

分类号: TP242. 6

学校代码: 10363

密 级: 公开

学 号: 2220320124



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 面向大视角基于改进MSS网
络的视觉SLAM算法

论 文 作 者	杜鹏
校 内 导 师	陈孟元 教授
校 外 导 师	袁学超 高级工程师
专业学位类别	电子信息
研究方向 (领域)	控制工程与自动化技术

论文提交日期: 2025 年 5 月 15 日

单位代码: 10363

学 号：2220320124

目 录

A Visual SLAM Algorithm Based on Improved MSS Network for Large Viewpoints

研究方向（领域）： 控制工程与自动化技术

论文答辩日期: 2025 年 05 月 12 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：杜鹏

日期：2025 年 5 月 15 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：杜鹏

指导教师签名：陈吉云

日期： 2025 年 5 月 15 日

日期： 2025 年 5 月

面向大视角基于改进 MSS 网络的视觉 SLAM 算法

摘 要

同步定位与地图构建(Simultaneous Localization And Mapping, SLAM)是指移动机器人在面对复杂场景中能够独立自主的构建周围环境的地图并且精准定位自己所处的位置。近年来,由于视觉 SLAM 在各种复杂场景中具有较好的鲁棒性,逐渐成为研究主流。传统的视觉 SLAM 算法通过匹配和跟踪相邻帧之间的特征点来重建相机的位姿轨迹,并利用闭环检测技术对全局轨迹进行校正。然而,移动机器人在面对大视角场景时,点特征提取算法往往难以实现相邻两帧之间稳定的特征匹配,导致轨迹出现漂移,甚至引发跟踪丢失等问题。因此,如何消除视角变换对视觉 SLAM 定位和建图的影响,仍然是亟待解决的问题。

针对移动机器人在大视角运动下因图像模糊导致特征匹配准确率低的问题,设计了一种基于改进 MSS 网络与误匹配剔除的视觉 SLAM 算法。首先,该网络对图像进行下采样操作得到不同尺度的图像,通过特征注意模块增强图像特征信息提取的能力。其次,利用监督注意力模块对各阶段的特征信息进行监督和融合,使模糊图像逐步恢复为原尺寸的清晰图像。最后,通过局部信息熵对图像进行区域划分,根据对极几何约束判断相机是否发生大视角变换,分别采用十字五宫格和对角五宫格统计方法进行误匹配剔除。

针对传统视觉 SLAM 算法在大视角场景下关键帧选取不准确的

问题，设计了一种基于改进关键帧选取策略的视觉 SLAM 算法。首先，通过计算相邻帧之间旋转矩阵与平移向量，构建共视区域，并结合特征点在连续帧中的跟踪帧数筛选出稳定的有效特征点，为关键帧选取提供可靠判断依据。其次，利用共视区域占比率与运动连续性相结合，建立关键帧选取条件，从而有效避免因视角变换而导致的误判。最后，通过融合运动微分熵与信息熵，构建熵增益函数对新选取的关键帧进行信息评估，实现冗余关键帧的高效剔除。

为验证本文所提算法有效性，在公开数据集 TUM、EUROC 和 KITTI 中对基于改进的 MSS 网络与误匹配剔除的视觉 SLAM 算法和基于改进关键帧选取策略的视觉 SLAM 算法进行验证，并与主流的 ORB-SLAM3、PL-SLAM、OV²-SLAM 算法对比分析，验证本文所提算法在大视角场景下的定位精度与构图效果。在搭建的硬件平台上验证本文算法的可行性，为大视角场景下视觉 SLAM 研究提供参考依据。

关键词：同步定位与地图构建，大视角，特征匹配，共视区域，关键帧

A VISUAL SLAM ALGORITHM BASED ON IMPROVED MSS NETWORK FOR LARGE VIEWPOINTS

ABSTRACT

Simultaneous Localization And Mapping (SLAM) refers to the ability of a mobile robot to independently and autonomously construct a map of its surroundings and accurately locate its position in a complex scene. In recent years, visual SLAM has gradually become the mainstream of research due to its better robustness in various complex scenes. Traditional vision SLAM algorithms reconstruct the camera's positional trajectory by matching and tracking the feature points between neighboring frames, and correct the global trajectory using closed-loop detection. However, when a mobile robot faces a large viewing angle scene, it is often difficult for the point feature extraction algorithm to achieve stable feature matching between two neighboring frames, resulting in trajectory drift and even triggering problems such as tracking loss. Therefore, how to eliminate the influence of view angle transformation on visual SLAM localization and map building remains an urgent problem to be solved.

Aiming at the problem of low feature matching accuracy due to image blurring under large view angle motion of mobile robots, a visual SLAM algorithm based on an improved MSS network with false matching rejection is designed. First, the network performs downsampling

operations on images to obtain images at different scales, and enhances the ability of image feature information extraction through the feature attention module. Secondly, the supervised attention module is utilized to supervise and fuse the feature information at each stage, so that the blurred image is gradually restored to a clear image of the original size. Finally, the image is divided into regions by local information entropy, and the cross-five grid and diagonal-five grid statistical methods are used for mismatch rejection based on the pair of polar geometric constraints to determine whether the camera undergoes a large viewing angle transformation, respectively.

Aiming at the problem of inaccurate key frame selection of traditional visual SLAM algorithms in large viewing angle scenes, a visual SLAM algorithm based on improved key frame selection strategy is designed. Firstly, by calculating the rotation matrix and translation vectors between neighboring frames, we construct the co-visual region, and screen out the stable and effective feature points by combining the number of tracking frames of the feature points in consecutive frames, so as to provide a reliable basis of judgment for key frame selection; secondly, we use the ratio of the co-visual region to the continuity of the motion to establish the conditions for the key frame selection, so as to effectively avoid misjudgment due to the change of the viewpoints; and lastly, we integrate the geometric differential entropy and the information entropy into the key frame selection strategy. Finally, by integrating geometric differential

entropy and information entropy, an entropy gain function is constructed to evaluate the information of the newly selected keyframes and realize the efficient elimination of redundant keyframes.

In order to verify the effectiveness of the algorithms proposed in this paper, the visual SLAM algorithm based on the improved MSS network with false match rejection and the visual SLAM algorithm based on the improved keyframe selection strategy are verified in the public datasets TUM, EUROC, and KITTI, and compared and analyzed with the mainstream ORB-SLAM3, PL-SLAM, and OV²-SLAM algorithms, to validate the effectiveness of the proposed algorithms. Compare and analyze with mainstream ORB-SLAM3, PL-SLAM, and OV²-SLAM algorithms to verify the localization accuracy and framing effect of the proposed algorithm in the large-viewing angle scene. The feasibility of this paper's algorithm is verified on the built hardware platform, which provides a reference basis for the research of visual SLAM in large view angle scenes.

Du Peng(Electronic Information)

Supervised by Professor Chen Mengyuan

Keywords: Synchronized Localization and Map Construction, Large Viewing Angle, Feature Matching, Co-view Vegin, Key Krame

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 研究现状.....	3
1.2.2 大视角下视觉 SLAM 研究现状.....	5
1.3 研究内容与章节安排	7
第 2 章 视觉 SLAM 与深度学习基础理论.....	9
2.1 视觉 SLAM 系统框架	9
2.1.1 传感器数据.....	9
2.1.2 视觉里程计	11
2.1.3 后端优化.....	13
2.1.4 闭环检测.....	13
2.1.5 建图.....	14
2.2 深度学习基础结构	15
2.2.1 卷积神经网络.....	15
2.2.2 U-Net 基本结构.....	17
2.2.3 去模糊算法.....	18
2.3 本章小结	18
第 3 章 基于改进 MSS 网络与误匹配剔除的 VSLAM 算法.....	19
3.1 改进 MSS 网络的去模糊算法	19
3.1.1 特征注意力模块 FAM 结构设计	21
3.1.2 监督注意力模块 SAM 结构设计	22
3.1.3 损失函数.....	23
3.2 改进的误匹配剔除算法	23
3.2.1 局部信息熵网格划分	24
3.2.2 五宫格统计法.....	25
3.3 实验及分析	26

3.3.1 不同视角下图像去模糊对比实验	27
3.3.2 不同视角下误匹配剔除对比实验	28
3.3.3 轨迹对比实验	29
3.4 本章小结	31
第 4 章 基于改进关键帧选取策略的视觉 SLAM 算法.....	32
4.1 关键帧补偿	32
4.1.1 共视区域构建.....	33
4.1.2 有效特征点选取.....	35
4.2 冗余帧剔除	37
4.3 实验及分析	38
4.3.1 关键帧选取实验.....	39
4.3.2 室内场景轨迹对比实验.....	40
4.3.3 室外场景轨迹对比实验.....	42
4.4 移动机器人真实场景实验.....	44
4.5 本章小结	45
第 5 章 全文总结展望	46
5.1 全文总结	46
5.2 展望	47
参考文献	48
攻读学位期间发表的学术论文和科研成果	53
致谢	54

第 1 章 绪论

1.1 研究背景及意义

随着科技的不断的创新和发展,机器人技术在各个领域得到广泛的应用。机器人是一种可以独立自主完成任务的智能设备,因为其高度智能化和优秀的决策能力,可以轻松模仿人类一些简单的动作和行为,替代人们做一些危险的工作^{[1][2]}。机器人技术的出现对于人类来说是一种挑战和机遇,通过机器人技术不仅使人们的生产和生活方式发生翻天覆地的变化,也为工业、医疗和军事等领域产生了深远的影响。

在工业领域中,机器人技术已得到广泛使用,工业机器人能够完成高精度、高强度和繁琐度高的任务,大幅度提升生产效率,同时因为机器人技术的发展催生了很多就业的机会。在医疗领域中,手术机器人能够提高手术的精度,实现微创手术,并减少患者的痛苦,从而使社会的医疗条件大大提高。在军事领域中,机器人凭借其高度自主能力,执行高危的军事任务,从而大幅度减少人员的伤亡。此外,机器人在教育、服务,娱乐和农业等领域也有广泛的应用,给人们提供便捷和多样化的服务。由于目前机器人技术具有较高的发展前景,世界各国都投入大量的人力和物力对机器人技术进行创新和发展。随着计算机视觉和机器学习技术的快速发展,移动机器人的视觉 SLAM 技术取得了重要突破^[3]。现在 SLAM 技术已被广泛应用于无人驾驶、智能家居、工业自动化等领域,为人们的生活和工作带来了诸多便利和效益。

近年来,随着移动机器人应用场景逐渐向技术要求更高的复杂环境扩展,传统移动机器人技术面临更大的挑战,因此复杂环境下的移动机器人技术已成为当前研究的热点方向^{[4][5]}。

视觉 SLAM 技术为移动机器人提供了一种在未知环境下自主导航和理解环境的能力,主要依赖于机器人的视觉传感器如摄像头来实现。这项技术的核心在于能够同时处理两个紧密相连的任务:确定机器人的位置与构建周围环境地图。定位过程涉及准确识别机器人在其所处环境中所在的位置和方向,而地图构建则是基于初始空白状态逐步描绘出周围环境的详细图景。通过连续追踪机器人的运

动轨迹，并运用概率方法与优化算法，SLAM 技术能够不断改进对机器人当前位置和环境布局的理解^[6]。借助从视觉传感器获取的数据，系统可以实时解析信息，以精确估算机器人的姿态并更新其对环境的认知。因此即使在复杂的场景中，机器人也能够自我定位并探索新的领域，为其执行更复杂的任务奠定了基础。这种方法不仅提升了机器人在动态环境中的适应能力，还增强了其对未知空间的探索效率。

视觉 SLAM 技术在自主导航过程中发挥着至关重要的作用，通过传感器捕捉周围环境信息，可以实时构建车辆行驶环境的地图并精确定位车辆位置^[7]。根据传感器的不同主要分为相机和激光雷达两大类，两种传感器各有优缺点，相较于激光雷达，相机的优势在于能够提供丰富的视觉信息，如颜色和纹理等，有助于识别和分类物品，应用更广泛，现如今视觉 SLAM 所使用的相机主要为单目、双目、RGB-D 相机。

在视觉 SLAM 发展的初期主要使用概率滤波技术，如卡尔曼滤波、扩展卡尔曼滤波和粒子滤波等^{[8][9][10]}。这些滤波方法基于概率模型，忽略了更早的历史数据，只认为当前状态和前一刻状态相关。这导致滤波方法对相邻图像帧有严格要求，难以较好的完成闭环，并且滤波算法对计算和储存有着较高的要求，这会导致该算法的性能受到限制。为了提升 SLAM 系统的实时性和精度，研究重点逐渐转向如何更高效地处理图像数据。基于特征点提取与匹配的技术为视觉 SLAM 提供了可靠的前端视觉里程计方案，但这种方法也存在明显的局限性。例如，在光照剧烈变化、场景纹理稀疏或相机快速运动的情况下，单一视觉传感器采集的图像出现模糊或失真，导致特征点提取不准确。此外，当相机视角发生显著变化时，特征匹配的精度会大幅下降，进而影响相机运动轨迹的恢复效果。因此，优化视觉 SLAM 系统面临的核心问题是如何在复杂环境下实现稳定且精确的特征提取与运动轨迹恢复。特别是在光照条件不佳或纹理缺失的场景中，确保特征点能够准确表征环境信息至关重要^[11]。同时，解决大视角变化下的动态模糊问题也是提升系统鲁棒性的关键。通过改进特征提取算法和匹配策略，可以有效应对这些挑战，从而显著提升视觉 SLAM 系统的整体性能。

1.2 国内外研究现状

1.2.1 视觉 SLAM 研究现状

SLAM 技术的核心目标是通过传感器数据实现实时定位与地图构建。根据传感器类型可分为激光 SLAM 和视觉 SLAM 两大主流方案。激光 SLAM 依赖激光雷达获取精确的三维点云数据，虽然定位精度高但成本昂贵；视觉 SLAM 则利用摄像头采集图像，通过特征提取与匹配实现定位建图，具有成本低、信息丰富等优势，在机器人导航和室内定位等领域广泛应用。

近年来，基于视觉的 SLAM 系统因其能感知丰富的环境信息并实现稠密地图重建而受到重视。视觉 SLAM 的研究方法主要分为直接法和特征点法的两类。直接法是一种视觉 SLAM 技术，通过直接利用图像的像素灰度信息进行跟踪，无需提取特征点^[12]。基于图像帧之间的光度误差来解算相机的位姿变换，避免了特征计算，具有较高的鲁棒性。由于直接法依赖于像素亮度信息，因此在纹理稀疏的场景中仍能有效完成相机跟踪，但对相机运动速度和光照条件有一定要求。直接法在处理图像时，根据涉及像素的密度不同，可分为三类：稀疏直接法、稠密直接法和半稠密直接法。稀疏直接法专注于图像中信息量丰富的关键点，通过最小化点的光度误差来估计相机的运动和场景结构，稠密直接法则是直接处理整个图像的所有像素，利用高度的深度信息和估计相机的运动，生成高质量的三维地图，其优势在于在复杂的场景中也能提供可靠的定位和建图，但计算成本较高^[13]。半稠密直接法则介于疏密和稠密之间，通过选择性地处理图像中的重要像素，既保留了稠密方法的高质量地图生成能力，又具备了稀疏方法的高效性，适用于那些对实时性和地图质量都有较高要求的应用场景^[14]。

随着直接法和视觉 SLAM 结合的应用越来越广泛，2011 年，Newcombe 等人提出的 DTAM^[15]是首个基于直接法的稠密视觉 SLAM 系统。DTAM 通过最小化像素亮度误差来估计相机运动，并构建稠密三维地图。尽管 DTAM 在稠密重建方面表现出色，但其高计算需求限制了实时应用。2014 年，Jakob Engel 等人提出的 LSD-SLAM^[16]是一种半稠密直接法 SLAM 系统。LSD-SLAM 在普通 CPU 上通过提取梯度显著的像素并利用归一化的光度误差进行跟踪，实现了大规模场

景的半稠密重建，具有较高的实用性。同年 Forster 等人提出的 SVO 是一种基于稀疏直接法的视觉里程计。SVO 无需计算描述子，仅依赖图像中的少量关键点便可对局部像素进行特征匹配，计算需求较低，在资源受限的环境中表现出色。

基于特征点法的视觉 SLAM 是视觉 SLAM 领域中最经典和广泛应用的技术路线之一。其核心思想是通过提取图像中的特征点，并利用这些特征点进行图像帧匹配，从而估计相机的运动轨迹和构建环境地图^[17]。在 2007 年，Davison 等人提出的 Mono-SLAM^[18]是首个结合卡尔曼滤波器的单目视觉 SLAM 系统，通过对图像特征点进行提取和跟踪，虽然实现了对移动机器人的实时定位和周围环境的地图构建，但计算复杂度较高，难以应用在大规模的场景。为了解决这些问题，Klein G 等人提出了 PTAM。作为视觉 SLAM 发展历程中的重要突破，PTAM 创新性地将定位与建图过程分解为两个并行线程，使二者能够独立运行且互不干扰，大幅提升了系统的实时性和运行效率。此外，其通过全局优化相机位姿和地图点的方式有效降低了系统运行时的误差累积，通过引入关键帧机制选取特征丰富且具有代表性的图像帧来构建地图显著减少了计算复杂度。然而，该算法仍存在明显局限，由于缺乏闭环检测模块，系统在长时间运行时会产生持续累积的误差，特别是在复杂场景下，这种误差累积会导致轨迹无法闭合和地图失真等问题，严重制约了系统的整体性能表现。2015 年，西班牙 Mur-Artal 及其团队提出的 ORB-SLAM^[19]是一种单目视觉 SLAM 算法，该算法通过跟踪线程对 ORB 特征进行提取和匹配并估计相机运动，通过词袋模型识别两帧之间的相似程度，判断系统是否回环并进行全局优化，极大提高了系统的定位精度和跟踪成功率。为了进一步提高 ORB-SLAM 系统的稳定性和定位精度，2017 年，Mur-Artal 等人提出了 ORB-SLAM2^[20]，该算法在原有的 ORB-SLAM 基础上进行改进，适用于各种传感器如单目、双目、RGB-D 相机，在面对复杂场景时，能够处理更丰富的特征信息，进一步提高系统的适用性和鲁棒性。2020 年 Mur Artal 等人为了实现惯性和视觉的融合，在 ORB-SLAM2 的基础上提出 ORB-SLAM3^[21]，增加了 IMU 传感器，使系统能够利用多种传感器的数据进行更加精准的定位和地图构建。ORB-SLAM3 系统框架如图 1-1 所示。

