strong environmental adaptability, and provide a reference basis for the application research of visual SLAM in indoor scenes.

Gong Guangqiang (Electronic Information)

Supervised by Associate Professor Sun Xinzhu

KEY WORDS: Simultaneous Localization and Mapping, Structured scenes, Structural features, Manhattan world hypothesis, Planar segmentation

目 录

第 1 章 绪论	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 视觉 SLAM 技术研究现状	3
1.2.2 室内结构化场景的视觉 SLAM 研究现状	5
1.3 论文研究内容与组织结构.....	7
第 2 章 视觉 SLAM 与结构特征基础理论.....	9
2.1 视觉相机模型.....	9
2.2 图像特征提取算法.....	12
2.2.1 点线特征提取.....	12
2.2.2 平面分割方案.....	14
2.3 结构特征与曼哈顿世界假说.....	17
2.3.1 结构特征.....	17
2.3.2 曼哈顿世界假说模型.....	18
2.3.3 曼哈顿主方向提取.....	19
2.4 本章小结.....	20
第 3 章 具有结构特征辅助的 RGB-D 视觉里程计	21
3.1 改进边缘拟合的线特征提取.....	21
3.2 基于线相交关系的结构特征提取.....	23
3.2.1 结点-线结构特征构建	23
3.2.2 结点-线结构特征提取与匹配	24
3.3 相机位姿估计.....	24
3.3.1 结点-线观测模型	24
3.3.2 帧间位姿计算.....	26
3.4 实验结果与分析.....	27
3.4.1 改进线特征提取与匹配实验.....	27
3.4.2 结点-线结构特征提取与匹配实验	28

3.4.3 定位精度评估实验.....	30
3.5 本章小结.....	33
第 4 章 基于曼哈顿假设的多特征视觉 SLAM 方法.....	34
4.1 基于改进 PlaneRecNet 网络的平面分割.....	34
4.2 方向引导的主平面筛选.....	36
4.3 曼哈顿坐标系跟踪及位姿估计.....	37
4.4 实验结果与分析.....	39
4.4.1 平面分割实验.....	40
4.4.2 ICL-NUIM 数据集轨迹对比实验	41
4.4.3 TUM RGB-D 数据集轨迹对比实验	43
4.4.4 真实场景实验结果与分析.....	45
4.5 本章小结.....	48
第 5 章 全文总结与展望	49
5.1 全文总结.....	49
5.2 研究展望.....	50
参考文献	51
攻读学位期间发表的学术论文与取得的科研成果	56
致 谢	57

第 1 章 绪论

1.1 研究背景及意义

自动化技术发展的目的在于提高生产效率的同时，减轻劳动者的负担，将人类从繁重和危险的任务中解放出来，使人们能够专注于更具创造性、智力密集性的工作。随着自动化技术的进步，机器人的理论研究得到了蓬勃发展，早期的工业机器人主要用于重复简单的生产任务。而随着人工智能、传感器和机器视觉等技术的不断进步，现代机器人已经具备了更加强大的感知、决策和执行能力^[1]。目前，机器人技术正面临着前所未有的机遇。在医学领域，随着人口老龄化问题的日益突出，机器人在医疗保健、护理等方面的需求不断增加，医疗机器人可以用于手术辅助和康复治疗等任务；在智能驾驶领域，无人驾驶汽车将推动物流运输和城市交通系统的数字化与智能化，为人们带来更加便捷的出行方式和物流服务；在家用机器人领域，随着人们生活水平的提高，室内清洁和教育培训等场景为机器人的应用提供了新的机会^[2]。随着相关技术的不断进步和政策法规的逐渐完善，这些领域将迎来更广阔的发展前景，常见的移动机器人如图 1-1 所示。



图 1-1 智能移动机器人

一直以来，自主定位始终是机器人技术领域亟待解决的重要问题之一。机器

人之所以能够完成各项复杂的工作,得益于机器人能够利用各类传感器像人类一样感知周围环境。机器人同步定位与地图构建是智能机器人在陌生场景中通过自身携带的各类传感器感知周围信息来估计自身位置,并建立所处场景三维点云地图的技术^[3]。当前移动机器人主要利用激光雷达和相机观测场景信息数据,惯性测量单元和轮速计作为辅助,以达到较高的定位精度^[4]。SLAM 技术使得移动机器人能够在未知环境中实现自主导航,不依赖于预先构建的地图或外部定位系统,提高了机器人的自主性和灵活性,在实现复杂场景中移动机器人高智能和快决策方面起着至关重要的作用。

近年来,随着视觉传感器的发展,一般相机已具有低成本和应用场景丰富等特点,搭载相机的移动智能设备逐渐成为研究的热点和难点,目前已经有一些优秀的开源算法在无人驾驶领域得到了应用^[5]。但是,通过跟踪图像特征进行运动估计不可避免地会受到环境纹理和光照变化的影响,现有的纯视觉 SLAM 算法大多仅利用点特征,且适用于纹理丰富、光照充足的场景^[6]。随着生活水平的提高,现代室内装修设计通常倾向于简约、清晰的线条和几何形状,减少了不必要的装饰和繁杂的细节,更多的是注重简洁而富有节奏感的空间布局,存在明显的轮廓分界、空旷的房间和单调的墙壁。移动机器人长时间在结构简约、纹理稀疏或光照不均匀的现代室内结构化场景中运行,往往很难找到足够可靠的点特征来进行状态估计,会造成误差累积,导致轨迹紊乱甚至跟踪失败^[7]。为应对该问题,除了采用额外传感器作为辅助,充分利用图像中可检测到的其他特征也是一种有效方法,如线特征、语义特征和平面特征^[8]。点特征在纹理稀疏和光照变化的环境中缺乏鲁棒性,与之相比,线特征具有清晰的几何形状和方向,包含了丰富的几何信息,可以提供更多的约束条件,帮助机器人更准确地定位和建立地图。同时,线特征的提取过程比点特征更加高效,更容易被识别和提取。现有的一些算法已经证明,合理利用线特征可以有效提高系统的定位精度和地图构建效率。在视觉 SLAM 技术中,平面是比点和线具有更多约束的结构特征,具有较强的空间连续性,可以更直观地表示室内环境的空间结构,有助于提高 SLAM 系统的稳定性和准确性^[9]。曼哈顿世界(Manhattan World, MW)假设^[10]是一种描述特定空间结构的模型。在室内结构化场景中,规整的平面通常会呈现相互正交的关系,例如地面、墙壁和天花板。这些平面法线矢量分布在空间中三个相互正交的

方向。将沿着这三个方向建立的坐标系定义为曼哈顿坐标系（Manhattan Frame, MF），大多数室内结构化场景也可称作曼哈顿场景，可以被形式化为 MW 假设模型，该模型简化了对室内环境的建模和理解。目前该方法已应用于包括视觉 SLAM 在内的多种技术层面，例如相机校准和三维场景重建。

随着视觉 SLAM 技术的进步，面对纹理稀疏以及光照不均匀的室内结构化场景，除了采用其他传感器进行辅助，充分利用点、线和平面等基础特征也是一种较好的方案。本文以室内场景的移动机器人视觉 SLAM 技术为研究目标，提出一种多结构特征融合的室内视觉 SLAM 方法。针对传统视觉 SLAM 在室内纹理稀疏以及光照不均匀环境下存在特征退化和定位失准这一核心问题，提出具有结构特征辅助的 RGB-D 视觉里程计和基于曼哈顿假设的多特征视觉 SLAM 方法。重点研究室内场景的线特征和平面特征提取问题，以及如何构建结构特征观测模型和曼哈顿坐标系跟踪方法。

1.2 国内外研究现状

1.2.1 视觉 SLAM 技术研究现状

SLAM 是使机器人能够在没有外部定位系统（如全球导航卫星系统 GNSS）的情况下，根据内置传感器定位自身位置同时绘制未知环境地图的技术。这项技术最早可以追溯到 1986 年 Smith 等人发表的系列论文《空间位置的表达与估计》所研究的未知环境中机器人导航问题，在这一时期，统计概率的方法刚开始被引入到机器人和人工智能领域，SLAM 还需要依赖价格昂贵或专门定制的传感器，这项技术并未走入市场。2006 年，Durrant-Whyte 和 Bailey 在发表的论文中首次使用了“SLAM”这一名词^[1]，探讨了信息传递、运算效率和收敛性等问题，为定位与导航问题构建了最初的理论分析框架，从此 SLAM 技术真正走向了系统性研究阶段。传统移动机器人通常使用激光雷达采集点云信息进行定位与建图，虽然在精度和稳定性方面表现优异，但能够满足高精度定位的激光雷达价格太过昂贵，限制了实际应用中的普及和推广。近年来，基于视觉传感器的 SLAM 方法备受关注。相比于激光雷达，视觉传感器具有成本低、体积小以及能够获取更丰富环境信息等优点^[12]。基于视觉的 SLAM 方法可以使用单目、双目以及 RGB-

D 相机等传感器，通过图像处理和计算机视觉技术实现对环境的感知和机器人定位。目前视觉 SLAM 系统在硬件、算法和应用方面的研究都取得了重要进展，视觉 SLAM 经典框架如图 1-2 所示。

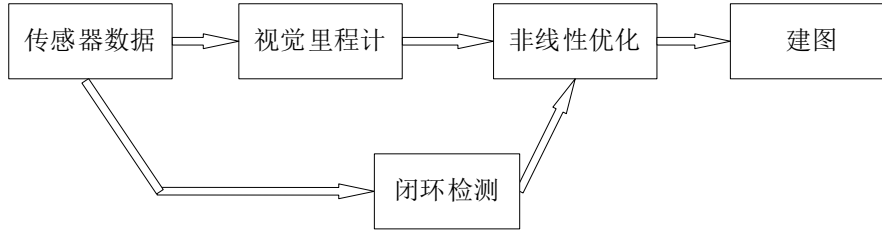


图 1-2 视觉 SLAM 经典框架

在过去二十年现代机器人不断发展的过程当中，视觉 SLAM 已经成为一种强大的工具。视觉 SLAM 研究主要分为概率滤波和图优化两类方法，早期基于滤波的方法通过状态估计算法（如卡尔曼滤波和粒子滤波）进行对机器人传感器数据的观测和运动模型的预测，并且计算机器人在不同时刻的状态概率分布^[13]。在这个过程中，通常会依赖系统的 Markov 性假设，即当前时刻的状态仅与前一时刻的状态和控制输入相关，只考虑相邻两帧图像之间的关系。这就意味着很容易受到传感器噪声、模型误差或环境变化等因素的影响，估计的误差会随着机器人的运行大量累积，导致最终定位和建图的结果不准确。

Davison 等人^[14]提出了使用扩展卡尔曼滤波进行位姿估计和稀疏地图重建的视觉 SLAM 系统—MonoSLAM，使用单目相机获取的图像信息对机器人的位置和地图进行估计和更新，通过特征提取和匹配算法提取图像序列中的关键特征点，并跟踪这些特征点在不同图像帧之间的运动。然后，通过三角化方法确定特征点的空间位置，并估计机器人的位姿和环境地图。由 Klein 等人^[15]于 2007 年提出的 PTAM 首次实现了并行跟踪与建图，即同时在两个线程中进行相机的位置跟踪和环境地图的构建。这种并行化的设计使得系统能够在实时性和鲁棒性之间取得平衡，有效提高了 SLAM 系统的性能。Mur-Artal 等人^[16]在 PTAM 系统的基础上进行改进，开源了 SLAM 系统的标杆之作 ORB-SLAM，首次实现了跟踪、局部地图构建和闭环检测三个线程的数据关联。在团队后续的研究中，ORB-SLAM2^[17]可应对多样化场景，增加了双目和 RGB-D 模式。2020 年，由 Campos 等人开源的 ORB-SLAM3^[18]作为首个基于特征的紧耦合 VIO 系统，可以在单目、双目和 RGB-D 相机上利用针孔或者鱼眼模型运行，在视觉相机基础上增加惯性测量单元（Inertial measurement unit, IMU），以解决纯视觉模式或者视觉惯导融合

模式的完整混合地图 SLAM 技术问题, ORB-SLAM3 比上一代 ORB-SLAM2 在精度、鲁棒性和实时性等方面都有了显著提高, 已成为视觉 SLAM 技术领域的杰出代表, 但仅利用点特征, 难以在纹理稀疏场景持续稳定运行。

随着视觉 SLAM 技术研究的深入, 越来越多的研究者将视觉 SLAM 技术投入应用当中。Lajoie 等人^[19]开源了多机协作 Swarm-SLAM 系统, 在没有外部定位系统的环境中成功进行多机器人协同, 该系统具备群体机器人技术的关键特性, 即具有较强的可扩展性和灵活性, 并且支持激光雷达、双目和 RGB-D 相机, 目前算法已具有基于 ROS2 的实现。Kachurk 等人^[20]提出了一种可穿戴式视觉惯性多传感器协同定位方法 WeCo-SLAM, 该方法是一种在室内/室外环境下具有定位应急响应机制的实时跟踪系统, 由多个 SLAM 方法辅助的姿态估计可为穿戴者提供准确的自身位置信息。苏黎世联邦理工学院团队^[21]提出一种辅助无人机导航的双目视觉惯性 SLAM 方法, 使用两台配备单目摄像机和 IMU 的无人机, 通过视角重叠和相对距离测量结果来实现协作式视觉惯性里程计, 实现了姿态估计低延迟, 同时通过基于共识度的优化方法确保每个智能体的同步。以上是视觉 SLAM 技术应用的案例, 但并未考虑特征退化可能带来的定位问题。

1.2.2 室内结构化场景的视觉 SLAM 研究现状

通过跟踪图像特征进行运动估计不可避免地会受到环境纹理和光照变化的影响, 而不同类型的图像特征对光照变化的敏感度不同。点特征在低纹理和光照变化的室内结构化场景中缺乏稳定性, 与之相比, 线特征则更具鲁棒性, 并且可以更直观地表示室内环境的空间结构。2019 年, Gomez-Ojeda 等人^[22]开源了经典的双目 PL-SLAM, 利用双目相机的视差信息感知深度, 能够更准确地估计相机的运动轨迹, 同时重构环境地图。采用 LSD 直线检测算法 (Line Segment Detector, LSD)^[23]提取图像线特征, 通过线特征的几何约束提高 SLAM 系统在点特征退化情况下的稳定性和鲁棒性。2021 年, 于雅楠等人^[24]通过信息熵大小对图像块进行筛选, 对信息熵较大的图像块进行亮度自适应调整, 增强细节, 从而增加局部特征点提取数量。尽管已经提出大量改进方法, 但基于点的定位方案仍然难以在室内场景保持较高性能。

室内结构化场景具有丰富的边缘和明显的线性特征, 但现有的大多数点线定

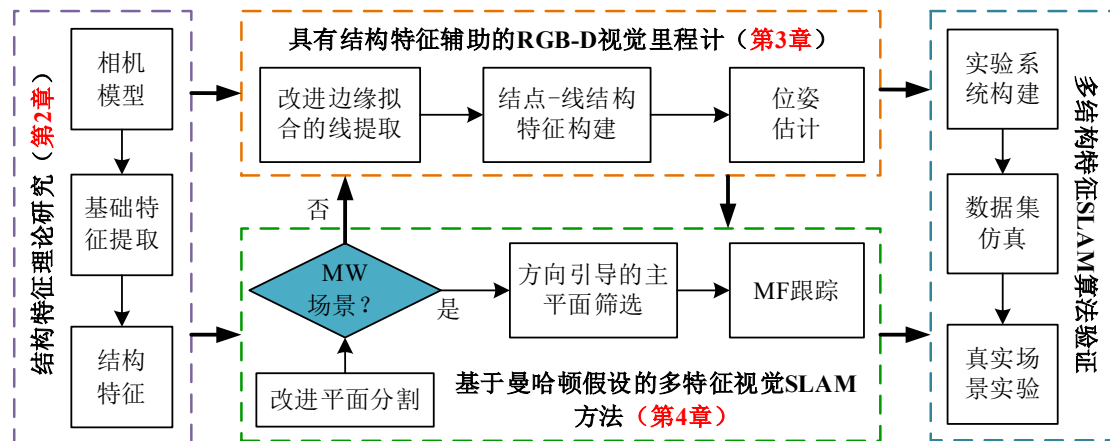
位方法并未充分利用线特征之间的空间结构关系,例如平行、正交、共面和相交,这些结构关系提供的几何约束能为 SLAM 系统带来更高的效率 and 更精准的定位^[25]。其实,在诸如曼哈顿世界和亚特兰大世界这样的结构化场景中,可以利用直线之间的平行关系减少直线的参数,避免优化过程的超参数问题^[26]。2021 年,Corcoles 等人开源的 MSC-VO^[27]提出了一种局部地图优化方法,在局部地图优化阶段结合了曼哈顿假设以构建点和线重投影误差函数,将线的正交性约束与重投影误差和曼哈顿轴相结合,通过局部优化来细化相机姿态。邹丹平教授团队提出的 StructVIO^[28]系统仅使用反深度和角度这两个参数来表示特定方向的直线,此外,直线之间共面性约束也被引入到 BA 优化中,使线特征之间的结构关系得到了进一步利用。闫莉萍等人^[29]针对传统的仅依赖点特征的视觉里程计鲁棒性较差的问题,提出了一种结合点特征、线特征以及结构特征的视觉-惯性里程计。使用结构特征描述结构物体所包含的平行与正交信息,为视觉惯性里程计提供了全局约束,可以有效减少旋转误差的累计。2022 年,Lim 等人提出的 UV-SLAM^[30]以 ORB-SLAM2 为基础,在没有约束的情况下使用结构规律,利用线特征的平行关系获得消失点,将通过图像中的线特征与通过方向向量计算的消失点之间的差异被定义为残差,并添加到基于优化的 SLAM 代价函数中。虽然线特征的结构关系正在被逐渐用于各种定位方法,但并没有得到充分开发。

近年来,平面特征也逐渐被应用于 SLAM 系统。Kim 等人^[31]提出基于线和平面特征的视觉里程计,提出了一种在结构化环境中的正交平面检测方法,通过 SO(3)流型约束均值漂移算法跟踪 MF 来降低旋转误差。2019 年,Yang 等人^[32]将平面特征引入到 SLAM 系统中,并采用统一的优化框架同时处理点、面特征。孙沁璇等人^[33]提出针对 SLAM 中平面特征位姿求解易出现退化问题的解决方案,采用多特征关联方法,将 MW 中涉及的结构特征融合到 SLAM 中。通过直线和平面特征增强系统的稳定性和鲁棒性,发挥位姿计算过程中面特征的主导作用。Li 等人^[34]针对室内场景的定位与建图问题,于 2020 年提出了 Structure-SLAM,该算法一改传统平面提取方式,率先采用卷积神经网络,从单帧彩色图像中直接提取平面特征及其参数,并结合当前帧的平面特征和线特征计算出 MF,从而实现对当前旋转矩阵的计算。2021 年,Li 等人的课题组开源了支持 RGB-D 模式、基于结构化约束的 Planar-SLAM^[35],首先提出平面实例化网格方法,在稀疏点云

中生成轻量化环境表征。基于曼哈顿假设原理，对位姿进行解耦估计，结合点、线和平面三种特征来提高系统在结构化弱纹理场景下的稳定性。Planar-SLAM 在追踪线程中还包含了对位姿的优化过程，即利用平行或垂直结构规律来优化姿态。但解耦方案对于每一帧都要求符合 MW 假设，这在室内环境中将导致方法的应用范围受限，性能可能会下降，甚至导致跟踪失败。同样在 2021 年，Yunus 等人也提出了基于 RGB-D 相机的 Manhattan-SLAM^[36]，该算法利用点、线和面特征的融合来构建室内场景模型，并且在 MW 假设的基础上，采用了点、线特征跟踪和 MF 跟踪两种策略，有效改善了 SLAM 系统的定位和建图性能。然而，所采用的聚集层次聚类（Agglomerative Hierarchical Clustering, AHC）^[37]点云分割易受深度信息噪声的影响，这可能造成 MF 的高误差。

