分类号：TP242.6　　学校代码：10363
密　级：公开　　学　　号：2210330113

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于感知增强与特征约束的视觉 SLAM 算法研究

论文作者　程浩
校内导师　陈孟元　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：2210330113

题　目　基于感知增强与特征约束的视觉 SLAM 算法研究

题　目　Research on Visual SLAM Algorithm Based on Perceptual Enhancement and Feature Constraints

学生姓名： 程浩

校内导师： 陈孟元（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

论文答辩日期：2024 年　5 月　15 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：程 浩

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：程 指　　指导教师签名：

日期：2024 年 5 月 30 日　　日期：2024 年 5 月 30 日

基于感知增强与特征约束的视觉 SLAM 算法研究

摘　要

基于视觉的同时定位与地图构建（Simultaneous Localization and Mapping，SLAM）是指在没有先验信息的情况下，通过相机传感器获取图像来确定机器人所处的位置，并同时构建环境地图。视觉 SLAM 主要依据图像特征匹配来实现自身定位与地图构建，传统方法严重依赖图像特征，然而相机曝光期间，快速移动的物体会在图像上留下轨迹造成模糊，从而可能导致视觉 SLAM 算法失效，同时环境中运动的物体也会对算法产生较大影响。本文通过结合深度学习和几何特征的方法来解决上述问题，利用语义信息和特征约束来判断物体的实际运动状态，增强系统在运动模糊环境下的鲁棒性同时提高位姿精度。

针对移动机器人在识别模糊图像中易出现漏检测和错误检测的情况，传统视觉 SLAM 算法在运动模糊场景下特征点提取数量不足的问题，本文提出一种基于感知增强的视觉 SLAM 方法，设计一种单阶段去模糊化识别网络，由基于模糊区域关注模块的去模糊前端和基于增强识别机制的检测后端组成。将相机采集到的图像输入到模糊检测模块，该模块对输入图像进行清晰度计算，判断为模糊的图像输入至去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，剔除动态物体所在区域，提升了移动机器人的感知能力。

针对传统动态 SLAM 方法无法有效判断物体的运动状态以及剔除动态物体后特征点数量较少的问题，在感知增强策略基础上提出一

种基于特征约束的改进 SLAM 方法，利用重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场的物体运动判断模型，将动态点和静态点的筛选转化为特征点标签的二分类问题，并根据分类的动态点数量判断物体的实际运动状态，剔除筛选出的动态点集合，提升了 SLAM 系统的定位精度。

针对传统 SLAM 方法特征提取匹配耗时，全局计算描述符导致系统运行效率不高的问题，本文采取特征和光流相融合的跟踪策略，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算，改善了跟踪线程中由全局特征匹配引起的计算耗时情况，提高了 SLAM 系统的运行效率。

为验证本文所提算法的有效性，在公开的 TUM 数据集分别对基于感知增强的视觉 SLAM 方法与基于特征约束的改进 SLAM 方法进行测试，并与主流的 ORB-SLAM3、Crowd-SLAM、DS-SLAM、Dyna-SLAM 算法进行对比，验证了所提方法在运动模糊场景下的定位准确性与鲁棒性。在 Husky 移动机器人平台上进行真实场景测试，结果表明本文算法具有较强的适应性，为运动模糊环境下视觉 SLAM 研究提供参考依据。

关键词：同时定位与地图构建，运动模糊，感知增强，特征约束，运动判断

RESEARCH ON VISUAL SLAM ALGORITHM BASED ON PERCEPTUAL ENHANCEMENT AND FEATURE CONSTRAINTS

ABSTRACT

Visual Simultaneous Localization And Mapping (VSLAM) refers to the determination of the robot's location and the simultaneous construction of an environment map from images acquired by the camera sensor without a priori information. Visual SLAM mainly relies on image feature matching to achieve its own localisation and map construction. Traditional methods rely heavily on image features, however, during camera exposure, fast moving objects will leave trajectories on the image causing blurring, which may lead to the failure of the visual SLAM algorithm, and moving objects in the environment will also have a large impact on the algorithm. In this paper, we solve the above problems by combining deep learning and geometric features, using semantic information and feature constraints to determine the actual motion state of the object, and enhancing the robustness of the system in motion blurred environments while improving the positional accuracy.

Aiming at the mobile robot's susceptibility to missed detection and false detection in recognising blurred images, and the problem of insufficient number of feature points extracted by traditional vision SLAM

algorithms in motion blurred scenes, this paper proposes a vision SLAM method based on perceptual enhancement to design a single-stage deblurring recognition network, which consists of deblurring front end based on blurred region attention module and detection back end based on enhanced recognition mechanism Composition. The image captured by the camera is input to the blur detection module, which calculates the clarity of the input image, and the image judged to be blurred is input to the deblurring recognition network, otherwise the feature point extraction is carried out directly, and the data association is carried out according to the semantic information to eliminate the region where the dynamic object is located, which enhances the perceptual ability of the mobile robot.

Aiming at the problems that traditional dynamic SLAM methods cannot effectively judge the motion state of objects and the number of feature points is small after excluding dynamic objects, an improved SLAM method based on feature constraints is proposed on the basis of perceptual enhancement strategy, which constructs a feature constraint function by using the reprojection error, the polar distance, and the number of covariant frames, and fuses the semantic information and the feature constraints to construct the object motion judgement model of the global conditional random field. model, transforming the screening of dynamic and static points into a binary classification problem with feature point labels, and judging the actual motion state of an object based on the number

of classified dynamic points, eliminating the set of screened dynamic points, and improving the positioning accuracy of the SLAM system.

Aiming at the problems of time-consuming feature extraction and matching and inefficient operation of the system due to global computation of descriptors in traditional SLAM methods, this paper adopts the tracking strategy of fusion of features and optical flow, which divides the input image into matching frames and optical flow frames according to the need of supplementary features, extracts feature points in matching frames only, and tracks these feature points in subsequent optical flow frames, which reduces the descriptor computation in the nonessential image and improves the tracking threads. This reduces the descriptor computation in non-essential images, improves the computational time-consumption caused by global feature matching in the tracking thread, and improves the operational efficiency of the SLAM system.

In order to verify the effectiveness of the proposed algorithm in this paper, the visual SLAM method based on perceptual enhancement and the improved SLAM method based on feature constraints are tested in the publicly available TUM dataset, respectively, and compared with the mainstream ORB-SLAM3, Crowd-SLAM, DS-SLAM, and Dyna-SLAM algorithms, which demonstrate the proposed method's motion blur scenario's positioning accuracy and robustness. The real scene test on Husky mobile robot platform shows that the algorithm in this paper has

strong adaptability, which provides a reference basis for the research of visual SLAM in motion blur environment.

Keywords: Simultaneous Localization and Mapping, Motion Blur, Perceptual Enhancement, Feature Constraints, Motion Judgment

目 录

第 1 章 绪论

1.1 本文研究背景及意义

随着科技的不断发展，机器人技术在各个领域中得到了广泛的应用。机器人是一种能够自主执行任务的机械设备，它可以模拟和替代人类的动作和行为，具有高度智能化和自主决策能力[1][2]，机器人技术的出现不仅改变了人们的生活方式，也给工业、医疗、军事等领域带来了巨大的变革。机器人技术的应用范围非常广泛，在工业领域，机器人可以代替人工完成重复、危险或繁琐的工作，提高生产效率和质量；在医疗领域，机器人可以进行精确的手术操作，减少手术风险和恢复时间；在军事领域，机器人可以执行军事任务，减少人员伤亡和风险，此外，机器人还可以应用于家庭、教育、娱乐等领域，为人们提供更加便利和丰富多样的服务。近年来，机器人技术的发展在全球范围内都受到了广泛关注和重视，各个国家都在加大对机器人技术的研发和应用力度，以提升自身在科技领域的竞争力。随着计算机视觉和机器学习技术的发展，移动机器人视觉 SLAM 技术取得了长足的进步，伴随该技术研究的不断深入，移动机器人在未知环境下进行自主导航和任务执行的能力不断增强。现代 SLAM 技术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便利和效益。移动机器人应用范围近年来逐渐扩大到技术要求更高的复杂场景，这就对传统移动机器人技术提出了更高的要求，因此复杂环境下的移动机器人技术已经成为热门研究领域[3][4]。

移动机器人视觉 SLAM 是一种利用机器人的视觉感知能力同时实现自主定位和地图构建的技术[5][6][7]，作为机器人导航和环境理解的关键技术之一，该技术的核心是同时解决定位和地图构建两个互相依赖的问题。定位是指确定机器人在已知地图中的位置和姿态，而地图构建是指从头开始构建环境的地图。SLAM 技术通过不断地观测和更新机器人的位置和地图信息，利用概率滤波、优化算法等方式实现对机器人状态的估计和地图的更新[8][9]。通过移动机器人的传感器获取环境中的特征信息，并对这些信息进行处理和分析，实时地估计机器人的位置和姿态，并生成环境地图。

在自动驾驶领域，SLAM 技术具有关键作用，有助于实现车辆的实时定位和环境地图构建。通过传感器数据，无人车能够在未知环境中确定自身相对于周围环境的准确位置[10]并实时构建和更新车辆周围环境的地图。相较于激光雷达，相机传感器价格低廉且能够捕捉丰富的图像信息，包括颜色、纹理和形状等，相机在 SLAM 应用中能够提供更全面的环境感知，视觉 SLAM 以相机作为主要传感器，包括单目相机、双目相机以及 RGB-D 相机。在传统的视觉 SLAM 系统中，物体的运动通常被视为噪声，对于移动的物体，可能引起系统的定位误差。现实道路场景通常比较复杂，环境中存在较多的运动的目标，如走动的行人、马路上运动的车辆等，同时由于被拍摄对象在相机曝光期间内移动时，物体会在图像上留下一条或多条轨迹，从而导致图像模糊[11]，模糊现象通常出现在拍摄运动场景、低光条件下或相机不稳定等情况。视觉 SLAM 系统通常依赖从图像中提取关键点和特征，然后通过特征匹配来定位和建图，模糊图像往往导致关键点提取数量不足，目标检测精度较低等问题，同时环境中的动态物体不仅会造成运动模糊，还会引起错误的特征匹配，进而导致轨迹漂移。

随着深度学习理论的迅速发展，视觉 SLAM 领域也经历了许多创新和变革，深度学习在视觉 SLAM 中的应用取得了显著的进展，提升了系统的鲁棒性、精度和适应性。例如通过卷积神经网络（CNN）提取图像特征[12]，这些深度特征适用于各种环境条件下的定位和地图构建，尤其是在有动态物体、光照变化或复杂纹理的情况下，常规图像特征可能效果较差。现阶段结合深度学习的 SLAM 方法仍有一定的局限性，当机器人识别动态场景时，相机和拍摄物体发生相对运动导致获取的图像模糊，此时机器人无法获取模糊图像中的语义信息，特征点数量不足并造成错误的特征匹配，进而导致定位误差和轨迹漂移[13][14]。

综上所述，针对传统视觉 SLAM 系统在面对运动模糊场景时出现的问题，本次研究以高效获取复杂场景下的语义信息，提高系统在实际应用中的鲁棒性，更好地适应各种场景和条件为目的，提出一种面向运动模糊场景下基于感知增强和特征约束的视觉 SLAM 方法，结合深度学习和几何约束来解决复杂环境下机器人的定位精度不高的问题，有效提高了视觉 SLAM 系统感知能力和鲁棒性。

1.2 国内外研究现状

1.2.1 视觉 SLAM 研究现状

SLAM 的基本目标是在环境中通过使用传感器数据来获取有关环境的信息，包括三维空间的地图和设备的当前位置，SLAM 有多种不同的实现方式，主要根据使用的传感器类型、地图表示方法和定位技术等方面的差异来进行分类。例如激光 SLAM 使用激光雷达作为传感器，通过测量激光束在环境中的反射来获取三维点云数据，从而构建环境地图并实时定位设备[15]，但由于激光雷达较为昂贵且视觉传感器可以感知更加丰富的环境信息，因此视觉 SLAM 广泛应用于自动驾驶、机器人导航、室内定位等领域。

视觉 SLAM 可以根据其实现方法、使用的技术和应用领域分类，通常包括直接法和特征点法两种类型[16][17]。基于直接法的视觉 SLAM 根据图像的像素亮度信息估计相机的运动，无需提取计算特征点和描述符，避免了特征计算的同时具有一定的鲁棒性，直接法要求相机运动变化较小并满足灰度一致性的假设，根据处理的像素数量的不同被分为稀疏直接法、稠密直接法和半稠密直接法三种。稀疏直接法仅对图像中的小部分像素进行处理，通常先选取关键点，然后通过最小化这些关键点之间的光度误差来估计相机的位姿或深度信息，稀疏直接法的优势在于计算效率相对较高，更适用于需要实时应用的场景。稠密直接法处理图像中每个像素的亮度变化，而不仅仅是选择性的关键点，通过构建整个图像的稠密运动场估计相机的运动、深度信息或场景的几何结构，该方法通常生成稠密地图，图像中的每个像素都包含有地图信息。稠密直接法的优势在于能够获得更为详细和准确的结果，但计算成本相对较高，尤其是在处理高分辨率图像时，需要较大的内存空间。半稠密直接法介于稀疏和稠密之间，选择图像中的一部分像素进行处理，这样可以在获得相对较为详细的结果的同时，减少计算负担，平衡了精度和计算效率。根据任务的性质、数据的特点以及系统的要求，可以灵活选择合适的方法。

近年来，基于直接法的视觉 SLAM 不断涌现。Newcombe 等人[18]于 2011 年提出 DTAM 系统，该系统作为单目视觉 SLAM 系统，直接计算关键帧中的每个像素的信息并构建环境的稠密三维地图，系统利用图像序列中连续帧之间的像素

匹配估计相机位姿，根据图像灰度信息，通过最小化当前帧与参考帧之间的像素亮度误差来优化相机运动轨迹，针对特征缺失、图像模糊等情况适应性较高，但无法做到实时运行。2014 年 Jakob Engel 等人[19]提出了 LSD-SLAM，这是一个大尺度半稠密的视觉 SLAM 系统，并成功的移植到了双目和 RGB-D 相机上，该系统提出了一种用来直接估计关键帧之间相似变换、尺度感知的图像匹配算法，在 CPU 上实现了半稠密场景的重建，提取梯度较为明显的像素，然后使用归一化后的光度误差对图像的梯度进行跟踪，但依然无法解决尺度漂移问题。同年 Forster 等人[20]提出了 SVO（Semi-direct Visual Odometry）算法，该算法仅依赖于图像中的一些稀疏关键点，而无需对其描述子进行计算，对一定范围内像素进行特征匹配，作为基于稀疏直接法的视觉里程计，因此即使计算平台条件较差，系统也能有理想的速度。