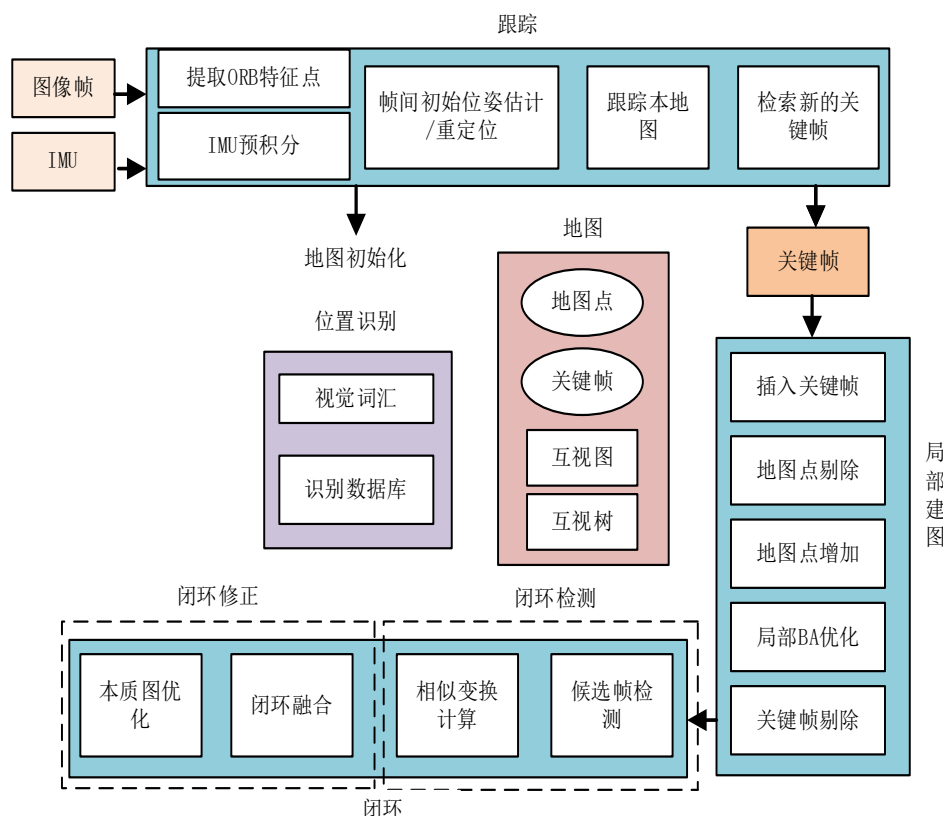


图 1-1 ORB-SLAM3 系统框图

1.2.2 大视角下视觉 SLAM 研究现状

ORB-SLAM3 是当前最先进的基于特征点的视觉 SLAM 系统。该系统采用 ORB 特征描述子进行特征提取与匹配，通过关键帧技术和词袋模型实现运动估计与闭环检测。相比前代版本，ORB-SLAM3 优化了多地图融合机制，显著提升了系统的鲁棒性和定位精度，已成为最广泛应用的视觉 SLAM 解决方案^[22]。然而，在实际应用中，受相机帧率 and 环境因素影响，系统所使用的相机传感器难以实现较高的拍摄帧率，在相机快速运动中易出现模糊图像帧。这些模糊图像帧通常包含较大的像素误差，影响了真实物体图像局部纹理的灰度值梯度，使传统的特征提取算法因图像特征信息缺乏而出现难以提取大量的图像特征点的问题，进而对后续的视觉 SLAM 系统特征跟踪与匹配线程造成了很大干扰，严重影响了特征提取与匹配的准确性和算法的鲁棒性。因此，解决大视角下图像运动模糊问题成为了视觉 SLAM 领域的研究重点^{[23][24]}。

2017 年由 Pumarola 等人提出的 PL-SLAM^[25]创新性地将点线特征相结合，有效改善了传统视觉 SLAM 在低纹理环境中因特征点不足导致的性能下降问题。该算法通过同时利用点特征和线特征，显著提升了系统在稀疏场景下的鲁棒性，减少了相机轨迹漂移现象。然而，PL-SLAM 在处理大视角场景时存在明显缺陷，其过于严格的关键帧选取策略容易造成跟踪线程中关键帧丢失或数量不足，进而影响相机的局部位姿估计精度，导致重建轨迹出现断裂。值得注意的是，由于算法框架本身的局限性，单纯优化关键帧选取策略难以从根本上解决这一问题。同年 Lee 等人提出反卷积去模糊 SLAM 方法，该算法利用相机多视图几何与空间 3D 点来估计图像模糊核，再通过反卷积技术提高特征跟踪的能力和地图构建的丰富度，解决了视觉 SALM 系统如何处理模糊图像的问题，但是该算法也有本身的局限性，在面对图像出现旋转模糊和图像出现非均匀模糊，该算法的性能显著下降，甚至可能完全失效。Russo 等人^[26]提出了一种基于卡尔曼滤波的去模糊 SLAM 算法，该方法基于 Mono SLAM，通过在系统中引入卡尔曼滤波来对图像的模糊核进行处理，并将模糊区域进行特征匹配，虽然该方法能够提高视觉 SLAM 处理模糊图像的鲁棒性，但是该算法假设的模糊场景是线性不变的，不符合现实情况。2019 年，Yu 等人为了视觉 SALM 在低纹理和模糊环境中鲁棒性较低的问题，其提出一种基于 Haar 小波变换的去模糊 SLAM 算法，该算法结合 Haar 小波变换和特征匹配相关性加权优化，但是，在面对大视角环境下由于视角变化过大该算法容易出现跟踪丢失的现象。

随着深度学习在图像领域的迅速发展，一些研究人员提出了基于神经网络的 SLAM 算法，以应对图像处理中的挑战。例如，李维鹏等人提出一种基于场景显著性区域的卷积神经网络框架，为解决视觉 SLAM 系统中闭环检测时视角方向不一致的问题提供了一种创新思路。虽然该算法通过利用深度学习提取显著性区域的特征信息提高了闭环检测的鲁棒性和准确性，但其计算复杂度和对训练数据的依赖仍需进一步的优化。Tang 等人^[27]提出了基于神经网络的 GCNv2-SLAM 算法，该算法使用卷积网络生成二进制描述符向量代替 ORB 特征，提高了图像特征计算的效率，降低了算法对外界纹理特征的敏感度，增强了算法的鲁棒性。另外，Guo 等人首次提出一种将 DeblurGANv2 与深度学习相结合的视觉 SLAM 算法，该算法能够有效地减轻运动模糊对图像质量的影响并且通过改进特征点的

提取和匹配过程，提高了对模糊图像的处理能力，增加了可用于相机位姿估计的特征点数量，从而提高相机位姿的准确度。然而，该算法受限于 DeblurGANv2 的网络结构，导致该算法在光照变化场景下图像的灰度值发生显著的变化，使特征算子难以稳定检测特征点、描述符之间的相似程度降低，图像之间匹配错误率大幅增加。Li 等人提出的 DX-SLAM^[28]是一种基于卷积神经网络的视觉 SLAM 系统，通过卷积神经网络 CNN 来对图像进行提取特征信息，从而使生成的词袋来进行闭环检测，增强 DX-SLAM 在复杂的环境中实现高精度的定位和地图构建，尽管该算法通过深度学习技术提升了系统在复杂环境中的鲁棒性和精度，但是，在大视角快速运动下容易出现特征匹配差，相机运动的轨迹差等问题。

1.3 研究内容与章节安排

本文主要研究移动机器人在视角转变较大情况下，视觉 SLAM 系统性能降低问题并提出一些改进的方法，研究基于改进 MSS 网络与误匹配剔除的 VSLAM 算法，以及结合改进关键帧选取策略的视觉 SLAM 算法。本文框架如图 1-2 所示。

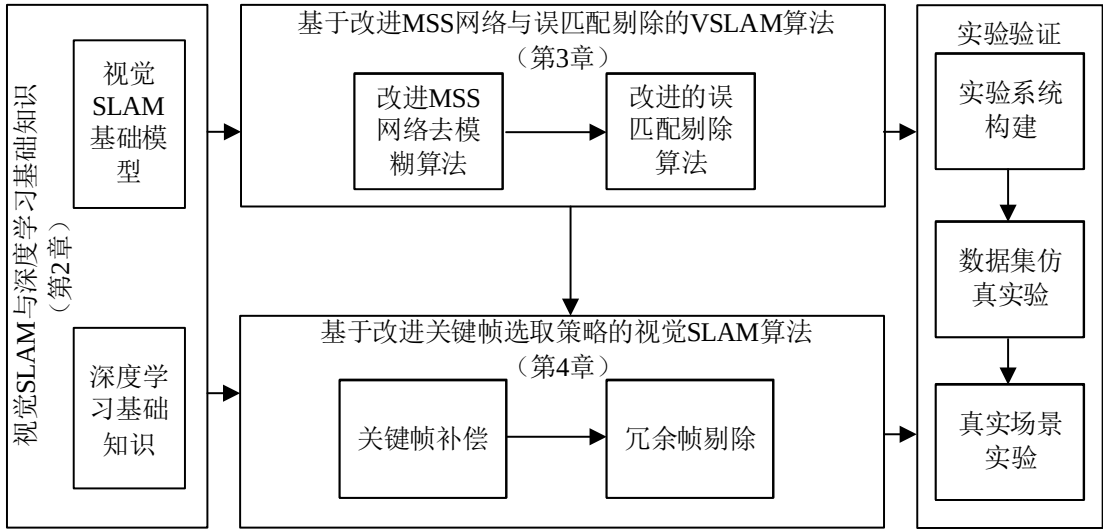


图 1-2 论文总体技术总线

本文研究内容总共五个章节，其安排如下：

第 1 章：绪论。介绍研究的背景、意义以及现状，首先，先介绍移动机器人技术的发展和应用领域，并着重介绍移动机器人的视觉 SLAM 技术的发展、创

新以及实际应用，简单分析了国内外视觉 SLAM 发展的现状，旨在介绍大视角下视觉 SLAM 算法的研究现状，最后对本论文研究的内容进行详细的安排。

第 2 章：视觉 SLAM 与深度学习的基础理论。首先对视觉 SLAM 系统的首先分别对视觉 SLAM 系统的传感器数据、视觉里程计、后端优化、闭环检测和建图这几部分的基础理论和流程进行分析和模型构建。最后对深度学习的基础知识和算法进行简要分析，为接下的研究奠定坚实的基础。

第 3 章：基于改进的 MSS 网络与误匹配剔除的 VSLAM 算法。本章针对移动机器人在大视角运动下图像模糊的问题，设计了一种基于改进 MSS 网络的去模糊算法以恢复高质量的清晰图像。该网络对图像进行下采样操作得到不同尺度的图像，通过特征注意模块增强图像特征信息提取的能力，随后利用监督注意力模块对各阶段的特征信息进行监督和融合，使模糊图像在逐步恢复过程中保持有效信息传递，恢复为原尺寸的清晰图像。针对 VSLAM 算法在大视角场景下特征匹配准确率低的问题，本章提出了一种改进的 GMS 算法，该方法通过局部信息熵对图像进行区域划分，然后根据对极几何约束判断相机是否发生大视角变换，根据视角变换大小分别采用十字五宫格和对角五宫格统计方法进行误匹配剔除。

第 4 章：基于改进关键帧选取策略的视觉 SLAM 算法。针对传统视觉 SLAM 算法在大视角场景下因关键帧选取不准确导致的定位精度下降问题，本章提出了一种基于改进关键帧选取策略的视觉 SLAM 算法。首先，通过计算相邻帧之间旋转矩阵与平移向量，构建共视区域，并结合特征点在连续帧中的跟踪帧数筛选出稳定的有效特征点，为关键帧选取提供可靠判断依据；其次，利用共视区域占比率与运动连续性相结合，建立关键帧选取条件，从而有效避免因视角变换而导致的误判；最后，通过融合几何微分熵与信息熵，构建熵增益函数对新选取的关键帧进行信息评估，实现冗余关键帧的高效剔除。在保证系统实时性的同时，进一步优化关键帧选取策略。

第 5 章：全文总结展望。对本文所有的研究内容进行分析与总结，并展望研究过程中新出现的待解决问题。

第 2 章 视觉 SLAM 与深度学习基础理论

本章主要介绍经典视觉 SLAM 算法和深度学习的基础理论。首先对视觉 SLAM 传感器数据、视觉里程计、后端优化、闭环检测和建图这五个部分的基础知识进行介绍，着重分析了传感器数据、视觉里程计中的特征提取和匹配，最后对深度学习的基础理论和算法进行阐述，包括卷积神经网络、U-Net 和去模糊网络，为大视角下视觉 SLAM 的研究奠定基础。

2.1 视觉SLAM系统框架

随着视觉 SLAM 的快速发展，其整体框架已经趋于成熟，主要包括传感器数据、视觉里程计、后端优化、闭环检测和建图这五大模块。传感器数据通过采集周围环境信息，经过预处理输入视觉 SLAM 系统中。视觉里程计是视觉 SLAM 的前端部分，通过跟踪相机运动轨迹，实现连续帧之间的特征匹配与位姿估计。后端优化是视觉 SLAM 的核心，根据视觉里程计输出的位姿和地图信息进行全局优化，从而减少累计误差，提高位姿估计准确性和地图构建的精度。闭环检测通过识别当前环境与历史场景的相似程度，从而校正移动机器人的位姿和轨迹。本章节主要对传感器数据和视觉里程计进行重点介绍，并对后端优化、闭环检测及建图进行简要分析。视觉 SLAM 基础框架如图 2-1 所示。

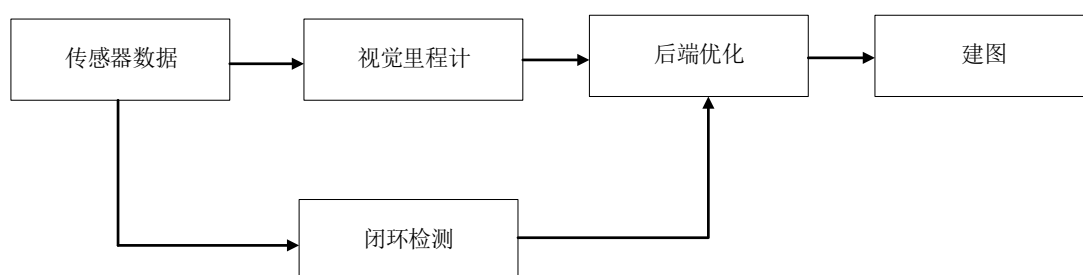


图 2-1 视觉 SLAM 基础框架图

2.1.1 传感器数据

视觉 SLAM 系统中，传感器通常固定在机器人本体上，主要用于捕捉外部环境的视觉信息。根据工作原理的不同，相机主要分为三类：单目相机、双目相机和 RGB-D 相机。单目相机仅配备一个摄像头，结构简单且体积小巧，但其只

能捕获二维图像，无法直接获取场景的深度信息。双目相机则配备了两个摄像头，通过左右图像的视差计算物体的距离，从而间接获取深度信息，但其配置和标定过程较为复杂，且深度测量范围受限于相机基线长度和分辨率。RGB-D 相机则是一种能够同时捕获彩色图像和深度信息的传感器^{[29][30]}，与传统相机不同，其能够直接提供每个像素点到物体表面的距离信息。RGB-D 相机的深度信息通常通过红外技术获取，常见的方法包括结构光法、飞行时间法和双目立体视觉等。这些技术使得 RGB-D 相机在深度感知方面具有显著优势。

视觉 SLAM 系统通常利用相机将三维坐标系的坐标点映射到二维坐标系中一般使用几何投影模型，最经典的投影模型就是针孔模型，利用针孔背面投影成像的原理将坐标进行投影，因此下面通过介绍针孔投影模型来阐述相机投影模型^{[31][32]}。在视觉 SLAM 系统中可以将每帧图像看成不同坐标信息的像素点组成，相机投影过程是将三维的坐标点投影到二维坐标系。针孔相机投影模型如图 2-2 所示。

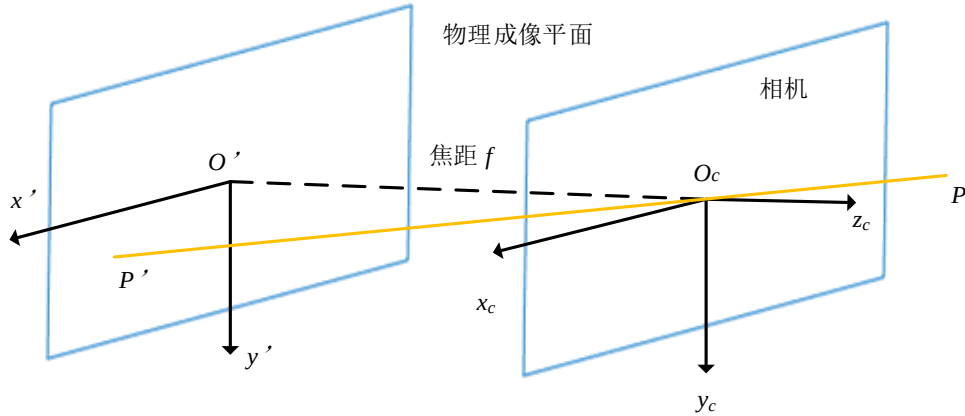


图 2-2 针孔成像模型

三角相似的几何关系公式如 (2-1) 所示：

$$\frac{Z_c}{f} = \frac{Y_c}{y} = \frac{X_c}{x} \quad (2-1)$$

通过式 (2-1) 可以得到每个像素的坐标，假设物理成像平面有一个像素坐标系 $o-u-v$ ，像素坐标系和物理成像平面之间具有几何关系^[33]，像素坐标在 u 、 v 轴上缩放了 α 和 β 倍，从而原点坐标也随之进行了改变，相较于之前原点平移了 $[c_x, c_y]^T$ 。为了计算方便，对 p 点像素坐标引入其次坐标，可得公式如 (2-2) 所示：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_a} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \frac{1}{Z_a} KP \quad (2-2)$$

其中， 3×3 矩阵为相机的内参数 K ， $f_x = \alpha f$ ， $f_y = \beta f$ 。

为了让移动机器人运行时能够更好的确定像素坐标系下像素点的坐标，将世界坐标系下的点 $P_w[X_w, Y_w, Z_w]^T$ 带 Z_w 入公式 (2-2) 中可得公式如 (2-3) 所示：

$$\begin{cases} Z_a P_{uv} = K T_1 P_w \\ T_1 = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \end{cases} \quad (2-3)$$

其中， R 、 t ，表示为相机在此刻的旋转矩阵和平移向量，通过这两个参数可以得到相机的位姿， T_1 表示相机的外参矩阵。

2.1.2 视觉里程计

视觉里程计是视觉 SLAM 系统的核心模块之一，其主要功能是通过分析传感器采集的连续图像帧之间的特征匹配信息，估计相机的运动状态^{[34][35][36]}。作为系统的前端模块，视觉里程计无需依赖预先构建的地图，而是通过计算相邻帧之间的相对运动，为后端优化提供初始的位姿估计值。根据实现原理的不同，传统视觉里程计可以分为特征点法和直接法。相比较于直接法，特征点法提取的特征点具有尺度不变性和旋转不变性，因此特征点法依旧是目前 SLAM 系统前端的主流算法^[37]。

特征点是图像中具有显著性和代表性的关键点，理想的特征点通常具备可重复性、高效性和局部性等特点。即使相机发生旋转或距离变化，特征点仍能保持稳定，从而在不同视角下被准确识别和匹配。在计算机视觉领域，研究者们提出了多种用于表示局部图像特征的特征点算法，例如 SIFT、FAST 和 ORB 等。其中，ORB 特征点因其提取速度快，对光照、尺度和旋转变化的鲁棒性，成为当前视觉里程计中最常用的特征点算法^{[38][39]}。

通过计算相邻图像帧中特征描述子之间的距离或相似性度量，可以实现特征点的匹配。在成功匹配的特征点集合基础上，利用特征点的空间位置信息，可以进一步估计相机的位姿^[40]。根据所使用的相机类型和获取的数据形式，位姿估计

方法通常分为三类：2D 点到 2D 点的对极几何法、3D 点到 3D 点的 ICP 法以及 3D 点到 2D 点的 PnP 法^{[41][42]}。其中，重投影误差和对极几何是视觉 SLAM 中的重要概念，不仅用于相机位姿的求解，还作为几何约束条件帮助判断相机的运动状态^[43]。因此，本节将重点介绍对极几何法。