1.3 论文研究内容与组织结构

本文主要研究室内结构化场景下，视觉 SLAM 系统因特征退化导致性能下降的问题及改进方法，提出了具有结构特征辅助的视觉里程计算法以及基于曼哈顿假设的多特征视觉 SLAM 方法。在基础特征提取阶段，改进边缘拟合线特征提取和平面分割方法，克服传统方法在室内场景的线特征和平面特征提取准确性差的问题；在多结构特征融合位姿估计阶段，利用平面的结构关系判断 MW 假设条件，在非曼哈顿场景构建结点-线结构特征，采用具有结构特征辅助的定位方法；在曼哈顿场景，首先通过方向引导策略筛选主平面，再跟踪 MF 估计三自由度旋转矩阵。在算法验证部分，通过公开数据集和真实场景对本文算法进行测试。本文整体框架内容如图 1-3 所示。



本文研究内容共分为五个章节，安排如下：

第一章，介绍本文的研究背景、意义和 SLAM 技术的发展现状。首先从机器人技术出发，根据视觉 SLAM 在机器人技术中的重要地位，分析总结了当前研究视觉 SLAM 技术的重要意义；然后介绍了国内外视觉 SLAM 技术的发展现状；最后对本文主要内容进行了概述，并对各章节内容进行安排。

第二章，介绍多结构特征 SLAM 方法的相关理论基础。首先介绍了相机传感器的相机投影模型以及畸变模型；然后针对本文方法，对当前图像点特征、线特征以及面特征提取算法进行了简单分析；最后介绍了 MW 假说以及结构特征的背景以及理论，为本文后续结构特征的构建做好理论研究基础。

第三章，介绍具有结构特征辅助的 RGB-D 视觉里程计算法。在弱纹理以及光照不均匀的室内环境下，针对传统视觉里程计存在特征退化和定位失准的问题，提出一种具有结构特征辅助的 RGB-D 视觉里程计算法。首先，提出改进边缘拟合的线特征提取方法，通过引入水平与垂直方向约束的像素连接规则，缓解因室内光照不均匀导致线特征提取稳健性差的问题。然后，充分利用环境中线特征之间的相交关系，构建结点-线结构特征，减少累计误差，提高定位精度。最后，构造最小点线重投影误差函数，估计六自由度位姿。

第四章，介绍基于曼哈顿假设的多特征视觉 SLAM 方法。针对室内结构化场景下，基于点特征的视觉 SLAM 算法出现定位精度差甚至跟踪失败的问题，提出一种基于曼哈顿假设的多特征视觉 SLAM 方法。首先，使用改进的 PlaneRecNet 网络提高平面特征提取准确性，克服传统面特征提取算法检测精度差的问题。然后，利用平面正交结构关系判定 MW 假设条件，改善仅利用 MW 假设估计位姿时的应用场景受限问题，在非曼哈顿场景采用具有结构特征辅助的位姿估计方法；对于满足 MW 假设的场景，采用方向引导方法进行主平面筛选，通过平面结构的正交性进行三自由度旋转估计，从而降低旋转累积误差带来的漂移。最后，构造最小重投影误差函数，进行非线性优化，提升算法性能。

第五章，总结与展望。对本文的工作和研究成果进行总结，并提出完善方向与展望。

第 2 章 视觉 SLAM 与结构特征基础理论

本章主要介绍视觉 SLAM 与结构特征相关理论基础。首先叙述视觉相机投影模型以及畸变模型；然后针对本文方法，对当前主流图像点特征和线特征提取算法进行对比，重点分析了当前平面分割方案；最后对 MW 假说以及结构特征的背景以及理论进行介绍。为后续对本文方法的详细阐述做好基础理论准备。

2.1 视觉相机模型

相机是机器人进行视觉感知的传感器，当前在机器人领域主要应用的有三种类型相机，即单目相机、双目相机和 RGB-D 相机。单目相机的构造比较简单，仅由镜头和图像传感器（CMOS 或 CCD）构成，成像的本质是将三维世界物体点通过坐标系的转换投影到二维平面当中，成像过程可以通过简单有效的针孔模型来实现。针孔模型原理如图 2-1 所示：

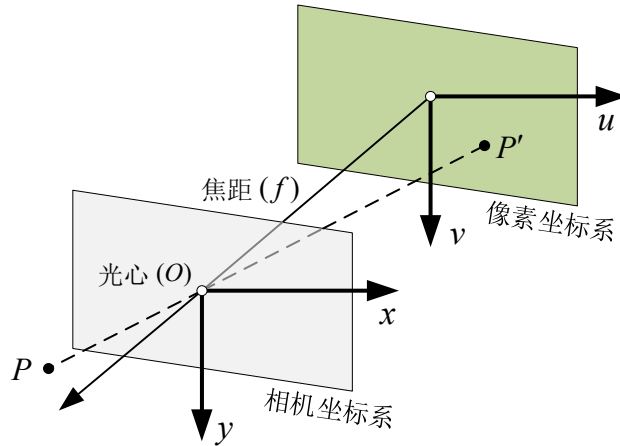


图 2-1 针孔相机模型原理

世界坐标系中的 P 点在相机坐标系下的坐标为 (X, Y, Z) ，当光线经过相机的光学中心 O 投影到图像传感器上时，在像素坐标系下形成的点为 P' ，设点 P' 在像素坐标系下的坐标值为 (U, V) ，通过相机的焦距，可以确定点 P 在图像传感器上的投影位置，从而建立相机与像素坐标系点之间的关系，如式 (2-1) 所示：

$$\begin{cases} U = \alpha \left(\frac{f}{Z} \cdot X \right) + c_x \\ V = \beta \left(\frac{f}{Z} \cdot Y \right) + c_y \end{cases} \quad (2-1)$$

式中 α 和 β 分别代表图像传感器的纵向和横向量化尺度， c_x 和 c_y 分别表示两个坐标系原点之间的偏移量。利用相似三角形的性质，可推导出成像点 P' 在像素坐标系下的坐标，如式（2-2）所示：

$$P' = \left[\left(\frac{f}{Z} \cdot X \right), \left(\frac{f}{Z} \cdot Y \right) \right] \quad (2-2)$$

由于 α 和 β 不同，并且在相机制造和安装过程中存在机械误差，导致相机坐标系原点与像素坐标系原点不完全对齐。在计算最终的像素值时，设 $f_x = \alpha \cdot f$ ， $f_y = \beta \cdot f$ ，将计算转换为全部在像素单位下进行，相机的无畸变内参模型为：

$$\begin{bmatrix} U \\ V \\ 0 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \frac{1}{Z} \mathbf{K} \mathbf{P} \quad (2-3)$$

式中矩阵 $\mathbf{K}$ 代表相机的内参数。在针孔相机模型中，物体点 P 直接沿直线透过光心 O 形成图像点 P' 。尽管相机镜头可以增加进光量，加快曝光速度，但也会引起光线的折射现象，导致图像产生变形。这种由镜头折射引起的图像畸变叫径向畸变，如图 2-2 所示。

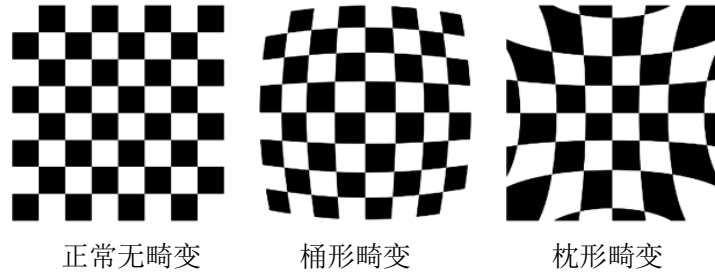


图 2-2 径向畸变

除了径向畸变外，还有相机制造过程中，因安装不当导致镜头与图像传感器不平行产生的切向畸变，如图 2-3 所示。

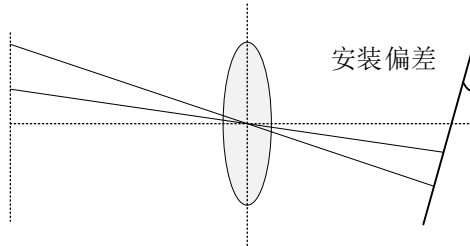


图 2-3 切向畸变示意图

对于切向畸变，其畸变程度和图像传感器安装偏差大小有关，可以用系数 p_1

和 p_2 描述。用像素点距中心距离 r 的三阶泰勒级数来近似描述畸变的程度，级数中的系数用 k_1 、 k_2 和 k_3 表示，对于径向畸变不明显的镜头，只需要考虑 k_1 和 k_2 两个系数，部分情况下甚至只需要考虑 k_1 。综合相机的径向和切向畸变，内参模型如式（2-4）所示：

$$\begin{cases} r^2 = X^2 + Y^2 \\ X_{distort} = X \cdot (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 XY + p_2 (r^2 + 2X^2) \\ Y_{distort} = Y \cdot (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 XY + p_1 (r^2 + 2Y^2) \\ U = \frac{f_x}{Z} \cdot X_{distort} + c_x \\ V = \frac{f_y}{Z} \cdot Y_{distort} + c_y \end{cases} \quad (2-4)$$

在实际应用中，空间点 P 给出的往往是世界坐标系下的坐标值 $P^w = (X^w, Y^w, Z^w) \sqrt{a^2 + b^2}$ 。但是在相机内参模型中，通常给定的是相机坐标系下的坐标 $P^c = (X^c, Y^c, Z^c)$ 。因此需要根据相机在世界坐标系下的位姿信息 (R, t) 对空间点 P 进行坐标转换，如公式（2-5）所示：

$$\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \frac{1}{Z} K P^c = \frac{1}{Z} K (R P^w + t) = \frac{1}{Z} K T P^w \quad (2-5)$$

式中的矩阵 T 代表相机外部参数，描述了相机在世界坐标系中的位姿，也是 SLAM 问题中待求的参数之一。从公式(2-4)可以看出，已知物体成像信息 (U, V) ，但是无法唯一确定点物体点 P 在相机坐标系下的坐标值 (X, Y, Z) 。如图 2-1 所示， P' 沿光心 O 方向上出现的所有物体点，在相机中的成像位置都在 P' ，但缺少深度信息。

RGB-D 相机是一种集成了彩色图像传感器和深度传感器的设备，通过在 RGB 图像每个像素上使用深度信息来增强传统图像。RGB 色彩模式是目前工业和计算机视觉领域运用最广的颜色标准，几乎涵盖了人类视力所能感知的所有色彩。深度图是将结构光等深度传感器采集到的场景中各点距离，作为像素值形成的灰度图像，已成为计算机视觉领域最重要的图像表示方式之一。基于结构光技术的深度传感器通常还包括红外发射器和红外相机，红外发射器通过发射结构化的红外光线，而红外相机则用于接收和解析经过物体反射的红外光线，从而计算出物体的深度信息。RGB 彩色图像和深度图像的像素点之间需要保证对应关系，

因此, RGB-D 相机通过图像处理单元将彩色图像和深度信息融合在一起, 形成一个完整的 RGB-D 图像。图像处理单元还可以对图像进行校准、对齐和后续处理, 以提高数据质量和可用性^[38]。

RGB-D 相机促使计算机视觉领域研究者开始探索基于 RGB-D 图像的新颖解决方案。深度信息可以解决或简化许多具有挑战性的任务, 例如目标检测、姿态估计和语义分割等。

2.2 图像特征提取算法

图像特征提取是视觉 SLAM 技术中的关键步骤之一, 同时, 特征提取的数量和质量也会对系统产生最直接的影响。视觉 SLAM 初步的相机位姿通常需要使用相邻图像帧的特征对应关系来确定。首先利用相机获取环境的图像信息, 然后通过一系列图像特征提取算法获取具有代表性的特征和描述子, 目前在视觉 SLAM 中主要使用的是点特征、线特征和面特征。

2.2.1 点线特征提取

(1) 点特征提取算法

尽管现在线特征和面特征在针对弱纹理环境的视觉 SLAM 方法中被大量使用, 但点特征几乎从未被摒弃。FAST 是经典的特征点检测算法之一^[39], 但并不涉及特征点的特征描述。尺度不变特征转换 (Scale Invariant Feature Transform, SIFT) 算法^[40]通过在不同尺度空间和方向上检测图像中的局部极值点, 可以快速提取具有旋转不变性和尺度不变性的特征描述子, 但是不具备旋转不变性。加速稳健特征 (Speeded Up Robust Features, SURF) 算法^[41]是对 SIFT 的改进, 利用积分图像等技术加速了特征点的检测和描述子的计算, 具有更高的计算效率和速度。SURF 相较于 SIFT 算法运行速度提高了 3~7 倍, 但同时降低了鲁棒性。FAST 和 SURF 两种特征提取算法在视觉 SLAM 中均得到了广泛应用, 适用于纹理丰富的复杂环境。Rublee 提出的 ORB (Oriented FAST and Rotated BRIEF) 算法^[42]结合了 FAST 特征检测和 BRIEF 描述子构建。BRIEF 是具有旋转不变性的二值化特征描述子, ORB 算法所创建的是二元特征向量, 且会跟随特定关键点及其周围像素区域的变化而变化, 该向量表示关键点周围的强度模式。ORB 的特点

是速度超快，且计算量更小，在实时 SLAM 系统中得到了广泛应用。FAST、SIFT 和 SURF 特征提取效果如图 2-4 所示。



(a) FAST 方法

(b) SIFT 方法

(d) SURF 方法

图 2-4 点特征提取效果

(2) 线特征提取算法

线特征提取算法用于检测图像中的线状结构，例如边缘、直线、曲线等，并提取出其位置、方向和形状等信息。目前在视觉 SLAM 中使用较多的主要有三种方法：Hough 变换^[43]、以 LSD 算法为代表的局部梯度统计以及 EDLines 算法^[44]所代表的边缘拟合方法。

传统直线段检测算法利用 Canny 边缘检测器和 Hough 变换来识别图像中的直线，Hough 变换基于图像空间和参数空间点与线的对偶性原理，将图像空间的直线变换为参数空间中的点，通过统计特性来解决检测问题，缺点是耗时较长。LSD 算法则可以在较短时间里获得较高精度的直线检测结果，其核心思路是将与梯度相近的像素点进行合并。首先对图像进行降采样，获得原图 64% 的像素点。用这些像素点进行梯度计算和排序，梯度幅值越大，表明该像素点越是显著的边缘点。通过筛选，由符合条件的像素点构成的线段支持域。最后将线段支持域矩形化，得到一个较为规整的矩形区域作为线段提取结果展示^[45]。LSD 算法的优势在于其快速的检测速度和无需手动调节参数的特点，并且利用错误控制的技术来提高直线检测的准确性。EDLines 算法在保证实时性的同时抑制了误检量，对比三种算法在提取质量和速度上的差异，该算法在保证较高准确性的同时，其速度比 LSD 算法快一个数量级，但 EDLines 算法提取线段的连续性较差。LSD 与 EDLines 算法的线特征提取效果如图 2-5 所示。

视觉 SLAM 中的图像线特征提取算法需要具备良好的稳定性、鲁棒性和计算效率，以适应不同场景下的定位和地图构建任务。选择合适的特征提取算法是视觉 SLAM 系统设计中的重要一环，对系统的性能和实时性也有着重要影响。

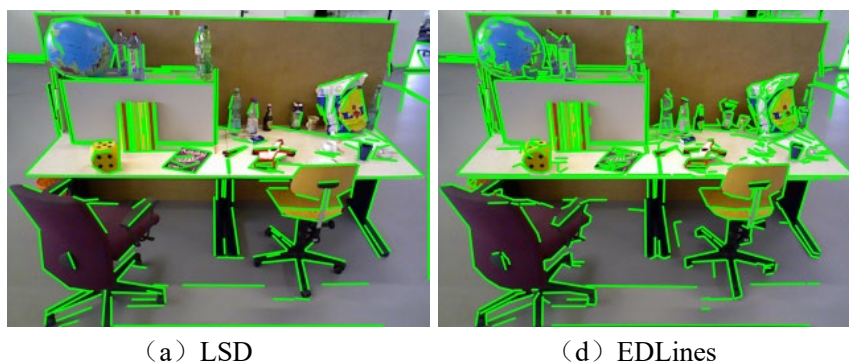


图 2-5 线特征提取效果

2.2.2 平面分割方案

平面在室内环境中无处不在，例如墙壁、地板、路面等表面，同时也不乏由平面组成的物体，如桌子、椅子和屏幕。因此，平面的提取与室内场景下的机器人应用高度相关。目前平面分割方法主要分为四大类：模型拟合、区域增长、特征聚类以及基于深度学习的方法。

(1) 模型拟合方法

随机抽样一致（Random Sample Consensus, RANSAC）^[46]是一种以假设性和随机性为核心思想的迭代性算法，目前该算法多用于样本数据剔除外点。假设性指的是将选取出的样本数据全都假设为正确数据，然后通过假定出的正确数据构建满足问题的模型来验证其他点，最后对每次迭代的结果得出评分。随机性则是基于大数定律的原理，意味着需要在数据中根据正确的概率去随机选择样本，这种抽样方式可以接近正确结果。RANSAC 除了可以剔除异常数据，还是早期主流的直线拟合与点云平面分割方法。

RANSAC 首先对点云进行初始化，随机选择点云集中的小部分样本作为初始内点集合，并利用最小二乘法从这些内点拟合出一个平面模型。然后计算剩余数据点到拟合平面模型的距离，并将距离小于阈值的点视为内点，这一步是 RANSAC 算法的关键。最后计算拟合模型与内点之间的一致性度量，包括内点数目或内点的平均距离。如果一致性度量超过预设最小阈值，则认为当前模型是一个较好的模型。多次重复执行上述步骤，直到满足最大迭代次数，或是已经找到了满足一定条件的模型。满足停止条件后，选择具有最多内点的模型作为最终的平面模型，并利用所有内点重新拟合该模型，从而实现点云平面分割。RANSAC

算法需要进行大量的随机抽样和迭代过程,因此随着数据规模的增加,算法的执行时间也会相应增加,导致效率较低。此外,RANSAC 未充分利用图像二维结构特点,在由 RGB-D 相机生成的有序点云中分割效率低下,复杂的室内环境下容易过度分割生成伪平面,因此更适合缺乏拓扑关系的无序点云平面分割。模型拟合方法分割效果如图 2-7 (b) 所示。

(2) 区域生长方法

区域生长 (Region Growing) ^[47] 是一种基于相似性度量的区域划分方法,是根据提前规定的基准将像素或小区域归类划分成更大区域的过程。基本思想是从一组生长点开始,根据像素或区域之间的相似性(如灰度值、纹理、颜色等)来判断是否将一个像素或区域合并到当前生长点中,从而确保最终分割结果中的区域具有相似的特征。区域生长分割平面时,是从种子点开始,逐步合并与当前区域相邻且满足相似性条件的像素,直到不能再合并为止。合并的条件通常是相邻像素与当前区域内像素之间的相似性度量来确定。当区域生长结束后,对生成的区域进行平面拟合,以获得最佳拟合的平面参数。平面拟合可以采用最小二乘或主成分分析等方法最小化区域内像素与拟合平面之间的误差。区域生长方法平面分割效果如图 2-7 (c) 所示。

区域增长算法适用于大规模和复杂场景下的点云处理,具有高效性和计算成本低的优势,其性能高度依赖于增长策略的制定和种子点的选择,如果选择的种子点不合适或不具有代表性,可能导致平面过分割或者欠分割的情况。区域生长方法通常需要主动设置相似性度量的阈值和生长停止条件等参数,这些参数的选择对算法的效果具有较大的影响,且对不同的图像场景可能需要不同的参数设置,因此需要不断进行调整和优化,不具有普适性。