基于特征点法的视觉 SLAM 是一种通过提取、描述和匹配图像中的特征点进行同时定位与地图构建的方法，由于特征点法具有尺度不变性，能在不同尺度的图像中检测相似的特征点，因此基于特征点法的视觉 SLAM 是当下主流的方法。2007 年 Davision 等人[21]提出了 Mono-SLAM，该系统通过在每帧图像中提取稀疏的特征点来进行位姿估计，并使用单目相机完成实时的定位建图，能够在概率框架内在线运行，但该系统只适用于小规模运动，逐帧进行位姿估计和建图，计算复杂度高。为了解决计算复杂度的问题，同年 Gerorg Klein 等人[22]提出 PTAM，不再使用单线程框架，提出了跟踪和建图分离的前后端框架，选取图像中一些具有代表性的作为关键帧，仅处理关键帧中的特征点，减少数据冗余，提高了计算速度，但该方法缺少闭环检测模块，因此会导致累计误差，最终影响定位和建图的准确性。2015 年 Mur-Artal 等人[23]在 PTAM 的基础上提出了基于单目相机的 ORB-SLAM，该系统使用了跟踪、建图和闭环检测三线程并行的方案，跟踪阶段提取图像中的特征点并初步估计当前相机位姿，在闭环检测模块中将关键帧进行相似比对，更新相机位姿，极大提高了计算效率。2017 年 Mur-Artal 等人推出了 ORB-SLAM2[24]，该系统在 ORB-SLAM 的基础上扩展了双目和 RGB-D 相机，并加入了全局重定位功能，提高了系统闭环检测的性能。2020 年 Mur-Artal 等又在 ORB-SLAM2 的基础上提出了 ORB-SLAM3[25]，增加了惯性测量单元，实现了视觉和惯性的的结合。

1.2.2 动态场景下视觉 SLAM 研究现状

传统视觉 SLAM 在静态环境下取得了较高精度，但现实环境中物体的运动会导致图像模糊，造成特征点提取数量不足以及误匹配等问题，从而影响地图构建和机器人定位。由于运动造成的模糊等一系列问题对 SLAM 发展提出了极大的挑战，基于静态场景的假设很大程度上限制了视觉 SLAM 的应用，目前针对动态场景最直接的方法就是识别环境中的动态目标，并剔除在这些物体上提取的特征点，减少动态点对系统的影响，从而提高系统鲁棒性。当下解决动态 SLAM 问题的方法主要分为两类：基于特征的方法和依赖语义信息的方法[26][27]。

基于特征的 SLAM 方法主要依赖图像中的几何约束来检测运动情况异常的特征点，进而实现对异常特征点的剔除。Wang 等人[28]提出一种针对室内动态场景的 RGBD-SLAM，该方法利用几何约束实现运动目标的检测，通过光度误差来判断动态点。Dai 等人[29]提出利用地图点间的关联，将特征地图中静态点和动态点实时分离，并排除动态物体的干扰，使得算法有较高的鲁棒性。Gao 等人[30]提出一种集成运动检测算法的半直接 SLAM 算法，通过最小化光度误差实现稀疏图像匹配并估算相机位姿。Yang 等人[31]将特征点剔除分为两阶段，第一阶段利用几何约束初步过滤动态点，第二阶段结合 RANSAC 算法进一步剔除动态点。

随着深度学习技术的不断发展，越来越多的学者尝试引入深度学习方法来处理动态 SLAM，依赖语义信息的视觉 SLAM 方法不断提出。相较于利用传统特征的方法，系统可以获取高层次的语义信息，并直接通过语义信息初步分割动态区域。在计算机领域，语义分割网络和目标检测算法是常用的两种获取场景语义信息的方式，在 SLAM 系统中加入语义分割网络，将图像中的不同物体或区域分割成语义上有意义的区域，标识出不同物体和场景的位置，提高系统的感知和理解能力。Bescos 等人[32]结合 Mask RCNN 对场景中运动的目标进行分割，利用多视图几何对特征点进行运动判断，剔除动态区域的同时进行背景修复，有效提升了系统在动态环境下的定位精度。Yu 等人[33]在 ORB-SLAM2 中引入 SegNet 网络，并将语义分割网络和运动一致性相结合，分割出图像中的人形区域的同时构建对极几何约束，剔除动态特征点后建立全局一致的静态八叉树地图。Fan 等人[34]提出 Blitz-SLAM，利用 BlitzNet 获取语义信息，然后结合极线约束过滤环境中的动态特征点，在构建稠密语义地图的同时消除了运动物体造成的噪声影响。

基于目标检测算法的视觉 SLAM 系统根据目标检测网络的不同可以分为单阶段和两阶段两种。单阶段的常见网络有 YOLO 和 SSD[35][36]，而两阶段模型代表有 FasterRCNN[37]，两阶段网络会先产生目标候选区域，再进行分类和位置修正，由于 SLAM 系统对实时性要求较高，所以研究人员一般选择单阶段的目标检测网络。Yang 等人[38]提出的 SGC SLAM 利用 YOLOv3 检测并剔除运动目标，特征点的动态性则根据基础矩阵进行判断。Zhong 等人[39]提出基于 SSD 目标检测网络的 Detect SLAM，为了进一步提高系统速度，仅在关键帧中获取图像语义信息，使用跟踪算法追踪目标框以传递语义信息。Soares 等人[40]提出的 Crowd-SLAM 通过改进的 YOLO Tiny 目标检测网络在拥挤环境中检测行人，通过剔除目标框中的特征点获得稳定静态特征，但现实场景中往往有丰富的语义信息，只识别人语义信息无法充分利用且拥挤场景下人的检测框过多，导致人占据视角过大的情况下也会导致跟踪丢失的情况。

1.3 研究内容与章节安排

由于运动导致的模糊等问题严重影响了 SLAM 系统的鲁棒性，因此本文的研究内容主要针对运动模糊场景下移动机器人的定位问题，提出一种基于感知增强和特征约束的视觉 SLAM 方法。首先，将相机采集到的图像输入到模糊检测模块，该模块对输入图像进行清晰度计算，判断为模糊的图像输入至设计的单阶段去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，提升了移动机器人的感知能力；其次，在获取可靠语义信息的基础上，利用重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束相融合以构建全局条件随机场物体运动判断模型，将动态点和静态点的筛选转化为特征点标签的二分类问题，并根据分类的动态点数量判断物体的实际运动状态；最后，采取特征和光流相融合的跟踪策略，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算。论文的总体框架如图 1-1 所示。

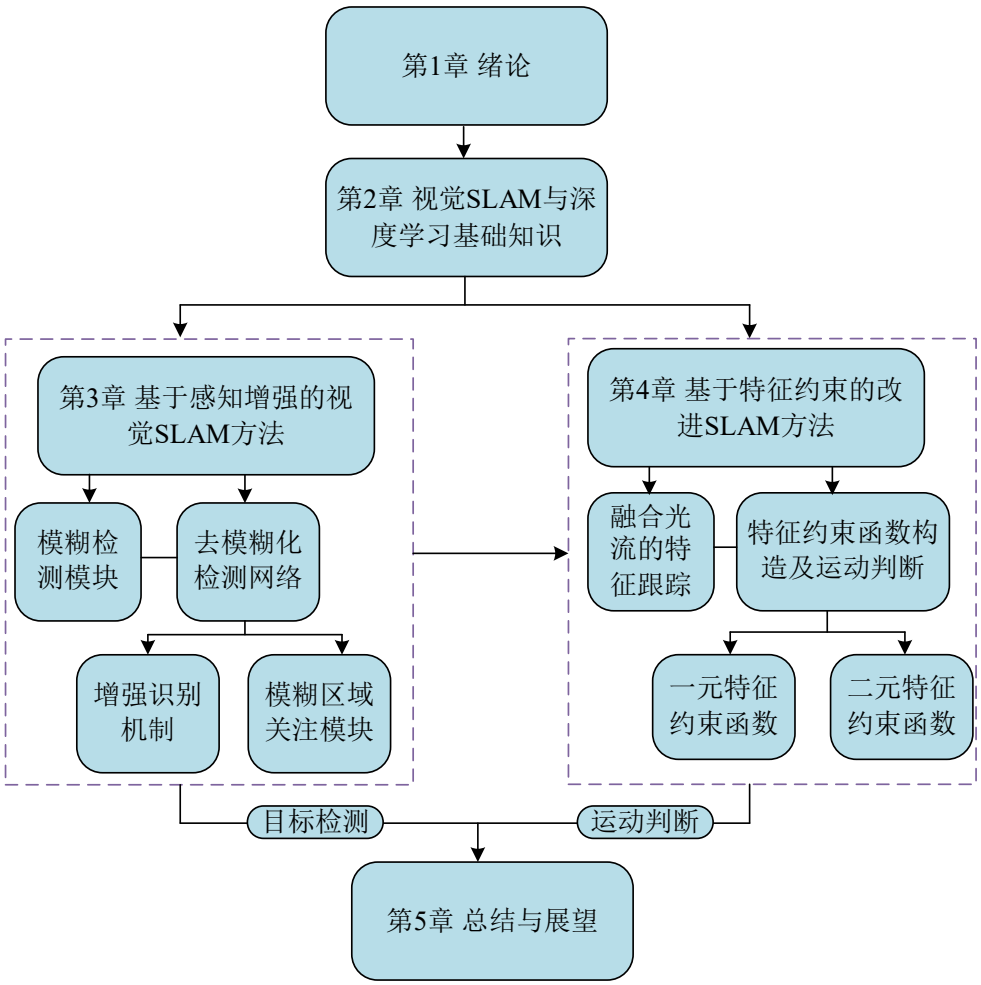


图 1-1 论文总体框架

各章节内容安排如下：

第 1 章，绪论。首先介绍了移动机器人技术的背景和现状，重点分析了移动机器人 SLAM 技术的应用和核心问题，然后阐述了传统视觉 SLAM 技术发展的局限性和国内外研究现状，特别是动态场景下基于深度学习的视觉 SLAM 技术发展，最后对本文的研究内容与后续章节安排进行了简要介绍。

第 2 章，视觉 SLAM 与深度学习基础知识。首先将视觉 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、视觉里程计等部分，最后阐述了深度学习的基础知识和经典算法，包括生成对抗网络和目标检测算法，为后续的研究打下了基础。

第 3 章，基于感知增强的视觉 SLAM 方法。针对移动机器人在识别模糊图像中易出现漏检测和错误检测的情况，传统视觉 SLAM 算法在运动模糊场景下特征点提取数量不足的问题，本章设计一种单阶段去模糊化识别网络，由基于模糊区域关注模块的去模糊前端和基于增强识别机制的检测后端组成。将相机采集

到的图像输入到模糊检测模块，该模块对输入图像进行清晰度计算，判断为模糊的图像输入至去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，剔除动态物体所在区域，提升了移动机器人的感知能力。

第 4 章，基于特征约束的动态 SLAM 方法。针对传统动态 SLAM 方法无法有效判断物体的运动状态以及剔除动态物体后特征点数量较少的问题，本章基于重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场的物体运动判断模型，将动态点和静态点的筛选转化为特征点标签的二分类问题。针对传统 SLAM 方法特征提取匹配耗时，全局计算描述符导致系统运行效率不高的问题，采取特征和光流相融合的跟踪决策，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算。

第 5 章，总结与展望。总结了本文的工作内容和创新之处，在此基础上讨论了当前研究不足之处并展望了未来工作，指出了下一步可以研究的方向。

第 2 章 视觉 SLAM 与深度学习基础知识

本章首先将视觉 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，重点讨论了视觉里程计中特征提取、匹配和相机位姿估计部分。最后阐述了深度学习的基础知识和经典算法，包括生成对抗网络和目标检测算法，为后续的研究打下了基础。

2.1 视觉SLAM数学模型

SLAM 中最基本的定位和建图问题，可以假设为未知环境下移动机器人携带相机传感器运动的过程，根据相机在离散时间采集到的图像进行机器人位姿的求解。在离散时间 $t\in[1,k]$ 的各时刻机器人自身的位置也在不断变化，分别记为 $\boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_k$，相机在 k 时刻观测到部分路标，而地图是由许多个路标组成，这些观测到的路标可以记为 $\boldsymbol{y}_1,\boldsymbol{y}_2\cdots,\boldsymbol{y}_n$。基于此可以得到机器人位置的变化以及某个时刻的观测，也就完成了定位和建图。数学语言描述上述过程如式（2-1）所示：

$$\begin{cases}\boldsymbol{x}_k=f\left(\boldsymbol{x}_{k-1},\boldsymbol{u}_k\right)+\boldsymbol{w}_k & k=1,\cdots,N\\ \boldsymbol{z}_k=h\left(\boldsymbol{y}_i,\boldsymbol{x}_k\right)+\boldsymbol{v}_{k,i} & i=1,\cdots,M\end{cases} \tag{2-1}$$

其中，$\boldsymbol{x}_k$ 和 $\boldsymbol{x}_{k-1}$ 分别为第 k 时刻和 k-1时刻机器人的位置，$\boldsymbol{u}_k$ 为传感器传入数据，$\boldsymbol{w}_k$ 为符合零均值高斯分布函数的噪声项，$\boldsymbol{z}_k$ 为观测数据，$h\left(\boldsymbol{y}_i,\boldsymbol{x}_k\right)$ 为 $\boldsymbol{x}_k$ 位置对路标 $\boldsymbol{y}_i$ 的观测方程，$\boldsymbol{v}_{k,i}$ 为此次观测含有的外部噪声。

当获取到传感器数据和运动测量读数时，SLAM 可以转化为状态估计问题，该问题受到运动方程形式和噪声模型的影响。早期，以 EKF 滤波器为基础的方法占主导地位，但目前普遍采用图优化技术进行状态估计。

2.2 经典SLAM系统框架

经过几十年的发展，视觉 SLAM 形成了学术界普遍公认的框架，具体如图 2-1 所示，整体框架包括五个部分，分别为获取传感器数据、视觉里程计、非线

性优化、闭环检测和建图。传感器数据是 SLAM 的关键组成部分，视觉 SLAM 通过相机获取的图像估计相机位姿及地图构建。视觉里程计作为系统前端部分，通过分析相邻帧间的特征匹配信息来估计相机的运动。非线性优化在 SLAM 中通常被称为后端优化，该部分利用前端处理结果和传感器数据优化整个系统状态，提高轨迹的全局一致性。闭环检测模块是在建立环境地图和机器人定位的过程中，检测机器人是否返回到先前访问过的位置，一旦检测到路径闭环，系统会利用闭环检测结果来校正地图和机器人的位姿。