对极几何通过分析特征点的匹配关系和几何约束得到同一空间点在不同成像平面上的投影规律^[44]。当相机从不同的角度拍摄两帧图像 I_1 和 I_2 ，相机的中心位置分别为 O_1 和 O_2 。空间中的某一点 P 在图像 I_1 上有一个与之对应的点 p_1 ，同理，在图像 I_2 上有一个点 p_2 ，这两个特征点 p_1 和 p_2 通过特征匹配确定为一对匹配点，向量 $\overrightarrow{O_1P_1}$ 和 $\overrightarrow{O_2P_2}$ 在三维平面中相交于同一个空间点 P ，相机中心 O_1 和 O_2 的连线于 I_1 和 I_2 的交点分别为极点 e_1 和 e_2 ，极点于 p_1 和 p_2 的连线 l_1 和 l_2 称为极线。对极几何模型如图 2-3 所示。

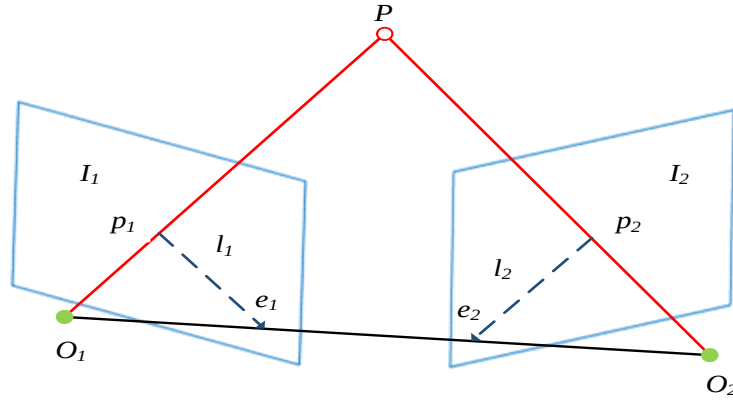


图 2-3 对极几何约束

假设三维空间点 $P = [X, Y, Z]^T$ 在第一帧图像中的投影为 $p_1 = [x_1, y_1]^T$ ，第二帧图像投影的坐标为 $p_2 = [x_2, y_2]^T$ 。根据相机的投影模型和特征点匹配对关系，可得公式如 (2-4) 所示：

$$\begin{cases} s_1 p_1 = K P \\ s_2 p_2 = K (R P + t) \end{cases} \quad (2-4)$$

其中， K 是相机的内参矩阵， R 和 t 分别表示相机从第一帧到第二帧的旋转矩阵和平移向量。根据公式 (2-4) 可以得到对极几何约束的数学表达式，可得公式如 (2-5) 所示：

$$p_2^T K^{-T} t^{\wedge} R K^{-1} p_1 = 0 \quad (2-5)$$

为了简化表达和估计相机位姿，通过引入本质矩阵 E 和基础矩阵 F 来估计对上述公式（2-5）进行变形，可得公式如（2-6）所示：

$$\begin{cases} E = t^{\wedge} R \\ F = K^{-T} E K^{-1} \end{cases} \quad (2-6)$$

2.1.3 后端优化

在 SLAM 系统中，为了能够得到相机初始的位姿信息，前端视觉里程计进行相邻帧之间的特征匹配和运动估计。然而，由于传感器噪声、特征匹配误差以及运动模型的局限性，视觉里程计所估算的位姿误差会不断累积，导致地图构建的精度产生显著影响。为了解决这个问题，SLAM 系统引入了后端优化模块，通过对全局模型进行构建使前端计算的位姿进行校正，从而提高系统的整体精度和鲁棒性。后端优化方法主要分为两类：基于滤波的方法和图优化的方法^[45]。早期 SLAM 系统多采用基于滤波的方法，这类方法通过不断更新状态估计，但这种方法仅仅考虑相邻帧之间的数据关系，适合对实时性要求较高的场景。随着视觉 SLAM 系统对定位精度和全局一致性的不断增加，基于滤波的方法逐渐暴露出局限性，尤其是在处理大规模环境时，其计算复杂度和内存消耗会显著增加。

图优化方法通过调整节点的取值，在满足约束条件的同时，最小化成本函数，从而获得全局最优解。图优化方法凭借其强大的全局优化能力和灵活性，逐渐成为视觉 SLAM 后端优化的主要算法。

2.1.4 闭环检测

在 SLAM 系统中，尽管后端优化能够显著减少位姿估计的累积误差，但由于传感器噪声、运动模型的不精确以及环境复杂性等因素，优化后的位姿仍可能与真实位姿存在偏差。为了进一步提高系统的全局一致性，SLAM 引入了闭环检测模块。闭环检测通过判断机器人是否返回到先前访问过的位置，并在检测到闭环时，利用历史信息对当前位姿和地图进行校正，从而有效减少漂移误差，提升地图的精度和一致性^{[46][47]}。如图 2-4 所示，图 2-4（a）展示了未检测到闭环的情况，由于系统无法识别机器人回到同一地点，导致轨迹误差逐渐累积；而在图 2-

4 (b) 中，系统成功检测到闭环消除了累积误差，显著提高了地图的一致性和准确性。

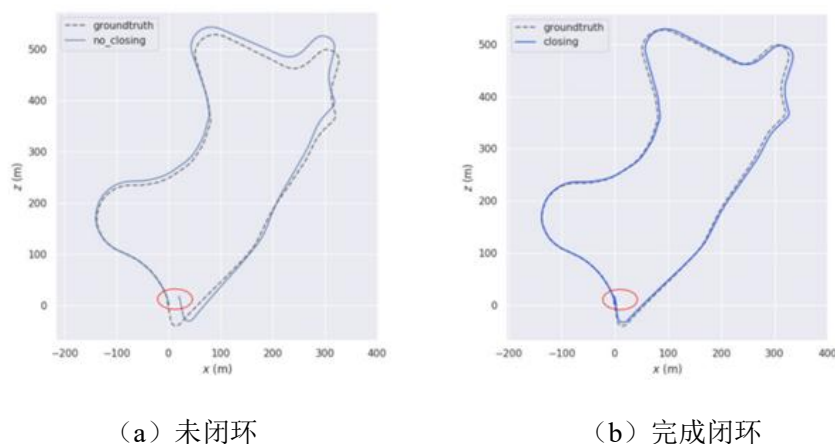


图 2-4 闭环检测

2.1.5 建图

建图是 SLAM 系统的核心目标之一。无论是前端视觉里程计还是后端优化，其最终目的都是为了更精确地确定路标点的位置，从而构建出环境的地图。地图的质量和表达形式直接影响 SLAM 系统的性能和应用场景^{[48][49][50]}。在视觉 SLAM 中，建图的主要目的是服务于定位，但在实际应用中，地图的功能远不止于此，SLAM 系统的设计需要根据具体应用场景的需求，在地图的稀疏性、精度和实时性之间找到平衡。根据地图的密度和表达方式，地图可以分为稀疏地图和稠密地图两大类。稀疏点云地图和稠密建图在表达形式和信息密度上存在显著差异。稀疏地图更适合实时性要求高的任务，而稠密地图则更适合需要高精度环境建模的场景。如图 2-5 所示。

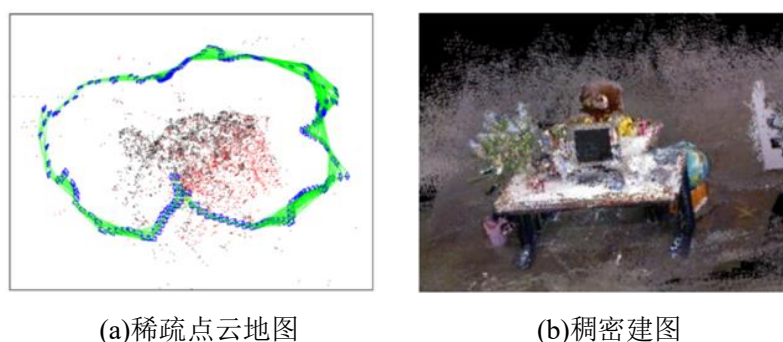


图 2-5 地图类型

2.2 深度学习基础结构

深度学习作为人工智能的一个重要分支，自 20 世纪 80 年代被提出以来，经历了多个发展阶段，并在近年来取得了显著的进展。深度学习的核心在于通过大规模神经网络自动从数据中学习复杂的特征表示，从而实现高效的任务处理。其成功依赖于强大的计算资源、丰富的数据集以及先进的算法设计。随着技术的不断进步，深度学习在各个领域的应用前景也愈发广阔。下面主要对卷积神经网络、U-Net 基本结构和去模糊算法进行介绍。

2.2.1 卷积神经网络

卷积神经网络（Convolution Neural Network，CNN）是一种用于处理网络结构数据的深度学习模型。CNN 通过模仿人类视觉系统的工作方式，能够自动学习图像中的层次化特征，并广泛应用于计算机领域，如图像分类、特征提取、目标检测和语义分割等任务。卷积神经网络主要由卷积层、激活函数、池化层和全连接层组成，每个卷积层后面都连接一个激活函数，全连接层位于卷积神经网络的末端，卷积神经网络结构图如图 2-6 所示。

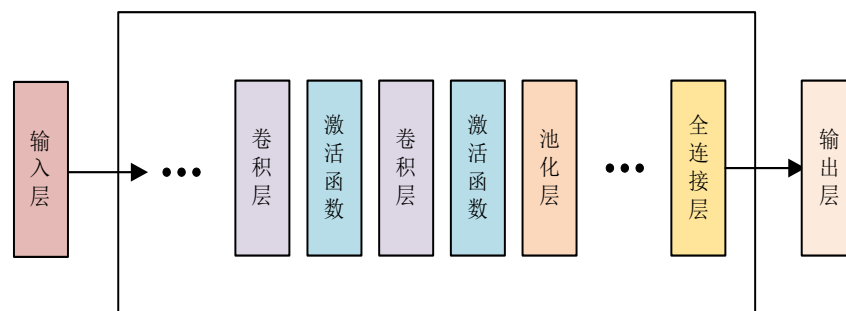


图 2-6 卷积神经网络结构图

卷积层是卷积神经网络中核心模块，负责从输入数据中提取特征。卷积层可以通过使用不同尺寸的卷积核，可以捕捉到不同尺度的特征，除了卷积核尺寸外，卷积层的数量极大的影响了特征提取的能力，浅层卷积层位于网络的前端通常复杂提取基础特征那个，如颜色、纹理和简单的几何形状，深层卷积层能够学习到更加抽象和复杂的特征，通过多不同尺寸的卷积核和多层次的卷积层设计，提高

了卷积神经网络的特征提取能力。采用 3×3 的卷积核，步长为 1，卷积运算过程如图 2-7 所示。

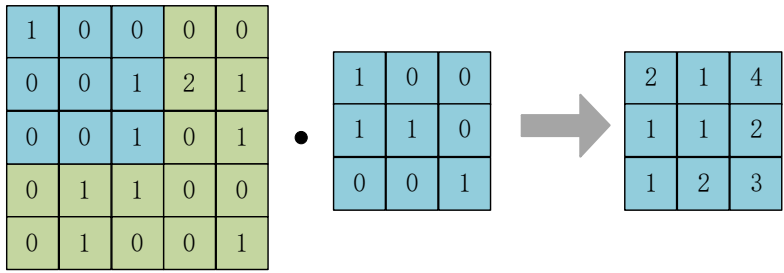


图 2-7 卷积运算过程

激活函数在卷积神经网络中起着至关重要的作用，在每个卷积层后面都会使用激活函数，用于引入非线性因素到模型中争抢模型的表现能力，激活函数能够使卷积神经网络学习和表示复杂的非线性模式，而不是仅仅简单的线性变换。例如，在图像识别任务中，低级特征如边缘和纹理是线性的，但更高级的特征如形状、对象类别等则需要非线性映射才能够有效捕捉。其次，激活函数增强了模型的层次化特征提取能力。每一层卷积之后应用激活函数，使该层能够学习到更加复杂的特征表示。常见的激活函数包括 ReLU、Sigmoid、Tanh 等，特别是 ReLU 函数因其简单性和高效性在卷积神经网络中被广泛使用，这三种常见的激活函数公式如 (2-7)、(2-8) 和 (2-9) 所示：

$$\text{ReLU}(x) = \max(0, x) \tag{2-7}$$

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{2-8}$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2-9}$$

池化层是卷积神经网络中的一个重要组成部分，通常位于卷积层之后，其主要作用是对输入数据进行下采样，从而减少特征图的空间维度。池化操作分成最大池化和平均池化两大类。最大池化选取局部区域内最大值，保留关键特征信息，因为最大值往往表示该区域内最显著的特征。平均池化通过对每个局部区域取平均值作为输出，能够得到较为平滑的特征图，减少噪声的影响。采用步长为 2 和 2×2 的卷积核对输入的特征图进行池化操作。最大池化和平均池化运算过程如图 2-8 所示。

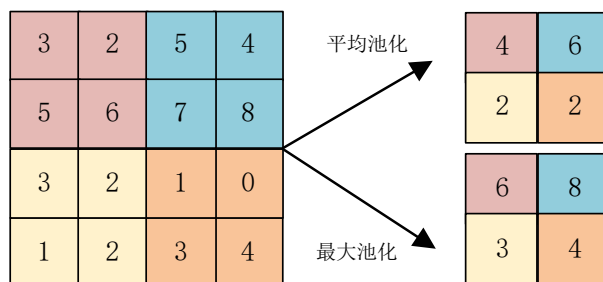


图 2-8 池化运算过程

2.2.2 U-Net 基本结构

U-Net 网络由 Olaf Ronneberger 等人于 2015 年提出，最初是为解决生物医学图像分割任务而设计的深度学习架构。该网络凭借其独特的 U 型对称结构和创新的特征融合机制，不仅在医学图像处理领域取得突破性成果，也逐渐扩展到遥感图像分析、自动驾驶场景理解等多个计算机视觉领域。

其网络架构采用经典的编码器-解码器设计，编码器部分通过多级卷积和下采样操作逐步提取图像的深层语义特征。每一编码层级包含两个连续的 3×3 卷积层用于特征变换，后接 2×2 最大池化层实现空间维度缩减。随着网络深度的增加，特征图分辨率逐渐降低而通道数相应增加，最终形成包含丰富全局语义信息的低维特征表示。解码器部分通过系统的上采样和特征融合操作实现精确定位。每个解码层级首先采用 2×2 转置卷积进行特征图上采样，随后通过跳跃连接将对应编码层的高分辨率特征与上采样特征进行通道拼接。这种跨层特征融合机制能够将编码阶段捕获的细粒度空间信息与解码阶段恢复的语义信息有机结合，有效解决了传统编解码网络中存在的细节丢失问题。每个融合后的特征再经过两个 3×3 卷积层进行特征精炼，逐步重建出具有精确边界的分割结果。

U-Net 的核心创新在于其密集的跳跃连接设计，这些连接在编码器和解码器对应层级之间建立起直接的信息通路。通过将早期编码层的高分辨率、低层次特征与后期解码层的高层次语义特征进行多尺度融合，网络能够在保持全局上下文理解的同时，精确恢复目标的局部细节和空间结构。这种独特的架构设计显著提升了网络对复杂图像模式的建模能力，特别是在需要精确定位的像素级预测任务中展现出卓越性能。

2.2.3 去模糊算法

去模糊技术旨在从模糊图像中恢复清晰的图像，是计算机视觉和图像处理领域的重要研究方向。随着深度学习的发展，基于神经网络的去模糊方法逐渐成为主流，并取得了显著的进步。在深度学习兴起之前，去模糊技术主要依赖于传统的图像处理技术，这些方法通常基于频域变换或优化算法。去模糊网络通过多层卷积神经网络逐步提取图像中的低级特征如边缘、纹理等，并重建高级语义信息。例如，SRN 将卷积神经网络和循环神经网络结合，以处理图像的不同尺度信息。该网络首先通过多尺度特征提取模块，提取不同尺度上的模糊图像特征图。随后，引入空间递归机制，使不同尺度之间能够传递和更新信息，从而在保证去模糊效果的同时，降低了计算的复杂度。该方法有效增强模糊图像的清晰度，使原本模糊的部分变得清晰、边缘锐利和细节丰富。并且 SRN 还能够不同尺度上聚焦于关键区域，提升去模糊的效果的同时应对复杂的现实场景。SRN 去模糊效果如图 2-9 所示。

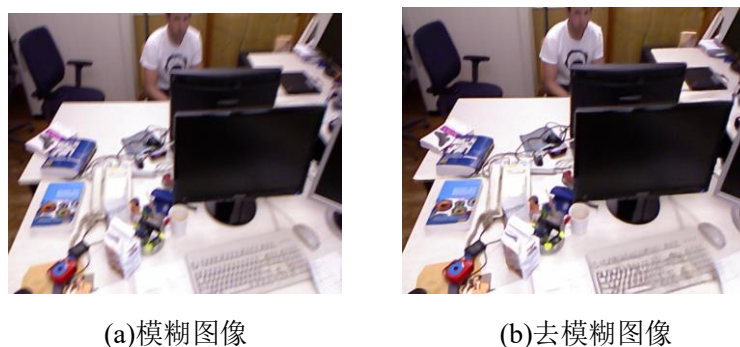


图 2-9 去模糊前后对比图

2.3 本章小结

本章主要介绍了视觉 SLAM 与深度学习的基础知识。将视觉 SLAM 模型按定位和建图分别进行数学描述并介绍了整体框架，在大框架下分别介绍了传感器数据、视觉里程计、后端优化、闭环检测和建图五个部分。最后介绍了深度学习的基础理论，并分析了基于深度学习的卷积神经网络、U-Net 结构、去模糊算法，为深度学习和视觉 SLAM 的结合提供了理论基础。

第3章 基于改进 MSS 网络与误匹配剔除的 VSLAM 算法

针对 VSLAM 算法在大视角运动场景下出现图像模糊和特征匹配准确率低的问题，本章构建了一种基于改进多尺度多阶段（Multi Scale Stage，MSS）网络与误匹配剔除的 VSLAM 算法。首先，对输入图像进行特征响应方差计算评估其模糊程度。若该值较大，则表明图像存在较为明显的运动模糊，采用改进的 MSS 网络进行去模糊处理，输出去模糊图像到特征匹配环节，反之则直接进入特征匹配。在特征匹配时通过对极几何约束分析相机视角变换，若相机视角变化较小（角度 θ 小于等于 45° ）时，采用十字五宫格统计法进行误匹配剔除，当检测到存在大视角变换时（角度 θ 大于 45° ）时，则采用对角五宫格统计方法进行误匹配剔除，使其能更好的适应大视角情况下的特征匹配关系，并通过优化后的匹配点对后续进行位姿估计，提升 VSLAM 系统整体定位精度。整体框架如图 3-1 所示。