(3) 聚集层次聚类方法

聚集层次聚类方法 (Agglomerative Hierarchical Clustering, AHC) 是一种根据相似度对数据点进行分组的方法,能实时可靠地从深度传感器获取的有序点云中提取出多个平面。其基本思想是根据相似性度量,迭代合并具有相似性的两类数据,直到达到停止条件为止。主要流程如图 2-6 所示。

首先,将图像中的点云划分为大小均匀的非重叠区域。在聚类之前,需移除几类节点和它们对应的边:均方误差很高的节点、深度不连续的节点以及两个平

面边界上的节点，此外，为了减少由于图像噪声造成信息感知误差的影响，还需要将缺失数据的节点去除。值得注意的是，初始节点在图像空间中应尽可能接近正方形，可以避免初始节点上生成错误的平面法线方向。

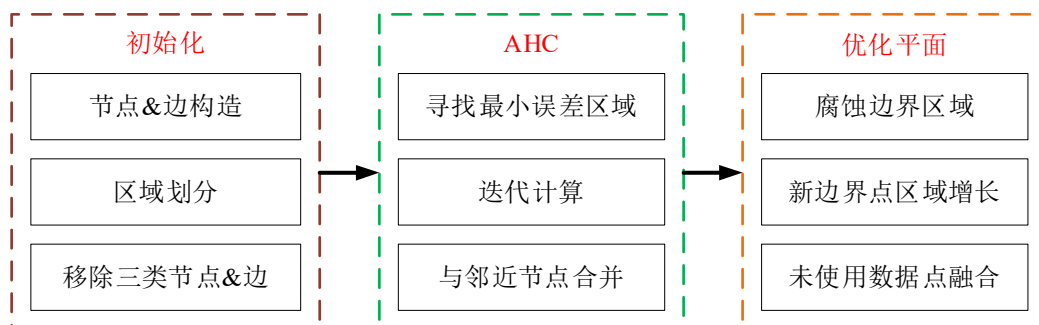


图 2-6 AHC 方法平面分割主要流程

然后，在构造的区域上执行层次聚类。迭代计算平面拟合均方误差，找到具有最小均方误差（Mini Mean Squared Error, MMSE）的区域。如果 MMSE 超过一个预先设定的阈值，则产生一个平面的分割区域。最后，优化因为抛弃较多节点而造成的锯齿状分割。层次聚类只是粗糙地提取了平面，在平面边缘处会产生锯齿状的分割，使得少量异常点包含在其中，从而影响平面分割的效果。可以通过腐蚀边界区域优化平面。最后再用区域增长的方法对新边界点区域进行扩张，将未使用的数据点融合至已提取的最近平面中。聚集层次聚类方法分割效果如图 2-7（d）所示。

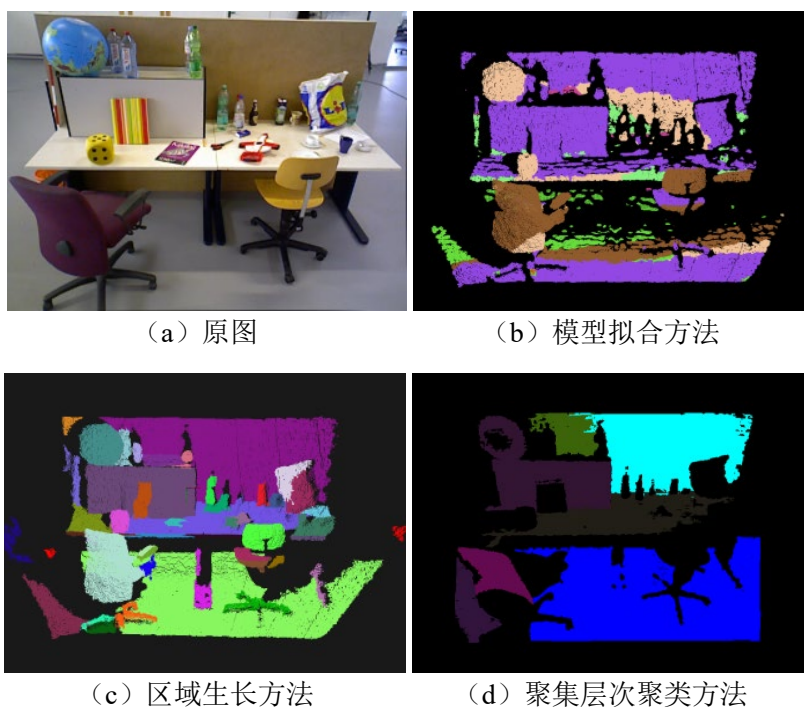


图 2-7 传统方法平面分割效果

(4) 基于深度学习的方法

随着深度学习的兴起,卷积神经网络(Convolutional Neural Networks, CNN)在计算机视觉领域的地位越来越高,已经被广泛应用于特征提取与图像分割。PlaneNet^[48]是一种用于室内场景平面分割和估计的深度学习模型,由华盛顿大学的研究团队提出,旨在解决室内场景理解中的平面分割问题。该模型通过从单个 RGB 图像中学习场景的平面几何结构分段重建平面深度图。尽管在平面分割方面具有重要意义,但在存在噪声的场景下,平面提取的效果并不理想。此外,该方法的场景训练数据跨区域泛化能力也欠佳。由 Mask R-CNN 改进而来的卷积神经网络 PlaneRCNN^[49]则是针对单幅 RGB 图像进行三维平面检测与重建,使用新的损失函数联合细化分割掩码。PlaneRCNN 在小平面分割方面具有较好的表现,但是同样会误检测到小区域假性平面。

与 PlaneRCNN 方法不同,PlaneRecNet^[50]从增强多任务卷积神经网络的跨任务一致性开始,集成了单级实例分割网络用于分段平面分割和深度解码器,从单个 RGB 图像重建场景,引入了一系列基于几何约束的损失函数,共同提高分段平面分割精度。同时,该网络利用平面实例的感知,提出了一种新的平面优先注意力模块来指导深度估计。PlaneRecNet 网络在平面检测、分割和重建度量方面的表现较优异,但是,在纹理稀疏且相似的环境中可能会因信息丢失造成的平面漏检测和错分割问题。

2.3 结构特征与曼哈顿世界假说

在移动机器人领域,曼哈顿世界假说被用来简化定位和导航问题,因为在曼哈顿世界中,机器人可以更容易地识别和跟踪结构特征,从而更准确地估计自身的位姿。本节主要介绍结构特征和曼哈顿世界假说的相关概念,以及二者之间的关联,为后续研究多结构特征融合的室内视觉 SLAM 方法提供理论基础。

2.3.1 结构特征

在计算机视觉领域中,结构特征指的是场景中具有几何形状或拓扑结构的线条、平面和一些具有固定结构的组合特征。在视觉 SLAM 技术中,结构特征通

常用于描述建筑物、家具、街道等人造环境中物体或场景的轮廓，这些物体通常具有明显的几何结构，例如平行和正交的平面、线条和边缘等。室内移动机器人通过这些结构特征可以完成对物体或场景的识别、跟踪和匹配等任务。随着生活水平的提高和人们审美观念的变化，现代室内装修设计趋向于通过清晰的线条和几何形状作为装饰元素，室内轮廓由此得以突显，从而使空间更具有层次感和立体感，这种明确的轮廓分界使整体布局更加有序。这种布局在造就大量弱纹理环境的同时，也会保留有清晰的边界和规整的平面，因此这种环境必然包含大量结构特征。室内结构化场景示例如图 2-8 所示。

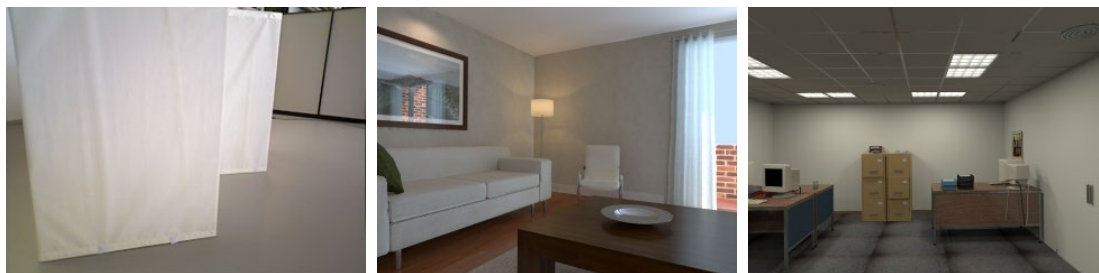


图 2-8 室内结构化场景示例

结构特征是对三维空间中具有特定结构对象的特征描述，同时也是一个满足曼哈顿世界假设的正交基，可用于描述结构物体包含的平行与正交信息，满足曼哈顿世界由三个相互正交的主导方向组成的假设。由这三个相互正交的主导方向组成的结构特征可以为 SLAM 技术提供全局约束。

2.3.2 曼哈顿世界假说模型

曼哈顿世界假说可以追溯到 1999 年 Coughlan 在第七届 IEEE 国际计算机视觉会议发表的一篇论文，在他的论文中首次提出了这个概念。曼哈顿世界假说可以将室内环境描述为假设性的世界模型，用于描述和模拟现实世界的几何结构。这个假设认为在虚构的世界中，物体和场景的几何结构主要由相互垂直和平行的直线或平面构成，并且形成了一个规则的网格结构，使得物体和场景轮廓呈现出类似城市街道交叉的模式，就像美国纽约市曼哈顿的街道网格一样。根据曼哈顿世界假说，室内环境中的物体主要分布在平行于地板和墙壁的表面上，而这些表面通常是垂直或水平的。因此，相机拍摄的图像中大部分线条都应该是水平或垂直的，而不是斜向的。这种假设简化了计算机视觉任务中对物体和场景的理解和分析，使得许多问题可以更容易地解决，例如物体检测、场景重建和三维定位等。

曼哈顿世界假说模型示例如图 2-9 所示。

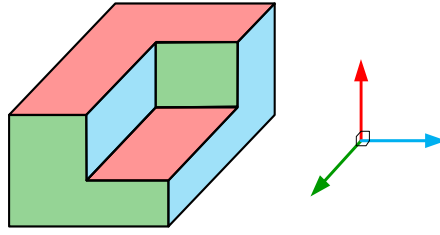


图 2-9 曼哈顿世界假说模型示例

曼哈顿世界假说将场景假设为由平行和垂直三个主方向的线条或平面组成的虚拟世界，因此在室内利用具有几何形状或拓扑结构的结构特征构建该假设无疑最为合适。随着研究的深入，一系列由曼哈顿世界假说演化的结构模型如雨后春笋般涌现，其中最具代表性的是亚特兰大世界（Atlanta World, AW）假说模型^[51]，这些理论模型的出现，极大的提升了计算机视觉领域智能设备对现代场景的理解能力和对结构特征的识别能力。近年来，曼哈顿世界系列假说在机器人和自动驾驶领域得到了广泛的应用和研究，成为了这些领域的重要理论基础之一。

2.3.3 曼哈顿主方向提取

对于早期利用单目相机采集的二维图像，通常采用相互平行的边缘特征检测图像中的消失点，以计算曼哈顿主方向。在曼哈顿场景，对于当前相机视线内提取到的平面集合 $H = \{h_1, h_2, \dots, h_n\}$ ，任何一个平面 h 的法向都可能与主体平面法线方向呈垂直关系或平行关系，因此，曼哈顿主方向的提取相当于从 H 中搜寻尽可能多相互垂直的三个平面法向，提取主要流程如图 2-10 所示。

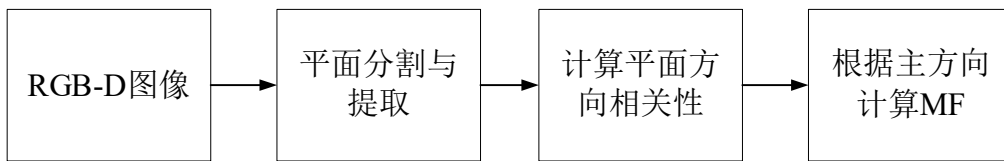


图 2-10 曼哈顿主方向提取主要流程

MW 假说建立在由三个相互正交主方向形成的曼哈顿坐标系上，若用室内平面来表示，则需满足三个平面的法向量分别平行于 MF 的三个主方向。对于服从 MW 假说的场景结构，平面特征法向量 $\mathbf{n}_i (i=3)$ 在虚拟单位球体上是一个有组织的分布，如图 2-11 所示。这种分布可以被假设为一个概率密度函数的样本，法向量 $\mathbf{n}_i$ 更有可能分布在 MF 的主轴上，用于识别曼哈顿主方向。因此，根据表

面法向量 $\mathbf{n}$ 确定曼哈顿主方向可以看作是在这个概率密度分布函数中寻找局部极大值的过程。

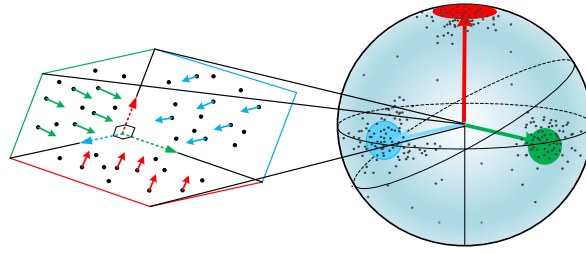


图 2-11 曼哈顿主方向分布

在 SLAM 技术中, 相机两帧之间的位姿转换矩阵可以分解为 MF 下的旋转矩阵和平移矩阵, 曼哈顿主方向被用来确定 MF, 并且通过匹配相邻帧计算位姿, 以此有效利用室内场景的结构特点来辅助视觉 SLAM 完成定位。该方法突显了曼哈顿世界假说在室内视觉 SLAM 技术中的辅助价值。

2.4 本章小结

本章主要介绍多结构特征 SLAM 方法的相关理论基础。首先介绍了当前视觉 SLAM 系统搭载的视觉相机模型。分别介绍了针孔模型和畸变模型, 同时阐述了无畸变和有畸变情况下的相机内参模型。然后针对本文方法, 对当前主流图像点特征、线特征进行了简单介绍, 并且分别对模型拟合、区域生长、层次聚类 and 深度学习四种类型的平面分割方案进行了剖析, 从各自的提取原理和所代表的算法出发进行了优劣分析和取舍说明。最后, 研究了结构特征和曼哈顿系列假说的背景、理论以及二者之间的关联, 并且从视觉 SLAM 系统应用层面探讨了曼哈顿主方向的提取过程及其重要意义, 为本文后续结构特征的构建和曼哈顿主方向的提取做好理论研究基础。

第 3 章 具有结构特征辅助的 RGB-D 视觉里程计

本章节算法包含改进边缘拟合的线特征提取环节、结构特征构建与跟踪环节和 6 自由度位姿估计环节。在线特征提取阶段，为克服因室内光照不均匀造成线特征提取稳健性差的问题，引入方向约束改进 EDLines 像素链连接规则，检测出更多连续线段。在结构特征构建与跟踪环节，通过检查线特征之间的交叉关系，构建更稳定的结点-线结构特征(Junctions Line Structure Feature, JLSF)，并以 JLSF 结点辅助特征点匹配。然后对 JLSF 进行帧到帧跟踪，估计旋转矩阵。在位姿估计环节，匹配确定的关键帧，并剔除冗余关键帧；然后融合点、线 BA 优化，构造最小化重投影误差函数，进一步提高位姿准确性。具有结构特征辅助的 RGB-D 视觉里程计系统整体框架如图 3-1 所示。

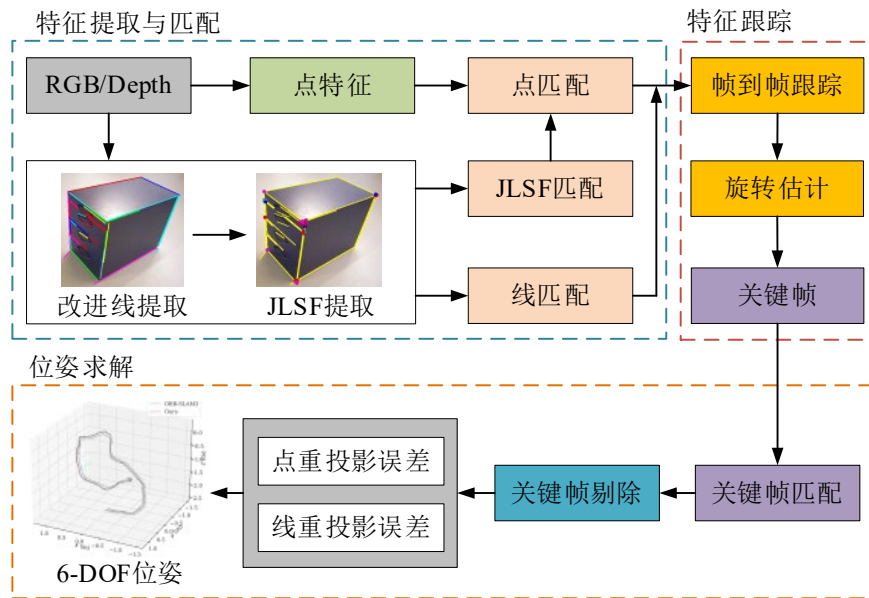


图 3-1 本章算法框架图

3.1 改进边缘拟合的线特征提取

线特征的提取与匹配是机器人定位技术中长期存在的问题，当前主流的线特征提取方式分为三类：全局霍夫变换、以 LSD 算法为代表的局部梯度统计以及 EDLines 算法所代表的边缘拟合方法。近几年的研究以速度较快的边缘拟合方法为主，其速度快的特点更适合应用于实时性要求较高的定位系统。LSD 算法虽然能得出亚像素级别的线特征检测结果，但非常耗时；EDLines 算法速度更快，但

对图像中的噪声或弱边缘非常敏感,检测过程中线特征容易断裂或产生不准确的连接。室内结构化场景的建筑结构边缘明显,且线特征大多分布在水平与垂直方向,即水平与垂直方向上的像素强度差异较大。如果按照传统全局搜索方法直接使用所有锚点拟合直线,不仅降低了效率,而且检测结果可能会受到弱边缘的干扰。基于此特点,本章采用一种改进边缘拟合的线特征提取方法,通过水平与垂直方向约束为 EDLines 算法前期边缘提取提供潜在的方向。

首先,通过中值滤波进行灰度图像平滑,在暗光环境中抑制噪声的同时有效保护边缘信息。然后,确定水平与垂直方向的锚点。水平线上下像素间的强度差异较大,而垂直线左右像素之间的强度差异较大。根据梯度方向的不同,找到属于水平和垂直方向的像素集合。像素梯度方向计算如式(3-1)所示:

$$\theta_o = \arctan \left[\frac{|I(x, y+1) - I(x, y-1)|}{|I(x+1, y) - I(x-1, y)|} \right] \quad (3-1)$$

式中 θ_o 代表梯度方向, x 和 y 分别表示像素的横纵坐标, $I(x+1, y)$ 表示 $(x+1, y)$ 位置预处理得到的像素强度, $|I(x+1, y) - I(x-1, y)|$ 表示 y 轴方向的梯度值。梯度值不为零且 $\theta_o \in (0, \pi/6)$ 的所有像素表示为水平方向像素集合 G_h ; 梯度值不为零且 $\theta_o \in (\pi/3, \pi/2)$ 的所有像素表示为垂直方向像素集合 G_v 。定义其中梯度值大于最小阈值的像素为锚点,锚点是与周围像素梯度值变化剧烈的像素,更有可能组成物体边界。集合 G_h 和 G_v 越大,表明属于水平与垂直方向的锚点越多。对于梯度方向 $\theta_o \in [\pi/6, \pi/3]$ 的像素,则按照传统方式进行锚点提取。