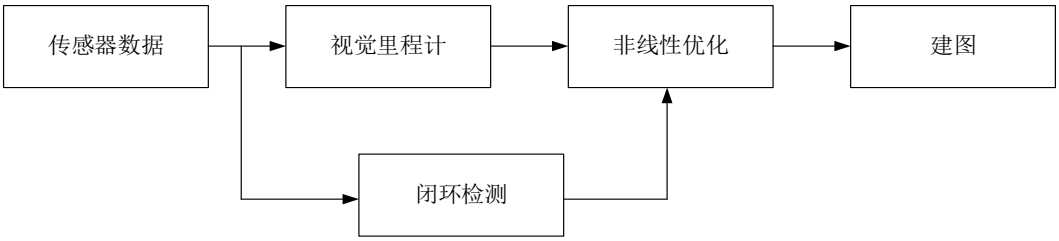


图 2-1 经典视觉 SLAM 框架

2.2.1 传感器数据

视觉 SLAM 中的传感器一般是固定在机器人本体上，主要通过相机直接获取外部环境图像信息，按照工作方式的不同，相机可以分为单目、双目、RGB-D 三类，相机模型如图 2-2 所示。单目相机仅含有一个摄像头，结构简单、体积小，但只能提供二维图像，无法获取深度信息。双目相机包含左右两个相机，可以利用视差来计算物体到相机的距离，从而获取深度信息，但双目相机的配置和标定较为复杂，且测量范围受限于基线和分辨率。RGB-D 相机是一种集成了彩色图像和深度信息的传感器，而传统相机只能获取彩色图像，深度信息通常以像素到物体表面的距离来表示，RGB-D 相机的深度信息通过使用红外或其他方法来获取，其中包括结构光法、飞行时间法和双目立体视觉等技术。

（a）单目相机　　（b）双目相机　　（c）RGB-D 相机

图 2-2 相机类型

2.2.2 视觉里程计

视觉里程计是视觉 SLAM 系统的关键模块，其主要任务是通过传感器获取的相邻帧间的特征匹配信息来估计相机的运动，作为整个系统的前端部分，视觉里程计不依赖于预先构建的地图，计算相邻帧间的运动，为后端优化提供初步估计值。传统视觉里程计基于实现原理可以分为特征点法和直接法，本文的整体框架围绕特征点法展开且直接法基于图像中的所有像素进行相机位姿估计，计算复杂度高且对光照敏感，故本节主要介绍基于特征点法的视觉里程计相关技术。

特征点是图像中选取的具有代表性的点，好的特征点一般具有可重复性、可区别性、高效率和本地性等特点，当相机旋转或者距离发生改变，特征点的性质仍能保持不变，在不同角度下可以被准确识别和匹配。在计算机视觉领域中，研究者设计了许多稳定表示局部图像特征的特征点，包括 SIFT、FAST、ORB 等[41][42][43]，特征提取如图 2-3 所示。由于 ORB 特征点提取速度快且对光照、尺度、旋转均具有较好的鲁棒性，是目前视觉里程计最常用的特征。

（a）SIFT　　（b）FAST　　（c）ORB

图 2-3 特征点提取

通过计算相邻图像帧特征描述子之间的距离或相似性度量进行匹配，在匹配成功的特征点集合上，利用这些特征点的空间位置信息来估计相机的位姿。由于选择的相机不同，获取的数据信息也不同，通常可以分为 2D 点到 2D 点的对极几何法[44]、3D 到 3D 的 ICP 法[45]和 3D 到 2D 的 PnP 法[46]，重投影误差和对极几何作为 SLAM 中重要特征，充分利用这些信息不仅可以用于相机位姿的求解，还可以作为几何约束进行运动判断，因此本节主要介绍对极几何法和 PnP。

对极几何是描述两帧图像之间几何关系的数学框架，通过对特征点的匹配和几何约束来获取同一空间点在不同成像平面上的映射，几何模型如图 2-4 所示。当相机在不同角度获取两帧图像 I_1 和 I_2，假设第一帧到第二帧的运动为 $\boldsymbol{R}$ 和 $\boldsymbol{t}$，

相机中心分别为O_1和O_2，空间点$\boldsymbol{P}$在图像I_1中有一个特征点$\boldsymbol{p}_1$，在图像I_2中对应特征点$\boldsymbol{p}_2$，则$\boldsymbol{p}_1$和$\boldsymbol{p}_2$是通过特征匹配得到的匹配对。$\overrightarrow{\boldsymbol{o}_1\boldsymbol{p}_1}$和$\overrightarrow{\boldsymbol{o}_2\boldsymbol{p}_2}$在空间中相交于点$\boldsymbol{P}$，$O_1$ O_2的连线与像平面I_1 I_2相交点$\boldsymbol{e}_1$ $\boldsymbol{e}_2$被称为极点，极点和像平面中特征点的连线$\boldsymbol{l}_1$、$\boldsymbol{l}_2$为极线。

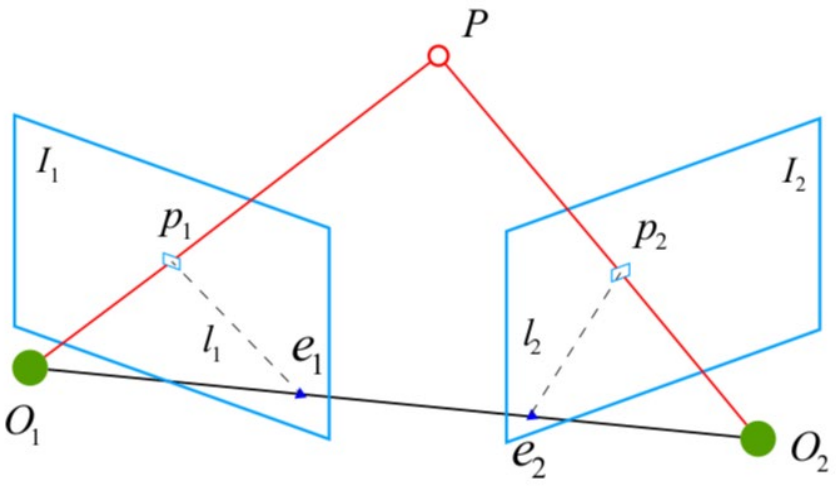


图 2-4 对极几何约束

假设$\boldsymbol{P}=[X,Y,Z]^{\mathrm{T}}$，$\boldsymbol{p}_1=[x,y]^{\mathrm{T}}$，$\boldsymbol{p}_2=[x',y']^{\mathrm{T}}$，根据匹配关系可以得到式（2-2）：

$$\begin{cases} s_1\boldsymbol{p}_1=\boldsymbol{K}\boldsymbol{P} \\ s_2\boldsymbol{p}_2=\boldsymbol{K}\left(\boldsymbol{R}\boldsymbol{P}+\boldsymbol{t}\right) \end{cases} \tag{2-2}$$

其中，$\boldsymbol{K}$代表相机内参矩阵，$\boldsymbol{R}$和$\boldsymbol{t}$表示相机的旋转和平移，进一步可以得到对极约束，具体如式（2-3）所示：

$$\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{K}^{-\mathrm{T}}\boldsymbol{t}^{\wedge}\boldsymbol{R}\boldsymbol{K}^{-1}\boldsymbol{p}_1=0 \tag{2-3}$$

若将约束中的旋转和平移部分转化为矩阵形式，具体如式（2-4）所示：

$$\begin{cases} \boldsymbol{E}=\boldsymbol{t}^{\wedge}\boldsymbol{R} \\ \boldsymbol{F}=\boldsymbol{K}^{-\mathrm{T}}\boldsymbol{E}\boldsymbol{K}^{-1} \end{cases} \tag{2-4}$$

其中，$\boldsymbol{E}$和$\boldsymbol{F}$分别命名为基础矩阵和本质矩阵，通过基础矩阵或本质矩阵来估计相机的位姿。

2D 点和 3D 点之间的运动通常采用 PnP 法，通过最小化重投影误差来估计相机位姿，使三维空间点经过相机变换后投影到图像平面上的像素位置与观测到的像素位置之间的误差最小化，具体计算如式（2-5）所示。重投影误差表示通过估计的相机位姿将三维空间点投影到图像平面上时，计算其对应的像素位置与实

际观测到的像素位置之间的差异，几何模型如图 2-5 所示。求解该最小二乘问题，一般优化算法将重投影误差作为损失函数，并通过调整相机的内外参数来使损失最小化进行求解。

$$\boldsymbol{T}^{*}=\arg\min_{\boldsymbol{T}}\frac{1}{2}\sum_{i=1}^{n}\left\|\boldsymbol{p}_{i}-\frac{1}{s_{i}}\boldsymbol{K}\boldsymbol{T}\boldsymbol{P}_{i}\right\|_{2}^{2} \tag{2-5}$$

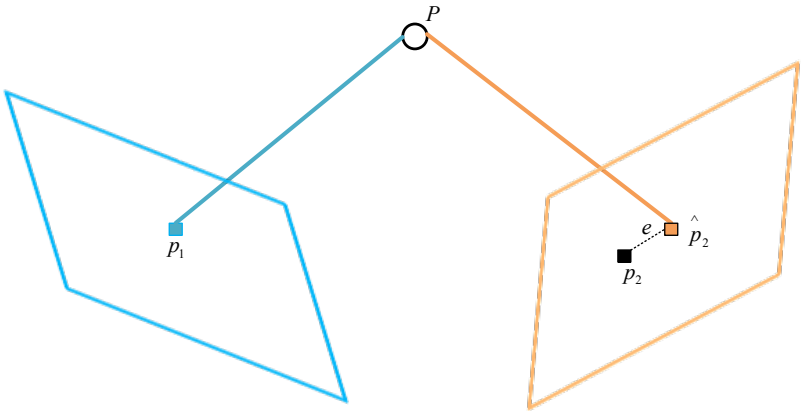


图 2-5 重投影误差

两组 3D 点间的运动估计可以通过 ICP 求解，其具体过程可以表示为两幅匹配的 RGB-D 图像中有一对 3D 点，须满足式（2-6）：

$$\boldsymbol{p}_{i}=\boldsymbol{R}\boldsymbol{p}_{i}'+\boldsymbol{t} \tag{2-6}$$

其中旋转矩阵 $\boldsymbol{R}$ 和平移矩阵 $\boldsymbol{t}$ 就是所需要优化求解的结果。通过构建最小二乘问题，求使误差平方和达到最小的 $\boldsymbol{R}$ 和 $\boldsymbol{t}$，具体计算如式（2-7）所示：

$$\min_{\boldsymbol{R},\boldsymbol{t}}\boldsymbol{J}=\frac{1}{2}\sum_{i=1}^{n}\left\|\left(\boldsymbol{p}_{i}-\left(\boldsymbol{R}\boldsymbol{p}_{i}'+\boldsymbol{t}\right)\right)\right\|_{2}^{2} \tag{2-7}$$

2.2.3 非线性优化

SLAM 前端视觉里程计部分进行初步位姿估计的同时会不断产生误差，并会对地图构建产生较大影响，通过构建大规模的优化模型，后端非线性优化改善了初始估计的位姿，从而减少累计误差。后端优化有基于滤波和图优化两类，由于基于滤波的优化方法只考虑相邻帧间数据关系以及图优化理论的兴起，图优化算法逐渐成为 SLAM 中的主流方案[47][48]。

图优化算法是一种在图形结构中寻找最优解的方法，其中图的节点表示待优化的变量，而边则表示这些变量之间的约束关系，通过调整节点的取值，优化算法尝试在满足约束条件的同时最小化成本函数，图优化模型如图 2-6 所示。顶点

通常代表着待估计的变量，例如相机的位姿或者环境中的路标点的位置，实线表示相机之间的位姿关系，虚线表示相机与观测到的路标点间的空间关系，两者构成了图优化的边。

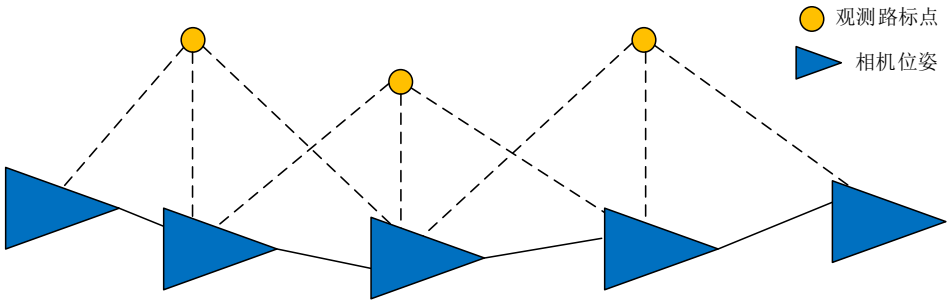
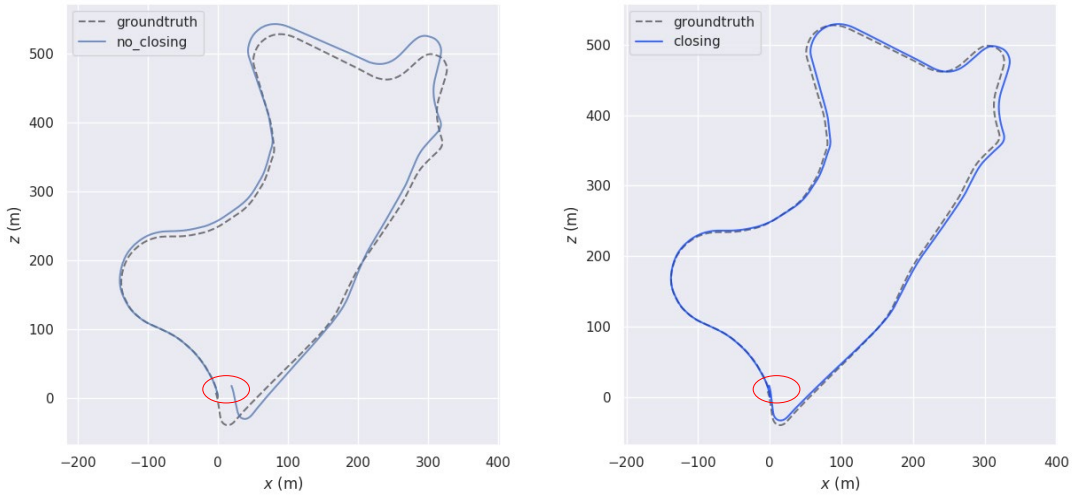


图 2-6 图优化模型

2.2.4 闭环检测

经过后端优化后位姿仍难与真实位姿保持一致，通过加入闭环检测模块修正轨迹，减少地图中的漂移误差。在建立环境地图和机器人定位的过程中，闭环检测通过判断机器人是否返回到先前访问过的位置，一旦检测到路径闭环，系统会利用闭环检测结果来校正地图和机器人的位姿，最终得到一个全局一致的地图，闭环检测如图 2-7 所示。图 2-7（a）中由于系统没有检测到机器人回到同一地点，无法形成闭环，导致算法与真实轨迹误差较大，由图 2-7（b）可知，当成功检测到闭环时，系统会进行闭环矫正消除累计误差，提高地图的一致性和准确性。