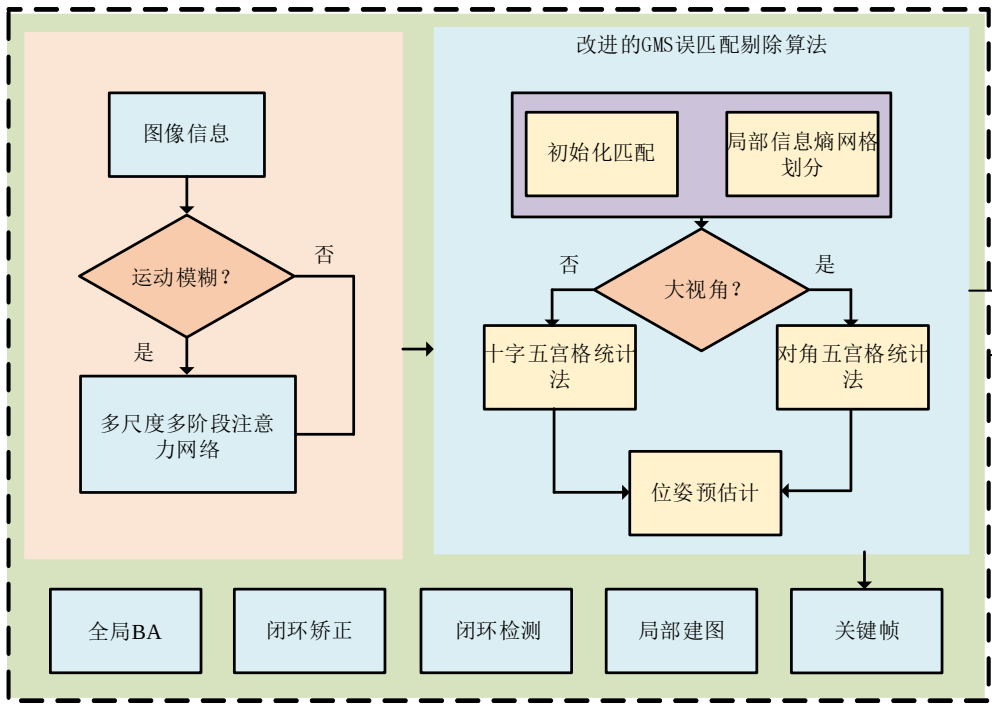


图 3-1 本章算法框架图

3.1 改进MSS网络的去模糊算法

针对传统算法在大视角场景下采集的图像易产生模糊的问题，本章在

MSS-Net 网络的基础上进行改进，通过构建更深层次的多尺度信息表达结构，以提升复杂场景下的图像去模糊性能。改进的 MSS 网络采用三级尺度架构，其中 S_3 为原始图像， S_2 为原始图像的 1/2 下采样， S_1 为原始图像的 1/4 下采样。网络在不同尺度上使用渐进式处理，分别包含 1、2 和 3 个处理阶段，每个处理阶段均由轻量级编码器-解码器结构的 U-Net 模块构成，记作 U_m^n ，其中上标 n 表示为尺度索引，下标 m 代表阶段索引。U-Net 模块在共享相同的网络架构的同时，又分别具有独立的权重参数，每个 U-Net 模块通过生成残差特征图来补偿模糊信息，并与模糊图像进行融合，从而逐步重建清晰图像。网络结构如图 3-2 所示。

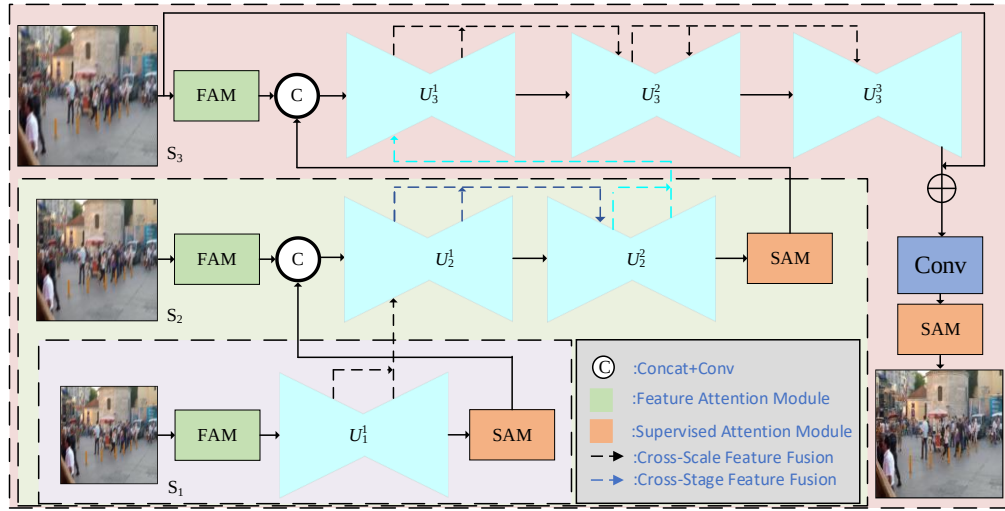


图 3-2 改进的 MSS 去模糊网络结构

为增强去模糊网络的特征提取能力，在网络各尺度的初始阶段引入了特征注意力模块（Feature Attention Module, FAM），该模块与 U-Net 模块级联并形成完整的处理单元。此外，在不同尺度的末端引入了监督注意力机制（Supervised Attention Module, SAM），以增强跨阶段之间的特征传递能力，确保网络在不同尺度上的信息能够更好的流通，从而提升网络整体性能和计算效率。其中特征注意力模块使用空间注意力与通道注意力相结合的机制，通过自适应地调整模糊物体的像素权重，增强模糊区域特征信息，从而提高对模糊区域的识别准确率。为解决网络间信息衰减问题，在 U-Net 模块内部以及跨模块间引入跨阶段和跨尺度特征融合机制。该机制通过构建多层次的特征连接，确保先前阶段的特征信息能够有效地传递到后续处理阶段，丰富了模糊图像的特征信息。通过多尺度和多阶段的协同处理机制，改进的 MSS 网络能够在太视

角下有效处理图像模糊问题，在保证网络轻量化的同时，显著提升去模糊效果。

3.1.1 特征注意力模块 FAM 结构设计

特征注意力 FAM 由 ResNet101、通道注意力模块和空间注意力模块组成。其中 ResNet101 用于提取原始图像的特征，生成一系列具有丰富语义信息的特征图。通道注意力模块通过全局平均池化和全局最大池化对特征图进行统计分析，自适应地分配各通道的权重以增强关键特征通道的响应，同时使用空间注意力机制通过卷积操作学习特征图在空间维度上的显著性分布，自适应地突出重要的空间信息。

为了确保特征图与输入图像的特征对齐，采用双线性上采样及 1×1 卷积操作，对特征图进行重采样，使其恢复到与输入图像相同的分辨率。通过空间注意力机制进一步调整和增强特征图，以提高网络的表达能力和关注度。通道注意力和空间注意力模块能够灵活地调整特征图的权重和响应，使其获得关键的特征信息。通过双线性上采样和 1×1 卷积等操作对特征图进行处理，使其调整为与输入图像相同大小的特征图，确保了特征图与输入图像的对应关系，有利于特征图与原始图像的对齐。特征注意力模块结合了通道和空间注意力机制，在抑制干扰信息影响的同时，生成更具判别能力的特征表达，突出模糊区域的有效特征信息，从而提高大视角场景下 VSLAM 系统的稳定性和性能。特征注意力模块 FAM 结构如图 3-3 所示。

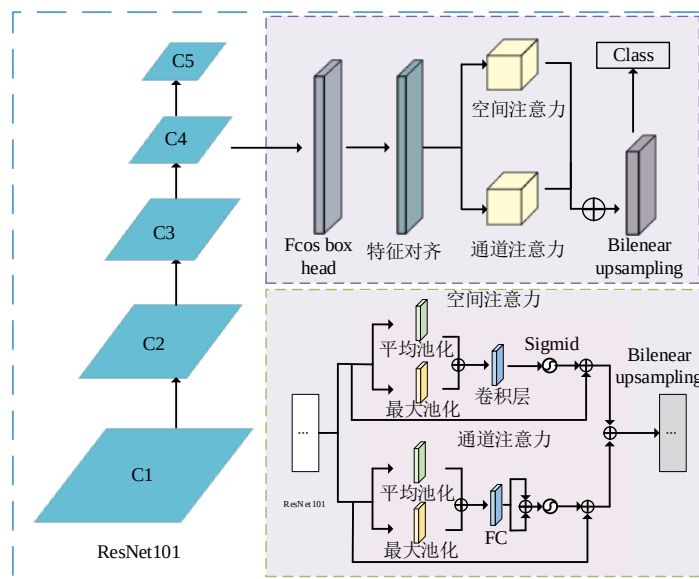


图 3-3 特征注意力模块

首先将由 ResNet101 提取的特征图输入通道注意力机制中，并对其分别进行最大池化和平均池化操作，从而得到特征图的各通道的信息。然后，将得到的结果经过全连接层（Fully Connected, FC）处理，以增加各通道之间的关联性，并对通道权重进行自适应重分配，从而增强网络对关键特征的关注度，提升对模糊特征的学习能力。最终，生成优化后的通道注意力图。

空间注意力机制主要作用为增加特征图中模糊区域像素值的权重，使网络在特征学习过程中将权重值进行更好的调整，从而更加关注受模糊影响的关键区域。空间注意力机制将 ResNet101 得到的特征图作为输入，分别通过最大池化层和平均池化层后，将得到的结果进行通道拼接融合，随后，通过 3×3 的卷积层对融合的特征进行局部感受野建模，以捕捉局部空间特征。最终，经过 sigmoid 激活函数函数得到所需要的空间注意力图。

3.1.2 监督注意力模块 SAM 结构设计

为了有效的传递关键信息，使后续阶段能够充分利用前一阶段的特征并提升去模糊效果，本章在去模糊网络的每个阶段末尾引入了监督注意力模块 SAM，以解决每个阶段图像带来的信息传递不足的问题，SAM 原理图如图 3-4 所示。

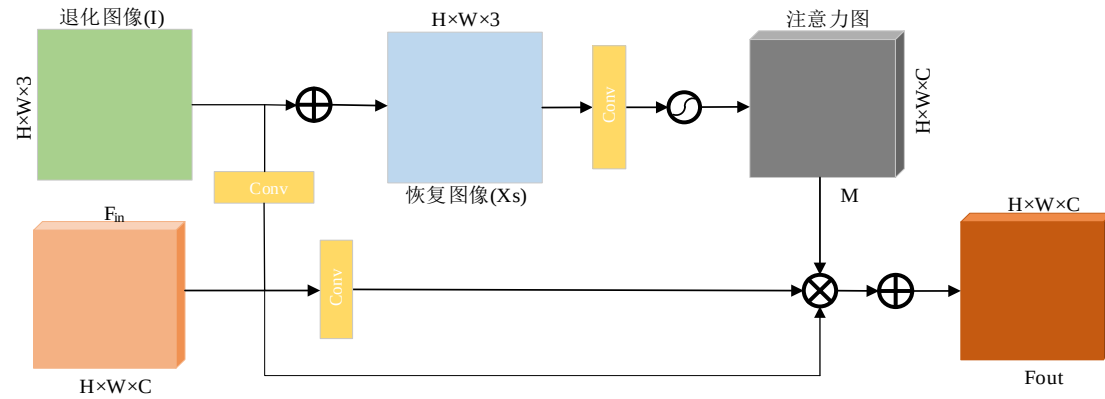


图 3-4 监督注意力模块

SAM 通过对恢复图像的每个阶段提供有效的监控作用，使网络下个阶段可以根据上一阶段的结果进行自适应调整，从而逐渐提升去模糊图像的质量和准确性。其次，通过引入监督注意力模块生成特定的注意力图，使网络关注输入数据中的关键信息，并抑制次要和无关特征，这样不仅提高网络去模糊的效率，还确保每段生成的关键信息能够有效地传递到下一尺度。

监督注意力模块接收由子网络 U-Net 提供的输入特征 F_{in} ，其中 $H \times W$ 表示

空间通道， C 是通道数。首先， F_{in} 首先通过 1×1 的卷积层生成残差图像 R_S ，并将得到的残差图像和退化图像 I 结合，从而得到恢复图像 X_S ，以抑制当前阶段信息不丰富的特征，仅将有用的特征传递给下一阶段。再通过 1×1 的卷积层和 sigmoid 激活函数生成注意力图 M ，该注意力图 M 用于重新校准输入特征图 F_{in} 。经过注意力图 M 的引导下，监督注意力模块对重要区域的关键特征进行保存和增强，确保重要信息得到有效的强化。最终，生成的注意力增强的特征 F_{out} 被传递到下一阶段。

3.1.3 损失函数

在去模糊网络的某一阶段中，模型通过预测残差图像 R_S 对模糊图像进行恢复，而不是直接预测恢复图像 X_S 。将残差图像 R_S 加到退化输入图像 I 上以获得恢复后的图像，整个去模糊网络通过以下损失函数进行端到端的优化。总损失 L 是由三个阶段的损失之和组成的，并且每个阶段的损失包括 Charbonnier 损失函数 L_{char} 和边缘损失函数 L_{edge} ，具体公式如（3-1）所示：

$$L = \sum_{S=1}^3 \left[L_{char}(X_S, Y) + \lambda L_{edge}(X_S, Y) \right] \quad (3-1)$$

其中， Y 为真实图像， λ 是控制两个损失函数重要的参数。

L_{char} 主要作用是用来处理图像重建，衡量预测图像 X_S 和真实图像 Y 之间的像素级差异。相比于传统的均方误差， L_{char} 处理噪声时更加稳健，因为其对误差采取平方根形式，并通过减少极端值对网络的影响，其计算公式如（3-2）所示：

$$L_{char} = \sqrt{\|X_S - Y\|^2 + \varepsilon^2} \quad (3-2)$$

边缘损失函数通过拉普拉斯算子提取图像边缘信息，能够有效的捕捉到图像中的细微结构变化，边缘损失函数对高质量图像的恢复起着至关重要的作用。对于所有的实验，常数 ε 设置为 0.3，边缘损失函数计算公式如（3-3）所示：

$$L_{edge} = \sqrt{\|\Delta(X_S) - \Delta(Y)\|^2 + \varepsilon^2} \quad (3-3)$$

3.2 改进的误匹配剔除算法

GMS 算法是一种基于网格运动统计的快速鲁棒特征匹配方法，用于解决因相机视角剧烈变化导致的相邻帧间特征匹配失效问题。正确的匹配通常需要满足

运动平滑性约束，即正确匹配点对的局部邻域内会聚集大量具有一致运动特性的支撑匹配点，与之相对的是随机错误难以在空间上形成一致性分布，因此错误匹配邻域内则很少有支撑的匹配点。基于这一原理，GMS 通过统计特征点邻域网格内的匹配数量来评估匹配的可靠性，从而提升匹配的鲁棒性。

为提升特征匹配的准确性和效率，本章在 ORB-SLAM3 的特征匹配框架基础上对 GMS 算法进行优化。首先引入基于局部熵的自适应网格划分策略，通过分析图像区域的信息熵动态调整网格尺寸，确保特征点密集区域的边界能被精确划分，从而提高网格统计的准确性。其次，提出改进的五宫格统计策略，同时考虑中心网格及其相邻或对角的网格共五个区域的匹配统计量，并针对大视角变化场景分别设计十字五宫格和对角五宫格两种模式，增强算法对视角变化的适应性。

3.2.1 局部信息熵网格划分

信息熵通常用于描述一个随机变量的不确定度。随机变量是指在一组可能的取值中以一定概率分布出现的变量。信息熵量化了这个随机变量的不确定性。如果一个随机变量的概率分布越均匀，其信息熵就越大；反之，如果概率分布更加集中，那么信息熵就越小。在图像处理方面，图像熵的信息熵能够准确的表示图像上含有的信息量，为了更好对网格划分利用信息熵来对图像上的信息进行筛选，信息熵的公式如（3-4）所示：

$$H(X) = -\sum_{i=1}^n P(x_i) \cdot \log_2 P(x_i) \quad (3-4)$$

其中 $P(x_i)$ 表示随机变量 x 取值为 x_i 的概率， n 为 x 的取值的个数，范围在（0，255）之内。当图像中包含较多的特征信息时，其信息熵较大，表明图像具有丰富的纹理信息丰富和显著的像素信息。反之，图像灰度变化不明显且信息熵较低，说明图像的纹理信息较少且像素分布区域均匀。

当一维熵 $H(X)$ 接近于 0 时，表明灰度变化越小，图像内部的的纹理和结构特征信息缺乏。因此，通过局部信息熵对图像进行划分，对于信息熵较小的地方则进行较大的网格划分，这样网格内的特征点数就会更多且更易于剔除误匹配，对于信息熵叫大的区域则对其进行更细的网格划分，使之前处于网格分界线的特征点也包含于网格内，从而在进行误匹配剔除时能够对特征信息密集的区域更好地判断网格内那些匹配对需要剔除。

3.2.2 五宫格统计法

GMS 算法使用 3×3 的网格代替原本的圆形中心领域。在统计特征点所在网格的特征分数时，还会统计环绕其四周的 8 个相邻网格的特征分数。对于每一对初始匹配点，这一过程有助于更准确地评估匹配的可靠性。对于每一对初始匹配点 $(P_{1,j}, P_{2,j})$ ，都有各自的邻域为 $N_{1,i}$ ， $N_{2,i}$ 。对于每个点 $P_{1,i}$ ，在其邻域 $N_{1,i}$ 都能找到匹配点集合 $\{P_{1,i,1}, P_{1,i,2}, \dots, P_{1,i,w_i}\}$ ，其中 w_i 是这个邻域内粗匹配特征点数量的均值。如果在每个 $N_{1,i}$ 中的特征点 P_{1,i,w_i} ，在 $N_{2,i}$ 中有对应的匹配点 $P_{2,i,1}$ ，则标志分数 $s_{i,m}$ 为 1，反之则为 0，最后通过累加标志分数得到特征分数 S_i ，其计算公式如 (3-5) 所示：

$$S_i = \sum_{m=1}^{w_i} s_{i,m} - 1 \quad (3-5)$$

在保证网格分布的对称性和大视角场景保持鲁棒性的基础上，提出了一种新的计算方法，将 GMS 九宫格分解成十字五宫格和对角五宫格，通过对极几何约束计算相机的角度变化从而使用不同的统计方法来计算网格的特征分数。设定一个阈值 δ ，若相机视角变换大于阈值 δ ，则通过对角五宫格来计算网格的特征分数；反之，若小于阈值 δ 则用十字五宫格。通过这种方法在面对大视角场景时，对角五宫格统计算法可以更好地捕捉到沿对角线方向的运动信息，提高匹配的鲁棒性，对于相机视角大幅度变化的误匹配剔除尤其有利。五宫格只统计当前网格和与之相邻且对角对称的四个网格的特征分数，将这五个网格的特征分数之和称为五宫格特征分数 S ，并且根据每个网格中特征点正确匹配对的数量为网格分配权重。如图 3-5 所示。