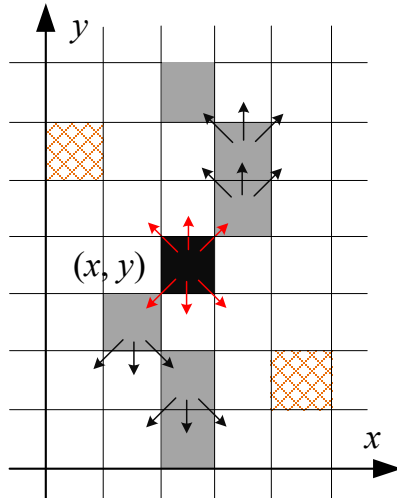


图 3-2 垂直方向的锚点连接

最后，通过锚点连接像素链，并拟合出线段。在集合 G_l 和 G_v 中选择局部最大梯度值的像素作为初始锚点。以垂直方向为例，从集合 G_v 的初始锚点开始按梯度值大小进行连接， (x, y) 为 G_v 中的初始锚点（黑色），从 (x, y) 开始在垂直方向上依次链接其他锚点（灰色），如图 3-2 所示。在得到一系列像素链集合 $C_v \subseteq G_v$ 之后，以像素链为对象依次经过像素格，用最小二乘法线拟合出线段，直到误差超过特定阈值。用算法迭代处理像素链上的剩余像素直到所有的像素都拟合完成，输出提取线段，水平方向也是如此。

3.2 基于线相交关系的结构特征提取

3.2.1 结点-线结构特征构建

JLSF 是指两条线及其交点的简单结构，与广泛使用的点特征不同，JLSF 由物理存在的线特征构造，包含了两条线之间的空间关系，JLSF 结构示例如图 3-3 所示。在具有准确线特征的基础上，JLSF 的提取效率会有所提高，但是两条线的交点不能仅考虑二维图像中简单的交叉。因为在投影变换下，二维图像中相交的直线在 3D 空间中并非一定共面，因此也不一定相交，如图 3-4 所示。JLSF 在投影变换下是几何不变的，众所周知，二维图像中两个共面线段在三维空间中也一定共面。因此，提取 JLSF 首先需要寻找组成交点的共面线。RGB-D 相机已具有较高的红外灵敏度，在室内暗光环境依然可以获得良好的深度信息，可以利用深度信息来确定线特征的空间位置，并通过判断在空间中的投影是否近似共线来推断它们是否共面，去除二维图像中非共面线形成的 JLSF。

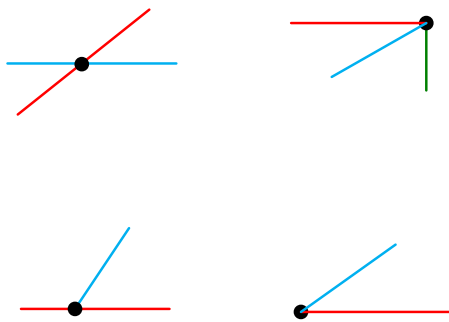


图 3-3 四种 JLSF 结构

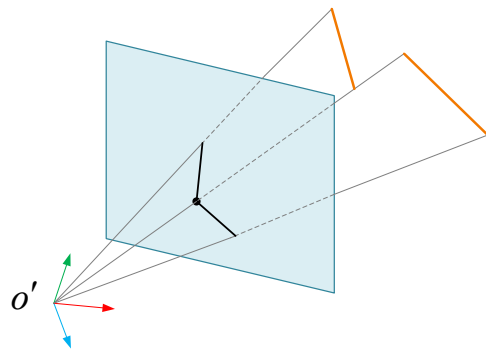


图 3-4 错误相交

3.2.2 结点-线结构特征提取与匹配

在构造 JLSF 后, 借鉴 SIFT 关键点的方法, 在结点周围选取不同与 SIFT 方法的邻域来构建梯度方向直方图, 利用交点周围区域的梯度方向直方图生成描述符。以交点为中心的三个局部区域同心圆, 两条相交线将同心圆分为两个部分, 每部分由小圆扇区、小环形区域和外围大环形组成, 如图 3-5 所示。将外围大环形区域均匀划分为两个子区域, 保持每个子区域的面积和对应内环形和内扇区的面积相等。因此, 三个同心圆总共被划分为 16 个区域, 即两组 8 个相等大小的区域, 从每个子区域和扇区计算 8 个方向的梯度直方图, 产生 128 维向量来表征描述子, 提取的 JLSF 如图 3-6 所示。

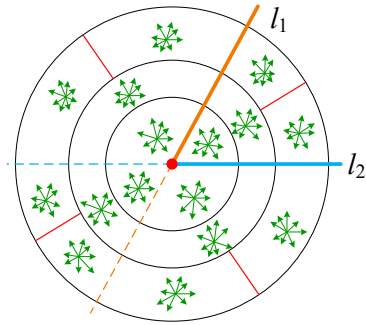


图 3-5 描述符示例

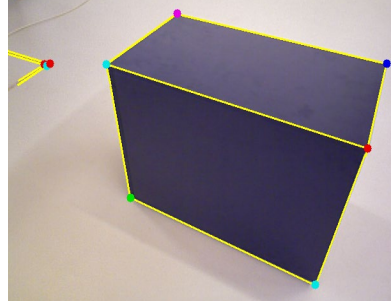


图 3-6 JLSF 提取示例

根据 128 维向量欧氏距离实现两组 JLSF 的匹配, 欧氏距离越小, 则相似度越高, 当欧氏距离小于设定的阈值时, 可以判定为匹配成功。JLSF 依赖于线特征, 因此其结点比 ORB 特征点更加鲁棒。在纹理稀疏和光照不均匀场景下拥有更高匹配率的结点作为辅助, 可以提供充足匹配点特征 p , 进行后续重投影误差计算, 进而缓解位姿估计过程中定位失准甚至跟踪丢失的问题。

3.3 相机位姿估计

3.3.1 结点-线观测模型

在 RGB-D 视觉里程计中, 由于尺度信息已确定, 世界坐标系到相机坐标系的相对姿势存在 6 个自由度: 包括 3-DOF 旋转 $R_{cw} \in SO(3)$ 和 3-DOF 平移 $t_{cw} \in \mathbb{R}^3$, $SO(3)$ 为旋转矩阵的李群, 用于相机位姿优化。

首先, 对 JLSF 结构进行参数化表示。用三元组表示 JLSF, 设 l_1 、 l_2 为 JLSF

两条相交直线的单位向量，以两条线的交点 P 为原点构建一个坐标系 J ， x 轴沿着 l_1 方向， l_1 和 l_2 所在的两条直线确定坐标系的 x 和 y 轴平面。世界坐标系中 JLSF 结构的参数化可表示为：

$$\begin{cases} F_w = (\theta, P_w, R_{wj}) \\ \theta = \arccos(\frac{l_1^T l_2}{\|l_1\| \|l_2\|}) \end{cases} \quad (3-2)$$

其中， $\theta \in \mathbb{R}^1$ 为 l_1 和 l_2 之间的角度， $P_w \in \mathbb{R}^3$ 是交点， $R_{wj} \in \text{SO}(3)$ 表示 J 坐标系相对于世界坐标系的旋转矩阵，由 JLSF 两个单位向量构造。同时，空间单位方向矢量在球坐标系上参数化时有两个自由度，因此空间中的 JLSF 有包括角度在内的七个自由度；当 $l_1^T l_2 = 0$ ，即两条线垂直时，JLSF 只有六个自由度。正交 JLSF 对应坐标系 J 也必定正交，相对于非正交 JLSF 更能为相机旋转估计提供方向约束。室内结构化场景包含大量的正交结构，设由众多正交 JLSF 共同的三个相互正交主方向构建的虚拟坐标系为 J_v ，对正交的 JLSF 进行统一的 l_1 和 l_2 向量积计算，如式（3-3）所示：

$$R_{wj_v} = \begin{bmatrix} j_x & j_y & j_z \end{bmatrix} = \begin{bmatrix} l_1 & l_2 & l_1 \times l_2 \end{bmatrix} \quad (3-3)$$

式中 j_* 代表坐标系 J_v 对应三个轴的方向向量。 R_{wj_v} 代表正交 JLSF 对应坐标系 J_v 与世界坐标系之间的旋转矩阵。

然后，根据前后帧的相机坐标系与 J_v 坐标系的旋转计算得出相机之间的旋转矩阵。设 $R_{c_i w}$ 为第 i 帧相机与世界坐标系之间的旋转矩阵，初始值为单位矩阵，将坐标系 J_v 各轴相对于相机坐标系进行转换得到 $R_{c_i j_v}$ ，如式（3-4）所示：

$$R_{c_i j_v} = R_{c_i w} R_{wj_v} \quad (3-4)$$

随着相机运动过程的帧到帧跟踪， J_v 坐标系及其各轴方向向量不断迭代更新。更新后的第 $i+1$ 帧 J_v 坐标系到相机坐标系旋转矩阵为 $R_{c_{i+1} j_v}$ ，则下一帧相机与世界坐标系之间的旋转矩阵为 $R_{c_{i+1} w}$ ，如式（3-5）所示：

$$\begin{cases} R_{c_{i+1} j_v} = \begin{bmatrix} \hat{j}_x & \hat{j}_y & \hat{j}_z \end{bmatrix} \\ R_{c_{i+1} w} = R_{c_{i+1} j_v} R_{c_i j_v}^T R_{c_i w} \end{cases} \quad (3-5)$$

坐标系 J 与相机坐标系之间的关系，如图 3-7 所示。

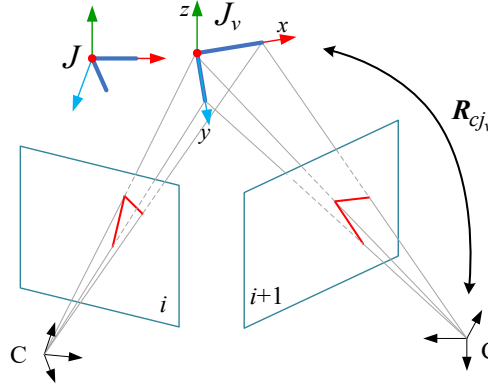


图 3-7 JLSF 结构与相机之间的关系

最后，在旋转估计完成后，利用点线特征构建重投影误差求解 3-DOF 平移运动。平移运动估计的本质是将上帧三维点线投影到当前帧。以 JLSF 匹配的结点作为辅助，产生充足的匹配点特征 p ，归入 ORB 点进行统一优化，将上一帧的 3D 点重新投影到当前帧 i 的重投影误差函数定义为：

$$e_{p+j}^i = p^i - \Pi(R_{cw}P_w^i + t_{cw}) \quad (3-6)$$

式中 $\Pi(*)$ 代表投影函数， t_{cw} 是平移矩阵。 p^i 代表匹配的二维特征点和结点， P_w^i 是二维点特征与结点相对应世界坐标系中的 3D 地图点。

3.3.2 帧间位姿计算

图像平面中的线段可以由两个端点 $(u_{start}, v_{start}, 1)^T$ 和 $(u_{end}, v_{end}, 1)^T$ 表示。在 3D 空间中，用 6-DOF 普吕克坐标表示法进行坐标变换和投影，普吕克坐标中的 3D 线 $L = (n, d)^T \in \mathbb{R}^6$ ，其中 n 为直线与坐标原点确定的平面法向量， $d \in \mathbb{R}^3$ 是直线的方向向量。在上一帧相机坐标系中表示一条直线后，可以将其投影到当前帧相机图像平面，在图像平面中建立 3D 线和 2D 线段之间的关系：

$$l = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = K n_c = \begin{bmatrix} f_y & 0 & 0 \\ 0 & f_x & 0 \\ -f_y c_x & -f_x c_y & f_x f_y \end{bmatrix} n_c \quad (3-7)$$

式中 K 是投影矩阵，由相机标定参数决定，根据式 (3-7) 可知，3D 线投影的线段坐标仅与普吕克坐标的法向量 n_c 有关。2D 线重投影误差可以定义为当前帧线段的两个端点到投影线的距离：

$$e_l^i = \left[\frac{u_{start}l_1 + v_{start}l_2 + l_3}{\sqrt{l_1^2 + l_2^2}}, \frac{u_{end}l_1 + v_{end}l_2 + l_3}{\sqrt{l_1^2 + l_2^2}} \right]^T \quad (3-8)$$

式中 e_l^i 表示 3D 地图线在第 i 帧的投影误差。将得到的点和线结合起来，构造最小化重投影误差函数：

$$t_{cw} = \arg \min \sum_{i \in K} \left[\sum_{(p+j) \in P}^N H_{p+j}(e_{p+j}^i{}^T \varphi_{p+j}^{-1} e_{p+j}^i) + \sum_{l \in L}^M H_l(e_l^{iT} \varphi_l^{-1} e_l^i) \right] \quad (3-9)$$

式中 K 、 P 、 L 分别为局部地图中的关键帧、路标点以及路标线的集合。 N 为局部地图中 ORB 点和 JLSF 结点的匹配数量， M 代表线匹配的数量， φ_{p+j}^{-1} 和 φ_l^{-1} 分别是与检测到的临界点和线端点尺度相关的协方差矩阵， H_{p+j} 与 H_l 是点线 Huber 损失函数。采用 g2o 中的 L-M 算法优化 (3-9) 式，获得相机位姿。

3.4 实验结果与分析

本章实验平台对应硬件配置均使用搭载 Intel Core i5-11400F CPU (6 核 12 线程, 2.6GHz) 和 16GB RAM 的计算机进行, GPU 为 NVIDIA RTX3060 (显存 12GB), 算法运行软件平台为 Ubuntu20.04。

3.4.1 改进线特征提取与匹配实验

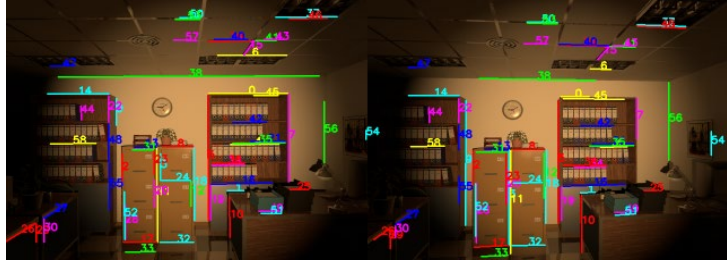
为验证改进后线特征提取方法的效果, 在模拟光照变化的 ICL-NUIM 数据集上比较两种具有开源实现的方法, 提取 syn1_flash 序列中的 RGB 单帧图像进行测试。该序列每张图像分辨率是 640×480 , 包含 880 张办公室照片, 具有结构化场景与光照不均匀的特点, 满足本章实验需要。该序列是在原始 ICL-NUIM 数据集上做了光照变化处理, 恶劣光照条件给特征提取算法带来了很大挑战。

LSD 算法、EDLines 算法以及本章算法的单帧图像线特征提取效果对比, 如图 3-8 所示。为便于观察, 三种算法均使用相同的线宽和随机彩色线条, 具有相同数字标签的线段为匹配线段。LSD 算法采用缩小比例的方式检测线段, 会丢失一些图像梯度信息; 同时对暗光区域鲁棒性较差, 不能保证图像线特征充分提取和匹配, 如图 3-8 (a) 和图 3-8 (b) 所示。EDLines 算法采用边缘拟合的方式提取线特征, 前期通过高斯滤波抑制图像噪声, 在暗光区域的效果有所进步; 但提

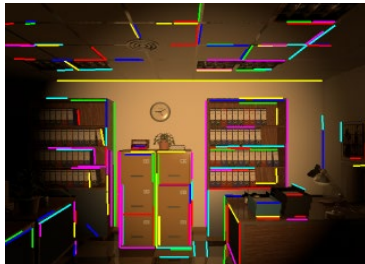
取过程很大程度上依赖于线的方向，容易出现较多短线、断线，如图 3-8（c）所示。本算法在像素链生成过程中采用水平与垂直方向引导策略，保证了室内结构线提取的完整性，实现了更好的匹配效果，如图 3-8（e）和图 3-8（f）所示。



（a）LSD 线提取结果



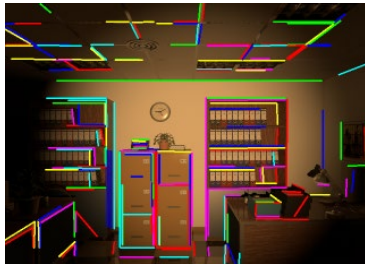
（b）LSD 线特征匹配效果



（c）EDLines 线提取结果



（d）EDLines 线特征匹配效果



（e）本章算法线提取结果



（f）本章算法线匹配效果

图 3-8 改进线特征提取与匹配效果

3.4.2 结点-线结构特征提取与匹配实验

为验证改进后线特征提取方法的效果，在模拟光照变化的 ICL-NUIM 数据集上比较两种具有开源实现的方法，提取 syn1_flash 序列中的 RGB 单帧图像进行测试。该序列每张图像分辨率是 640×480 ，包含 880 张办公室照片，具有结构化场景与光照不均匀的特点，满足本章实验需要。该序列是在原始 ICL-NUIM 数据集上做了光照变化处理，恶劣的光照条件给特征提取算法带来了很大挑战。

LSD 算法、EDLines 算法以及本章算法对应的 JLSF 提取与匹配效果对比，如图 3-9 所示。为方便观察，线特征采用统一黄颜色表示，结点采用不同颜色标

出。JLSF 匹配采用绿色线条将结点链接表示匹配关系，用黄色细线表示相机运动偏移方向。由实验可知，LSD 算法的 JLSF 提取并不充分，作为点特征的辅助作用并不突出，LSD 提取与匹配效果分别如图 3-9 (a) 和 3-9 (b) 所示。EDLines 算法采用边缘拟合的方式提取线特征，前期通过高斯滤波抑制图像噪声，在暗光区域的效果较 LSD 有所进步，但也会产生 JLSF 误识别，匹配过程中容易发生误匹配，如图 3-9 (c) 和 3-9 (d) 所示。本章算法采用水平与垂直方向引导策略，提取的线特征较完整，保证了识别结点的准确率如图 3-9 (e) 所示，匹配效果较 LSD 和 EDlines 有很大提升，如图 3-9 (f) 所示。

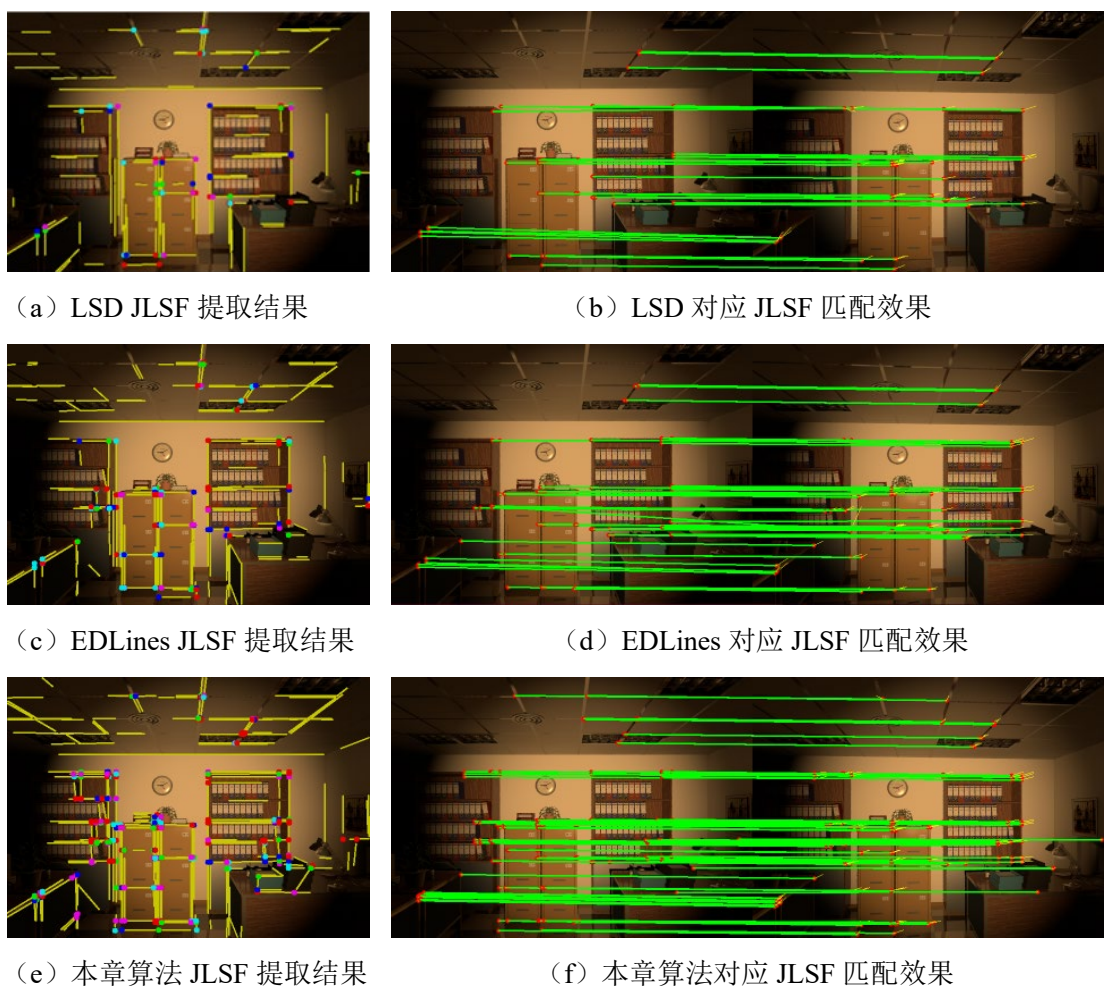


图 3-9 结点-线结构特征提取与匹配

目前对于线段检测，没有足够的评估标准和性能比较基准，因此本实验采用包括运行时间、提取与匹配数量三种标准。三种算法在 syn1_flash 序列前十帧提取时间与提取数量的平均量化结果，如表 3-1 所示；本章算法（图、表中用 Our 表示）和 EDLines 算法检测的线特征数量较 LSD 算法有少量增加，运行时间也