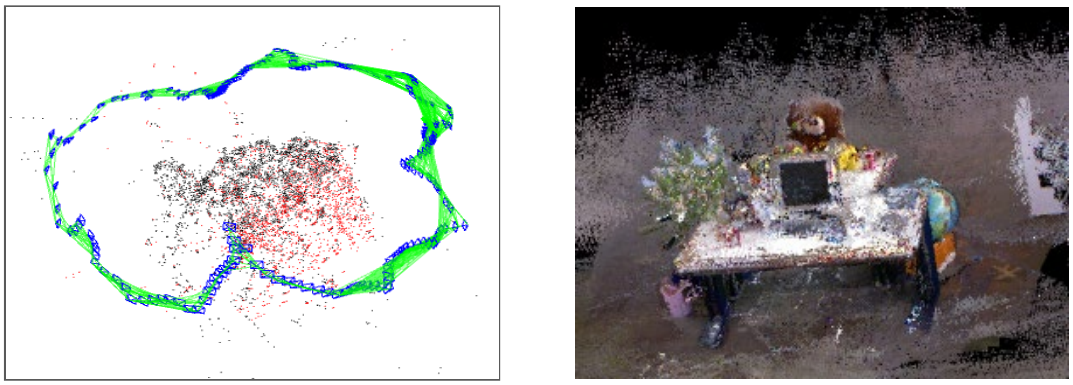


（a）未闭环　　（b）形成闭环

图 2-7 闭环检测

2.2.5 建图

建图作为 SLAM 主要目标之一，前端和后端都是为了更准确地建立路标点位置，所有已观测到的路标点集合组成了 SLAM 中的地图。根据地图的密度和表达方式可以分为稀疏地图和稠密地图，稀疏地图是由一组选定的关键点或地标组成的地图，其中只包含环境的显著特征信息，该地图适用于计算资源有限或实时性要求较高的应用，例如移动机器人导航和实时定位。稠密地图则包含了环境的更多细节，通常是通过网格、像素或者点云的形式来表示，由于该地图每个单元都包含环境的信息，可以提供更丰富的地图表达和感知，所以，稠密地图适用于需要更精细的环境建模和场景理解的应用，例如虚拟现实、自动驾驶和三维重建，稀疏点云地图和稠密地图如图 2-8 所示。



（a）稀疏点云地图　　（b）稠密地图

图 2-8 地图类型

2.3 深度学习基础

近年来，随着计算机性能的提升和大数据时代的到来，深度学习已经成为人工智能领域的重要技术，并在各种应用中取得了成功。深度学习是机器学习的一种分支，其核心思想是利用多层次的神经网络结构来学习和表示数据的复杂特征，以实现各种任务的自动化学习和预测。本节主要介绍生成对抗网络和目标检测算法[49][50]。

2.3.1 生成对抗网络

生成对抗网络是一种由生成器和判别器两个模块组成的无监督学习的深度

网络模型，该网络的核心原则为零和博弈，生成器接收一个随机向量作为输入，并映射到与真实数据相同的数据空间，其目标是学习生成逼真的假数据样本，以欺骗判别器。判别器接收来自真实数据集和生成器产生的假数据集的样本，并尝试将它们区分开，其目标是学习识别真实数据和生成数据之间的差异。在训练过程中，生成器和判别器相互对抗，生成器试图生成足够逼真的数据样本来欺骗判别器，而判别器则努力提高对真实和生成数据的识别准确性，如式（2-8）所示：

$$\min_{G}\max_{D}V(G,D)=\underset{x\sim P_r(x)}{E}\left[\log\left(D(\boldsymbol{x})\right)\right]+\underset{x\sim P_z(z)}{E}\left[\log\left(1-D\left(G(\boldsymbol{z})\right)\right)\right] \quad (2\text{-}8)$$

其中，$\boldsymbol{x}$ 和 $\boldsymbol{z}$ 分别表示输入的真实序列和随机噪声，$P_z(\boldsymbol{z})$ 和 $P_r(\boldsymbol{x})$ 分别表示噪声的概率分布和真实数据概率分布；$D(\boldsymbol{x})$ 和 $G(\boldsymbol{z})$ 分别表示判别器和生成器模型。

2.3.2 目标检测算法

目标检测算法可以在图像或视频中识别目标物体及其类别，同时确定目标在图像中的位置，可以分为单阶段检测器和两阶段方法两种，其中单阶段检测器相较于两阶段方法无需生成候选区域，通过单个神经网络直接预测目标的类别和位置，比如 SSD，YOLO 系列等，单阶段方法运行效率高，能满足视觉 SLAM 对实时性的要求。

YOLO 作为单阶段检测器的典型代表，采用直接回归方法获得目标信息，其网络部分主要分为卷积层、目标检测层和非极大抑制筛选层。卷积层主要对输入图像进行特征提取，以此提升模型的泛化能力，目标检测层将输入图像分割成多个网格，当目标中心处于某个网格中，则该网格负责在图像中定位对象的边界框，通过预测边界框的坐标、尺寸以及对象的类别概率，确定图像中目标的位置。目标检测层会输出目标多个重叠的边界框，非极大值抑制筛选层遍历所有检测到的边界框，并根据重叠程度和置信度分数来选择最佳的边界框，筛选后通常保留相邻领域里得分最高的边界框，而重叠较大的边界框则会被抑制。YOLOv5 算法目标检测图如图 2-9 所示，通过 YOLOv5 算法对图像中物体进行目标检测，充分获取图像中的语义信息，提升视觉 SLAM 系统对环境的感知能力。

（a）原始图像　　（b）目标检测图

图 2-9 目标检测

2.4 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生成对抗网络和目标检测算法，为深度学习和视觉 SLAM 的结合提供了理论基础。

第 3 章 基于感知增强的视觉 SLAM 方法

移动机器人在识别模糊图像中易出现漏检测和错误检测的情况，传统视觉 SLAM 算法在运动模糊场景下特征点提取数量不足的问题。针对这一问题，本章提出一种在运动模糊场景下基于感知增强的视觉 SLAM 方法，设计一种单阶段去模糊化识别网络，将相机采集到的图像输入到模糊检测模块，该模块对输入图像进行清晰度计算，判断为模糊的图像输入至去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，剔除动态物体所在区域并进行地图构建，本章技术路线如图 3-1 所示。

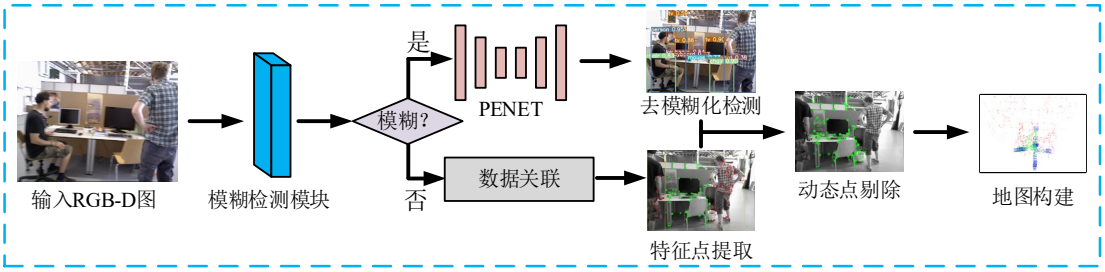
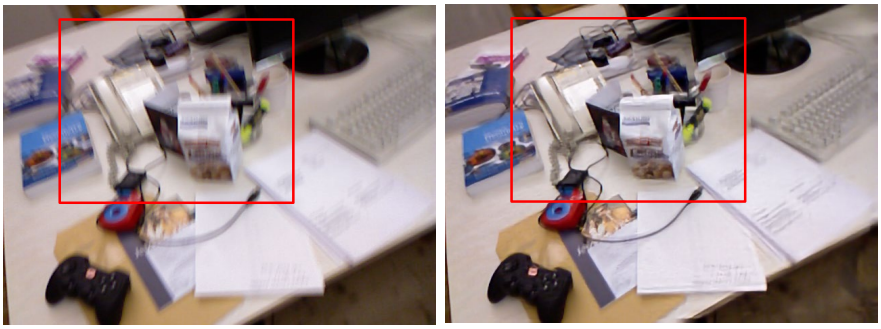


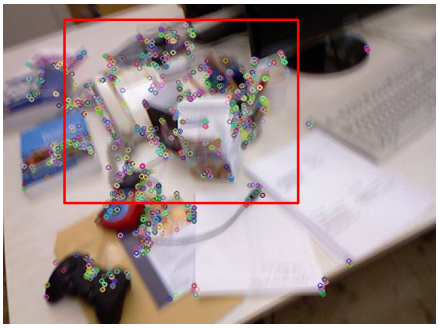
图 3-1 本章技术路线图

3.1 模糊图像与特征提取分析

为了直观展现模糊对于图像特征的影响，对真实场景中相机采集的对应清晰图像和模糊图像分别进行特征提取，如图 3-2 所示。由图 3-2（a）可知，由于在曝光期间相机抖动，导致图像存在明显模糊，图 3-2（b）是获取到的清晰图像，对比特征提取效果可知。当图像模糊时，角点的检测精度下降特征点提取数量会明显减少，同时模糊影响了图像局部灰度梯度，使得提取位置相较真实区域中特征点出现了偏移，错误的特征提取导致后续位姿估计产生较大误差。



（a）模糊图像　　（b）清晰图像


（c）模糊图像特征点

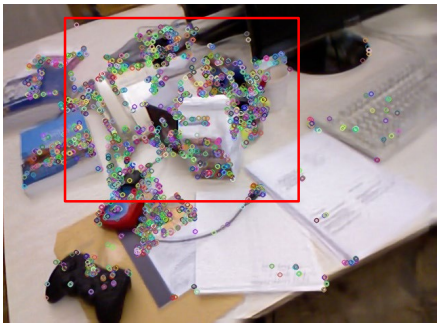
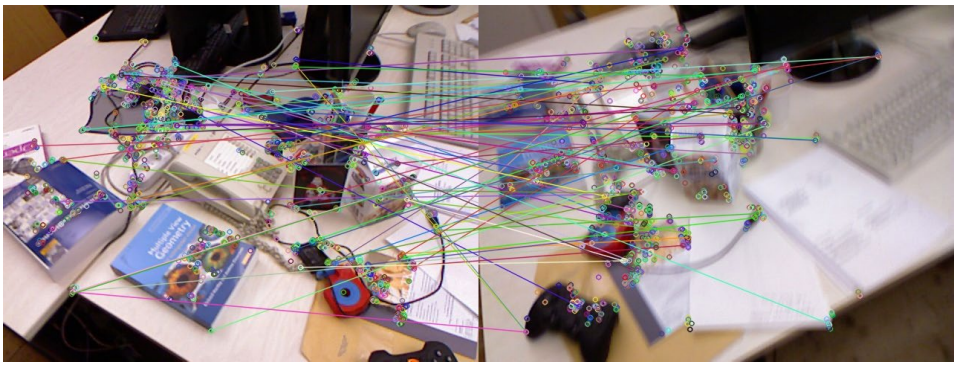

（d）清晰图像特征点

图 3-2 特征提取对比

图像模糊不仅对特征点提取造成较大误差，本章将进一步研究对特征匹配过程的影响，清晰图像和模糊图像特征匹配效果如图 3-3 所示。由图 3-3（a）可知，模糊图像在进行图像特征点匹配时，由于物体边缘振铃伪影与运动拖影像素影响了图像局部灰度梯度，导致前后图像帧特征点描述符相差极大，造成了图像特征点成功匹配数量较少且误匹配较多，这使得后续位姿估计易出现较大误差。


（a）模糊图像匹配

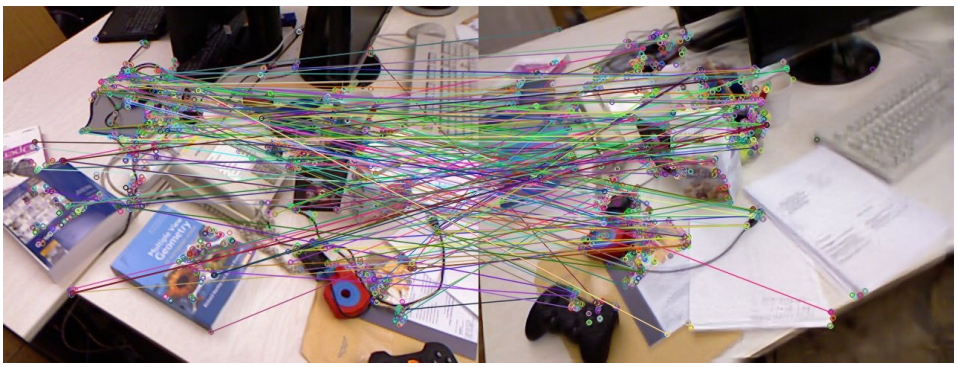

（b）清晰图像匹配

图 3-3 特征匹配对比

3.2 模糊检测模块

当传感器获取图像的过程中易受外界环境干扰造成图像模糊，模糊通常是由

于场景中物体与相机之间相对运动所造成，然而采集的图像并不都是模糊的。对每帧图像都进行去模糊会导致大量不必要的时间消耗，进而影响 SLAM 系统的实时性。本章提出一种模糊检测模块，该模块通过 3×3Laplacian 卷积核对输入图像进行卷积计算得到特征响应，通过计算特征响应方差判断图像是否模糊。

Laplacian 梯度函数采用拉普拉斯算子提取水平和垂直方向的梯度，定义如式（3-1）所示：

$$\boldsymbol{L}=\frac{1}{6}\begin{bmatrix}0 & 1 & 0\\ 1 & -4 & 1\\ 0 & 1 & 0\end{bmatrix} \tag{3-1}$$

基于 Laplacian 算子对每个像素进行卷积操作得到变换后的高频信息图像，然后对图像的高频分量求和，用高频分量和作为图像的清晰度评价标准。具体计算如式（3-2）所示：

$$D(\mathrm{f})=\sum_i\sum_j\left|G(i,j)\right| \quad \left(\left|G(i,j)\right|>T\right) \tag{3-2}$$

其中，$G(i,j)$是像素点(i,j)处 Laplacian 算子的卷积。对于一个 $M\times N$ 像素图像的清晰度评价函数如式（3-3）所示：

$$f=\frac{1}{M\times N}\sum_{(x,y)}z(x,y) \tag{3-3}$$

其中，$z(x,y)$是图像中某点(x,y)的灰度值。如果特征响应方差小于阈值则为模糊图像输入至本文设计网络，否则直接进行特征点提取。为验证模糊检测模块效果，本次实验从 TUM 数据集中分别选取一张清晰图像和一张模糊图像做验证，判断效果如图 3-4 所示。左侧图像由于计算得到的结果小于设定阈值，将该图像标记为模糊，否则为清晰图像，后续无需进行去模糊操作。