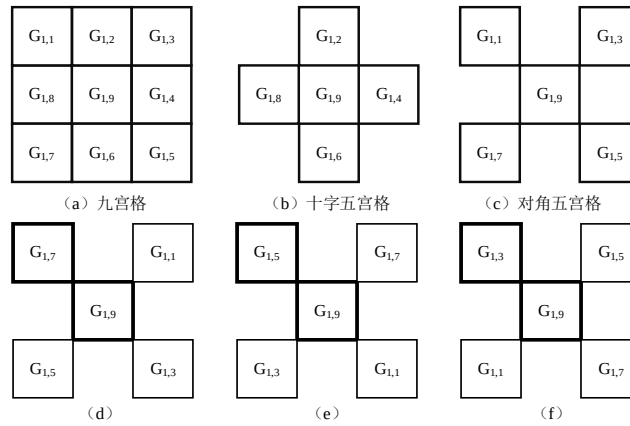


图 3-5 五宫格统计法原理图

在大视角场景下，由于相机视角的变化，会导致图像出现旋转。为此，使用旋转运动核对五宫格进行旋转操作，其状态如图 3-5 (c) ~ (f) 所示。处于五宫格中心的网格为 $G_{1,9}$ ，其相对位置始终不变，而以 $G_{1,9}$ 网格为对角线中心的另外四个网格会因为旋转而发生位置变化。初始状态下，对角五宫格左上角框内的网格为 $G_{1,1}$ ，经过第一次顺时针旋转后，网格由 $G_{1,1}$ 变为 $G_{1,7}$ 。同理，当顺时针旋转第 2 次时，网格由 $G_{1,7}$ 变成了 $G_{1,5}$ ，如图 3-5(e) 所示。但当顺时针旋转到第 4 次时，此时网格的状态恢复至初始未旋转状态。因此，只需统计 4 次不同情况下最大五宫格特征分数。

根据相机角度 δ 的变化，五宫格特征分数计算公式如 (3-6) 所示：

$$S = \begin{cases} \sum_{i=2,4,6,8,9} \sum_{m=1}^{w_i} \alpha_i \cdot (s_{i,m} - 1), & \text{其中 } \theta < \delta \\ \max\{ \sum_{i=1,3,5,7,9} \sum_{m=1}^{w_i} \alpha_i (S_{i,m}^n - 1), n = 1, \dots, 4 \}, & \text{其中 } \theta > \delta \end{cases} \quad (3-6)$$

其中， S 为五宫格的特征分数， w_i 表示五宫格内各网格种粗匹配特征数量的均值， n 代表旋转的次数， α_i 代表第 i 个网格的权重， θ 为相机在运动中所变换的角度， δ 为角度阈值； $s_{i,m}$ 为五宫格中第 i 个网格第 m 个特征点的标志分数， $S_{i,m}^n$ 表示旋转第 n 次的对角五宫格中第 i 个网格第 m 个特征点的标志分数。

网格经过信息熵处理，不同的区域网格划分不同阈值也应该随之改变，所以提出一种新的自适应阈值 T 的计算公式如 (3-7) 所示：

$$T = \frac{\sqrt{k w_i} + c}{1 + e^{H(X_i)}} \quad (3-7)$$

其中， k 是特征点数量均值 w_i 的权重系数， c 是基础阈值， $H(X_i)$ 为第 i 个网格信息熵。

3.3 实验及分析

本章实验主要分为三部分：不同视角下图像去模糊对比实验、不同视角下误匹配剔除对比实验、轨迹对比实验。本章实验所用平台配置为：CPU 为 i5-13400F，主机显卡 3060，内存 16G。在系统为 Ubuntu18.04 的基础上采用 PyTorch 深度学习框架，对 GOPRO 数据集进行去模糊网络模型训练，并使用公开数据集 TUM 与 GOPRO 对本章算法进行验证。

3.3.1 不同视角下图像去模糊对比实验

实验选取 TUM 数据集下大视角图像序列验证本章算法的去模糊性能，并与 DeblurGANv2、MSS-Net 算法对比。其中第一行为参考帧，其余行分别为对应参考帧下因不同角度变换造成的模糊图像，第一列为原始模糊图像，其余列为不同算法得到的去模糊图像，可以看出，随着视角变化越大图像越模糊，图可以看到因视角快速变化而产生的模糊图像在经过处理后得到了显著改善，相较于其他两种算法去模糊效果明显更好，图像局部区域去模糊后更加清晰。如图 3-6 所示。

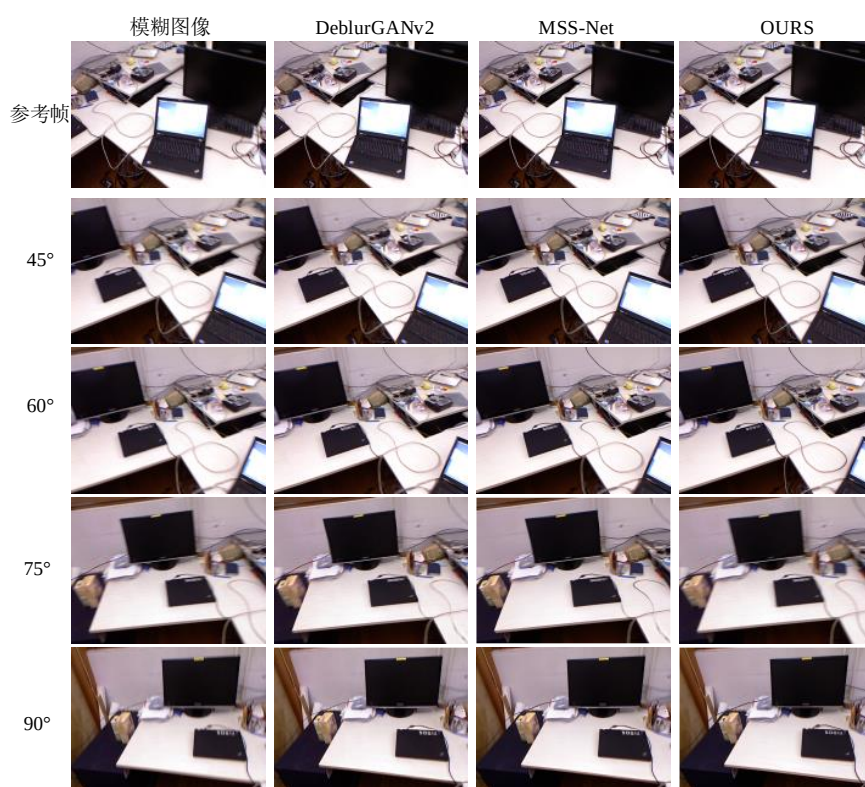


图 3-6 图像去模糊前后效果对比

本章算法根据峰值信噪比 PSNR、结构相似性值 SSIM、Time 这三项指标来表示去模糊算法的性能。由于 TUM 数据集不包含拍摄的模糊图与对应的清晰图，无法获得图像的 PSNR 和 SSIM 的数值。因此使用公共数据集 GOPRO 数据集进行实验并且与其他算法进行对比。

PSNR、SSIM、Time 这三项指标对比结果如表 3-1 所示，实验结果显示，本章算法在 GOPRO 数据集上获得的 PSNR、SSIM 明显高于另外几种算法。其中，相较于 MSS-Net 算法，本章算法的 PSNR 提升 0.75%，SSIM 提升了 1.56%。对于 DeblurGANv2 算法，本章算法 PSNR 和 SSIM 分别提升了 12.59%、2.95%。

实验结果表明，本章算法不仅在图像质量指标上有明显的提升，而且图像处理时间明显低于其他几种算法，展现了所提算法的优越性。

表 3-1 不同算法去模糊对比结果

算法	PSNR(dB)	SSIM	Time(s)
DeblurGAN ^[51]	28.69	0.862	11
DeblurGANv2 ^[52]	29.54	0.948	2.05
SRN ^[53]	30.26	0.934	0.736
MIMO-UNet ^[54]	31.73	0.951	0.263
MSS-Net ^[55]	33.01	0.961	0.255
Our	33.26	0.976	0.247

3.3.2 不同视角下误匹配剔除对比实验

为验证提出的基于改进 GMS 误匹配剔除算法有效性，本节实验选取 TUM 数据集中部分具有大视角变换的序列，并对不同视角下误匹配剔除效果进行对比实验。实验使用描述符平均准确率（mean Average Precision，mAP）作为判断误匹配剔除的性能指标，以量化不同算法在误匹配剔除实验中的准确性和鲁棒性。在大视角变化条件下，将本章提出的误匹配剔除算法与暴力匹配 BF、重投影匹配 Rep 和 GMS 这三种匹配方法进行对比，来验证本章设计的误匹配剔除算法的有效性和鲁棒性。不同视角下误匹配剔除对比实验如图 3-7 所示。

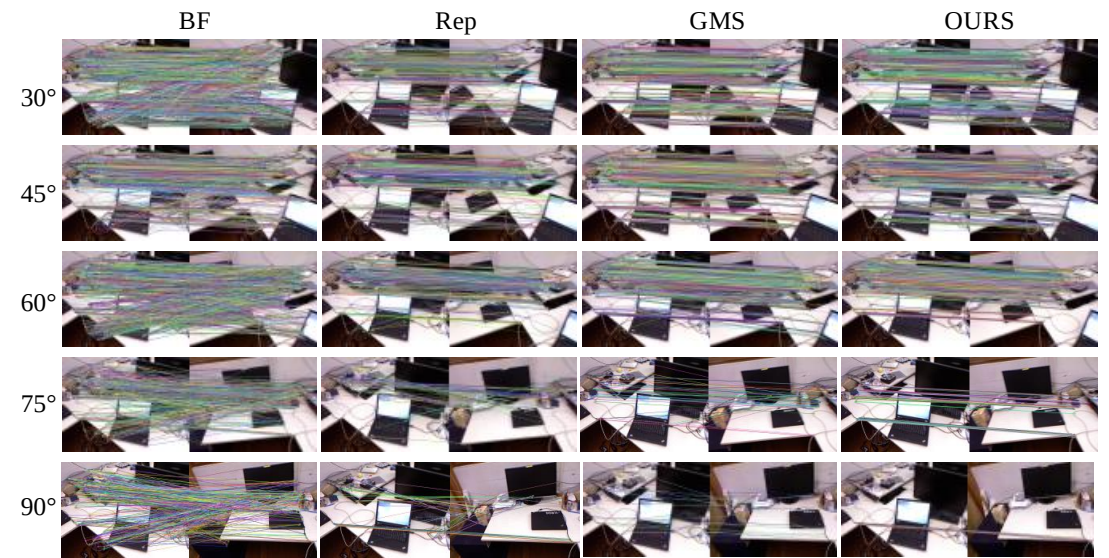


图 3-7 不同视角下不同算法误匹配剔除对比实验

实验结果表明，在不同视角下，本章算法在误匹配剔除方面效果最好，正确匹配对数相比于其他算法明显较多，同时特征匹配准确率更高。尤其是当视角变

化为 75° 时，Rep 和 GMS 算法均出现较多误匹配，而本章算法有效地避免了错误匹配对的产生，从而证明本章算法在大视角变化场景下误匹配剔除的优越性。

不同视角下误匹配剔除性能 mAP 对比结果如表 3-2 所示，该表展示了不同算法在多个 TUM 数据集下不同视角描述符的平均准确率。随着视角增大，所有算法的描述符准确率都普遍下降，并且在同一个序列中，不同视角变化对算法性能也有较大的差异性。其中，Rpe 算法误匹配剔除效果最差，GMS 算法在面对视角变化时误匹配剔除效果不稳定，而本章设计的误匹配剔除方法在不同视角下始终表现出较好的效果和稳定性。实验结果表明，本章节误匹配剔除算法的 mAP 在视角变化为 30° 、 60° 、 90° 场景中与 Rep 算法相比分别提升了 27%、57.8%、97.84%，相较于 GMS 算法分别提升了 16.2%、22.5%、37.4%。实验结果表明，本章算法在不同视角场景下具有较高误匹配剔除性能，角度变化越大误匹配剔除效果更明显。

表 3-2 不同视角下误匹配剔除性能 mAP 对比结果

数据集	RPE			GMS			OURS		
	30°	60°	90°	30°	60°	90°	30°	60°	90°
Fr1/floor	0.593	0.438	0.269	0.652	0.563	0.386	0.759	0.709	0.541
Fr1/desk	0.631	0.487	0.305	0.692	0.598	0.427	0.825	0.736	0.576
Fr1/room	0.579	0.425	0.254	0.621	0.552	0.375	0.716	0.674	0.525
Fr1/desk2	0.615	0.462	0.283	0.685	0.586	0.419	0.784	0.716	0.563
Fr2/desk	0.589	0.421	0.275	0.639	0.579	0.389	0.739	0.689	0.537

3.3.3 轨迹对比实验

为了验证本章算法在大视角场景的性能，从公开数据集 TUM 中选用五个序列进行实验，将实验的结果与 PL-SLAM、OV²-SLAM 和 ORB-SLAM3 这三种算法进行对比。实验通过平均绝对轨迹误差（absolute trajectory error, ATE）来对实验结果进行性能对比，并采用 `evalute_ate.py` 评测脚本对轨迹进行评估。

绿色、红色和黑色线条分别表示为真实轨迹、轨迹误差以及各算法运行的轨迹。从图中可以得到，PL-SLAM 和 OV²-SLAM 在大视角场景中都出现了跟踪丢失的情况。在 Fr1/desk2 序列中，这两个算法运行时由于特征点跟踪丢失导致无

法进行闭环，最终仅生成了部分轨迹。此外，在 Fr1/room 和 Fr1/360 序列中，PL-SLAM 和 OV²-SLAM 的轨迹误差较大，表明这两种算法在大视角场景下的鲁棒性较弱，难以保证精准的位姿估计。相比之下，ORB-SLAM3 和本章改进算法均能在四个序列中完成跟踪，整体性能比另外两种算法高。然而，ORB-SLAM3 在大视角场景下特征匹配较差，导致轨迹误差明显高于本章改进的算法。尤其是在角度变化较大的情况下误差更为显著。本章算法通过改进的 MSS 网络增强特征点在大视角变化下的鲁棒性，使其能够稳定匹配。同时通过改进的误匹配剔除算法有效的剔除了错误的匹配，提高特征点的跟踪的准确性。实验结果表明，本章改进的算法，能够更好的适应大视角场景，使特征点能够稳定跟踪，并获得了较低的绝对轨迹误差。不同算法在 TUM 数据集的运行轨迹如图 3-8 所示。

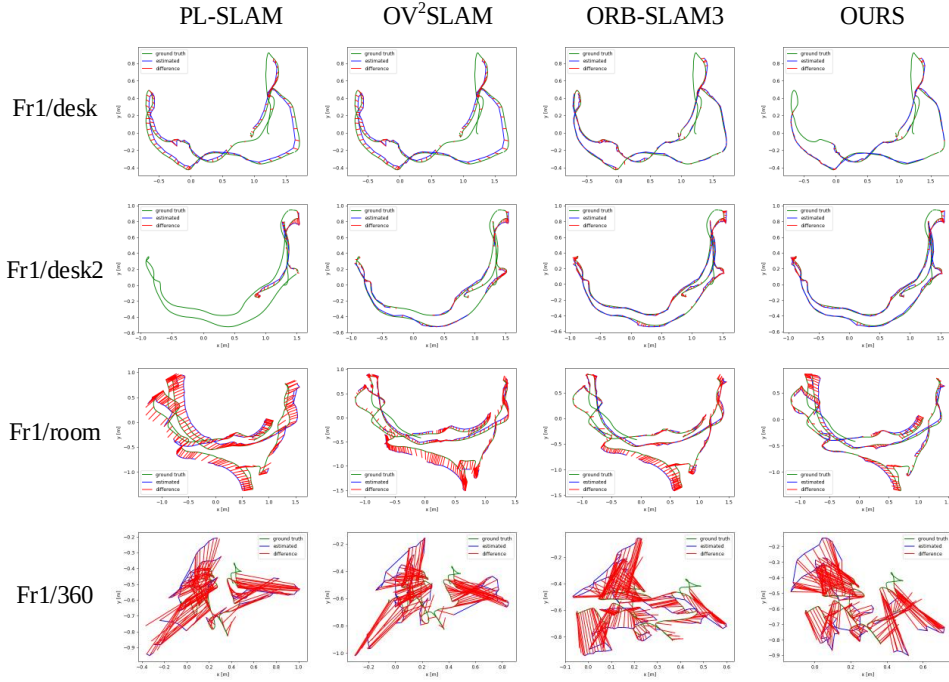


图 3-8 TUM 数据集轨迹对比

本章算法与 ORB-SLAM3、OV²-SLAM 和 PL-SLAM 这三种算法的通过运行 TUM 数据集所得到的平均绝对轨迹误差 ATE 对比如图表 3-3 所示。本章算法相较于其他三种算法在每个序列都取得较好的结果，从整体来看，PL-SLAM 算法在四个序列平均绝对轨迹误差相对其他算法明显最高，OV²-SLAM 次之。本章算法的平均绝对轨迹误差与 ORB-SLAM3 相比平均降低了 45%，相较于 OV²-SLAM 平均绝对轨迹误差平均降低了 59%，实验结果表明本章算法在大视角场景中性能较好。

表 3-3 TUM 数据集绝对轨迹误差 ATE 对比结果（单位：m）

TUM 数据集	PL-SLAM	OV ² -SLAM ^[56]	ORB-SLAM3	OURS
	ATE	ATE	ATE	ATE
Fr1/desk	0.443	0.158	0.095	0.018
Fr1/desk2	0.079	0.064	0.041	0.028
Fr1/room	0.182	0.129	0.118	0.081
Fr1/360	0.398	0.265	0.205	0.126

3.4 本章小结

本章针对移动机器人在大视角运动下图像模糊的问题，设计了一种基于改进 MSS 网络的去模糊算法以恢复高质量的清晰图像。该网络对图像进行下采样操作得到不同尺度的图像，通过特征注意模块增强图像特征信息提取的能力，随后利用监督注意力模块对各阶段的特征信息进行监督和融合，使模糊图像在逐步恢复过程中保持有效信息传递，恢复为原尺寸的清晰图像。针对 VSLAM 算法在大视角场景下特征匹配准确率低的问题，本章提出了一种改进的 GMS 算法，该方法通过局部信息熵对图像进行区域划分，然后根据对极几何约束判断相机是否发生大视角变换，根据视角变换大小分别采用十字五宫格和对角五宫格统计方法进行误匹配剔除。为验证本章算法的有效性，在公开数据集 TUM 进行轨迹误差实验，结果表明本章算法在大视角场景下依然能够保持较高的鲁棒性。

第 4 章 基于改进关键帧选取策略的视觉 SLAM 算法

针对传统视觉 SLAM 在大视角场景下关键帧选取不准确导致定位精度下降的问题，本章在第三章研究的基础上，提出一种基于改进关键帧选取策略的视觉 SLAM 算法。首先，通过计算相邻两帧之间的旋转矩阵和平移向量，得到相邻两帧之间的共视区域。然后，根据特征点连续跟踪的帧数计为有效值，并筛选出稳定的有效特征点。其次，分析当前帧与上一帧共视区域内有效特征点的数量及共视区域占比，综合判断当前帧是否具备关键帧选取的条件。最后，结合几何微分熵与信息熵，判断新选取的关键帧与普通帧之间的熵函数增益，对冗余关键帧进行剔除，从而提高系统整体定位精度。本章技术路线如图 4-1 所示。