比 LSD 更短。值得注意，EDLines 算法效率较 LSD 提高近十倍。本章算法比 EDLines 耗时更长，但对暗光环境鲁棒性更好，线段连续性更佳，实现了较高的匹配率，为后续准确定位奠定了良好基础。

表 3-1 syn1_flash 序列线特征与 JLSF 提取与匹配测试结果

算法	LSD	EDLines	Our
线段提取数量	106	153	185
线段匹配数量	59	87	128
JLSF 提取数量	62	80	109
JLSF 匹配数量	55	78	102
平均时间/ms	139.5	14.6	18.3

3.4.3 定位精度评估实验

为验证算法对室内暗光环境具有鲁棒性，本章节使用具有模拟照明变化的 ICL-NUIM 合成数据集^[52]。该数据集由苏黎世联邦理工学院研究者根据 ICL-NUIM 数据集^[53]提供的工具和轨迹生成，包含静态照明、变化光照、闪光灯和局部暗光等不同光照的室内场景序列，满足本章实验需求。场景如图 3-10 所示。

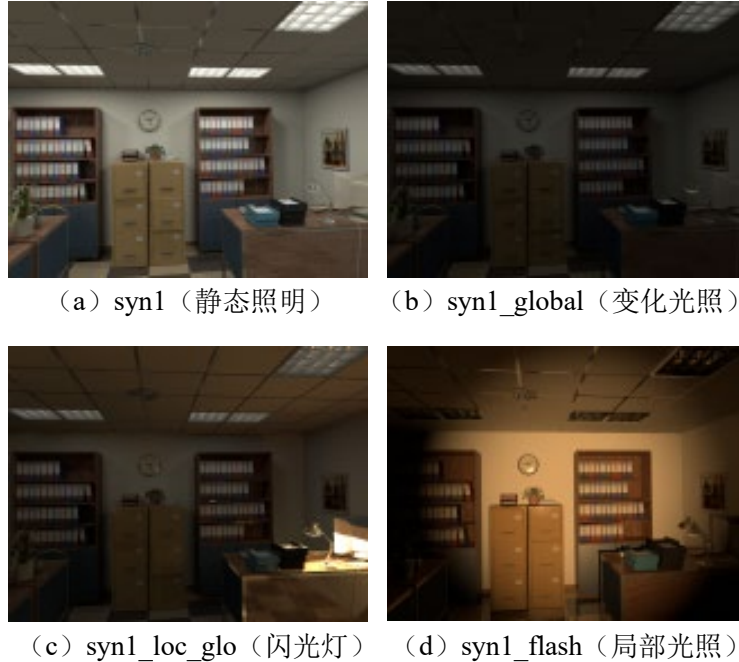


图 3-10 模拟光照变化的 ICL-NUIM 数据集示例

绝对轨迹误差 (Absolute Trajectory Error, ATE) 的均方根误差 (Root Mean Squared Error, RMSE) 是验证 VO 系统的常见指标，表示轨迹真值与系统估计路

径之间的差值。通过 ORB-SLAM3、MSC-VO 以及本章算法对模拟光照变化的 ICL-NUIM 数据集各序列进行实验。其中，为了避免闭环校正对 VO 测试产生影响，实验之前将 ORB-SLAM3 代码的闭环模块禁用。实验得到的 ATE RMSE 如表 3-2 所示，最佳结果以粗体进行突出。syn1 和 syn2 序列分别模拟了同一室内场景的不同光照情况。由图 3-10（d）可知，局部光照 syn1_flash 序列对 VO 系统的挑战性更高，但从表 3-2 可以看出本章算法在 syn1_flash 和 syn2_flash 两个序列上的误差较其他算法 ATE RMSE 降低明显。总体而言，本章算法的 ATE RMSE 相比于 ORB-SLAM3 降低了 11.48%，相对于 MSC-VO 降低了 36.36%，本章算法在暗光条件下的定位精度更高。

表 3-2 ICL-NUIM 数据集测试数据 ATE RMSE 对比（单位：m）

算法	ORB-SLAM3 (VO)	MSC-VO	Our
syn1	0.044	0.056	0.041
syn1_global	0.054	0.062	0.043
syn1_loc_glo	0.041	0.070	0.042
syn1_flash	0.060	0.082	0.051
syn2	0.042	0.038	0.040
syn2_global	0.040	0.047	0.041
syn2_loc_glo	0.054	0.067	0.044
syn2_flash	0.049	0.051	0.042

通过 evo 测评工具进行轨迹可视化，为方便观察，将轨迹投影到 X-Z 轴所在的二维平面，三种算法在模拟光照变化的 ICL-NUIM 数据集 syn1_flash 序列轨迹对比如图 3-11（a）所示，在 syn2_loc_glo 序列轨迹对比如图 3-11（b）所示。图中黑色实线轨迹代表本章算法，灰色短虚线代表该序列真实轨迹。与其他算法相比，本章算法生成的轨迹更接近对应序列真实轨迹。

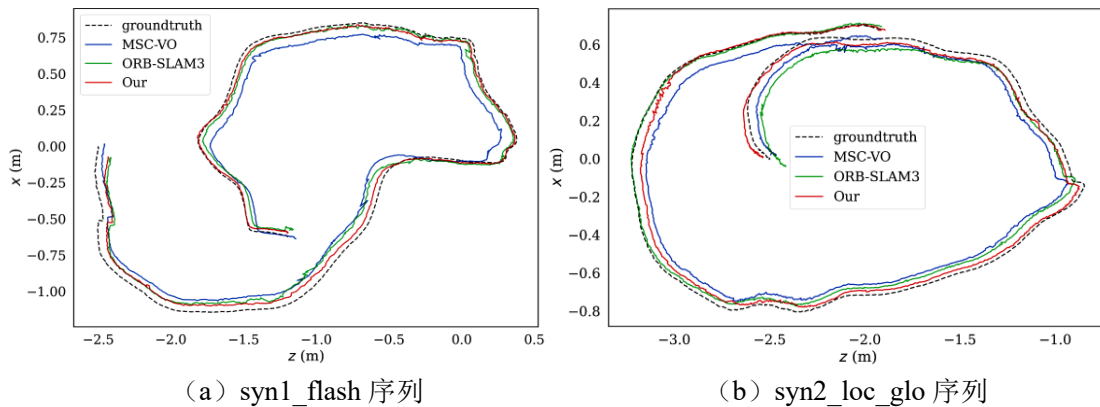


图 3-11 部分序列在 X-Z 轴向投影轨迹对比

TUM RGB-D 数据集^[54]由微软 Kinect RGB-D 相机在真实室内环境中采集的 RGB 和深度图像序列组成,采集频率是 30 帧/秒,分辨率为 640×480,用 100Hz 的高精度动作捕捉系统记录地面真实轨迹,被广泛应用于测试各种 VO 系统定位精度。不同序列具有不同的结构和纹理,为证明所提出方法的有效性,使用 fr1/floor、fr1/room 和 fr2/360_hemisphere 等具有不同纹理和光照强度的序列测试算法,ATE RMSE 具体量化结果如表 3-3 所示。值得一提的是,TUM 数据集中的 fr3/cabinet 和 fr3/large_cabinet 序列是典型纹理稀疏环境,物体表面点特征信息少,采集过程中相机做环绕运动,传统基于点特征的定位算法精度较差甚至跟踪丢失,而本章算法可稳定运行。fr2/360_hemisphere 序列展现的是光照较差的室内场景,相机移动速度快而且光照不均匀,本章算法精度较 ORB-SLAM3 和 MSC-VO 均有所提高。所有实验都运行了 3 次以上并且取均值,本章算法 ATE RMSE 相对于 ORB-SLAM3 降低了 29.83%,相对于 MSC-VO 降低了 31.01%。

表 3-3 TUM RGB-D 数据集测试数据 ATE RMSE 对比 (单位: m)

序列	ORB-SLAM3 (VO)	MSC-VO	Our
fr1/floor	0.102	0.099	0.065
fr1/room	0.118	0.114	0.067
fr2/plant	0.056	0.042	0.031
fr2/360_hemisphere	0.181	0.196	0.097
fr3/cabinet	-	0.111	0.122
fr3/large_cabinet	0.083	0.135	0.076

在 TUM 数据集的 fr1/room、fr2/360_hemisphere 和 fr2/plant 三个序列对 ORB-SLAM3 (图 3-12 中用 ORB3 表示)、MSC-VO 和本章算法 (图、表中用 Our 表示) 进行实验验证,轨迹误差对比如图 3-12 所示。fr1/room 序列是室内书房的场景,线特征相对丰富,但相机运动速度快,本算法通过快速提取线特征获得了更高的定位精度。fr2/360_hemisphere 序列是室内厂房的场景,从第 960 帧开始的连续 50 帧都是严重的暗光,特征退化现象严重,本章特征提取算法在暗光条件下的鲁棒性较强,且通过 JLSF 辅助特征点匹配,缓解了特征退化带来的影响。fr2/plant 序列展现的是由 RGB-D 相机拍摄的绿植,线性纹理较丰富但光线较暗。本章算法通过改进边缘拟合的线特征提取方法,提高了暗光环境的线特征提取鲁棒性,为视觉里程计提供更多连续性线段。由轨迹重合度与量化条可知,本章算法的定位精度相对良好。

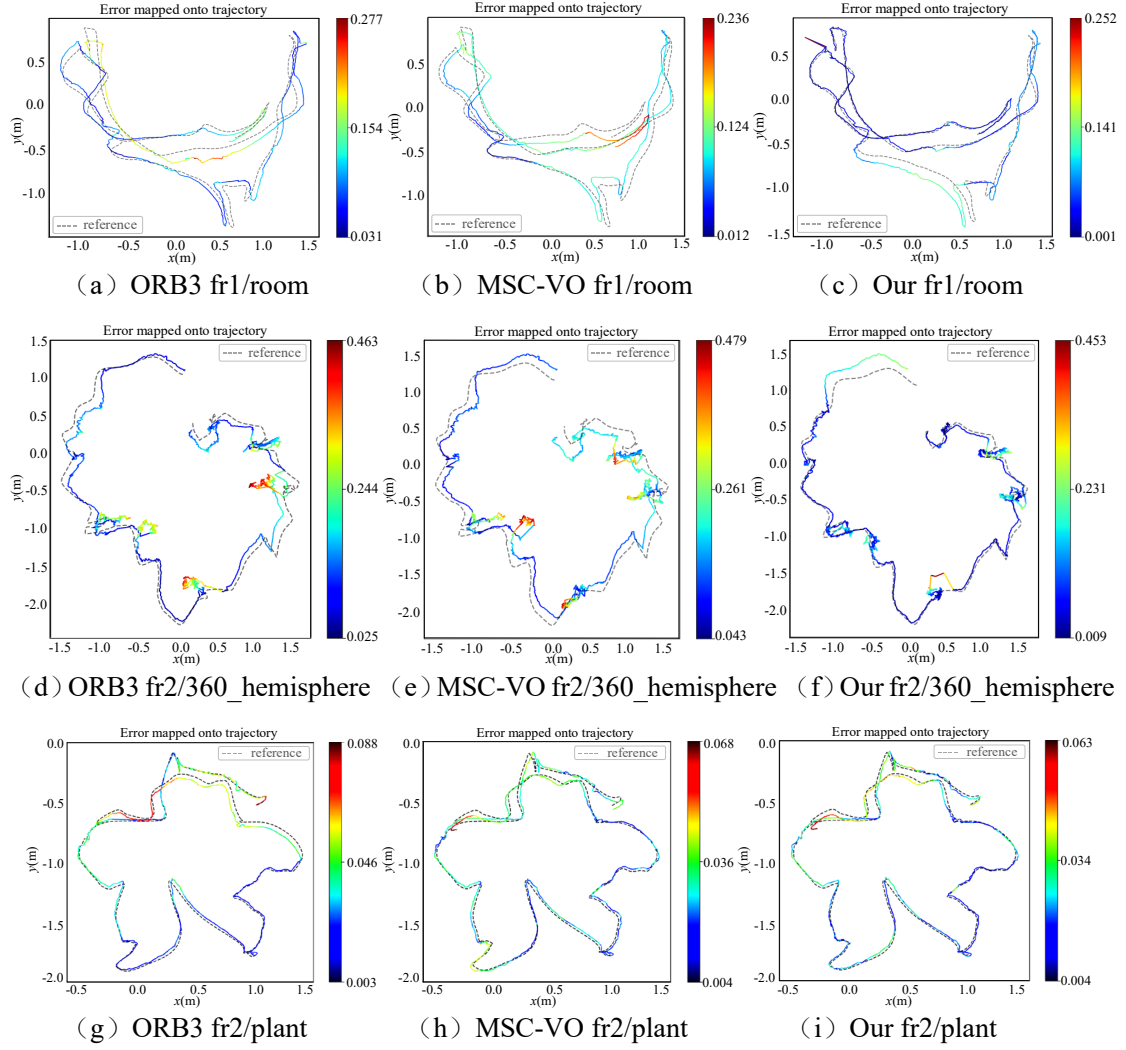


图 3-12 部分序列轨迹对比图

3.5 本章小结

本章节提出了一种具有结构特征辅助的 RGB-D 视觉里程计算法，有效缓解了室内结构化场景下视觉里程计定位精度差的问题。为减少室内纹理稀疏以及光照不均匀对视觉里程计的影响，本算法先引入水平与垂直方向约束的方法改进 EDLines 算法边缘提取过程中的像素连接方式，检测出更多连续线段。然后通过检查线特征的交叉关系，构建更稳定的结点-线结构特征，减少旋转误差，提高定位精度。最后，将匹配结点归入 ORB 点统一优化，构造最小点线重投影误差函数，进一步提高算法性能。用 TUM RGB-D 和模拟光照变化的 ICL-NUIM 数据集进行了对比实验，结果表明本章提出的方法定位精度更高。

第 4 章 基于曼哈顿假设的多特征视觉 SLAM 方法

针对基于点特征的视觉 SLAM 方法在室内结构化场景中位姿估计性能下降的问题，提出基于曼哈顿假设的多特征视觉 SLAM 方法。在前端多特征处理阶段，引入特征增强模块对 PlaneRecNet 网络进行改进，改善平面特征提取算法的分割效果。在基于 MW 假设的位姿估计阶段，利用平面结构的正交关系判断曼哈顿假设条件，在非曼哈顿场景，采用基于点、线结构特征辅助的位姿估计方法，在符合 MW 假设的场景，首先采用方向引导的方法筛选主平面，然后基于 MW 假设原理，比较和跟踪平面结构的正交性估计三自由度旋转矩阵，降低旋转累积误差。本章算法系统框架如图 4-1 所示。

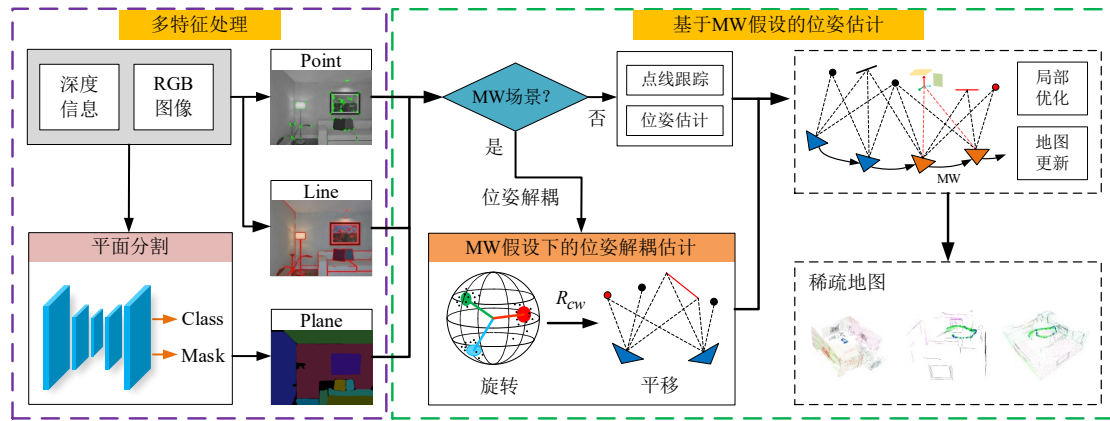


图 4-1 本章算法系统框架图

4.1 基于改进 PlaneRecNet 网络的平面分割

目前平面分割主要有模型拟合、区域增长、特征聚类以及深度学习等四大类方法，前三者传统平面特征算法是从 3D 点云中分割平面，分割效果易受环境噪声的影响，且在镜像表面和非平面中较难检测其几何性质，在视觉 SLAM 系统中效果并不理想。近年来，深度学习算法在计算机视觉技术中的广泛应用，为语义分割和目标检测技术的研究注入了新的活力。PlaneRecNet 是用于平面分割和重建的多任务卷积神经网络，采用具有平面优先注意力的深度解码器和共享主干来进行平面区域的估计和重建，且引入了一些新的损失函数，共同提高分段平面分割的效果。PlaneRecNet 网络因其独特的架构，在室内场景中对平面的检测效果较为出色。

特征金字塔网络（Feature Pyramid Network, FPN）是一种能有效处理多尺度特征的网络结构，在目标检测和语义分割技术中拥有举足轻重的地位，已成为大多数检测器的重要组成部分。在进行特征融合之前，FPN 的不同层级都会经过卷积步骤，将特征通道缩小到最小维度（256）。然而，骨干网络提取的原始特征往往包含更多的语义和特征信息。在进行特征融合过程中，FPN 的横向连接使特征通道维度逐步减少，可能会导致特征信息的流失，在高层特征中，这种问题尤其显著。在面对室内结构化场景中纹理相似的平面时，PlaneRecnet 往往会因为 FPN 信息丢失而产生平面特征漏检测等问题，因此引入特征增强模块（Feature Enhancement Module, FEM），如图 4-2 所示。