图 3-4 模糊检测

3.3 去模糊化检测网络设计

针对传统深度学习算法无法同时完成去模糊和目标检测，在识别动态场景时由于运动物体在图像中的边缘和细节丢失导致漏检测和错误检测，本章设计了一种融合模糊区域关注模块和增强检测的单阶段去模糊识别网络 PENET（Perceptual Enhancement Network）。该网络由基于模糊区域关注模块（Blurred Area Focus module，BAF-m）的去模糊前端和基于增强识别机制（Enhanced Recognition machine，ER-m）的检测后端组成。在网络前端加入 BAF-m 模块，通过获取和利用图像全局上下文先验信息，引导网络关注模糊像素区域，增强模糊区域边缘的细节信息，融合多尺度图像特征实现去模糊操作。在网络后端引入 ER-m，加强前端修复区域的像素权重并抑制背景的干扰，提升网络对于背景和前景的区分能力，提高模糊物体的识别率。本章提出的 PENET 网络结构如图 3-5 所示。

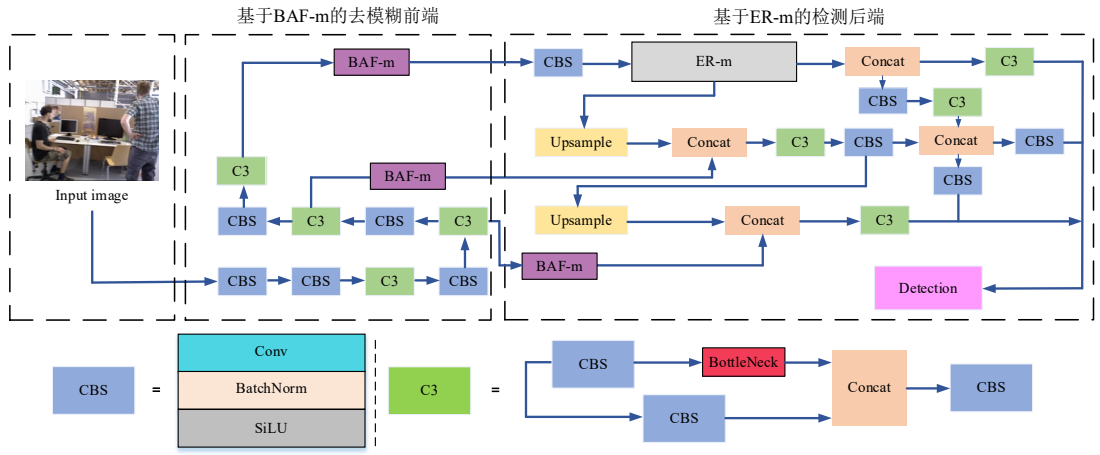


图 3-5 PENET

3.3.1 模糊区域关注模块

在网络初始阶段，前端提取的浅层特征图中所含语义信息少，未经充分处理且感受野受限，难以有效利用上下文信息，不利于处理复杂的模糊特征。在网络前端多个支路上引入 BAF-m 模块，通过扩大感受野获取全局上下文信息，BAF-m 模块由多尺度普通卷积和空洞卷积共同组成多分支结构，拼接后形成多通道特征图，通过横向扩展网络宽度，提升网络对图像中模糊区域的敏感性。BAF-m 模块结构如图 3-6 所示。

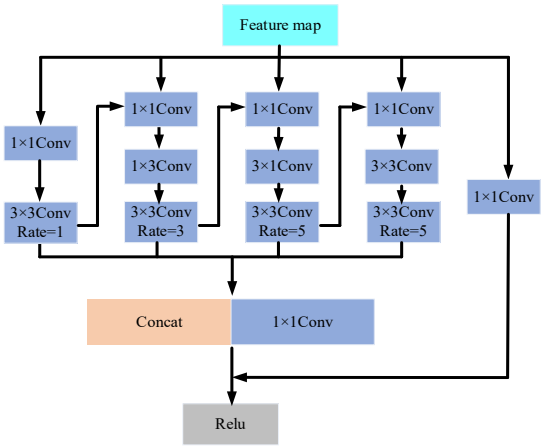


图 3-6 BAF-m 结构图

模块中每条支路均先对输入特征图进行 1×1 卷积进行降维，最右侧支路将输入端和输出端形成等价映射，增强模糊目标的特征表示。其余通路通过大小不同的普通卷积和 3×3 的空洞卷积并进行级联，加入的空洞卷积可以使特征图获取图像全局的上下文先验信息，增强模糊特征的表达能力，通过特征复用将每条支路的输出和下一分支的输入相连接，使得前面分支中的特征图再经过后面分支中的空洞卷积以获得不同尺度感受野，多尺度的感受野利用全局信息可以捕获复杂的模糊特征，最后将所有支路输出的结果进行拼接融合。模糊区域关注模块的计算如式（3-4）所示：

$$\begin{cases} D_1 = f_{dc=1}^{3\times3}\left[f_c^{1\times1}(F)\right] \\ D_2 = f_{dc=3}^{3\times3}\left[f_c^{1\times3}\left[f_c^{1\times1}(F)\right]\right] \oplus D_1 \\ D_3 = f_{dc=3}^{3\times3}\left[f_c^{1\times3}\left[f_c^{1\times1}(F)\right]\right] \oplus D_2 \\ D_4 = f_{dc=5}^{3\times3}\left[f_c^{3\times3}\left[f_c^{1\times1}(F)\right]\right] \oplus D_3 \\ D = Concat(D_1, D_2, D_3, D_4) \oplus f_c^{1\times1}(F) \end{cases} \tag{3-4}$$

其中，$f_c^{t\times q}$ 表示卷积核大小为 $t\times q$ 的普通卷积，$f_{dc=l}^{t\times q}$ 表示扩张率为 l 的空洞卷积，$Concat$ 是特征图拼接操作，F 表示输入特征图，$\oplus$ 表示特征图按位相加操作，D_v（$v=1\sim4$）表示第 v 条通路经过卷积后的特征图；D 表示融合后得到的总特征图。

为了更好地恢复图像原始信息，本章通过跨尺度实现不同特征图之间的融合，多尺度特征融合如图 3-7 所示，将 3 个 BAF-m 的输出特征进行 Resize 操作后拼接，然后使用卷积层对拼接后的特征进行融合，最后对将融合特征图进行上采样操作，将图像尺寸恢复到原始大小并得到清晰图像。

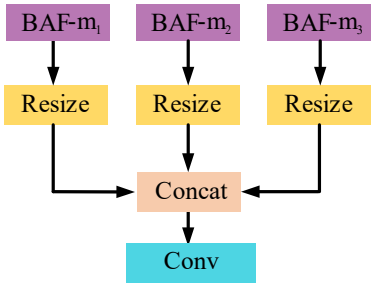


图 3-7 多尺度特征融合

3.3.2 增强识别机制

基于增强识别机制的检测后端将前端输出的去模糊特征作为输入，进行目标分类和边框回归。后端中高维特征图容易丢失目标的空间位置信息，因此在后端引入 ER-m，该模块不仅增强了目标区域中的特征通道权重，同时获取了感兴趣区域的空间位置，充分突出模糊区域的有效特征信息，以提高模型识别的准确率，增强识别机制结构如图 3-8 所示，对输入特征图分别沿 X 和 Y 方向进行平均池化，有效聚合空间坐标信息到特征图中，沿着两个空间方向对每个通道进行编码，得到第 c 个通道的输出如式（3-5）所示：

$$\begin{cases} y_c^h(h)=\dfrac{1}{W}\sum\limits_{0\le i\le W} x_c(h,i) \\ y_c^w(w)=\dfrac{1}{H}\sum\limits_{0\le j\le H} x_c(j,w) \end{cases} \tag{3-5}$$

其中，H 和 W 为池化核的大小，$y_c^h(h)$ 和 $y_c^w(w)$ 分别是高度为 h 和宽度为 w 的第 c 个通道的输出，$x_c(h,i)$ 和 $x_c(j,w)$ 分别是水平方向为 h 垂直方向为 i 的输入和水平方向为 j、垂直方向为 w 的输入。基于上述生成具有全局感受野和空间位置信息的特征图通过同一维度上的聚合，然后使用 1×1 卷积调整网络的通道数和维度，再沿空间维度将特征图分离为水平张量和垂直张量，中间特征映射如式（3-6）所示：

$$f=\sigma\left[K_1\left(\mathrm{y}^h,y^w\right)\right] \tag{3-6}$$

其中，K_1 表示 1×1 卷积变换函数，$\sigma(\cdot)$ 表示非线性激活函数，$f\in\mathbf{R}^{C/r(H+W)}$ 表示沿水平方向和垂直方向的中间特征图，C 为通道数，r 为采样比，用来控制模块

大小。再将 f 沿空间维度切分成水平注意张量 $f^h \in \mathbf{R}^{C\times H/r}$ 和垂直注意张量 $f^w \in \mathbf{R}^{C\times W/r}$，再通过两组 1×1 卷积 F_h 和 F_w 增加到与输入同样通道数，得到 g^h 和 g^w 的计算如式（3-7）所示：

$$\begin{cases} g^h = \sigma\left(F_h\left(f^h\right)\right) \\ g^w = \sigma\left(F_w\left(f^w\right)\right) \end{cases} \tag{3-7}$$

其中，g^h 和 g^w 分别为垂直方向和水平方向生成的注意力权重，将两个方向上的权重相乘后与输入特征进行融合，得到最终增强识别机制输出，具体计算如式（3-8）所示：

$$z_c = x_c\left(j,i\right) g^h g^w \tag{3-8}$$

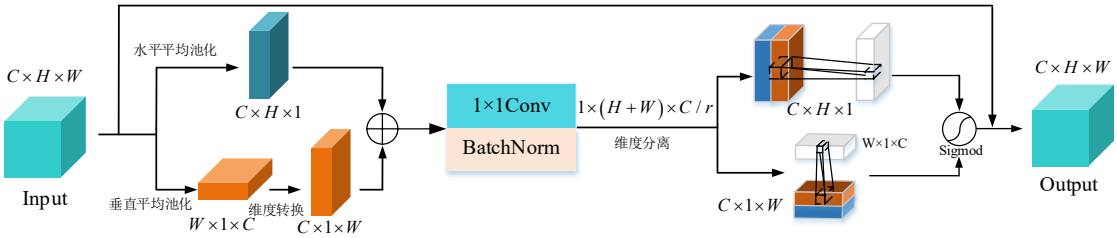


图 3-8 ER-m 结构图

3.4 实验结果与分析

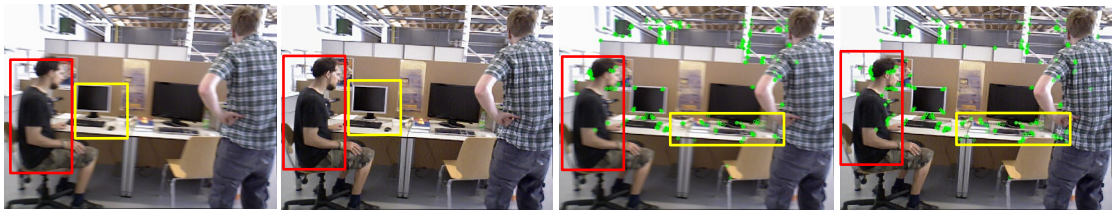
实验所用平台软硬件配置为：CPU 为 Inter i9-12900K 处理器，显卡为 RTX3090，显存 24 G，内存为 16 G，系统为 Ubuntu20.04，采用 PyTorch 深度学习框架。网络训练的超参数设置为：初始学习率设置为 0.01，衰减系数设置为 0.0005，batch size 设置为 16，动量设置为 0.97，采用了 Adam 优化器训练网络。本章选择 COCO 数据集进行训练，该数据集包含了 80 种不同类型的物体。为了评估本章 SLAM 算法性能，采用德国慕尼黑工业大学提供的 TUM 数据集对算法进行验证，其中包括模糊较多的 walking 序列并使用绝对轨迹误差和相对位姿误差对系统鲁棒性进行验证。

3.4.1 图像去模糊实验

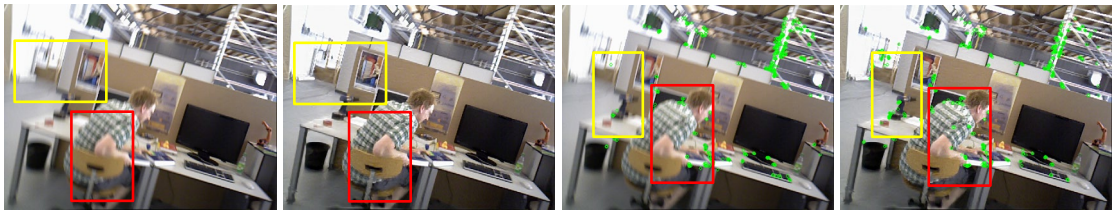
本章实验在图像预处理阶段先利用模糊检测模块对所选数据集进行清晰度

计算，筛选出模糊图像送入 PENET 中，否则无需去模糊操作。数据集中选取的序列包括相机旋转或拍摄对象移动导致的模糊图像和清晰图像，可以验证本章提出的模糊图像检测模块和去模糊算法，图像去模糊前后特征点提取对比如图 3-9 所示，红色框是本实验标注出来图像中前景对象的主要差别之处，黄色框为图像中背景的主要差别之处，特征点的数量、提取位置和图像质量均进行了对比。由图 3-9 可知，模糊图像存在物体拖影和像素梯度不明显问题并影响图像细节和边缘，特征点的位置可能会发生一定的偏移，同时也难以提取到足够数量的特征点，描述子也会因图像模糊化而失真或不准确，最终导致错误的特征匹配，进而引起轨迹漂移。本章算法有效提升了图像的纹理，改善了区域重叠造成的视觉拖影情况，恢复出物体准确结构，使得局部区域特征点数量明显增加，减少了因特征提取数量不足或特征匹配错误导致的累计误差。由于图像模糊导致几何结构失真，局部特征不够清晰，特征点发生位置偏移，同时前后图像描述符相差较大，导致存在较多误匹配。去模糊前后特征匹配对比如图 3-10 所示，从图 3-10（b）和（d）中可以看出去模糊后图像局部区域更加清晰，特征提取和匹配数量均明显增加，且更符合角点定义。