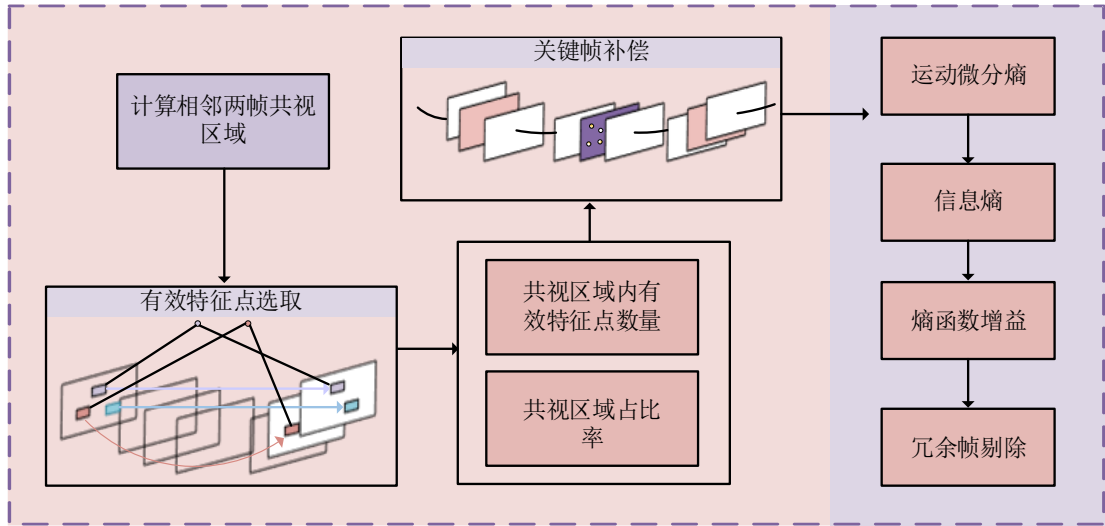


图 4-1 本章技术路线

4.1 关键帧补偿

大视角运动场景下，传统视觉 SLAM 算法由于特征提取方法的限制，导致关键帧质量下降。传统特征点提取方法在视角变化较大、物体遮挡、尺度变化或光照不均等复杂条件下，容易出现特征点检测不稳定、匹配点数减少甚至特征漂移的问题，进而影响关键帧的选取。此外，由于关键帧上的特征难以被稳定跟踪，系统容易发生位姿误差累积，导致回环检测失败和整体定位精度下降。因此，本章提出了一种关键帧补偿策略，该策略通过计算相邻两帧共视区域内的有效特征

点数量及其在图像中的占比，综合评估图像帧的信息有效性，决定是否将其选取为新的关键帧，进行关键帧补偿，提升系统在大视角运动场景下的鲁棒性和精度。

4.1.1 共视区域构建

首先，为计算相邻帧之间的共视区域，通过最小二乘法 and SVD 分解方法，计算相邻两帧之间的旋转矩阵和平移向量，公式如（4-1）所示：

$$\begin{cases} z_i = Rz'_i + t, \forall_i \\ \min_{R,t} \frac{1}{2} \|(z_i - (Rz'_i + t))\|_2^2 \\ e_i = z_i - (Rz'_i + t) \end{cases} \quad (4-1)$$

其中， z_i 和 z'_i 分别为相邻两帧图像中的第 i 点， e_i 为第 i 点的误差项， R 和 t 分别为相邻两帧间的旋转矩阵和平移向量。

其次，根据相机投影模型，以及旋转矩阵和平移向量，计算出时刻 t_0 到时刻 t_l 某一特征点在图像上的像素坐标，公式如（4-2）所示：

$$p_i = KRK^{-1}p'_i + Kt' \quad (4-2)$$

其中， p'_i 为 t_0 时刻图像上某一特征点的像素坐标， p_i 为 p'_i 在 t_l 时刻的投影坐标， K 为相机内参矩阵。

最后，将相邻两帧 I_l 和 I_2 的顶点坐标转换到同一坐标系中，设其对应的顶点集分别为 C_l 和 C_2 ，以获得共视区域，步骤如下：

1、将 I_l 和 I_2 的所有顶点及顶点之间的交点放入集合 L 中。为了确定相邻两帧之间的共视区域，可以将其视为求取两个多边形的交集，本质上是计算多边形边界之间的交点。设第一帧图像区域对应的多边形 I_l 和第二帧图像区域对应的多边形 I_2 ，定义所有顶点及顶点之间的交点组成的集合为 L 。设这两个多边形的顶点集分别为，公式如（4-3）所示：

$$C_1 = \{A_1, A_2, \dots, A_n\}, C_2 = \{A_1, A_2, \dots, A_m\} \quad (4-3)$$

每个多边形由边界组成边集，表达式公式如（4-4）所示：

$$\begin{cases} E_1 = \{(A_1, A_2), (A_2, A_3), \dots, (A_n, A_1)\} \\ E_2 = \{(B_1, B_2), (B_2, B_3), \dots, (B_m, B_1)\} \end{cases} \quad (4-4)$$

其中， E_l 和 E_2 为边集。为了计算多边形交点，需要逐一检测多边形 I_l 与 I_2 的所有边是否相交。假设两条线段分别为 I_l 和 I_2 ，其中 I_l 的端点以向量表示 $A = (x_1, y_1)$ ，

$A'=(x_2, y_2)$ ，线段 I_2 的两个端点为 $B=(x_3, y_3)$ ， $B'=(x_4, y_4)$ ，线段 I_1 和 I_2 中的任一点以向量形式可以表示为，公式如（4-5）所示：

$$\begin{cases} A_i = A + \mu \cdot r \\ B_j = B + \xi \cdot s \end{cases} \quad (4-5)$$

对式（4-5）分别进行两次向量叉乘法求解 μ 和 ξ ，公式如（4-6）所示：

$$\begin{cases} \mu = \frac{|(B-A) \times s|}{r \times s} \\ \xi = \frac{|(B-A) \times r|}{r \times s} \end{cases} \quad (4-6)$$

当两条线段出现相交时，出现有且仅有一个交点或共线的情况。若 $r \times s \neq 0$ ，且计算出来的 μ 和 ξ 都在 $[0,1]$ 范围内，则表示两个线段有且仅有一个交点，如果 $r \times s = 0$ 且 $(A-B) \times r = 0$ 或者 $(A-B) \times s = 0$ ，则两条线段共线，这两条线段有无数个交点，取共线四个顶点中的倒数第二个作为线段的交点。将所有交点构成无序点集 $L = \{A_1, A_2, A_3, A_n\}$ ，最终用于确定共视区域。

2、遍历集合 L ，筛选出同时属于 I_1 和 I_2 交点的点。遍历集合 L 中的每个点 $p(x, y)$ ，检查其是否属于 I_1 和 I_2 。这一判断可以通过多边形内部点检测方法来实现，其中射线法通过计算从点 P 出发的一条射线与多边形边界的交点数来确定点的位置，若交点数为奇数，则 P 位于多边形内部，若为偶数，则 P 在多边形外部。此外，若 P 位于多边形的边界上，则也满足筛选条件。凡是满足位于 I_1 和 I_2 之内或边界上的点，均被保留，并存入新的集合 L' 。

3、对筛选后的无序点集进行排序和整理，构建有序点集，从而得到共视区域。通过向量的夹角对无序点坐标进行排序，求取重心坐标计算公式如（4-7）所示：

$$\begin{aligned} x_o &= \frac{1}{m} \sum_{i=1}^m x_{A_i} \\ y_o &= \frac{1}{m} \sum_{i=1}^m y_{A_i} \end{aligned} \quad (4-7)$$

通过计算每个点 A_i 与重心 O' 的夹角，根据夹角对点集 L' 进行排序，将排序后的点依次连接，最终形成闭合多边形，表示相邻两帧之间的共视区域。不同位姿下相邻两帧之间的共视区域如图 4-2 所示。

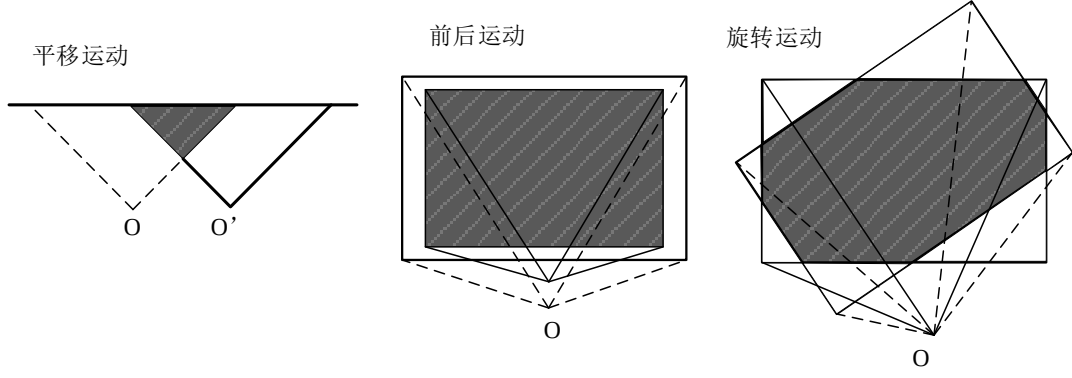


图 4-2 不同位姿的共视区域示意图

4.1.2 有效特征点选取

在视觉 SLAM 系统的运行过程中，部分特征点由于遮挡、视角变化等影响，导致无法稳定进行检测和匹配。通过共视区域内有效特征点进行关键帧选取，确保选取的关键帧包含对相机运动信息贡献较大的有效特征点。共视区域内的有效特征点由于能够在多帧之间持续跟踪，因此更具稳定性和可靠性。相比之下，那些由于消失或不稳定而无法在共视区域中持续跟踪的特征点应被剔除。因此，本章采取基于共视区域内特征点的稳定性来进行关键帧选取，系统能够优先保留那些能够提供有效运动估计信息的关键帧。

设当前帧为 K_d ， K_n 为第 n 帧，帧集合表示为 $\{K_1, K_2, K_3, \dots, K_d\}$ ，当前帧 K_d 与 K_3 特征点集合分别为 $\{K_d^1, K_d^2, K_d^3, \dots, K_d^N\}$ 和 $\{K_3^1, K_3^2, K_3^3, \dots, K_3^N\}$ ，并且二者特征点相互对应。为了衡量特征点的稳定性和持续性，对特征点赋予有效值属性，对于初次检测到的特征点，其有效值计为 0，若特征点连续出现在帧集合中，则有效值累加 1。特征点 $K_m^\wedge$ ，公式如 (4-8) 所示：

$$K_n^\wedge = K_n^m, m \in [1, N], M(K_n^m) \geq h \quad (4-8)$$

其中， $M(K_n^m)$ 表示为特征点的有效值， h 为判断特征点是否有效的阈值，特征点 $K_n^\wedge$ 在帧集合 $\{K_1, K_2, K_3, \dots, K_d\}$ 中跟踪，若大于 h ，则认为该特征点为有效特征点。

特征点 K_d^1 和 K_3^1 相互对应，两帧之间相差的帧数大于 h ，且能够在地图上找到与之对应的地图点 Q_l ，则 K_d^1 可以判定为有效特征点。 K_3^5 和 K_d^5 虽然能够连续在 h 帧上跟踪，但是在地图上没有与之相对应的地图点，表明在系统运行过程中被移除，因此不能作为有效特征点。此外， K_3^2 和 K_{d-2}^2 虽然在地图上有相应的地

图点，但是两帧之间相差的帧数小于 h ，所以不能计为有效特征点。在大视角场景下，通过有效特征点判定机制，系统能够筛选出在多帧之间稳定、可靠的特征点，并基于有效特征点进行关键帧的选取，从而提高整个 SLAM 系统的精度和鲁棒性。有效特征点的判定机制如图 4-3 所示。

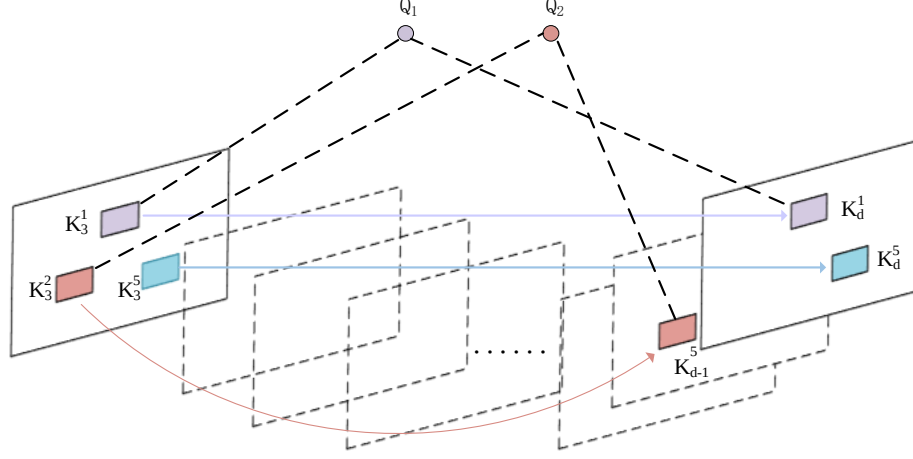


图 4-3 有效特征点的判定机制原理图

根据有效特征点的判定标准，将一帧中有效特征点的数量作为满足规定条件的所有特征点的总和。具体计算公式如（4-9）所示：

$$J = \sum_i \sum_{j \in N} (E(K_j^i, l)) \quad (4-9)$$

其中， J 表示当前帧中有效特征点的数量， i 表示特征点跟踪的帧数， $E(K_j^i, l)$ 表示特征点 K_j^i 的有效值。

在筛选出有效特征点后，通过计算当前帧共视区域内有效特征点的数量来评估图像质量。若有效特征点的数量大于设定的阈值 G ，则说明当前帧与上一帧图像质量发生了较大变化。虽然共视区域内的有效特征点数量远少于特征点提取的数量，但作为选取关键帧的条件，有助于提高位姿估计的准确性，并且能够减少由不稳定特征点引起的系统误差。然而，在大视角场景下，由于视角变化导致共视区域占比率发生变化，单纯依赖共视区域内有效特征点的数量不足以精准地选取关键帧。因此，本章进一步结合共视区域的占比率来决定是否选取当前帧为关键帧，确保系统在大视角场景下能够对关键帧进行有效补偿，并优先选取高质量的关键帧，从而提升系统的整体定位精度。共视区域占比率计算公式如（4-10）所示：

$$\begin{cases} R_c = \frac{S_c}{S_t} \\ S_c = \frac{1}{2} \left| \sum_{i=1}^n (a_i, b_{i+1}) \right| \end{cases} \quad (4-10)$$

其中， S_c 为共视区域的占比， S_t 为共视区域的面积， R_c 为图像帧的总面积， x_i 和 y_i 为共视区域内某一顶点的坐标。若 R_c 小于设定的阈值 R_c ，则说明两帧之间的视角变化较大，此时将选取当前帧为关键帧；若视角变化较小，则需要进一步判断共视区域内有效特征点的数量是否大于阈值 G 来决定是否选取当前帧为关键帧。关键帧选取策略原理图如图 4-4 所示。

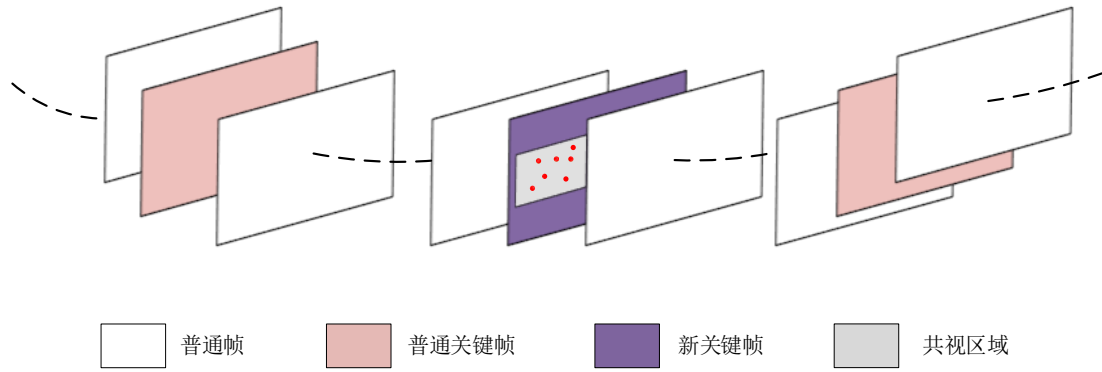


图 4-4 关键帧选取策略原理图

4.2 冗余帧剔除

传统的关键帧选取方法主要依赖信息熵，根据图像的灰度级分布衡量图像信息量，忽略了相机运动不确定性的变化。因此，为了避免补偿的新关键帧与已有的关键帧在图像信息上重复，导致关键帧冗余，本章提出一种结合运动微分熵与信息熵的方法，以更全面地评估关键帧的有效性，确保选取的关键帧不仅包含新的运动状态信息，还具有丰富的图像信息。

首先，计算运动微分熵，用于衡量相机运动估计中的不确定性变化。运动微分熵反映了相机在运动过程中的不确定性，较高的熵值表示相机的运动状态具有较高的不确定性。运动微分熵的计算公式如（4-11）所示：

$$h(x_0) = \frac{m(\log(2\pi e)) + \log\left(\left|\sum_{x_0}\right|\right)}{2} \quad (4-11)$$

其中， m 是维度数， $\left|\sum_{x_0}\right|$ 是误差分布的协方差矩阵。

其次，计算信息熵用于衡量图像内容的变化。信息熵反映图像的内容复杂度

或信息量，较高的信息熵表明图像内容具有较大的变化或复杂度。对于一幅图像 I_c ，其信息熵可以通过灰度直方图的统计特征来计算，公式如（4-12）所示：

$$H(I_c) = -\sum_{i=0}^{L_c-1} p_i \log_2(p_i) \quad (4-12)$$

其中， L_c 为图像的灰度级数（通常为 256）。 p_i 是灰度值 i 在图像中出现的概率。

然后，为了综合考虑相机运动状态和图像内容变化的双重影响，定义一个熵函数增益 M 。熵函数增益结合了相机运动的不确定性和图像信息的变化，其计算公式如（4-13）所示：

$$M = w_1 \cdot \alpha + w_2 \cdot \Delta H \quad (4-13)$$

其中， w_1 和 w_2 是权重系数，根据实际应用场景进行调整； α 是运动微分熵熵比，用于量化相机运动估计的不确定性变化。 ΔH 是信息熵差值，反映图像内容的变化。该函数的设计使得相机运动的变化和图像内容的变化能够综合评估，从而决定是否保留当前帧作为关键帧。运动微分熵熵比 α 计算公式如（4-14）所示：

$$\alpha = \frac{h(I_{k+1})}{h(I_k)} \quad (4-14)$$

其中， $h(I_k)$ 和 $h(I_{k+j})$ 分别上一个关键帧和当前帧的运动估计熵。信息熵差值 ΔH 计算公式如（4-15）所示：

$$\Delta H = |H(I_{k+1}) - H(I_k)| \quad (4-15)$$

其中， $H(I_{k+j})$ 是当前关键帧 I_{k+j} 的信息熵。 $H(I_k)$ 是上一个关键帧 I_k 的信息熵。

最后，通过比较关键帧 I_k 和 I_{k+1} 之间的熵函数增益 M 与设定的增益阈值 M_{th} ，决定是否剔除或保留新选取的关键帧。如果两帧之间的信息增益小于设定的阈值 M_{th} ，则剔除新选取的关键帧；否则，保留该关键帧，并继续筛选下一个关键帧。

4.3 实验及分析

本章实验主要分为三部分：关键帧选取实验、室内场景轨迹对比实验、室外场景轨迹对比实验。本章实验所用电脑配置为：系统为 Ubuntu18.04，主机显卡 3060，运行内存为 16G，选取 TUM 数据集和 KITTI 数据集，对本章所提出的算法进行验证进行对比分析。