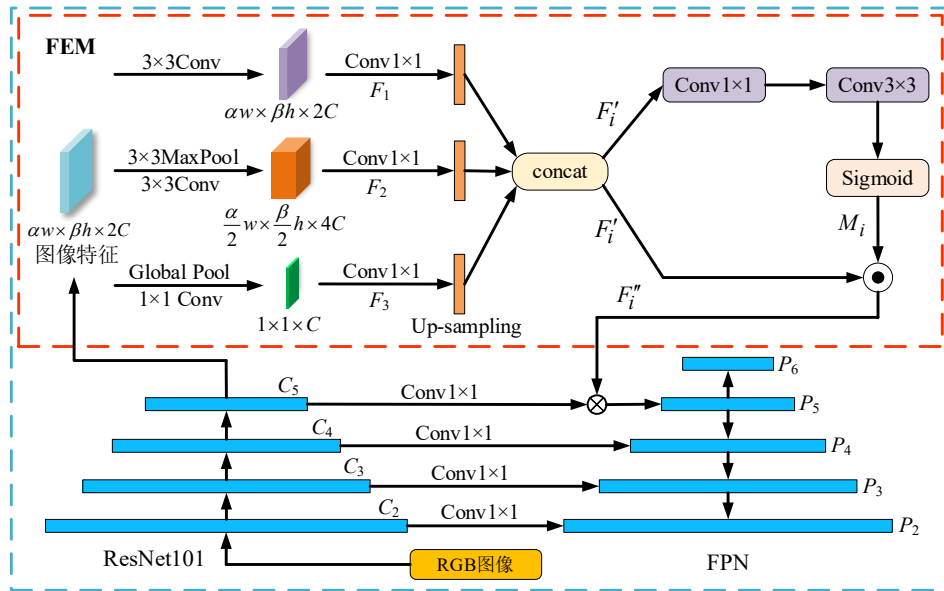


图 4-2 特征增强模块

将大量的局部和全局上下文信息融合在一起，产生更具区分度的特征，是 FEM 模块的核心思想。该模块弥补了 FPN 顶层横向连接的信息丢失，利用具有更大感受野的 C_5 层特征信息，提高检测特征的能力。假设通过 ResNet101 骨干网络对 RGB 图像通过下采样得到的特征图有 $\{C_2, C_3, C_4, C_5\}$ 。首先，FEM 对来自 C_5 层的信息进行池化和不同尺度卷积操作：（1）在 C_5 层应用 3×3 卷积来提取局部信息；（2）通过 3×3 最大池化将特征下采样到 $w \times h$ ，并利用 1×1 卷积层扩展通道维数；（3）为获取通道关联信息，对 C_5 进行全局平均池化。通过扩展三个尺度的特征表示，有效扩大了 C_5 的感受野，并细化了表示能力。然后，利用双线性插值将不同比例的特征 $F_i (1 \leq i \leq 3)$ 上采样到与 C_5 层相同的尺度，再进行 concat

融合获得 F'_i 。最后， F'_i 经过卷积与逐元素求和得到 M_i ， M_i 与 F'_i 融合进行张量对齐得到 F''_i ，再将得到的 F''_i 添加到 FPN。

$$M_i = \sigma(w_2(w_1(F'_i))) \quad (4-1)$$

$$F''_i = M_i \otimes F'_i \quad (4-2)$$

其中， σ 为 sigmoid 函数； w_1 和 w_2 是卷积层参数。

4.2 方向引导的主平面筛选

在满足 MW 假设的场景中，除了法向量集中在三个正交方向上的一系列主平面外，也不乏可能存在平行或垂直关系的非正交面和小平面区域，而且在相机运动过程中，每帧被探测的平面一般位置和面积都略有差异。这会导致 MF 估计的结果出现误差。因此，构建最佳 MF 面临一个关键问题，即如何从众多具有一致法向的平面中选取合适的平面。高斯球是三维欧式空间中表示直线、平面法向量等几何元素的一种简便方法，是一种在相机固有参数已知的情况下，以相机光学中心为球心的虚拟单位球体，曼哈顿场景的平面法向量 $\mathbf{n}$ 在单位高斯球面上的分布呈现一定的规律性，如图 4-3 所示。

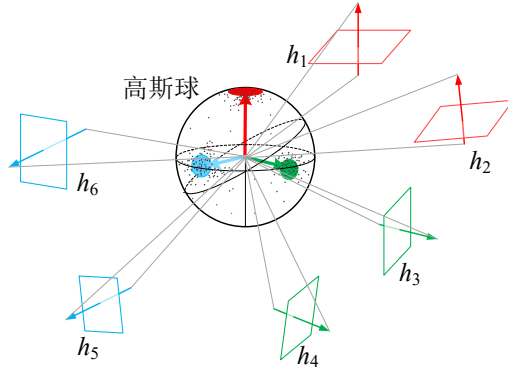


图 4-3 高斯球中的平面法向量分布

对于满足 MW 假设的室内场景，用集合 $H = \{h_1, h_2 \dots h_i\}$ 表示当前帧提取到的平面，其中 h_i 代表第 i 个平面 ($i \geq 2$)。用 θ 表示 H 中任意两个平面 h_j 和 h_k 之间的角度关系，夹角 θ 越接近 90° ， h_j 和 h_k 越接近垂直关系。考虑到满足 MW 假设场景主平面的稳定性，本章算法结合场景中提取平面之间的方向和面积关系，选择确定 MF 三个轴的优势平面。具体地，通过计算所有平面在 H 中的 $F_\perp(h_k)$ 值，比较不同平面的 $F_\perp(h_k)$ ，如公式 (4-3) 所示：

$$F_{\perp}(h_k) = \sum_{j=1, j \neq k}^i S(h_j) \sin(\theta) \quad (4-3)$$

其中, $F_{\perp}(h_k)$ 表示集合 H 中除 h_k 之外, 其它平面在垂直于 h_k 的平面上的投影面积之和, $S(h_j)$ 为平面的面积。 $F_{\perp}(h_k)$ 值越大, H 中与近似垂直于 h_k 的平面数量越多, 反之, $F_{\perp}(h_k)$ 值越小说明平行于 h_k 的平面越多。选取 $F_{\perp}(h_k)$ 值最大的前两个正交平面作为主平面, 第三个方向可由二者向量积求得。

4.3 曼哈顿坐标系跟踪及位姿估计

跟踪曼哈顿坐标系的核心思想是比较平面结构特征的正交性。使用 $\{\mathbf{R}_{cw}, \mathbf{T}_{cw}\}$ 表示六自由度的相机位姿, 其中 $\mathbf{R}_{cw} \in \text{SO}(3)$ 和 $\mathbf{T}_{cw} \in \mathbb{R}^3$ 分别表示从世界坐标系到相机坐标系的旋转和平移。MF 相对于相机坐标系的旋转矩阵表示为 $\mathbf{R}_{cm} = [\mathbf{r}_x \ \mathbf{r}_y \ \mathbf{r}_z] \in \text{SO}(3)$, 其中 $\mathbf{r}_x$ 、 $\mathbf{r}_y$ 、 $\mathbf{r}_z$ 分别表示 MF 的 x 、 y 、 z 轴对应的方向向量。将当前相机坐标系的平面法向量 $\mathbf{n}'$ 变换为在 MF 中的表示:

$$\mathbf{n}' = \mathbf{R}_{cm}^T \mathbf{n} \quad (4-4)$$

相机在满足 MW 假设的室内场景运动, 相当于交替地将当前帧检测到的平面法向量投影到上一帧 MF 的 x 、 y 和 z 轴对应的高斯球切平面上, 用 $\mathbf{n}'_x$ 表示与 MF 的 x 轴平行的平面相关法向量。

首先, 采用黎曼对数映射算法向量 $\mathbf{n}'_x$ 在高斯球切平面上的二维坐标位置 $\mathbf{m}'_x$, 如式 (4-5) 所示。

$$\mathbf{m}'_x = \frac{\text{sign}(\mathbf{n}'_x)}{\sqrt{(\mathbf{n}'_y)^2 + (\mathbf{n}'_z)^2} \sin(\sqrt{(\mathbf{n}'_y)^2 + (\mathbf{n}'_z)^2})} \begin{bmatrix} \mathbf{n}'_y \\ \mathbf{n}'_z \end{bmatrix} \quad (4-5)$$

然后, 使用具有高斯内核的均值偏移算法计算法向量在切平面上的偏移量。均值偏移是一种非参数的密度估计方法, 通过对数据点进行迭代处理来寻找数据点的密度最大值, 然后不断更新每个数据点的位置, 使其朝向密度梯度最大的方向移动, 最终收敛到密度估计的峰值位置。如式 (4-6) 所示:

$$\mathbf{s}'_x = \frac{\sum e^{-c\|\mathbf{m}'_x\|^2} \mathbf{m}'_x}{\sum e^{-c\|\mathbf{m}'_x\|^2}} \quad (4-6)$$

式中, c 代表内核的宽度, 根据法向量在切平面分散的稀疏程度来确定。

最后，用黎曼指数映射进行归一化，将 MF 各轴的偏移量从切平面更新到高斯球面上，得到变换后的法向量 $\mathbf{s}_x$ ，并将 $\mathbf{s}_x$ 相对于相机坐标系进行转换，获得相机坐标系中下一帧的方向向量 $\hat{\mathbf{r}}_x$ ，如式（4-7）所示：

$$\begin{cases} \mathbf{s}_x = \left[\frac{\tan(\|\mathbf{s}'_x\|)}{\|\mathbf{s}'_x\|} \mathbf{s}'_x{}^T \quad 1 \right]^T \\ \hat{\mathbf{r}}_x = \mathbf{R}_{c,m} \mathbf{s}_x \end{cases} \quad (4-7)$$

其中， $[\cdot]$ 代表向量的归一化。

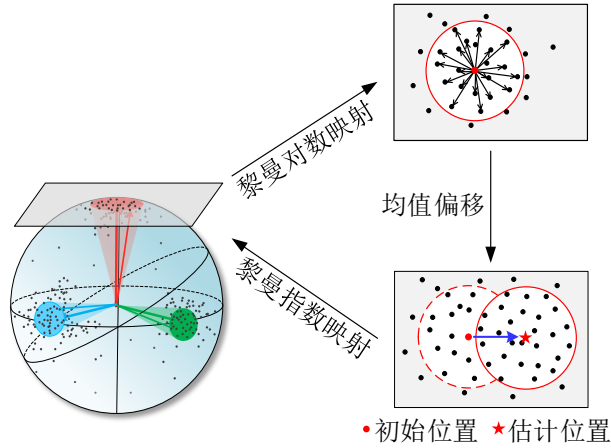


图 4-4 切平面上的均值偏移

在高斯球面上重复迭代计算当前第 i 帧 MF 的 x 、 y 、 z 三个轴的切平面和聚类中心，如图 4-4 所示，当旋转估计收敛，就可以获得 $i+1$ 帧的旋转矩阵 $\mathbf{R}_{c_{i+1}m}$ 。相机之间的旋转矩阵可根据前后两帧的相机坐标系与 MF 旋转计算得出，相机坐标系、世界坐标系和 MF 之间的关系如图 4-5 所示。

$$\mathbf{R}_{c_{i+1}m} = \begin{bmatrix} \hat{\mathbf{r}}_x & \hat{\mathbf{r}}_y & \hat{\mathbf{r}}_z \end{bmatrix} \quad (4-8)$$

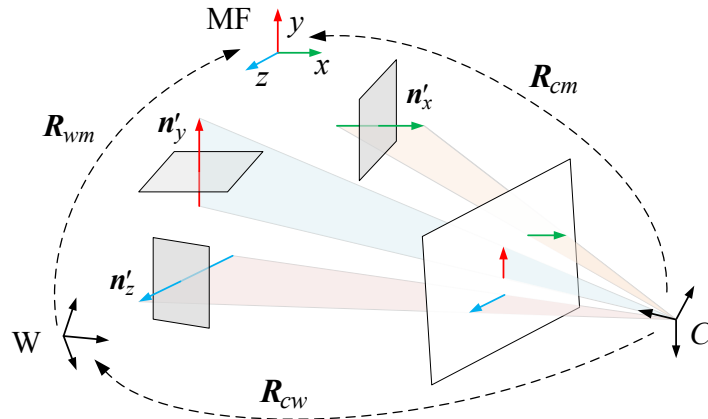


图 4-5 各坐标之间的转换

单独更新 $R_{c_{i+1}m}$ 的 x 、 y 、 z 轴，由于传感器噪声，会旋转矩阵的正交性受到影响，采用 SVD 奇异值分解（Singular Value Decomposition, SVD）来保证 $R_{c_{i+1}m}$ 的正交性，如公式（4-9）所示：

$$\begin{cases} SVD(R_{c_{i+1}m}) = UDV^T \\ \hat{R}_{c_{i+1}m} = UV^T \end{cases} \quad (4-9)$$

将相机坐标系和世界坐标系之间的旋转矩阵设为 R_{c_iw} ，初始值为单位矩阵，则下一帧的旋转矩阵为 $R_{c_{i+1}w}$ ，如式（4-10）所示。

$$R_{c_{i+1}w} = \hat{R}_{c_{i+1}m} R_{c_im}^T R_{c_iw} \quad (4-10)$$

为保证室内结构化场景中的基础特征被 SLAM 系统充分利用，在得到相机与世界坐标系之间的旋转矩阵之后，构建点、线重投影误差函数求解三自由度的平移运动，平移运动估计的本质是将上一帧的三维点、线特征重新投影到当前帧，定义相机测量的重投影误差函数如式（4-11）所示。

$$T_{cw} = \arg \min \left[\sum_{i \in P} \rho_p(\|e_i^p\|^2) + \sum_{j \in L} \rho_l(\|e_j^l\|^2) \right] \quad (4-11)$$

式中， P 、 L 分别表示匹配的所有点、线特征集合； M 和 N 分别是点、线特征数量； ρ_p 和 ρ_l 鲁棒核函数； e_i^p 和 e_j^l 分别为点误差和线误差，如式（4-12）所示。

$$\begin{cases} e_i^p = p_i - \delta(R_{cw} P_i^w + T_{cw}) \\ e_j^l = [l^T \delta(R_{cw} P_{k,start}^w + T_{cw}), l^T \delta(R_{cw} P_{k,end}^w + T_{cw})] \end{cases} \quad (4-12)$$

式中， $P_{i,start}^w$ 和 $P_{i,end}^w$ 分别表示线特征在世界坐标系中 3D 起点和终点； $\delta(\cdot)$ 代表投影函数， P_i^w 表示三维空间世界坐标系中的地图点。最终通过 Levenberg-Marquardt 算法进行优化求解。

4.4 实验结果与分析

本章实验系统平台为 Ubuntu20.04，计算机 CPU 处理器是 Inter Core i5-12400F，6 核 12 线程，主频 3.2GHz；GPU 为 RTX3060 显卡，显存 12GB；机身内存为 16GB。为了测试本章算法的效果，利用公开的合成 ICL-NUIM 数据集与真实世界 TUM RGB-D 数据集对改进的 PlaneRecNet 平面分割网络和定位精度进行测试。

试, ICL-NUIM 包含办公室和客厅场景的 RGB 图像和深度信息, 并且对传感器噪声进行了模拟, 全部八个序列均为室内结构化场景, 为室内场景下的视觉 SLAM 测试提供了重要依据。

4.4.1 平面分割实验

平面分割实验选取 TUM RGB-D 数据集的 fr3/s_t_near 序列场景以及 ICL-NUIM 数据集的 of-kt0 和 of-kt2 序列场景进行测试。对 PlaneRCNN、改进前后的 PlaneRecNet 进行了对比实验, 如图 4-6 所示。PlaneRCNN 算法基于两阶段分割方法和 ROI 内的局部掩模分割, 对小平面的分割能力较为出色, 但也会检测出很多假性小平面, 如图 4-6 (b)、图 4-6 (f) 和图 4-6 (j) 所示。图 4-6 (b) 中的电脑桌上检测出众多小平面和非平面; 图 4-6 (f) 和图 4-6 (j) 中均出现了错误识别的情况, 图 4-6 (f) 中的海报上还被分割出了小平面, 并且箭头所指的门框被识别为平面; 图 4-6 (j) 中的一个完整地面被识别为两个不同的平面。

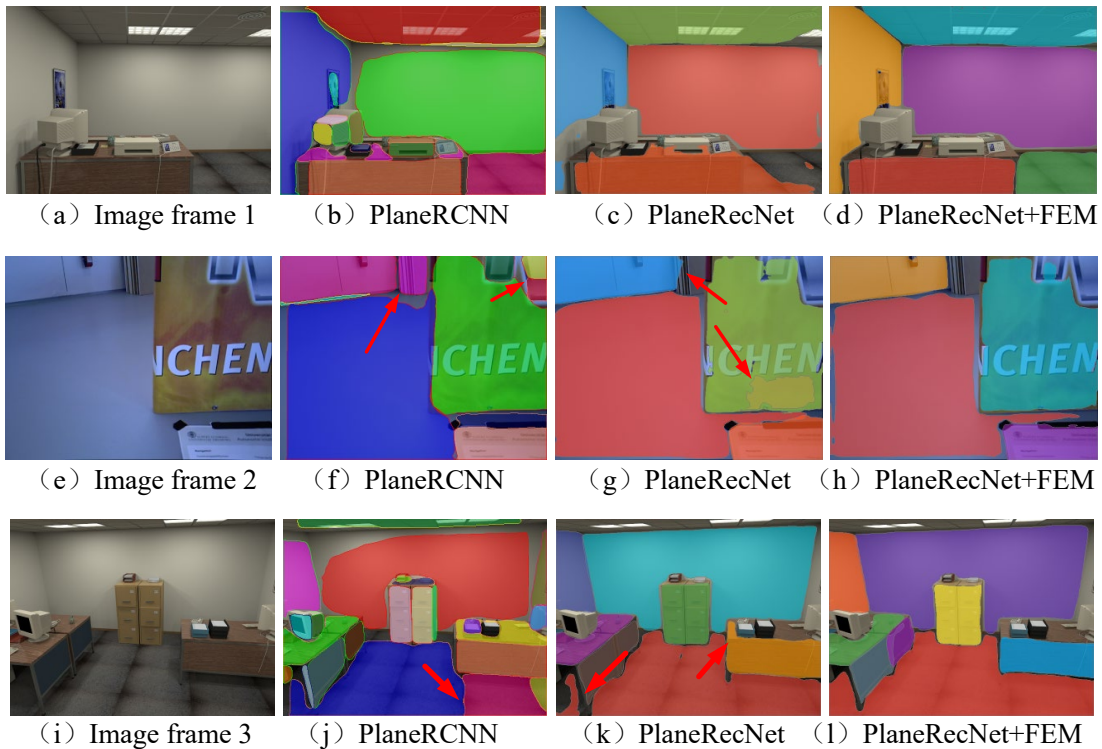


图 4-6 改进前后的平面分割效果对比

大尺寸平面的精确分割能使室内视觉 SLAM 定位精准度得到显著提升。PlaneRecNet 网络在大区域平面分割提取的完整性方面表现较为良好, 但 FPN 最高层级的特征信息丢失会导致平面检测精度降低。TUM RGB-D 的 fr3/s_t_near 序

列场景的测试结果表明, PlaneRecNet 存在检测错误的问题。海报实际上是一个整体平面, 而海报下方出现了额外的不规则平面分割(红色箭头所示区域), 如图 4-6 (g) 所示。PlaneRecnet 在 ICL-NUIM 数据集 of-kt0 和 of-kt2 序列的分割出现了严重错误, 电脑桌的正面和局部地面被辨认为一个整体平面, 但实际地面却未被正确识别, 如图 4-6 (c) 所示; 办公桌的正面和桌面被分割为一个平面, 如图 4-6 (k) 所示。本章引入 FEM 模块对 PlaneRecNet 进行改进, 减少了 FPN 高层图像语义信息过度损失。根据改进前后的分割效果对比可知, 本章算法的平面分割改进效果明显, 如图 4-6 (d)、图 4-6 (h) 和图 4-6 (l) 所示。