（a）原始图像 1　（b）去模糊图像 1　（c）模糊图像特征点 1　（d）去模糊图像特征点 1

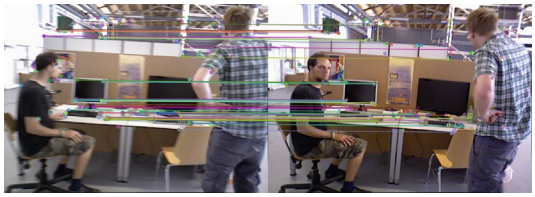


（e）原始图像 2　（f）去模糊图像 2　（g）模糊图像特征点 2　（h）去模糊图像特征点 2

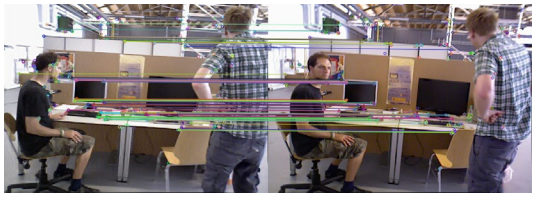


（i）原始图像 3　（j）去模糊图像 3　（k）模糊图像特征点 3　（l）去模糊图像特征点 3

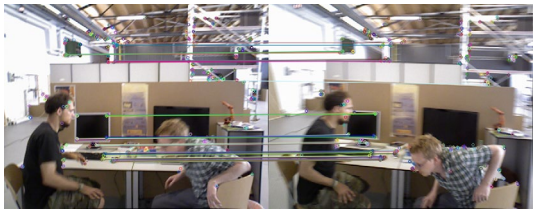
图 3-9 图像去模糊前后特征点提取对比



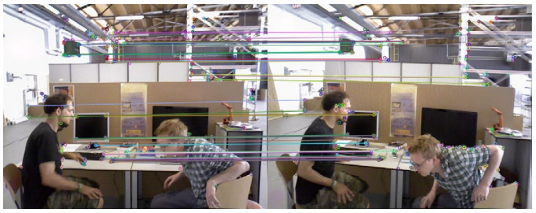
（a）模糊图像匹配 1



（b）去模糊图像匹配 1



（c）模糊图像匹配 2



（d）去模糊图像匹配 2

图 3-10 去模糊前后特征匹配对比

本实验分别从定性和定量的角度综合考察模糊对特征匹配的影响，去模糊前后匹配数量对比如图 3-11 所示。由于去模糊后图像细节和边缘明显增强，提取特征点数量更多也更为准确，因此，图像前后帧匹配点数增加，提升了位姿估计精度。

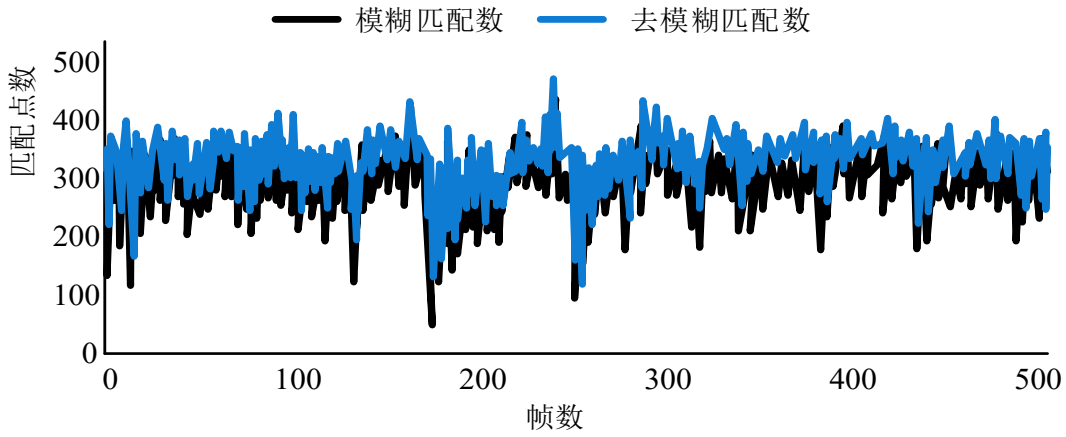
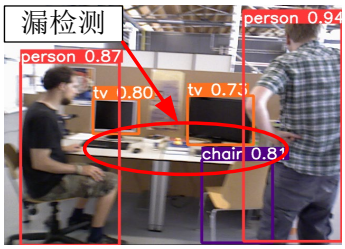


图 3-11 去模糊前后匹配数量对比图

3.4.2 物体检测对比实验

本章选取 TUM 数据集中模糊程度不同的序列对物体检测效果进行评估，图 3-12 图片从左往右模糊程度逐渐递增，本章算法与 YOLOv5s 算法检测结果对比如图 3-12 所示，其中红色矩形框为本章辅助框。由图 3-12（a）可知，YOLOv5s 在低模糊场景下容易出现物体的漏检测情况，如图中鼠标、键盘、书本等，同时在模糊区域下的检测精度也较低。由图 3-12（b）可知，在模糊较严重的场景下

会出现错误检测，当模糊程度提高不仅存在大量漏检测情况而且该算法错误地把背景识别成了椅子。由图 3-12（c）可知，由于相机旋转导致图像严重模糊情况下，该算法无法检测到任何目标。由于 PENET 算法包含模糊区域关注模块和增强识别机制恢复了模糊区域细节，提高了物体识别的精度和准确程度，由图 3-12（d）~（f）可知，目标检测成功率显著提升，部分遮挡的物体也能够成功识别，例如图 3-12（d）中人坐的椅子和别遮挡的键盘等。

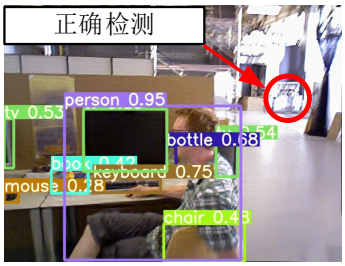


（a）YOLOv5s 漏检测

（b）YOLOv5s 错误检测

（c）YOLOv5s 无法检测

（d）PENET 更多检测



（e）PENET 正确检测

（f）PENET 完成检测

图 3-12 物体检测对比

为了定量分析提出算法的效果，分别与 YOLOv5s、DeepDeblur+YOLOv5s 以及 DeblurGAN-v2+YOLOv5s 用同样的数据训练进行参数对比，不同模型检测结果如表 3-1 所示。从表中数据可知，PENET 单阶段去模糊识别模型进行检测时，准确率和召回率相较于无模糊处理模型 YOLOv5s 和两阶段组合模型性能平均提高提升了 15%，同时网络参数量和检测速度也更优。

表 3-1 不同模型检测结果对比

去模糊模型	检测模型	准确率%	召回率%	模型内存/MB	检测速度/(f·s^{-1})
/	YOLOv5s	79.13	76.79	7.9	50.08
DeepDeblur	YOLOv5s	84.51	82.47	118.2	15.36
DeblurGANv2	YOLOv5s	87.42	86.35	24.6	28.46
PENET		90.12	89.97	18.7	37.62

3.4.3 位姿估计对比实验

为了验证所提算法有效性，本章采用绝对轨迹误差（ATE）和相对位姿误差（RPE）来验证算法性能，其中 ATE 通过计算真实轨迹与 SLAM 系统估计的轨迹对应点间的距离，该指标直观描绘出算法精度和轨迹全局一致性，RPE 表示描述相同两个时间戳内的位姿变化量的差，相当于直接测量里程计误差，适合评估系统的漂移。本章使用 ATE 与 RPE 的均方根误差（RMSE）来统计得到总体值，具体计算如式（3-9）所示：

$$\begin{cases} \mathrm{ATE} = \left(\dfrac{1}{m} \sum\limits_{i=1}^{m} \left\| \mathrm{trans}\left(Q_i^{-1} S P_i \right) \right\|^2 \right)^{\frac{1}{2}} \\ \mathrm{RPE} = \left(\dfrac{1}{m} \sum\limits_{i=1}^{m} \left\| \mathrm{trans}\left(\left(Q_i^{-1} Q_{i+\Delta} \right)^{-1} \left(P_i^{-1} P_{i+\Delta} \right) \right) \right\|^2 \right)^{\frac{1}{2}} \end{cases} \tag{3-9}$$

其中，P_i 和 Q_i 分别为第 i 帧的估计位姿和真实位姿，trans 表示位姿误差中的平移部分。

实验采用的 TUM 数据集由德国慕尼黑工业大学的计算机视觉与人工智能团队提供，涵盖了不同的场景和条件，包括室内、室外、不同季节和天气条件等，适合视觉 SLAM 不同场景下性能评估。本次实验选用 walking 序列对比本章算法与 ORB-SLAM3 轨迹误差，该序列包含一个行人在不同的环境中移动的视频记录，会有不同程度模糊的图像，可用于验证模糊场景对 SLAM 系统位姿的影响，不同 SLAM 系统的位姿误差如表 3-2 所示。从表中数据可以看出，本章方法增加了去模糊化检测处理，特征提取匹配增加的同时减少了动态目标的影响，ATE 和 RPE 相较于 ORB-SLAM3 有较大提升。

表 3-2 不同 SLAM 系统位姿误差

序列	ORB-SLAM3		本章算法	
	ATE	RPE	ATE	RPE
w_static	0.235	0.009	0.007	0.004
w_xyz	0.814	0.019	0.019	0.009
w_rpy	0.528	0.024	0.063	0.015
w_halfsphere	0.607	0.021	0.039	0.011

ORB-SLAM3 和本章方法在 walking_xyz 序列上的 APE 比较以及与真值的轨迹拟合如图 3-13 所示，从图中可以看出，ORB-SLAM3 在跟踪过程中出现轨迹漂移的情况，错误地跟踪特征点导致累计误差。本章方法因为正确识别并剔除动态特征点，减少了累计误差，轨迹与真值拟合一致提高了系统的鲁棒性。

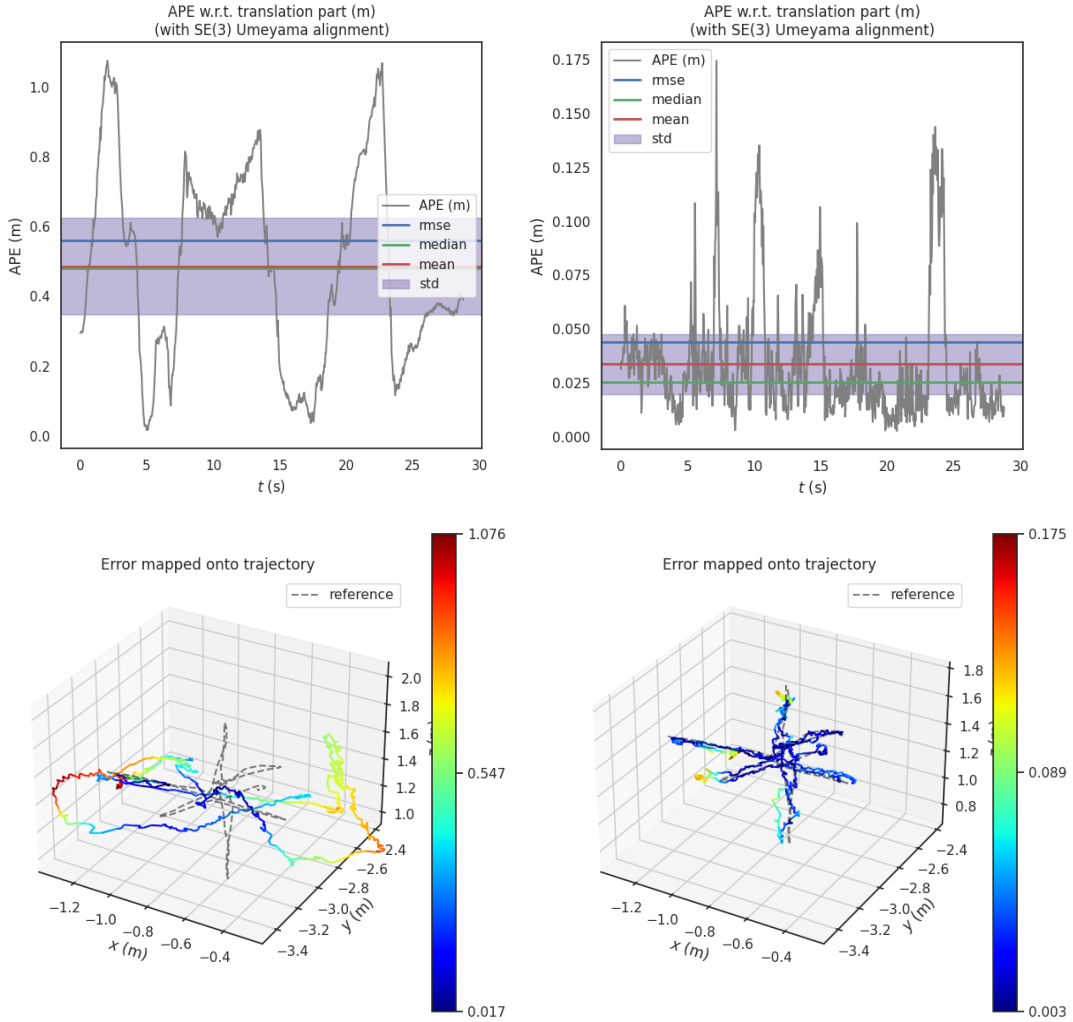
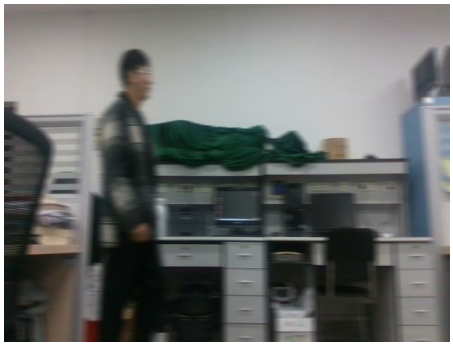


图 3-13 ORB-SLAM3 和本章方法 APE 及轨迹拟合图

3.4.4 真实场景实验

移动机器人在运动过程中获取的某帧图像及去模糊图像如图 3-14（a）和（b）所示，模糊图像检测和去模糊化检测结果如图 3-14（c）和（d）所示，从图中可以看出本章算法的 PENET 网络能有效地识别出运动模糊场景中的物体，减少了传统目标检测存在的漏检和误检情况，提升了 SLAM 系统在复杂环境下的感知能力。



（a）模糊图像

（b）去模糊图像

（c）模糊图像检测

（d）去模糊化检测

图 3-14 真实场景实验

3.5 本章小结

针对移动机器人在识别模糊图像中易出现漏检测和错误检测的情况，传统视觉 SLAM 算法在运动模糊场景下特征点提取数量不足的问题，本章提出一种基于感知增强的视觉 SLAM 方法。将相机采集到的图像输入到模糊检测模块进行清晰度计算，根据计算的结果与阈值进行比较，判断图像是否模糊，若为模糊图像则输入至去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，剔除动态物体所在区域。通过 TUM 数据集与真实场景实验表明，本章节算法减小了运动模糊环境对视觉 SLAM 系统的影响，提升了移动机器人在复杂环境下的感知能力。

第 4 章 基于特征约束的改进 SLAM 方法

针对传统动态 SLAM 方法无法有效判断物体的运动状态以及剔除动态物体后特征点数量较少的问题，本章在感知增强策略基础上提出一种基于特征约束的改进 SLAM 方法。基于重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场物体运动判断模型。针对传统 SLAM 方法特征提取匹配耗时，全局计算描述符导致系统运行效率不高的问题，本章采取特征和光流相融合的跟踪策略，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算。本章技术路线如图 4-1 所示。