4.3.1 关键帧选取实验

为验证本章算法面对大视角场景下快速运动过程中的关键帧选取策略的有效性,本节实验采用公开数据集 TUM 中的 6 个大视角场景较多的序列进行实验。将本章算法与 ORB-SLAM3 和 PL-SLAM 这两种算法进行对比实验,使用红色线条标记相机在各数据集中发生视角变换的区域。同时统计这些区域内采取关键帧补偿后的关键帧数量 f_k 以及计算此区域绝对轨迹均方根误差 RMSE 的平均值。绿色方框表示相机的运动轨迹,蓝色方框则表示系统在运行时选取的关键帧。不同算法的关键帧选取结果如图 4-5 所示。

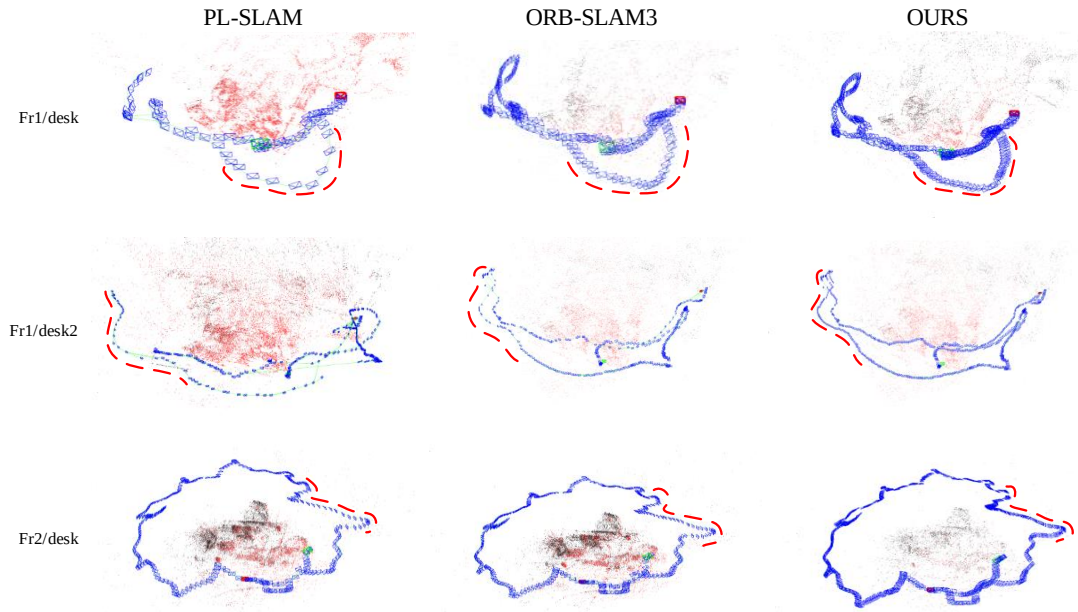


图 4-5 关键帧选取对比图

从实验可以看出,不同算法在大视角场景下关键帧的选取存在显著差异。PL-SLAM 算法在大视角弯道处关键帧选取的数量明显偏少,在 fr1/desk2 和 fr2/desk 这两个数据集中,由于相机视角变化幅度较大,PL-SLAM 算法在关键帧选取的过程中出现特征点跟踪丢失,导致部分运动轨迹未能被准确的还原。相比之下,ORB-SLAM3 算法在面对大视角运动时,虽然未出现明显特征点跟踪丢失的现象,但其关键帧的选取较为稀疏。特别是在连续弯道运动时,轨迹细节还原不足,影响了整体轨迹精度。本章提出的关键帧选取方法在所有实验场景均能实现关键帧均匀分布。在大视角快速运动场景较多的 fr1/desk2 数据集中,不仅能够稳定选

取高质量的图像作为关键帧，还更准确地还原相机运动轨迹。实验结果表明，本章算法在大视角下关键帧选取的质量更准确，提升了 SLAM 系统的定位精度。

PL-SLAM、ORB-SLAM3 以及本章提出的算法在大视角场景下（如图 4-5 红色虚线标记的区域）的关键帧选取数量及其绝对轨迹均方根误差 RMSE 对比结果如图表 4-1 所示。实验结果表明，在六个测试序列中，PL-SLAM 算法在关键帧选取和绝对轨迹均方根误差方面表现较差。特别是在 Fr1/360 和 Fr1/room 这两个数据集中，PL-SLAM 所选取的关键帧数量明显较少，这表明该算法在大视角场景下适应性较弱，无法有效地选取关键帧并保证位姿估计的准确性。与 PL-SLAM 和 ORB-SLAM3 算法相比，本章算法不仅关键帧数量上显著增加，其绝对轨迹均方误差的 RMSE 平均值与 PL-SLAM 和 ORB-SLAM3 相比分别降低 51.78%和 38.21%，充分证明本文算法在提升位姿精度方面的有效性。实验结果综合显示，在面对大视角场景时，本章提出的关键帧选取算法能够有效减少轨迹误差，提升系统的定位精度。

表 4-1 关键帧选取对比结果

TUM 数据集	图像帧	PL-SLAM		ORB-SLAM3		OURS	
		关键帧 (个)	RMSE(m)	关键帧 (个)	RMSE(m)	关键帧 (个)	RMSE(m)
fr1/desk	613	10	0.009325	16	0.008592	21	0.004263
fr1/360	756	7	0.008517	12	0.006135	25	0.002674
Fr1/room	1362	19	0.008731	24	0.007179	64	0.004839
Fr1/desk2	640	17	0.011904	21	0.009562	32	0.002429
Fr2/desk2	2965	30	0.010825	38	0.007825	57	0.005341
Fr3/l/o/h	2586	23	0.009053	32	0.006251	51	0.005591
平均值	1487	18	0.009726	24	0.007591	42	0.004690

4.3.2 室内场景轨迹对比实验

为了验证本章算法在室内场景下的性能，实验选取了公开数据集 EUROC 中的五个典型序列进行测试，并将实验结果与 ORB-SLAM3、OV²-SLAM 和 PL-

SLAM 三种主流算法进行对比。实验采用平均绝对轨迹误差 RMSE 作为性能评估指标，以衡量不同算法在室内环境中的定位精度。

本章算法与其他三种主流算法在室内 EUROC 数据集进行轨迹对比，黑色轨迹线条表示数据集提供的真实轨迹，蓝色轨迹线条表示各算法估计得到的位姿轨迹，红色轨迹线条表示真实轨迹与估计轨迹之间的误差偏差如图 4-6 所示。

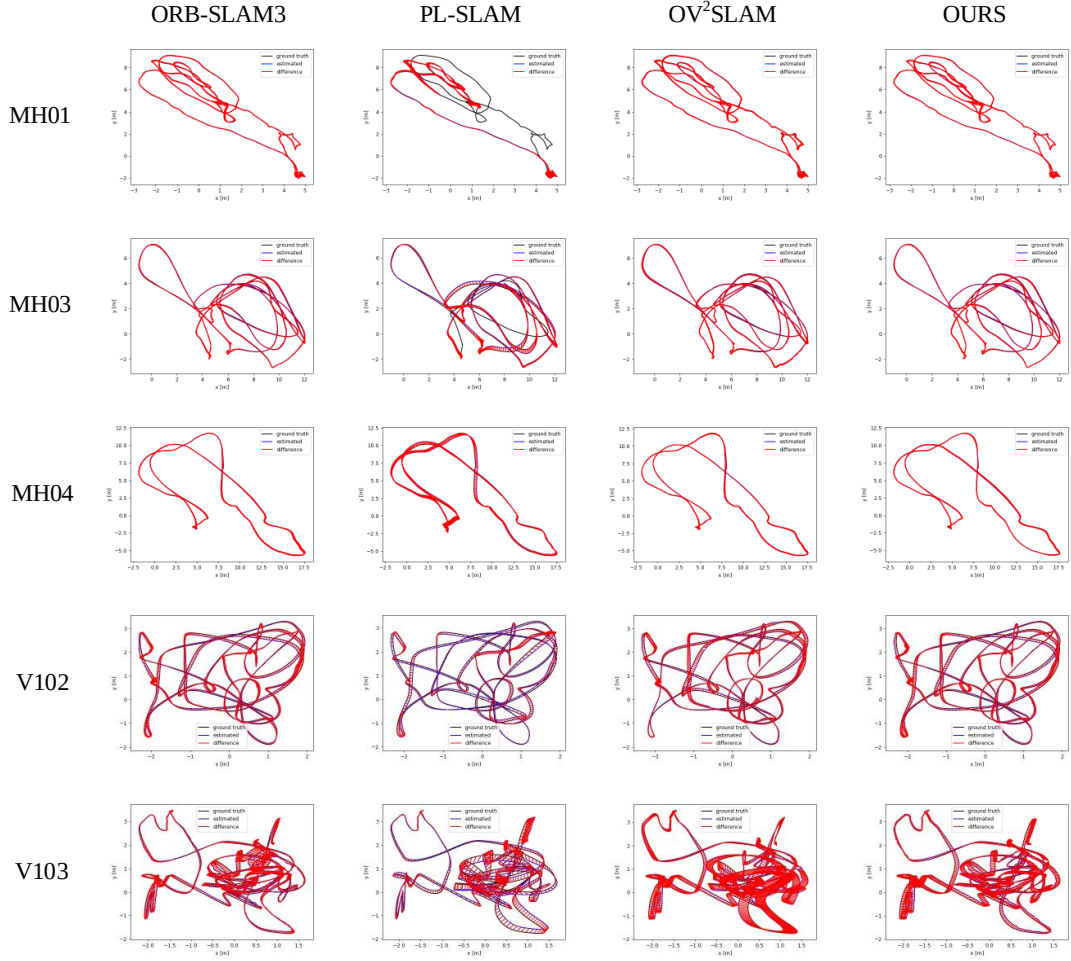


图 4-6 EUROC 数据集轨迹对比

从图 4-6 的对比结果可以看出，OV²-SLAM 和 PL-SLAM 算法的轨迹与真实轨迹相比存在较大偏差，尤其在 V103 数据集上，误差显著增大。此外，OV²-SLAM 算法在 MH01 数据集中出现了轨迹缺失，无法形成完整的位姿轨迹。而本章提出的改进匹配算法通过提高匹配准确率和增加匹配特征点对数，有效提升了全局轨迹优化效果，使得估计轨迹更接近真实轨迹。综上所述，本章算法在 EUROC 数据集上的实验结果表明，相较于 OV²-SLAM、PL-SLAM 等现有方法，

本章算法在关键帧匹配和轨迹优化方面表现更优，在提高 SLAM 系统定位精度的同时也展现出了更强的适应性和鲁棒性。

本章算法与 ORB-SLAM3、PL-SLAM、OV²-SLAM 三种算法在 EUROC 数据集上的轨迹误差对比如表 4-2 所示。根据表 3 中各个算法在 8 个序列上的 ATE 结果分析，本章算法在每个序列中均取得了优异的性能。相比之下，OV²-SLAM 和 PL-SLAM 的绝对轨迹误差较大，尤其是 OV²-SLAM 在 MH01 序列上未能成功构建地图，导致图像跟踪丢失，进一步影响了系统的定位稳定性。而本章算法在各个实验序列中均能保持较高的轨迹精度，有效降低了累积误差。本章算法相较于精度最高的 ORB-SLAM3，在部分序列上的 ATE RMSE 最高降低了 24.2%，相比 OV²-SLAM，ATE RMSE 最高降低了 76.8%。这一实验结果充分验证了本章算法在室内环境及运动模糊场景下的鲁棒性和准确性，并证明了其在提高视觉 SLAM 系统精度方面的显著优势。

表 4-2 EUROC 数据集平均绝对轨迹误差对比（单位：m）

数 据 集	ORB-SLAM3	PL-SLAM	OV ² -SLAM	OURS
	ATE	ATE	ATE	ATE
MH01	0.0384	0.1341	0.0504	0.0321
MH02	0.0479	0.0464	0.0603	0.3981
MH03	0.0455	0.1969	0.0518	0.0410
MH04	0.0986	0.3697	0.0812	0.0747
MH05	0.0629	0.2991	0.0899	0.0555
V101	0.0868	0.0869	0.0891	0.0859
V102	0.0658	0.0914	0.0770	0.0645
V103	0.1218	0.1766	0.1755	0.0939

4.3.3 室外场景轨迹对比实验

为了验证本章算法在室外场景下的性能，本文采用了 KITTI00-10 序列进行实验。KITTI 数据集是目前最广泛使用的室外场景 SLAM 数据集之一，包含大量视角变化较大的运动场景，因此非常适用于本章算法的性能验证。

本章提出的算法与其他两种算法在 KITTI 数据集进行轨迹对比，黑色线条表示真实的行驶路径，红色线条和蓝色线条分别表示实验中测得的最大误差和最

小误差。在 KITTI 02、07、09 这三个测试序列中，不同算法的最大轨迹误差如下：PL-SLAM：分别为 21.394m、6.773m、14.417m；ORB-SLAM3：分别为 19.690m、2.120m、13.061m；本研究提出的算法：分别为 18.345m、1.511m、5.899m。从实验数据可以看出，PL-SLAM 在大视角变化的室外环境下误差较大，而 ORB-SLAM3 的误差有所降低，但在部分序列仍然存在较大的偏差。相比之下，本章提出的改进算法在所有测试序列中均表现出更优的精度，特别是在 KITTI 09 序列中的误差减少幅度最为显著，验证了其在室外大视角变化场景下的鲁棒性和准确性。如图 4-7 所示。

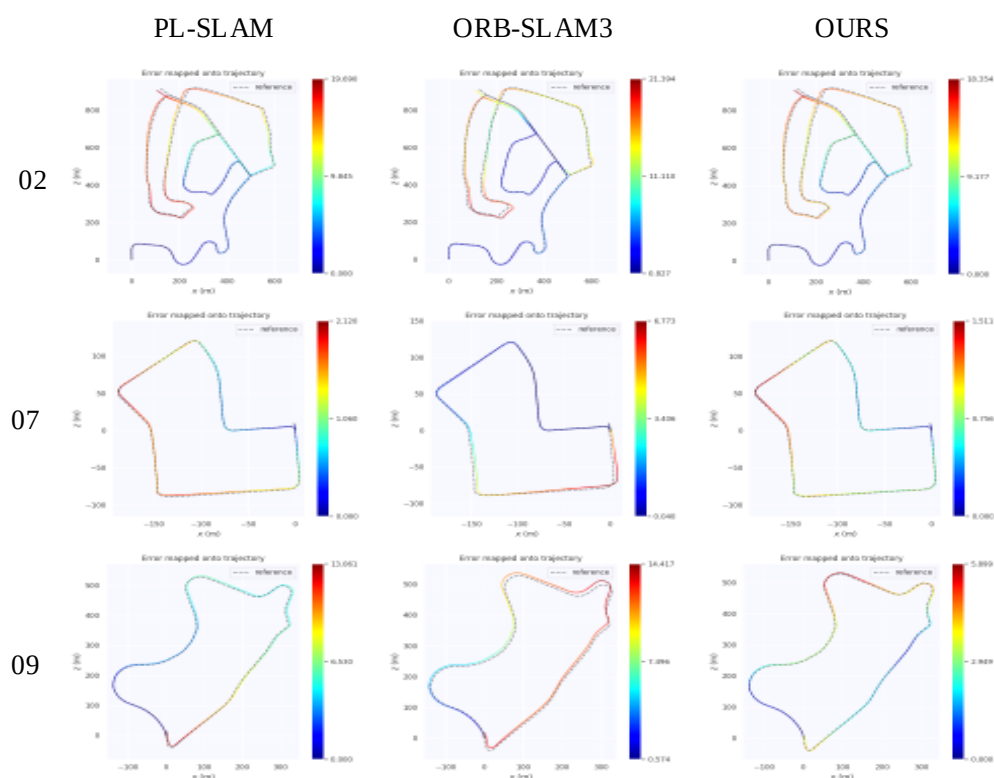


图 4-7 KITTI 数据集轨迹对比图

在 KITTI 数据集 00-10 序列三种算法进行绝对轨迹与相对位姿误差对比，相较于 ORB-SLAM3，本章算法将绝对轨迹误差平均减少了 24.81%，相对位姿误差降低了 9.01%，而与 PL-SLAM 相比，其绝对轨迹误差平均下降了 52.73%，相对位姿误差则降低了 28.02%。由此可见，本章所提出的算法在绝对轨迹和相对位姿误差方面均优于 ORB-SLAM3 及 PL-SLAM 两种方法，验证了在室外大视角移动场景时本章算法具有显著优势。如图 4-8 所示。

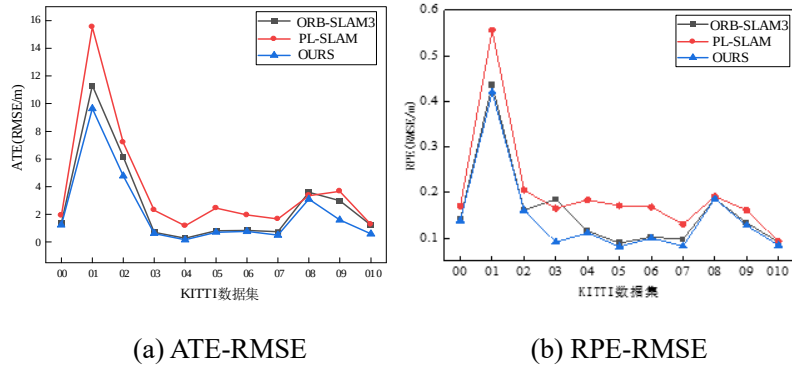


图 4-8 KITTI00-10 序列绝对轨迹和相对位姿误差对比

4.4 移动机器人真实场景实验

为了验证本文算法在实际场景的效果，构建面向大视角下基于改进 MSS 网络的视觉 SLAM 系统的实验平台。软件部分实现机器人的定位与建图，硬件部分确保移动机器人能够稳定运行，并且能够采集有效的图像数据。设置大视角运动的真实场景进行移动机器人对本章算法的定位轨迹精度进行分析，在实验中通过终端控制机器人从初始点 d 出发，依次经过障碍物 A 和 B，再避开障碍物 C，最终形成一个闭环路径。整个实验区域尺寸为 7 米×5.95 米。如图 4-9 所示。

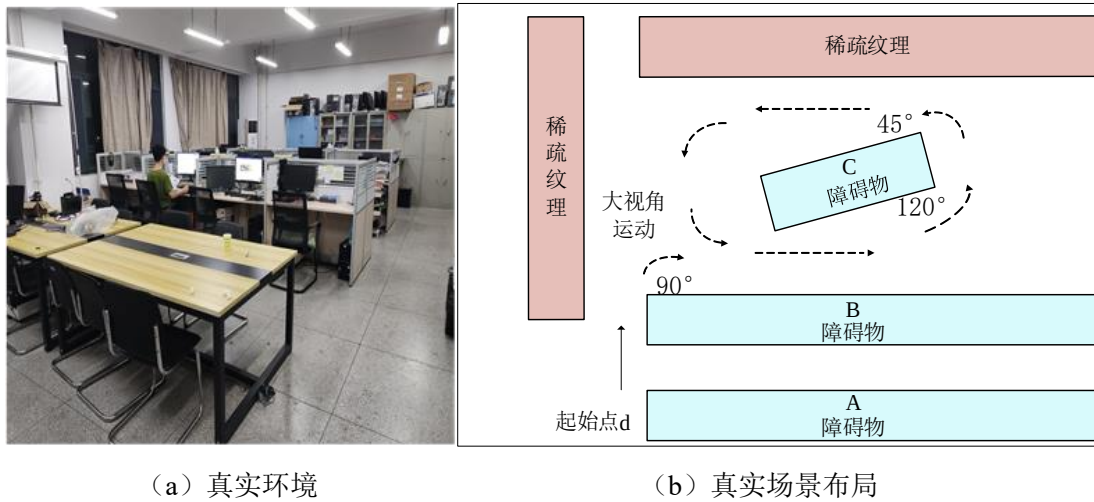


图 4-9 真实场景模拟图

分别使用 ORB-SLAM3、OV²-SLAM 以及本文算法对上述真实环境下的数据集进行机器人轨迹图构建。根据实验观察，当遇到视角显著变化时，ORB-SLAM3 算法在转角处表现出较差的轨迹跟踪性能，导致其运行路径出现较大偏差，并且难以实现精确的地图构建。同样地，在处理由大范围视角变动引起的关键帧选取