PlaneRCNN、PlaneRecNet 和 PlaneRecNet+FEM 的平均精准率 (Average Precision, AP) 数据如表 4-1 所示。交并比 (Intersection over Union, IoU) 是衡量预测区域和真实区域之间重叠比例的指标。本实验采用具有不同 IoU 的 AP、AP₅₀ (IoU $\geq$ 50%) 和 AP₇₅ (IoU $\geq$ 75%), PlaneRecNet+FEM 的平均精准率与改进前的 PlaneRecNet 相比提高了 5.77%, 比 PlaneRCNN 提高了 12.09%。

表 4-1 三种算法的 AP 值对比

算法	AP	AP ₅₀	AP ₇₅
PlaneRCNN	13.36	26.48	12.21
PlaneRecNet	14.19	28.31	12.78
PlaneRecNet+FEM	15.26	30.24	13.17

4.4.2 ICL-NUIM 数据集轨迹对比实验

在 ICL-NUIM 数据集全部八个序列对算法的整体定位性能进行评估。使用绝对轨迹均方根误差 (Root Mean Squared Error, RMSE) 和相对旋转的 RMSE 衡量 ORB-SLAM3、Planar-SLAM、Manhattan-SLAM 与本文算法(图、表中用 Our 表示) 轨迹精度, 并进行对比。绝对轨迹误差 (Absolute Trajectory Error, ATE) 适用于测量 SLAM 系统预测的相机运动轨迹与真实轨迹之间的差异, 而相对旋转误差则更适合用于轨迹漂移程度的衡量。

ICL-NUIM 数据集 lr-kt1、of-kt0 和 of-kt1 序列轨迹对比结果如图 4-7 所示。利用 TUM RGB-D 数据集官方的 RGBD_Benchmark 工具对轨迹误差进行可视化, 序列真实轨迹用黑色线条表示, 蓝色线条代表算法估计的轨迹, 轨迹误差用红色短线表示。lr-kt1 序列场景的点特征退化和室内光照变化剧烈导致 ORB-SLAM3

与 Planar-SLAM 误差较大, 如图 4-7 (a) ~ (d) 所示; Planar-SLAM 仅利用基于 MW 假设的原理估计旋转运动, 在面对 of-kt0 序列场景时有局限性, 从而导致轨迹精度不佳, 如图 4-7 (e) ~ (h) 所示; of-kt0 序列特征稀疏而且相机运动时转弯角域过大, 对 SLAM 系统的稳定运行提出了较大挑战。综合图 4-7 可看出, 本文算法预测的相机运动轨迹误差相对较小, 与序列真实轨迹吻合度较高。

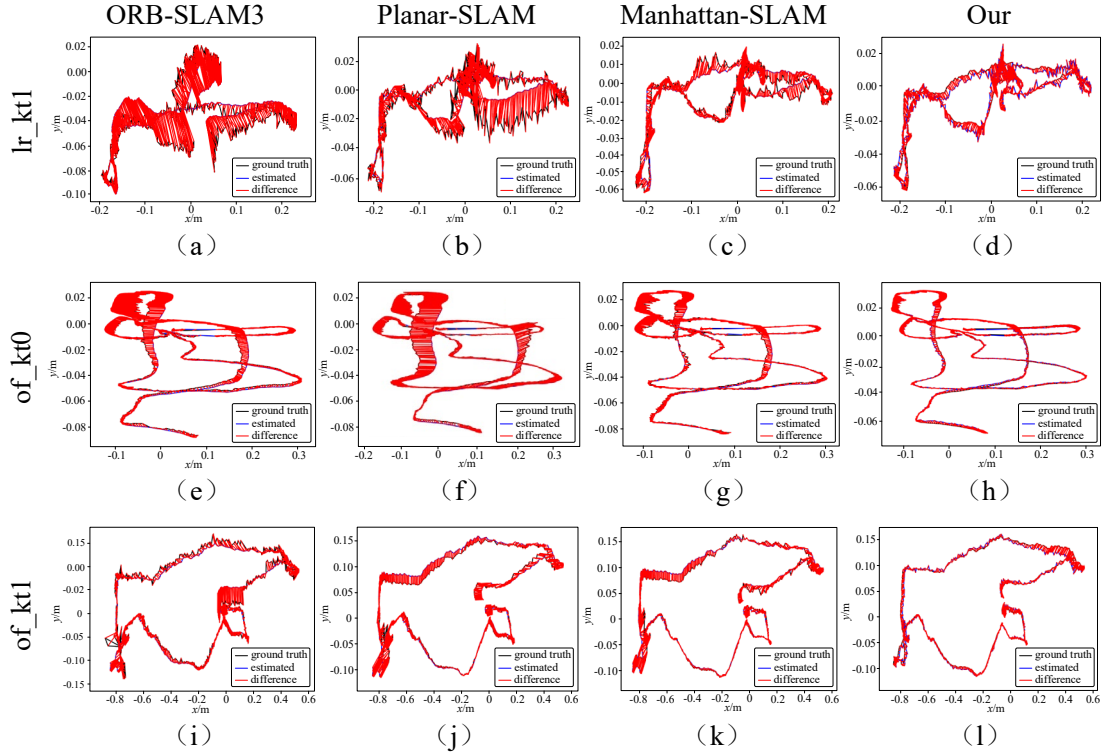


图 4-7 ICL-NUIM 数据集部分序列上的轨迹误差对比图

采用 ICL-NUIM 数据集的全部八个序列对算法进行测试, ATE RMSE 数据对比如表 4-2 所示, 其中每个序列的最优数据用粗体进行突出表示。ORB 特征提取算法在 ICL-NUIM 数据集场景中识别到的点特征数量有限且极不稳定, 导致基于点特征的方法出现特征退化; Planar-SLAM 和 Manhattan-SLAM 都是在符合 MW 假设的条件下, 将位姿估计解耦并结合平面特征估计旋转运动, Manhattan-SLAM 为改善仅基于 MW 假设应用的局限性, 引入了 MW 假设判断方法。然而, 这两种方法都使用传统 AHC 算法提取平面及其参数, 导致平面特征误差较大。本文采用改进的 PlaneRecNet 算法, 使得平面特征提取的效果有所提升, 并且采用方向引导策略, 选择场景优势平面对最优 MF 进行估算, 使跟踪效果得到提升。与 ORB-SLAM3、Planar-SLAM 和 Manhattan-SLAM 的实验数据进行对比, 本文算法的 ATE RMSE 分别降低了 35.53%、20.97%和 11.98%。

表 4-2 ICL-NUIM 数据集测试数据的 ATE RMSE 对比（单位：m）

序列	ORB-SLAM3	Planar-SLAM	Manhattan-SLAM	Our
lr-kt0	0.007	0.006	0.007	0.007
lr-kt1	0.015	0.015	0.011	0.008
lr-kt2	0.022	0.020	0.015	0.012
lr-kt3	0.021	0.012	0.011	0.013
of-kt0	0.036	0.041	0.025	0.021
of-kt1	0.028	0.020	0.013	0.011
of-kt2	0.016	0.011	0.015	0.012
of-kt3	0.015	0.014	0.013	0.011

在 ICL-NUIM 数据集上对算法进行了实验验证，通过 RGBD_Benchmark 工具获得预测轨迹相对旋转误差 RMSE，结果对比如表 4-3 所示。与 ORB-SLAM3、Planar-SLAM 和 Manhattan-SLAM 的测试数据进行对比，本文算法的相对旋转误差 RMSE 分别减少了 57.05%、17.08%和 12.47%。

表 4-3 ICL-NUIM 数据集测试数据相对旋转误差 RMSE 对比（单位：°）

序列	ORB-SLAM3	Planar-SLAM	Manhattan-SLAM	Our
lr-kt0	0.50	0.25	0.43	0.32
lr-kt1	0.63	0.38	0.37	0.33
lr-kt2	0.84	0.37	0.44	0.60
lr-kt3	0.80	0.46	0.25	0.31
of-kt0	1.24	0.40	0.55	0.29
of-kt1	1.08	0.85	0.31	0.23
of-kt2	0.92	0.59	0.62	0.43
of-kt3	0.78	0.43	0.25	0.20

4.4.3 TUM RGB-D 数据集轨迹对比实验

TUM RGB-D 数据集包含大量室内场景序列，因此本文选取该数据集中具有典型结构化场景特点的序列进行评估。使用 evo 测评工具进行轨迹可视化，为方便观察，将轨迹全部投影到 X-Y 轴的二维平面，四种算法在 fr2/desk、fr2/l_cabinet 和 fr3/s_t_near 序列轨迹对比如图 4-8 所示。本文算法在具有较多纹理的 fr2/desk 序列场景与其他算法轨迹精度区别较小，如图 4-8（a）所示。Planar-SLAM 虽能在 fr2/desk 序列生成完整轨迹，但与其他算法相比误差较大。ORB-SLAM3 在 fr3/l_cabinet 序列前期表现良好，但相机运动过程中场景纹理逐渐变少，轨迹漂

移越来越严重，如图 4-8（b）所示。fr3/s_t_near 序列是典型的弱纹理场景，特征点极少，ORB-SLAM3 算法失效，无法得出完整轨迹，如图 4-8（c）所示。

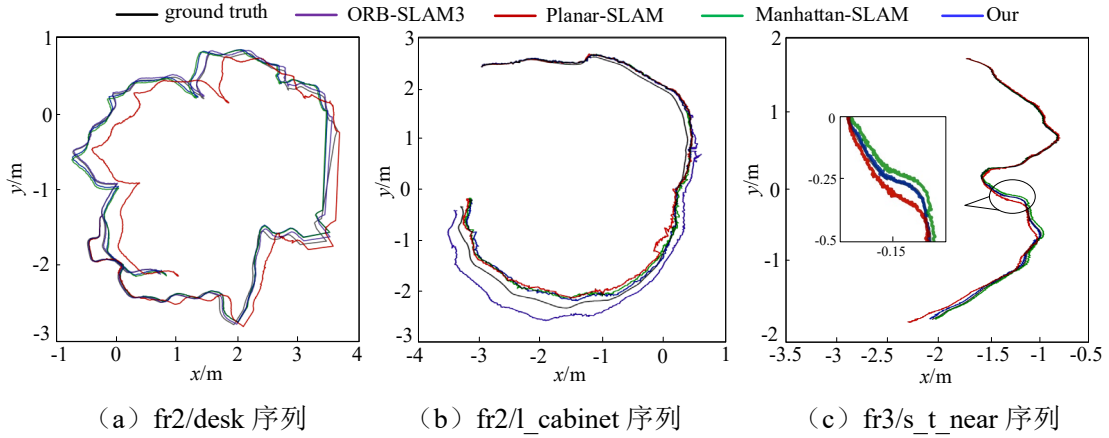


图 4-8 部分 TUM 序列的 X-Y 轴向轨迹对比

从 TUM RGB-D 数据集中选取部分具有结构化场景特点序列对算法进行测试，预测轨迹的 ATE RMSE 数据对比如表 4-4 所示。在以 fr3/s_t_near 序列为代表的弱纹理场景中，ORB-SLAM3 误差较大，甚至运行过程中跟踪丢失，无法预测完整的相机运动轨迹。本文算法采用 MF 跟踪与结构特征辅助定位两种策略，改善了基于 MW 假设方法的局限性，在 fr2/desk 序列轨迹精度略低于 ORB-SLAM3，但是高于 Planar-SLAM，与 ORB-SLAM3、Planar-SLAM 和 Manhattan-SLAM 的测试数据进行对比，本文算法的 ATE RMSE 分别减少了 45.94%、33.09% 和 18.47%。

表 4-4 TUM RGB-D 数据集测试数据 ATE RMSE 对比（单位：m）

序列	ORB-SLAM3	Planar-SLAM	Manhattan-SLAM	Our
fr2/desk	0.018	0.166	0.037	0.020
fr3/cabinet	0.083	0.035	0.023	0.019
fr3/l_cabinet	0.220	0.071	0.083	0.062
fr3/s_nt_far	-	0.022	0.040	0.036
fr3/s_nt_near	-	0.025	0.023	0.017
fr3/s_t_far	-	0.035	0.022	0.021
fr3/s_t_near	-	0.052	0.012	0.012

预测轨迹的相对旋转误差 RMSE 对比如表 4-5 所示。fr2/desk 序列场景包含室内杂乱的书桌，纹理丰富且平面较少，不满足 MW 假设，Planar-SLAM 仅基于 MW 假设进行位姿估计，在此序列上具有局限性，导致旋转估计的高误差；众所周知，视觉 SLAM 仅依赖特征点往往会产生较大的旋转累计误差，从而产生轨

迹漂移，而本文算法的轨迹漂移普遍小于其他三种算法。由 fr3/cabinet 和 fr3/l_cabinet 序列数据可知，基于 MW 假设的多特征方法可以有效减少旋转误差，与 ORB-SLAM3、Planar-SLAM 和 Manhattan-SLAM 的轨迹测试数据进行对比，本文算法相对旋转误差 RMSE 分别减少了 33.15%、24.13%和 18.51%。

表 4-5 TUM RGB-D 数据集测试数据相对旋转误差 RMSE 对比（单位：°）

序列	ORB-SLAM3	Planar-SLAM	Manhattan-SLAM	Our
fr2/desk	1.54	8.54	2.17	2.22
fr3/cabinet	5.21	1.47	1.38	0.97
fr3/l_cabinet	4.74	1.52	2.00	1.79
fr3/s_nt_far	-	1.39	1.52	1.56
fr3/s_nt_near	-	1.27	1.34	0.82
fr3/s_t_far	-	1.15	0.98	0.71
fr3/s_t_near	-	1.93	0.86	0.62

4.4.4 真实场景实验结果与分析

(1) 硬件实验平台

真实场景实验硬件为搭载 Intel RealSense D435i 相机的 Husky 轮式移动机器人，如图 4-9 所示。D435i 相机以每秒 30 帧的频率进行采样，机器人最大运动速度为 1m/s，主要参数设置如表 4-6 所示。



图 4-9 Husky 轮式移动机器人实验平台

表 4-6 主要参数设置

变量名称	参数	数值
相机采样帧率	h	30fps
最大速度	v	1m/s
工作环境温度	T	-10~40°C
尺寸	$L \times W \times H$	990mm × 670mm × 390mm

Intel RealSense D435i 是一款集成了 IMU 传感器的 RGB-D 深度相机, 将 RGB 模块和深度模块集成到 56 立方厘米的体积内, 内部构造如图 4-10 所示。处理器通过 USB 3.1 接口进行数据传输。RGB 模块用于拍摄环境中的彩色图像, 能够提供 1280×720 分辨率的拍摄质量, 速度为 30 帧/秒。深度模块由一组红外摄像头和一个红外激光发射器构成, 红外激光发射器向环境发射红外光的同时, 红外摄像头捕获反射回来的光信号, 并将其转换成数字信号, 以供深度传感器模块计算出物体到相机的距离, 并生成对应的深度信息。

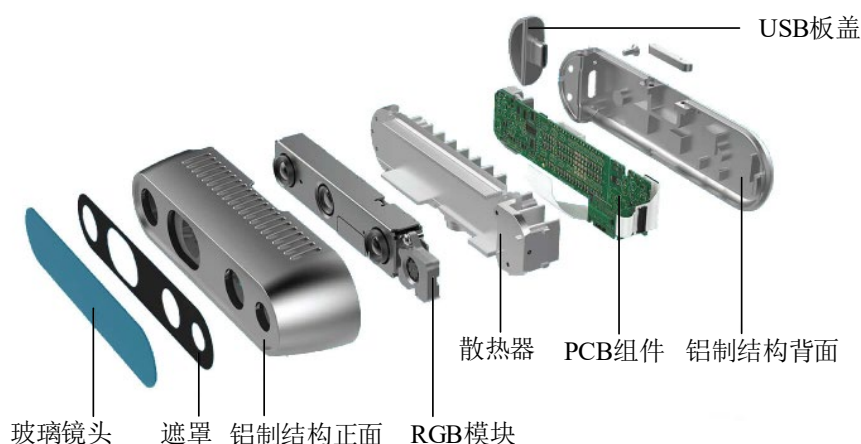
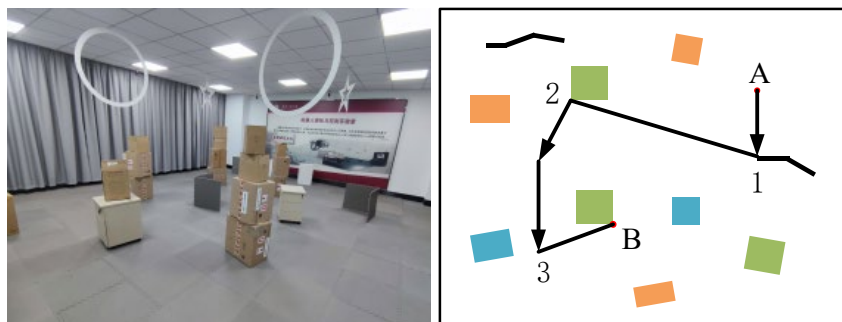


图 4-10 Intel RealSense D435i 相机结构图

(2) 室内自搭建场景实验

为了在真实环境中全面验证本文方法, 在 $7.8\text{m} \times 7.2\text{m}$ 的空旷室内实验室用纸箱和书柜搭建了具有众多平面的结构化场景。该场景符合 MW 假设, 但含有较多小平面和非正交平面, 会对 MF 的提取产生干扰, 如图 4-11 (a) 所示。手持 RealSense D435i 摄像头从 A 点沿箭头方向到达 B 点, 采集了实验场景数据集并且进行算法测试, 场景俯视布局如图 4-11 (b) 所示, 可以使用从 A 点开始, 途经 1、2 和 3 位置到达 B 点的折线路径, 来衡量算法轨迹的漂移程度。



(a) 室内自搭建场景

(b) 室内自搭建场景俯视布局图

图 4-11 室内自搭建场景及其俯视布局图

本文算法与 Manhattan-SLAM 在此场景的定位轨迹对比, 如图 4-12 所示。该场景特征稀疏并且相机运动弯道角域过大, ORB-SLAM3 在运行过程中无法正确跟踪并记录完整轨迹, Manhattan-SLAM 的 MF 构建过程中, 在面对众多杂乱平面的情况下会受到干扰, 从而使轨迹产生较大漂移, 本文算法通过方向引导策略, 使小平面和非正交面干扰问题得到缓解, 实现了更精确的旋转估计和更好的定位效果。自搭场景中数据对比如表 4-7 所示, 本文算法比 Manhattan-SLAM 的 ATE RMSE 降低了 34.18%。

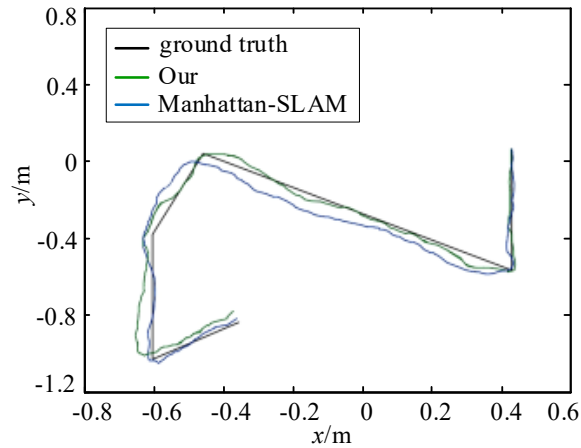
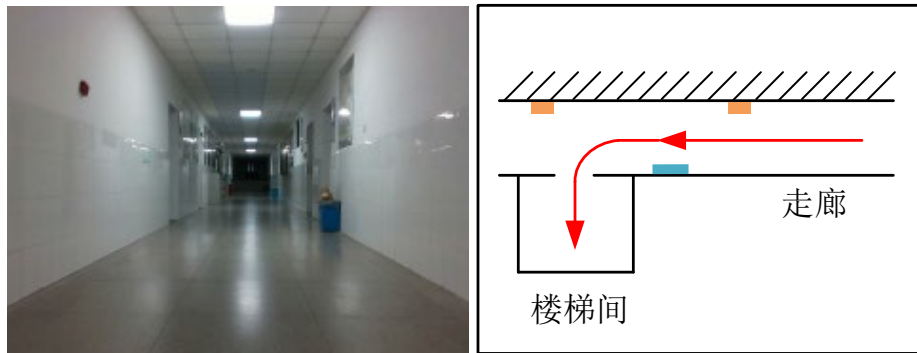