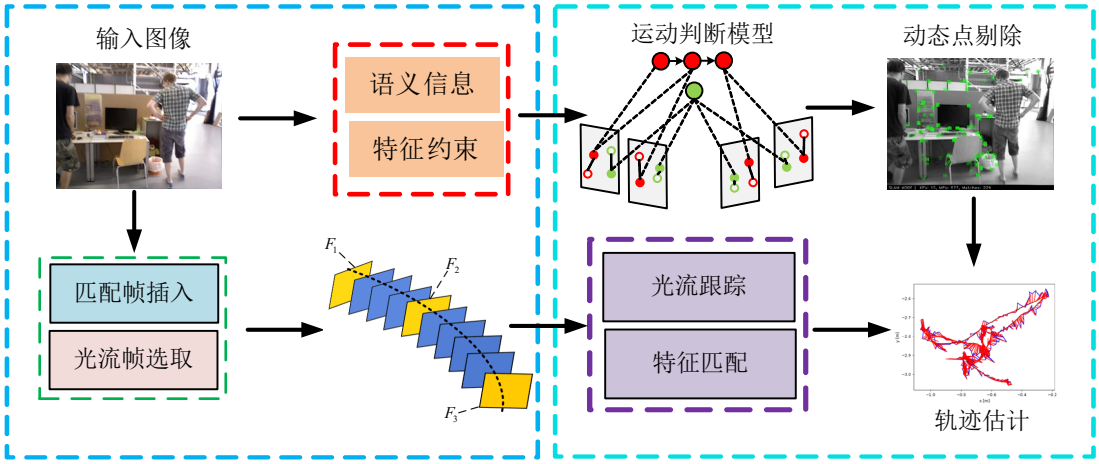


图 4-1 本章技术路线图

4.1 特征约束函数构造及运动判断

基于感知增强的策略提高了相机获取的图像质量，在优化过程中获取了场景的语义信息，该方法一定程度上提升了相机位姿估计精度，但仅靠语义信息无法完全解决动态物体对 SLAM 系统的影响，难以判断处于语义静止的目标实际运动状态，例如正在移动的椅子。因此，本章提出基于重投影误差、极线距离以及空间点共视帧构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场（Global Conditional Random Fields, GCRF）运动判断模型，将动态点和静态点的筛选转化为特征点标签的二分类问题，通过细化分类的标签结果，减少错误判断。当潜在动态物体上分类的动态点数量超过运动阈值，则认为目标是运动的，

将其中的动态特征点剔除以提高相机位姿的估计精度。运动判断模型如图 4-2 所示，根据全局范围内空间点在不同帧中的观测信息，全局条件随机场可以自主学习不同状态的特征信息，包括重投影误差、极线距离和空间点的共视帧数，并且利用学习到的特征约束和语义信息来对新的点进行分类。

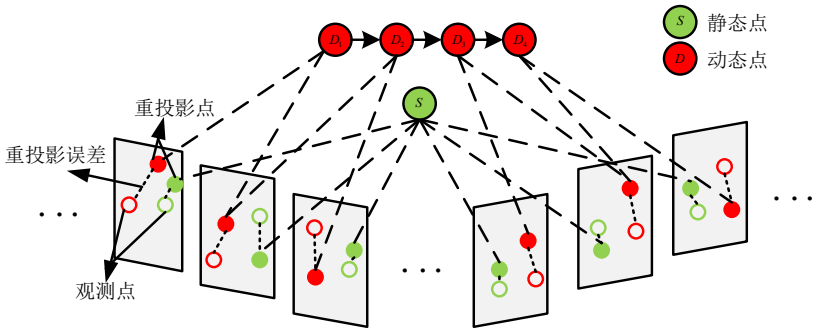


图 4-2 运动判断模型

本章中每一个空间点 $\boldsymbol{P}_i$ 都分配有一个标签 $s_i = L_i \in \{0,1\}$，其中 0 代表静态，1 代表动态。条件随机场是一种标签间相关性约束判别概率模型，通过将构建的概率模型转化为吉布斯能量函数来求解，最小化特征约束函数 $E(s)$ 得到所有点的最优标签分配，特征约束函数 $E(s)$ 如式（4-1）所示：

$$E(s) = \arg\min\left[\sum_{n\in N}\lambda_u(s_n) + \sum_{m,n\in N}\mu_p(s_m,s_n)\right] \tag{4-1}$$

其中：$E(s)$ 为总能量函数，$\lambda_u(s_n)$ 为一元特征约束函数；$\mu_p(s_m,s_n)$ 为二元特征约束函数；N 为模型的节点总数；m 和 n 表示不同的节点；s_m 和 s_n 分别为节点 m 和 n 的类别标签。构造的一元特征约束函数来拟合全局条件随机场中的每一个点的势能，二元特征约束函数来拟合顶点间相连的边，即节点间的相互势能。

4.1.1 一元特征约束函数

一元特征约束函数用来建模特征点集合与观测场之间的关系，用于衡量特征点属于动态类别标签的概率，本章定义的一元特征约束包含了三组观测特征变量，分别为空间点与其对应像素点之间的重投影误差项、观测范围内的共视帧数和双向极线约束，归一化特征变量后得到一元势约束函数，具体计算如式（4-2）所示：

$$\lambda_u\left(s_n\right)=-\ln\left(\left(\frac{\alpha_n-\delta_\alpha}{\eta_\alpha}\right)\left(\frac{\beta_n-\delta_\beta}{\eta_\beta}\right)\left(\frac{\gamma_n-\delta_\gamma}{\eta_\gamma}\right)\right) \tag{4-2}$$

其中：α_n、β_n和γ_n分别为节点n的重投影误差、共视帧数和匹配像素点到极线的距离，δ和η分别为相应变量的均值和标准差。双向极线距离计算如式（4-3）所示：

$$d=\frac{\boldsymbol{l}_1\boldsymbol{p}_1\sqrt{A_2^2+B_2^2}+\boldsymbol{l}_2\boldsymbol{p}_2\sqrt{A_1^2+B_1^2}}{2\sqrt{\left(A_1^2+B_1^2\right)\left(A_2^2+B_2^2\right)}} \tag{4-3}$$

其中：d为双向极线距离；$\boldsymbol{p}_1$和$\boldsymbol{p}_2$为节点n在前后两帧中的投影点，l_1和l_2为特征点与相机光心连线在平面上相交形成的极线，A_1和B_1为极线l_1直线方程的系数，A_2和B_2为极线l_2直线方程的系数。空间点的投影与图像中的特征点间一般存在着观测误差，相机最优位姿可以通过最小化重投影误差得到，具体计算如式（4-4）所示：

$$\boldsymbol{T}^*=\arg\min_{\boldsymbol{T}}\frac{1}{2}\sum_{i=1}^{s}\left\|\boldsymbol{u}_i-\pi\left(\boldsymbol{T}\cdot\boldsymbol{P}_i\right)\right\|_2^2 \tag{4-4}$$

其中：s为匹配点对总数，$\boldsymbol{u}_i$为投影像素坐标，$\pi(\cdot)$为投影函数，$\boldsymbol{T}$为相机位姿，$\boldsymbol{T}^*$为相机最优位姿。

4.1.2 二元特征约束函数

二元特征约束函数构建当前节点与其领域节点之间的关系，通过拟合特征点的空间相关性来提高检测性能，本章使用基于观测和定位的双重高斯核来构建二元特征约束函数。观测内核描述了观测信息相似的节点往往属于同一类标签，不同标签的节点在共视次数和平均重投影误差应该有明显差异；定位内核假设相邻空间点应该属于同一个物体，小区域范围内的点应该具有相同的标签，定位核中包含不同标签相邻点的惩罚因子。二元特征约束函数计算如式（4-5）所示：

$$\begin{aligned}\mu_p\left(s_m,s_n\right)=\theta\left(s_m,s_n\right)\Bigg(&k_1\exp\left(-\frac{\left(\omega_m-\omega_n\right)^2}{\sigma_p^2}\right)+\\&k_2\exp\left(-\frac{\left(\alpha_m-\alpha_n\right)^2}{\sigma_\alpha^2}-\frac{\left(\beta_m-\beta_n\right)^2}{\sigma_\beta^2}\right)\Bigg)\end{aligned} \tag{4-5}$$

其中，α_m 和 α_n 分别为节点 m 和 n 的平均重投影误差，β_m 和 β_n 分别为节点 m 和 n 的共视帧数，ω_m 和 ω_n 分别为节点 m 和 n 位姿参数，k_1 和 k_2 为权重参数对，σ_α、σ_β 和 σ_p 为常数，用来控制高斯核的形状和尺度，$\theta(\mathrm{s}_m, s_n)$ 为表示标签兼容函数，用于计算节点 i 和 j 属于同一类的概率。

4.1.3 运动判断

传统动态 SLAM 中动态区域和静态区域的边缘往往难以准确分割，影响动态点的筛选，通过融合语义信息和特征约束构建全局条件随机场实现精确的运动判断，当相邻区域的特征点分配不同标签时或观测信息差异较大的点分配同一标签都将导致较大的函数值。本章使用平均场近似法最小化 $E(s)$ 得到全局最佳标签，充分利用全局范围内的观测信息来细化标签分配结果，提高分类的准确性。运动判断方法如算法 1 所示。

算法 1 运动判断

输入： 连续帧图像 f_i
输出： 图像中运动物体与静止物体
参数： 运动阈值 τ
FOR new f_i available THEN
　获取图像物体目标框，提取图像特征点
　FOR 特征点属于图像 THEN
　　按照式（4-2）和式（4-5）计算一元特征约束函数和二元特征约束函数
　　按照式（4-1）最小化总特征约束函数
　END FOR
　IF $N > \tau$
　　目标为运动物体
　ELSE
　　目标为静止物体
　END IF
END FOR

4.2 融合光流的特征跟踪

传统的 SLAM 算法需要对输入图像提取特征点并执行特征匹配，通过获得的匹配关系来计算机器人运动并使用关键帧进行定位和地图构建，虽然关键帧能

够提高算法精确度以及降低系统内存消耗，但是在很多情况下存在着定位容易丢失、误差较大，造成关键帧稀疏或选取不当等问题。本章提出了一种新的帧选取方式，其选取策略流程如图 4-3 所示。

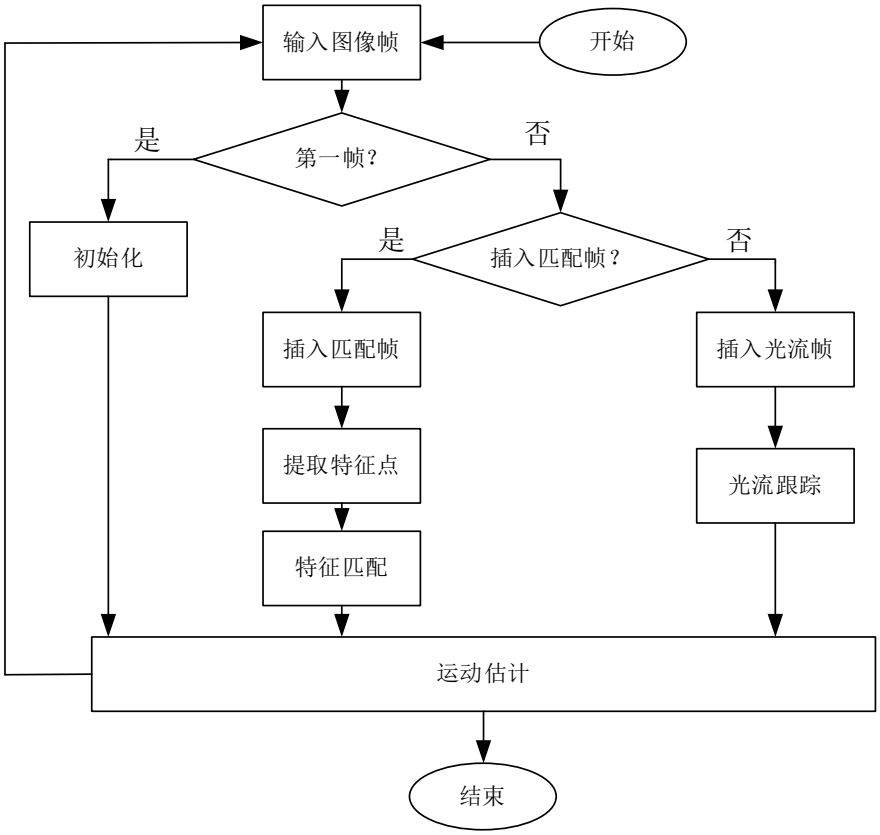


图 4-3 系统流程图

初始化后将第一帧设为匹配帧并提取特征点，然后连续 N 帧使用光流进行跟踪，获得特征匹配关系，特征点在光流跟踪过程中不断损失，同时特征点的不断减少造成光流跟踪难度的加大。因此本章提出匹配帧自适应插入方案，在合适的时间插入匹配帧并提取特征点执行特征匹配，特征匹配结束后使用 RANSAC 剔除误匹配，提高匹配精度，最后估计相机初始位姿，输出全局一致轨迹图。

光流跟踪帧数的增加可能导致跟踪失败，本章通过引入跟踪衰减系数φ来量化使用光流跟踪过程的难度，特征点的丢失率与光流跟踪的难度系数成正比。当从第一帧开始持续N帧都进行光流跟踪，具体计算如式（4-6）所示：

$$\varphi=\frac{\varepsilon_1-\varepsilon_N}{\varepsilon_1} \tag{4-6}$$

其中，ε_1和ε_N分别表示第一帧和第 N 帧中特征点个数，φ表示跟踪难度系数。跟踪难度系数是用来衡量成功跟踪特征点数量衰减的指标，其数值和跟踪效果成

反比，数值越小表示跟踪效果越好，成功跟踪的特征点数量也越多，随着跟踪的进行，每一轮迭代中特征点数量都会逐渐减少，当系数为 1 时跟踪失败。

本章提出自适应匹配帧插入方案，在跟踪困难时提前插入匹配帧，防止跟踪丢失。初始化后的第一帧作为匹配帧，进行特征点提取，之后选择连续 K 帧，将 F_2，…，F_{K+1} 分类为光流帧使用稀疏光流进行跟踪，并计算第一轮的跟踪难度系数，第 F_{K+2} 帧作为新一轮的匹配帧，重复上述步骤直至跟踪结束。本章根据跟踪难度系数自适应选择插入的光流帧数量 n，具体计算如式（4-7）所示：

$$n = m\frac{\varepsilon_N}{\varepsilon_1} \tag{4-7}$$

其中，n 为分类的光流帧数量，m 表示不提取新特征点的平均跟踪次数，后续跟踪过程将迭代此步骤直至跟踪结束。自适应匹配帧插入方案示意如图 4-4 所示，蓝色表示筛选出的光流帧，黄色表示匹配帧，K 为第一轮光流跟踪帧数，通过式（4-6）和（4-7）迭代计算后续每轮中光流跟踪帧数 n。