的问题上，OV²-SLAM 算法也显示出不足之处，该算法形成的运动轨迹与实际偏差较大，导致在闭环检测环节中对极几何的计算出现显著偏差，使轨迹无法形成闭环。但是，本文算法通过结合改进 MSS 去模糊网络和误匹配剔除算法，在面向大视角弯道运动较多的场景时，依旧能够得到准确的运行轨迹，同时通过改进的关键帧选取策略，实现准确定位与高精度建图，验证了本文算法的优越性和准确性。几种不同算法生成的移动机器人轨迹如图 4-10 所示。

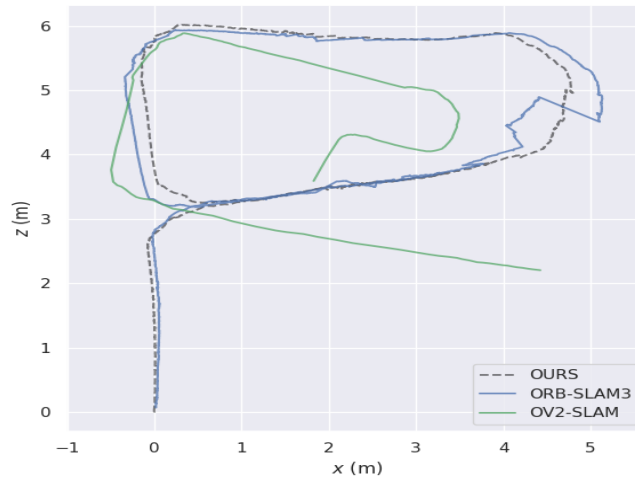


图 4-10 不同算法的轨迹对比

4.5 本章小结

针对传统视觉 SLAM 算法在大视角场景下因关键帧选取不准确导致的定位精度下降问题，本章提出了一种基于改进关键帧选取策略的视觉 SLAM 算法。首先，通过计算相邻帧之间旋转矩阵与平移向量，构建共视区域，并结合特征点在连续帧中的跟踪帧数筛选出稳定的有效特征点，为关键帧选取提供可靠判断依据；其次，利用共视区域占比率与运动连续性相结合，建立关键帧选取条件，从而有效避免因视角变换而导致的误判；最后，通过融合几何微分熵与信息熵，构建熵增益函数对新选取的关键帧进行信息评估，实现冗余关键帧的高效剔除。在保证系统实时性的同时，进一步优化关键帧选取策略。实验结果显示，相较于 ORB-SLAM3，本章算法将绝对轨迹误差平均减少了 24.81%，相对位姿误差降低了 9.01%，而与 PL-SLAM 相比，其绝对轨迹误差平均下降了 52.73%，相对位姿误差则降低了 28.02%，充分证明本章算法的有效性和优越性。

第5章 全文总结展望

5.1 全文总结

本文针对传统视觉 SLAM 系统在大视角场景下普遍存在的图像模糊、特征匹配率低及关键帧选取不准确等问题展开研究。在系统梳理经典视觉 SLAM 基础理论的同时,结合深度学习中的卷积神经网络,使视觉 SLAM 系统能够更好的适应各种复杂的场景。在多个公开数据集与真实场景中进行验证,有效证明在大视角场景下改善了视觉 SLAM 算法定位精度较低的问题,提升了系统的稳定性和实用性。论文主要研究内容于创新点概括如下:

(1) 针对移动机器人在大视角运动下图像模糊的问题,设计了一种基于改进 MSS 网络的去模糊算法以恢复高质量的清晰图像。该网络对图像进行下采样操作得到不同尺度的图像,通过特征注意模块增强图像特征信息提取的能力,随后利用监督注意力模块对各阶段的特征信息进行监督和融合,使模糊图像在逐步恢复过程中保持有效信息传递,恢复为原尺寸的清晰图像。本文算法在 GOPRO 数据集上相较于 MSS-Net 算法,PSNR 和 SSIM 分别提升了 0.75%、1.56%,相较于 DeblurGANv2 算法,所提算法的 PSNR 和 SSIM 分别提升了 12.59%、2.95%,图像质量指标上有明显的提升,实验结果充分证明了所提算法的优越性。

(2) 针对 VSLAM 算法在大视角场景下特征匹配准确率低的问题,提出了一种改进的 GMS 算法,该方法通过局部信息熵对图像进行区域划分,然后根据对极几何约束判断相机是否发生大视角变换,根据视角变换大小分别采用十字五宫格和对角五宫格统计方法进行误匹配剔除。实验选取 TUM 数据集中部分具有大视角变换的序列,并对不同视角下误匹配剔除效果进行对比实验,本章节误匹配剔除算法的 mAP 在视角变化为 30° 、 60° 、 90° 场景中与 Rep 算法相比分别提升了 27%、57.8%、97.84%,相较于 GMS 算法分别提升了 16.2%、22.5%、37.4%。实验结果表明,本章算法在角度变化越大时误匹配剔除效果越明显,在不同视角场景下均具有较高的误匹配剔除性能。

(3) 针对传统视觉 SLAM 算法在大视角场景下因关键帧选取不准确导致的定位精度下降问题,提出了一种基于改进关键帧选取策略的视觉 SLAM 算法。

通过计算相邻帧之间旋转矩阵与平移向量，构建共视区域，并结合特征点在连续帧中的跟踪帧数筛选出稳定的有效特征点。利用共视区域占比率与运动连续性相结合，进行关键帧补偿。然后通过融合运动微分熵与信息熵，构建熵函数增益对新选取的关键帧进行信息评估，实现冗余关键帧的高效剔除。实验结果显示，相较于 ORB-SLAM3，本文算法将绝对轨迹误差平均减少了 24.81%，相对位姿误差降低了 9.01%，而与 PL-SLAM 相比，其绝对轨迹误差平均下降了 52.73%，相对位姿误差则降低了 28.02%，验证了其在大视角移动场景时的显著优势。

5.2 展望

本文围绕“面向大视角下基于改进 MSS 网络的视觉 SLAM 算法”解决了视觉 SLAM 在大视角场景下图像模糊、匹配率低、关键帧选取差等问题，通过公开数据集和真实场景实验充分的验证了本文所提出的方法。然而，随着机器人技术的发展和应用场景的复杂化和多样化，所取得的成果仍然有提高的空间，主要可以从以下几个方面进行展开：

（1）本文着重研究了视觉 SLAM 的前端图像处理、特征匹配和关键帧选取方面，并未对视觉 SLAM 的后端优化、闭环检测和建图这三个方面进行涉及，因此后续的工作可以围绕这三个方面进行展开，设计出更适用于移动机器人定位和建图的视觉 SLAM。

（2）本文是在视觉 SLAM 的基础上进行改进和研究，在面对低纹理场景下本文算法无法获得足够的特征信息，可以结合激光雷达、点线特征的引入、语义信息的利用，避免因特征点提取少和特征匹配困难导致系统跟踪失败和定位精度差等问题。

参考文献

- [1] Cao Y, Beltrame G. VIR-SLAM: Visual, inertial, and ranging SLAM for single and multi-robot systems[J]. *Autonomous Robots*, 2021, 45(6): 905-917.
- [2] Soares J, Gattass M, Meggiolaro M. Crowd-SLAM: visual SLAM towards crowded environments using object detection[J]. *Journal of Intelligent & Robotic Systems*, 2021, 102(2):50-53.
- [3] Von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: A fast line segment detector with a false detection control[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 32(4): 722-732.
- [4] Gomez-Ojeda R, Moreno F A, Zuniga-Noël D, et al. PL-SLAM: A stereo SLAM system through the combination of points and line segments[J]. *IEEE Transactions on Robotics*, 2019, 35(3): 734-746.
- [5] Li Y, Yunus R, Brasch N, et al. RGB-D SLAM with structural regularities[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). Xian, China, IEEE, 2021: 11581-11587.
- [6] 王通典,刘洁瑜,吴宗收等.基于多尺度光流融合特征点视觉-惯性 SLAM 方法[J].*系统工程与电子技术*,2022,44(03):977-985.
- [7] Sun Y, Liu M, Meng M Q H. Motion removal for reliable RGB-D SLAM in dynamic environments[J]. *Robotics and Autonomous Systems*, 2018, 108(6): 115-128.
- [8] Kaneko M, Iwami K, Ogawa T, et al. Mask-slam: Robust feature-based monocular slam by masking using semantic segmentation[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Utah, USA, IEEE, 2018: 258-266.
- [9] Zhong F, Wang S, Zhang Z, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV, USA, IEEE, 2018: 1001-1010.
- [10] Bescos B, Fácil J M, Civera J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4):

4076-4083.

- [11]Yu C, Liu Z, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain, IEEE, 2018: 1168-1174.
- [12]Liu Y, Miura J. RDS-SLAM: real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9(23): 772-785.
- [13]张洪,于源卓,邱晓天.基于改进关键帧选择的 ORB-SLAM2 算法[J].北京航空航天大学学报,2023,49(01):45-52.
- [14]Wang L, Guo S, Huang W, et al. Places205-vggnet models for scene recognition[J]. 2015: 1508.01667.
- [15]Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C]IEEE International Conference on Computer Vision (ICCV). Utah, USA, IEEE, 2011: 2320-2327.
- [16]Engel J , Schps T , Cremers D . LSD-SLAM: Large-scale direct monocular SLAM[C]. Proceedings of the European Conference on Computer Vision (ECCV). Zurich, Switzerland, Springer, Cham, 2014: 834-849.
- [17]Ganti P, Waslander S L. Network uncertainty informed semantic feature selection for visual SLAM[C]. Proc of the 16th Conference on Computer and Robot Vision (CCRV). Piscataway, NJ: IEEE , 2019: 121-128.
- [18]Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 29(6): 1052-1067.
- [19]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.
- [20]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.
- [21]Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source

- library for visual, visual - inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [22]杜航原,郝燕玲,赵玉新.基于模糊自适应卡尔曼滤波的 SLAM 算法[J].华中科技大学学报: 自然科学版, 2012, 40(1):5-15.
- [23]刘华杰,梁冬泰,梁丹,等.基于生成对抗网络的运动去模糊视觉 SLAM 方法[J].传感器与微系统, 2020, 39(8):50-54.
- [24]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.
- [25]Pumarola A, Vakhitov A, Agudo A, et al. PL-SLAM: Real-time monocular visual SLAM with points and lines. 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 4503-4508.
- [26]CLEMENTEL , DAVISON A . Mapping large loops with a single hand-held camera[C]. Robotics Science and Systems (RSS 2007). Bei Jing,China, IEEE, 2007, 2(2) :297-304.
- [27]Tang J , Ericson L , Folkesson J , et al. GCNv2: Efficient Correspondence Prediction for Real Time SLAM[J]. IEEE Robotics and Automation Letters, 2019, 4(4):3505-3512.
- [28][33]Li Dongjiang, Shi Xuesong, Long Qiwei, et al. DXSLAM: A Robust and Efficient Visual SLAM System with Deep Features[C]2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2020: 4958-4965.
- [29]Pu H, Luo J, Wang G, et al. Visual SLAM integration with semantic segmentation and deep learning: A review[J]. IEEE Sensors Journal, 2023, 23(19): 119-138.
- [30]Sonugür G. A Review of quadrotor UAV: Control and SLAM methodologies ranging from conventional to innovative approaches[J]. Robotics and Autonomous Systems, 2023, 16(1):104-112.
- [31]Sheng X, Mao S, Yan Y, et al. Review on SLAM algorithms for Augmented Reality[J]. Displays, 2024,23(22):102-110.
- [32]Yuan C, Xu Y, Zhou Q. PLDS-SLAM: point and line features SLAM in dynamic environment[J]. Remote Sensing, 2023, 15(7): 189-195.

- [33]Lin Z, Zhang Q, Tian Z, et al. DPL-SLAM: enhancing dynamic point-line SLAM through dense semantic methods[J]. IEEE Sensors Journal, 2024, 24(9):596-607.
- [34]Yang H, Yuan J, Gao Y, et al. UPLP-SLAM: Unified point-line-plane feature fusion for RGB-D visual SLAM[J]. Information Fusion, 2023, 96(2): 51-65.
- [35]Wei Y, Zhou B, Duan Y, et al. DO-SLAM: research and application of semantic SLAM system towards dynamic environments based on object detection[J]. Applied Intelligence, 2023, 53(24): 30009-30026.
- [36]Islam Q U, Ibrahim H, Chin P K, et al. ARD-SLAM: Accurate and robust dynamic SLAM using dynamic object identification and improved multi-view geometrical approaches[J]. Displays, 2024, 82(22):102-111.
- [37]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.
- [38]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.
- [39]Xia Y, Cheng J, Cai X, et al. SLAM back-end optimization algorithm based on vision fusion IPS[J]. Sensors, 2022, 22(23): 93-102.
- [40]陈孟元,韩朋朋,刘金辉等.动态遮挡场景下基于改进 Transformer 实例分割的 VSLAM 算法[J].电子学报,2023,51(07):1812-1825.
- [41]Chen S, Zhou B, Jiang C, et al. A lidar/visual slam backend with loop closure detection and graph optimization[J]. Remote sensing, 2021, 13(14): 27-32.
- [42]Viset F, Helmons R, Kok M. An extended Kalman filter for magnetic field SLAM using Gaussian process regression[J]. Sensors, 2022, 22(8): 28-33.
- [43]陈孟元,丁陵梅,张玉坤.基于改进关键帧选取策略的快速 PL-SLAM 算法[J].电子学报,2022,50(3):608-618.
- [44]张志铭,张军,刘元盛,等.基于自身运动约束的 RGB-D 动态 SLAM 算法[J].计算机工程与应用,2025,61(08):173-182.
- [45]Smith C, Cheeseman P. On the representation and estimation of spatial uncertainty[J]. The International Journal of Robotics Research,2017,5(4):56-68.

- [46]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.
- [47]徐晓苏,王睿,姚逸卿.基于筛选策略的动态环境下激光 SLAM 算法[J].中国惯性技术学报,2024,32(07):681-689+695.
- [48]符强,腾先云,纪元法,等.基于改进特征描述的 SLAM 动态算法研究[J].系统仿真学报,2024,36(11):2712-2721.
- [49]李丹,李俊祥,赵文杰,等.基于改进的点线融合和关键帧选择的视觉 SLAM 方法[J].机器人,2025,47(02):200-212.
- [50]刘丰宇,程向红,曹毅.基于深度学习与特征点速度约束的室内动态 SLAM 方法[J].中国惯性技术学报,2023,31(05):438-443.
- [51]Kupyn O, Budzan V, Mykhailych M, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks[C]Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, Utah, USA, IEEE, 2018: 8183-8192.
- [52]Kupyn O, Martyniuk T, Wu J, et al. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better[C]//Proceedings of the IEEE/CVF international conference on computer vision. Lake Tahoe, NV, USA, IEEE, 2019: 8878-8887.
- [53]Ke W, Chen J, Jiao J, et al. SRN: Side-output residual network for object symmetry detection in the wild[C] IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii, US, IEEE, 2017: 1068-1076.
- [54]Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring[C].Proceedings of The IEEE/CVF International Conference on Computer Vision. Montreal, Canada, IEEE, 2021: 4641-4650.
- [55]Kim K, Lee S, Cho S. Mssnet: Multi-scale-stage network for single image deblurring[C].European Conference on Computer Vision. Proceedings of the European Conference on Computer Vision (ECCV). Tel Aviv, Israel, Cham: Springer Nature Switzerland, 2022: 524-539.
- [56]Ferrera M, Eudes A, Moras J, et al. OV2 -SLAM: A fully online and versatile visual SLAM for real-time applications[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1399-1406.

攻读学位期间发表的学术论文和科研成果

一、发表学术论文

- 【1】 杜鹏,陈孟元,徐奥.面向大视角下基于改进的 MSS 网络的视觉 SLAM 算法[J].安徽工程大学学报. (已录用, 本人第一, 导师第二)
- 【2】 陈孟元,李鹏飞,符乙,杜鹏, 等.低光条件下结合 FFE-Net 网络的视觉 SLAM 算法[J].中国惯性技术学报,2025,33(01):36-45.

二、发明专利

- 【1】 一种面向大视角下的机器人视觉 SLAM 方法、存储介质及设备(已授权, 导师第一, 本人第二)

三、参与的科技项目

- 【1】 安徽省重点研究与开发计划项目 (高新领域), 202304a05020073, 面向低纹理场景的增强感知视觉移动机器人关键技术与示范应用, 2023/08-2026/06, 100.0 万元。(皖科高秘〔2023〕250 号)
- 【2】 安徽省高校杰出青年科研项目, 2022AH020065, 基于视觉-惯性的移动机器人复杂环境仿生 SLAM 算法研究, 2022/09-2025/08, 100.0 万元。(皖教秘科〔2022〕241 号)
- 【3】 安徽省高校协同创新项目, GXXT-2021-050, 复杂环境下移动机器人位姿自适应优化算法研究及应用, 2021/01-2023/12, 40.0 万元。(皖教秘科〔2021〕58 号)

致谢

在毕业论文完成之际，我的硕士研究生学习生活即将结束，在三年的时间中我学习了许多知识，收获了许多技术能力，在此，我要特别感谢在三年里给予我生活和学习巨大帮助的老师，同学，家人。

我要感谢我的导师陈孟元教授，感谢陈老师在三年来对我的帮助。在科研方面，陈老师以身作则与我们一起探讨论文创新点，同时指导我们实验验证创新点的可行性，在论文撰写过程中陈老师对我们有极大的耐心，教导我们论文写作的注意事项，更改论文中的错误，为我们解决了一个又一个难题。老师对待工作与学术有着严谨的态度，对待学习有着刻苦钻研的精神，这些为我们树立了榜样。

我要感谢实验室已经毕业的师兄在学习生活中给了我很大的帮助，在刚开始对 SLAM 不了解的时候师兄们鼓励我们多探索，在生活中给了我巨大帮助。也感谢实验室师弟师妹们你们为实验室带来了活力，愿你们在今后工作学习中一帆风顺。同时，我也要感谢我同门李鹏飞、符乙、许瑞珩、杨苏鹏和我的室友李淼、于卓越、徐奥，我们在一起生活学习三年，在硕士研究生期间相互帮助，即使毕业后大家会走向祖国各地，但是我仍然相信这是我人生中的一段不可忘却的宝贵经历。另外，我要感谢我的父母和我的爱人殷琴，是你们对我无私的奉献与帮助才让我能够全身心投入到研究生的学习中，感谢你们默默无闻的付出。

最后非常感谢在珍贵的科研和教学时间中抽出时间来评审论文的各位老师。

尚德敏学 唯实惟新