图 4-12 室内自搭建场景轨迹对比

(3) 弱纹理走廊场景实验

为验证本文算法在纹理稀疏且光照不均匀室内场景的表现, 采用走廊场景进行算法测试, 走廊部分长 8m, 楼梯间部分长 4m。该场景主要由白色墙壁和天花板构成, 地面有少量纹理, 但出现了不同程度的反光, 并且由于灯光间距离较远, 整个走廊光照不均匀, 如图 4-13 (a) 所示。采用从走廊到楼梯间的红色路径对算法运行轨迹进行评估和比较, 场景俯视布局如图 4-13 (b) 所示。



(a) 弱纹理走廊场景

(b) 弱纹理走廊场景俯视布局图

图 4-13 弱纹理走廊场景及其俯视布局图

本文算法与 Manhattan-SLAM 在弱纹理走廊场景中的运动轨迹对比,如图 4-14 所示。走廊环境点特征稀少,导致 ORB-SLAM3 跟踪丢失,无法获得完整轨迹。传统平面特征提取算法在面对室内结构化场景时,提取的准确性会受到影响,造成 Manhattan-SLAM 算法定位精度降低;本文算法对平面特征的提取方式进行了优化,在面对稀疏且相似的纹理时具有更好的面特征提取效果,提高了机器人在弱纹理走廊场景的轨迹精度,如表 4-7 所示,本文算法比 Manhattan-SLAM 的 ATE RMSE 降低了 46.61%。

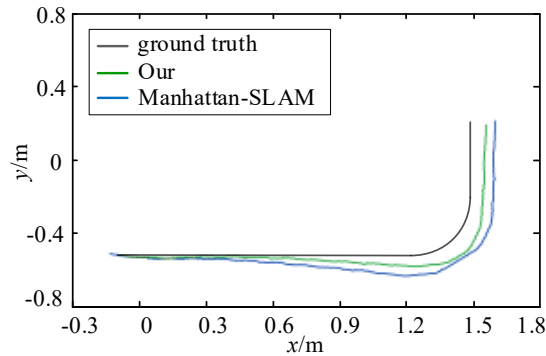


图 4-14 弱纹理走廊场景轨迹对比图

表 4-7 真实场景 ATE RMSE 对比 (单位: m)

算法	Manhattan-SLAM	Our
室内自搭建场景	0.079	0.052
弱纹理走廊场景	0.118	0.063

4.5 本章小结

在室内结构化场景下,针对传统视觉 SLAM 位姿估计性能下降的问题,提出一种基于曼哈顿假设的多特征视觉 SLAM 方法,针对传统平面特征提取算法存在检测错误的问题,引入 FEM 模块对 PlaneRecNet 网络进行改进,提高多特征处理阶段的平面分割效果。针对 MW 假设的场景局限性,判断曼哈顿假设条件,分别采用 MF 跟踪与点、线结构特征辅助的定位方法。在符合 MW 假设的场景,采用方向引导策略选取主平面,降低了小平面对 MF 准确度的影响,比较和跟踪平面结构的正交性获得旋转矩阵,降低旋转累积误差。在公开数据集上的实验结果表明,本章算法的平均 ATE RMSE 比 Manhattan-SLAM 减少了 15.23%,相较于 ORB-SLAM3 减少了 40.74%。

第 5 章 全文总结与展望

5.1 全文总结

本文以室内场景下的移动机器人自主定位与地图构建技术为研究目标,提出了一种多结构特征融合的室内视觉 SLAM 算法。聚焦基于点特征的视觉 SLAM 系统在纹理稀疏以及光照不均匀的室内结构化场景存在特征退化和定位失准这一核心问题,提出具有结构特征辅助的 RGB-D 视觉里程计和基于曼哈顿假设的多特征视觉 SLAM 方法。在基础特征提取部分,重点研究了光照不均匀情况下的边缘拟合线特征提取问题,和室内结构化场景的平面分割及主平面筛选策略;在多结构特征融合位姿估计阶段,重点研究了如何构建结构特征观测模型、结构特征辅助定位策略以及曼哈顿坐标系跟踪方法。

论文主要研究内容及创新点概况如下:

(1) 研究了视觉 SLAM 和结构特征相关基础理论,首先介绍了当前视觉 SLAM 系统搭载的相机模型,对针孔模型和畸变模型进行了分析。然后探索了传统图像点特征、线特征和平面特征常用提取算法的原理。最后,分析了结构特征和曼哈顿世界假说的背景和理论,以及二者之间的关联,并且从视觉 SLAM 系统应用层面探讨了曼哈顿主方向的提取过程及其重要性,为本文结构特征的构建和曼哈顿主方向的提取做好理论研究基础。

(2) 针对传统视觉里程计算法在室内纹理稀疏以及光照不均匀环境下存在特征退化和定位失准的问题,提出一种具有结构特征辅助的 RGB-D 视觉里程计算法。首先,为克服因室内光照不均匀造成线特征提取鲁棒性差的问题,通过引入水平与垂直约束的像素连接规则,提出改进边缘拟合线特征提取方法,提取更具连续性和稳定性的线段。然后,充分利用线特征之间的结构关系,构建更稳定的结点-线结构特征,减少旋转累计误差,同时通过结点辅助以提高特征匹配率。最后,构造点、线重投影误差函数,估计六自由度位姿。本章在 TUM RGB-D 和模拟光照变化 ICL-NUIM 数据集上的实验结果表明,本章算法平均 ATE RMSE 比 MSC-VO 降低了 33.69%。

(3) 针对室内结构化场景中传统视觉 SLAM 算法出现定位精度差甚至跟踪失败的问题,提出一种基于曼哈顿假设的多特征视觉 SLAM 方法。首先,考虑

到传统平面特征提取算法存在检测错误的情况，使用改进的 PlaneRecNet 网络提高平面特征提取准确性。然后，鉴于单纯利用 MW 假设方法的场景受限问题，通过平面间的正交关系判断 MW 假设条件，对于满足 MW 假设的场景，采用方向引导策略筛选主平面，比较和跟踪平面结构的正交性估计三自由度旋转矩阵，降低旋转累积误差；在非曼哈顿场景直接采用具有结构特征辅助的定位方法。最后，构造最小重投影误差函数，进行非线性优化，进一步提升算法性能。在公开 ICL-NUIM 与 TUM RGB-D 数据集的测试结果表明，本文算法的平均 ATE RMSE 比 Manhattan-SLAM 减少了 15.23%，相较于 ORB-SLAM3 减少了 40.74%，表现出了更优越的准确性和鲁棒性。

5.2 研究展望

本文以视觉 SLAM 系统在室内结构化场景因特征退化引起定位失准的问题为核心，通过公开数据集仿真和移动机器人硬件平台对本文方法进行了实验验证。结果表明，本文方法在室内结构化场景可以得到更准确的相机运动轨迹。鉴于现实应用的多样性，仍然有问题需要更进一步研究，主要有以下几个方向：

（1）对闭环检测线程作进一步改进。考虑到机器人运动过程中有时会存在轨迹闭环，后续研究可针对室内场景利用结构特征标注信息判断当前相机运动轨迹的闭环，并进行闭环矫正，提高轨迹精度与建图效果。

（2）对场景作稠密建图。鉴于结构特征对室内场景空间结构的表示，后续可以对建图模块进行改进，充分利用室内结构特征信息，构建较为准确的场景稠密地图。

参考文献

- [1]. 张继贤, 刘飞. 视觉 SLAM 环境感知技术现状与智能化测绘应用展望[J]. 测绘学报, 2023, 52(10): 1617-1630.
- [2]. Sousa R B, Sobreira H M, Moreira A P. A systematic literature review on long-term localization and mapping for mobile robots[J]. Journal of Field Robotics, 2023, 40(5): 1245-1322.
- [3]. Eyvazpour R, Shoaran M, Karimian G. Hardware implementation of SLAM algorithms: a survey on implementation approaches and platforms[J]. Artificial Intelligence Review, 2023, 56(7): 6187-6239.
- [4]. 袁国帅, 齐咏生, 刘利强, 等. 一种基于因子图消元优化的激光雷达视觉惯性融合 SLAM 方法[J]. 电子学报, 2023, 51(11): 3042-3052.
- [5]. Bescos B, Campos C, Tardós J D, etc. DynaSLAM II: Tightly-coupled multi-object tracking and SLAM[J]. IEEE robotics and automation letters, 2021, 6(3): 5191-5198.
- [6]. 夏琳琳, 崔家硕, 宋梓维, 等. 线特征优化配置在室内 RGB-D SLAM 系统中的应用[J]. 中国惯性技术学报, 2022, 30(06): 760-767+776.
- [7]. Yan J, Zheng Y, Yang J, etc. PLPF-VSLAM: An indoor visual SLAM with adaptive fusion of point-line-plane features[J]. Journal of Field Robotics, 2024, 41(1): 50-67.
- [8]. 王金科, 左星星, 赵祥瑞, 等. 多源融合 SLAM 的现状与挑战[J]. 中国图象图形学报, 2022, 27(02): 368-389.
- [9]. Wu F, Beltrame G. Direct sparse odometry with planes[J]. IEEE Robotics and Automation Letters, 2021, 7(1): 557-564.
- [10]. Guo H, Peng S, Lin H, etc. Neural 3D scene reconstruction with the Manhattan-world assumption[C]. 2022 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA. 2022: 5511-5520.
- [11]. Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: part I[J]. IEEE robotics & automation magazine, 2006, 13(2): 99-110.

- [12].刘庆运, 杨华阳, 刘涛等. 基于激光雷达与深度相机融合的 SLAM 算法[J]. 农业机械学报, 2023, 54(11): 29-38.
- [13].Saputra M R U, Lu C X, de Gusmao P P B, etc. Graph-based thermal-inertial SLAM with probabilistic neural networks[J]. IEEE Transactions on Robotics, 2021, 38(3): 1875-1893.
- [14].Davison A J, Reid I D, Molton N D, etc. MonoSLAM: Real-time single camera SLAM[J]. IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6): 1052-1067.
- [15].Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]. 2007 6th IEEE and ACM international symposium on mixed and augmented reality. Nara, Japan, IEEE, 2007: 225-234.
- [16].Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE transactions on robotics, 2015, 31(5): 1147-1163.
- [17].Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras[J]. IEEE transactions on robotics, 2017, 33(5): 1255-1262.
- [18].Campos C, Elvira R, Rodríguez J J G, etc. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [19].Lajoie P Y, Beltrame G. Swarm-slam: Sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems[J]. IEEE Robotics and Automation Letters, 2023, 9(1): 475-482.
- [20].Kachurka V, Rault B, Muñoz F I I, etc. WeCo-SLAM: Wearable cooperative SLAM system for real-time indoor localization under challenging conditions[J]. IEEE Sensors Journal, 2021, 22(6): 5122-5132.
- [21].Karrer M, Chli M. Distributed variable-baseline stereo SLAM from two UAVs[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi An, China, IEEE, 2021: 82-88.
- [22].Gomez-Ojeda R, Moreno F A, Zuniga-Noël D, etc. PL-SLAM: A stereo SLAM

- system through the combination of points and line segments[J]. IEEE Transactions on Robotics, 2019, 35(3): 734-746.
- [23]. Von Gioi R G, Jakubowicz J, Morel J M, etc. LSD: A line segment detector[J]. Image Processing On Line, 2012, 2: 35-55.
- [24]. 于雅楠, 卫红, 陈静. 基于局部熵的 SLAM 视觉里程计优化算法[J]. 自动化学报, 2021, 47(06): 1460-1466.
- [25]. 孙新柱, 龚光强, 陈孟元等. 室内场景下基于曼哈顿约束的多重特征视觉 SLAM 方法[J]. 中国惯性技术学报, 2023, 31(09): 890-899.
- [26]. Yang Y, Geneva P, Eckenhoff K, etc. Visual-inertial odometry with point and line features[C]. 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS). Macau, China, IEEE, 2019: 2447-2454.
- [27]. Company-Corcoles J P, Garcia-Fidalgo E, Ortiz A. MSC-VO: Exploiting manhattan and structural constraints for visual odometry[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2803-2810.
- [28]. Zou D, Wu Y, Pei L, etc. StructVIO: Visual-inertial odometry with structural regularity of man-made environments[J]. IEEE Transactions on Robotics, 2019, 35(4): 999-1013.
- [29]. 闫莉萍, 李程伟, 徐柏凯等. 一种基于结构特征的单目视觉惯性里程计[J]. 系统工程与电子技术, 2023, 45(10): 3207-3217.
- [30]. Lim H, Jeon J, Myung H. UV-SLAM: Unconstrained line-based SLAM using vanishing points for structural mapping[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 1518-1525.
- [31]. Kim P, Coltin B, Kim H J. Low-drift visual odometry in structured environments by decoupling rotational and translational motion[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane, Australia, 2018: 7247-7253.
- [32]. Yang S, Scherer S. Monocular object and plane SLAM in structured environments[J]. IEEE Robotics and Automation Letters, 2019, 4(4): 3145-3152.
- [33]. 孙沁璇, 苑晶, 张雪波, 等. PLVO: 基于平面和直线融合的 RGB-D 视觉里程

- 计[J]. 自动化学报, 2021: 1-13.
- [34]. Li Y, Brasch N, Wang Y, etc. Structure-SLAM: Low-drift monocular SLAM in indoor environments[J]. IEEE Robotics and Automation Letters, 2020, 5(4): 6583-6590.
- [35]. Li Y, Yunus R, Brasch N, etc. RGB-D SLAM with structural regularities[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi An, China, 2021: 11581-11587.
- [36]. Yunus R, Li Y, Tombari F. Manhattan SLAM: Robust planar tracking and mapping leveraging mixture of Manhattan frames[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi An, China, 2021: 6687-6693.
- [37]. Xia S, Chen D, Wang R, etc. Geometric primitives in lidar point clouds: A review[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 685-707.
- [38]. 曹力科, 肖晓晖. 基于卷帘快门 RGB-D 相机的视觉惯性 SLAM 方法[J]. 机器人, 2021, 43(02): 193-202.
- [39]. Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 32(1): 105-119.
- [40]. Ng P C, Henikoff S. SIFT: Predicting amino acid changes that affect protein function[J]. Nucleic acids research, 2003, 31(13): 3812-3814.
- [41]. Bay H, Ess A, Tuytelaars T, etc. Speeded-up robust features (SURF)[J]. Computer vision and image understanding, 2008, 110(3): 346-359.
- [42]. Rublee E, Rabaud V, Konolige K, etc. ORB: An efficient alternative to SIFT or SURF[C]. 2011 International conference on computer vision. Barcelona, Spain. IEEE, 2011: 2564-2571.
- [43]. Havel J, Herout A, Dubská M. Vanishing points in point-to-line mappings and other line parameterizations[J]. Pattern Recognition Letters, 2013, 34(6): 703-708.
- [44]. Akinlar C, Topal C. EDLines: A real-time line segment detector with a false detection control[J]. Pattern Recognition Letters, 2011, 32(13): 1633-1642.

- [45]. 许承权, 柳庆威. 基于线段检测算法的倾斜摄影直角像控点目标检测方法[J]. 地球信息科学学报, 2021, 23(03): 505-513.
- [46]. Barath D, Matas J. Graph-cut RANSAC: Local optimization on spatially coherent structures[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(9): 4961-4974.
- [47]. 刘海兵, 曲英杰, 颜青松, 等. 建筑物单体结构化重建的变尺度网格基元提取法[J]. 测绘学报, 2023, 52(11): 1953-1961.
- [48]. Liu C, Yang J, Ceylan D, et al. Planenet: Piece-wise planar reconstruction from a single rgb image[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA. 2018: 2579-2588.
- [49]. Liu C, Kim K, Gu J, etc. Planercnn: 3d plane detection and reconstruction from a single image[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 4450-4459.
- [50]. Xie Y, Shu F, Rambach J, etc. PlaneRecNet: Multi-task learning with cross-task consistency for piece-wise plane detection and reconstruction from a single RGB image[C]. 2021 British Machine Vision Conference (BMVC). 2021: 1-14.
- [51]. Liu Y, Chen G, Knoll A. Globally optimal vertical direction estimation in atlanta world[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 1949-1962.
- [52]. Park S, Schöps T, Pollefeys M. Illumination change robustness in direct visual slam[C]. 2017 IEEE international conference on robotics and automation (ICRA). Marina Bay, Singapore, IEEE, 2017: 4523-4530.
- [53]. Handa A, Whelan T, McDonald J, etc. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM[C]. 2014 IEEE international conference on Robotics and automation (ICRA). Hong Kong, China, IEEE, 2014: 1524-1531.
- [54]. Schubert D, Goll T, Demmel N, etc. The TUM VI benchmark for evaluating visual-inertial odometry[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain, IEEE, 2018: 1680-1687.

攻读学位期间发表的学术论文与取得的科研成果

一、发表的学术论文

- [1] 孙新柱*, 龚光强, 陈孟元, 程浩, 郭行荣. 室内场景下基于曼哈顿约束的多重特征视觉 SLAM 方法. 中国惯性技术学报 (EI 收录), 已检索 (EI: 20234114853521). (导师第一, 本人第二)
- [2] 孙新柱*, 龚光强, 陈孟元. 具有结构特征辅助的 RGB-D 视觉里程计. 传感技术学报 (CSCD 核心库收录), 已录用. (导师第一, 本人第二)
- [3] Hao Cheng, Guangqiang Gong, Mengyuan Chen, Xingrong Guo, Runbang Qian. A Localization Optimization Algorithm with Improved Feature Tracking Mechanism. The 38th Youth Academic Annual Conference of Chinese Association of Automation. 已检索. (本人第二)

二、参与的科研项目

- [1] 安徽省重点研究与开发计划项目 (参与)
- [2] 安徽省高校协同创新项目 (参与)
- [3] 安徽省高校杰出青年科研项目 (参与)

致 谢

时光荏苒，值此硕士学业即将完成之际，我要向许多人表达衷心的感谢。在求学之路上，他们给了我无私的关爱和帮助，让我能够顺利完成我的研究生课程。在此，我要向以下人士表示最真挚的谢意：

在研究生学习期间，首先衷心感谢我的导师孙新柱老师。孙老师在学术上表现出来严谨的追求和儒雅的作风都深深地影响到学生。感谢孙老师一直以来对我的关心和学术上的指导。同时感谢实验室的陈孟元老师，正是在陈老师的帮助与教导下，学生得以快速成长。

我还要感谢实验室于尧、郭俊阳、刘金辉、陈何宝、韩朋朋、方鑫等各位师兄给予我在实验和生活上的指导和帮助，让我受益良多。还要感谢我的室友和实验室的程浩、郭行荣、钱润邦等各位同门，他们在学术讨论中给予了我很多启发和帮助，我们共同探讨问题，一起解决困难。我也要感谢徐奥、李鹏飞等各位师弟和班级其他同学、朋友，在忙碌的学习和科研生活中，他们的陪伴给我带来了许多愉悦和难忘的时光。

在生活上，我要感谢我的父母。他们是我求学路上最坚实的后盾，是我前行路上最坚定的支持。感谢他们无私的支持和无条件的爱，是他们的鼓励和支持，让我能够顺利完成学业，愿他们健康、幸福。

最后，我要向评审委员会、学校的领导、教师和工作人员表达最真诚的谢意，感谢他们为我提供了良好的学习和生活环境，让我能够专注于我的学业，全身心地投入到我的研究中。

再次感谢所有给予我关心和帮助的人，你们的支持是我走向成功的动力源泉。
谢谢！

尚德敏学 唯实惟新