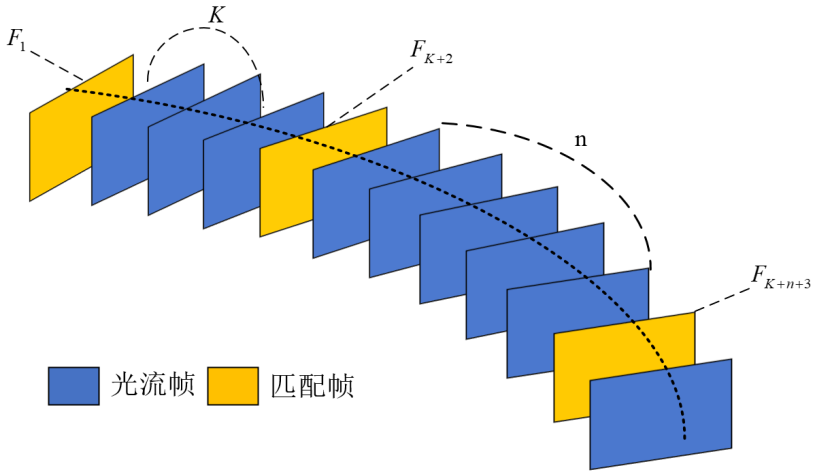


图 4-4 自适应匹配帧插入方案

针对已分类的匹配帧和光流帧采取不同的方式进行跟踪，在匹配帧中通过提取特征点来获得特征匹配关系，而在光流帧中通过双向光流来跟踪特征点。为了解决特征匹配耗时过长的问题，本章使用光流计算特征点在前一帧的位移，从而在当前帧中找到这些点并直接得到特征对应关系。为减少计算量，提高系统运行效率，本章采用稀疏光流跟踪特征点，在稀疏光流中可以把相机采集到的图当作关于时间的函数，图像中位于坐标 (x,y) 处的像素点在 t 时刻的灰度值可表示为 $I(x,y,t)$，t+1 时刻的同一特征点的灰度值应该相同，数学表达如式（4-8）所示：

$$\begin{cases} I(x,y,t)=I(x+dx,y+dy,t+dt) \\ I(x+dx,y+dy,t+dt)\approx I(x,y,t)+\dfrac{\partial I}{\partial x}dx+\dfrac{\partial I}{\partial y}dy+\dfrac{\partial I}{\partial t}dt \end{cases} \tag{4-8}$$

由式（4-8）可得式（4-9）：

$$\begin{cases} \dfrac{\partial I}{\partial x}dx+\dfrac{\partial I}{\partial y}dy+\dfrac{\partial I}{\partial t}dt=0 \\ \dfrac{\partial I}{\partial x}\dfrac{dx}{dt}+\dfrac{\partial I}{\partial y}\dfrac{dy}{dt}=-\dfrac{\partial I}{\partial t} \end{cases} \tag{4-9}$$

其中，$\dfrac{dx}{dt}$为像素水平方向的速度，而$\dfrac{dy}{dt}$为垂直方向的速度，分别用光流向量中的水平分量 u 和垂直分量 v 表示。同时$\dfrac{\partial I}{\partial x}$表示水平梯度，$\dfrac{\partial I}{\partial y}$表示垂直梯度，分别记为$I_x$，$I_y$，图像灰度对时间的变化量记为$I_t$，具体计算如式（4-10）所示：

$$\begin{bmatrix} I_x & I_y \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}=-I_t \tag{4-10}$$

迭代运算后可以找到当前帧中相同特征点位置，从而实现特征点的跟踪并完成特征匹配，同时本章采取双向光流跟踪去除错误匹配，双向光流跟踪如图 4-5 所示。

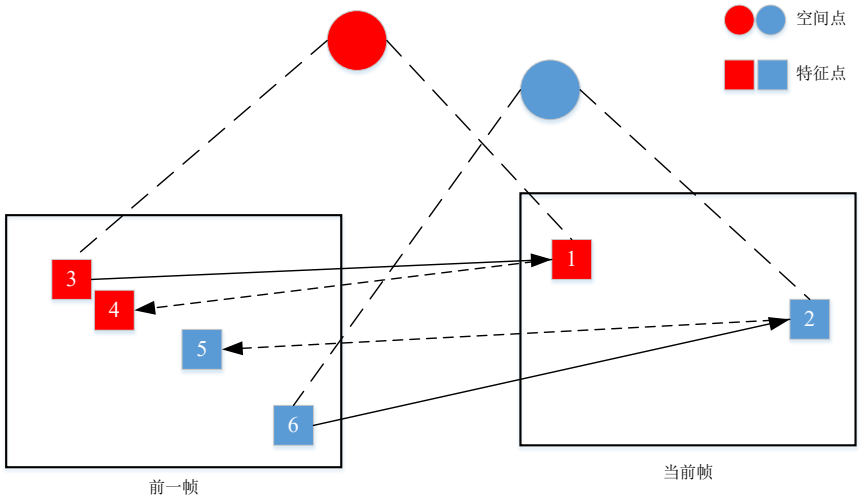


图 4-5 双向光流跟踪

图 4-5 中实线表示顺向光流跟踪过程，如 3 和 1 以及 6 和 2 的连线，虚线表示逆向光流跟踪过程，如 4 和 1 以及 5 和 2 的连线。如果 3 和 4 的像素距离小于一定范围，则可以认为是正确匹配，反之为错误匹配应该去除，经过双向光流跟踪筛选后，保留正确的匹配，从而提高光流跟踪的准确性。

当移动机器人在大角度运动下由于相机旋转等导致相机视角发生较大偏移

导致光流跟踪效果较差，极端情况下将直接跟踪丢失。为避免这一情况，引入复合运动幅值ξ来辅助移动机器人在大角度运动中的匹配帧插入，从而提升算法的鲁棒性。复合运动幅值描述了上一匹配帧和当前帧之间位姿变化情况，可以用来度量运动的大小以及两帧间相对运动距离的求解，具体计算如式（4-11）所示：

$$\xi=\min\left(2\pi-\|r\|,\|r\|\right)+\|\Delta t\| \tag{4-11}$$

其中，r和t是分别是两帧间的旋转向量和平移向量。在机器人在运动过程中复合运动幅值超过规定阈值时ξ_x说明两帧间运动距离较大，需要对当前帧重新提取特征点并设置为匹配帧以增强系统在大角度运动下的鲁棒性。

在每次插入匹配帧后，将当前的复合运动幅值初始化清零，再次插入匹配帧之前必须次达到一定的运动幅值，从而避免系统反复操作。在实际算法中，复合运动量是两个相邻匹配帧之间所有帧间的平移和旋转范数绝对值的和，其反映了相机姿态在相邻两个连续匹配帧中是否发生显著变化，相机的位姿和场景中的特征点是相互关联的，通过估计相机的位姿和特征点的位置，可以将相机观测到的特征点映射到三维空间中的坐标系中，并根据这些特征点构建场景地图。同时，特征点的位置信息也可以被用来辅助相机位姿的估计，从而实现更加准确定位。

融合光流的特征跟踪方法如算法 2 所示。

算法 2 融合光流的特征跟踪

输入：连续帧图像 f_i

输出：当前机器人的位姿估计

参数：光流跟踪自适应次数n，跟踪难度系数φ，复合运动阈值ξ_x

```
FOR new f_i available THEN
    初始化匹配帧并提取图像特征点
    执行光流跟踪
    按照式（4-6）计算第一轮光流跟踪难度系数 φ
    按照式（4-11）计算符合运动幅值ξ
    IF φ = 1 or ξ > ξ_x
        当前帧重新提取特征点，执行特征匹配，估计当前机器人位姿
    ELSE
        继续执行光流跟踪，估计当前机器人位姿
    END IF
END FOR
```

4.3 实验结果与分析

实验分为四个部分：运动判断实验、轨迹对比实验、光流跟踪实验和真实场景实验，针对不同部分分别与 ORB-SLAM3、Crowd-SLAM、Ds-SLAM、Dyna-SLAM 以及第三章所提方法进行了对比，实验平台软硬件配置均与上一章一致。

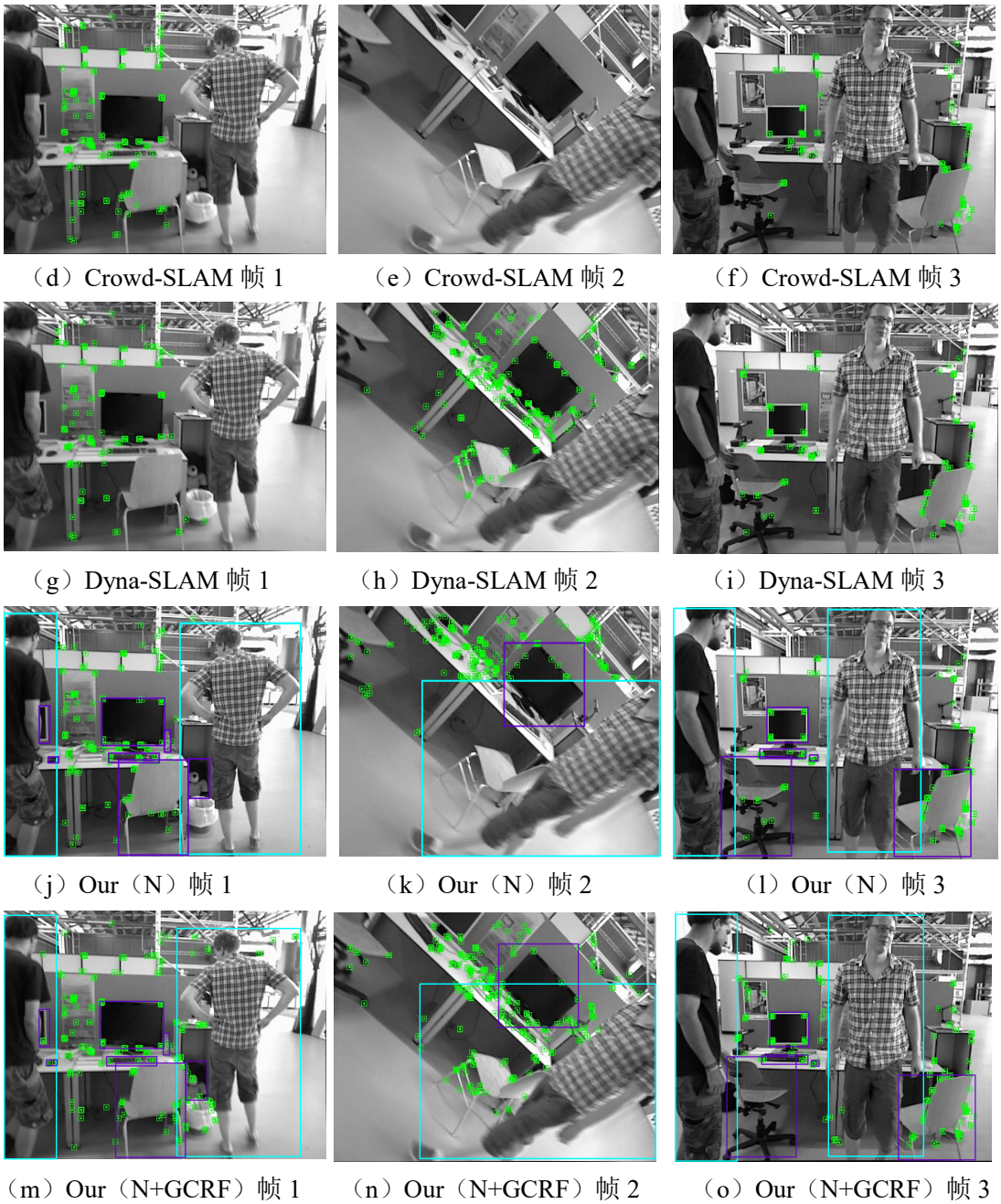
4.3.1 运动判断实验

为了验证基于特征约束的运动判断方法的有效性，设计了动态物体判断及剔除的对比实验，ORB-SLAM3、Crowd-SLAM、Dyna-SLAM、Our（N）和 Our（N+GCRF）算法特征提取对比如图 4-6 所示。其中，Our（N）表示使用第三章所提基于感知增强的方法，Our（N+GCRF）表示在第三章基础上加入基于特征约束的运动判断方法。由图 4-6（a）~（c）可知，ORB-SLAM3 无法处理高动态场景。Crowd-SLAM 算法使用基于 Yolov3-tiny 的目标检测网络来识别图像中的人，其缺点在于简单地将整个前景目标框全部剔除，而忽略了很多属于背景的静态特征点，由图 4-6（d）~（f）可知，该算法会出现大量误剔除，特别是当人占据图像较大面积时，会剔除大量的特征点导致系统跟踪失败。Dyna-SLAM 算法依据两帧中同一关键点的变化角度判断，但易受光线和角度的影响，由图 4-6（i）可知，该算法未能完全剔除被推动椅子上的动态点。Our（N）所示为仅使用感知增强策略剔除动态点，由图 4-6（j）~（l）可知，其效果与 Crowd-SLAM 性能相当，仅利用语义信息无法判断物体运动的实际情况，简单地将处于人目标框中的特征点全部剔除从而产生大量误剔除。由图 4-6（m）~（o）可知，本章算法将语义信息和重投影误差、极线距离以及共视帧等特征约束相融合构建全局条件随机场进行运动判断，因此相较于其余算法，本章算法能更准确地判断物体的实际运动状态，动态点的剔除也更为合理。

（a）ORBSLAM3 帧 1

（b）ORBSLAM3 帧 2

（c）ORBSLAM3 帧 3



（d）Crowd-SLAM 帧 1　（e）Crowd-SLAM 帧 2　（f）Crowd-SLAM 帧 3

（g）Dyna-SLAM 帧 1　（h）Dyna-SLAM 帧 2　（i）Dyna-SLAM 帧 3

（j）Our（N）帧 1　（k）Our（N）帧 2　（l）Our（N）帧 3

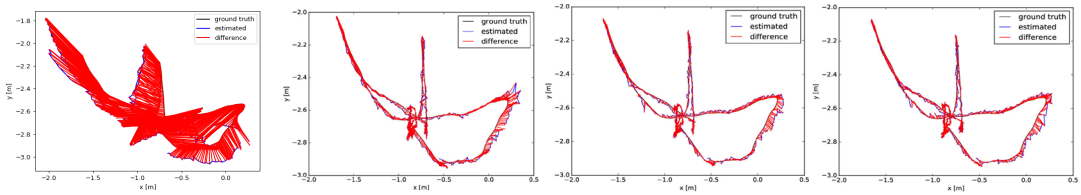
（m）Our（N+GCRF）帧 1　（n）Our（N+GCRF）帧 2　（o）Our（N+GCRF）帧 3

图 4-6 不同算法动态物体剔除对比

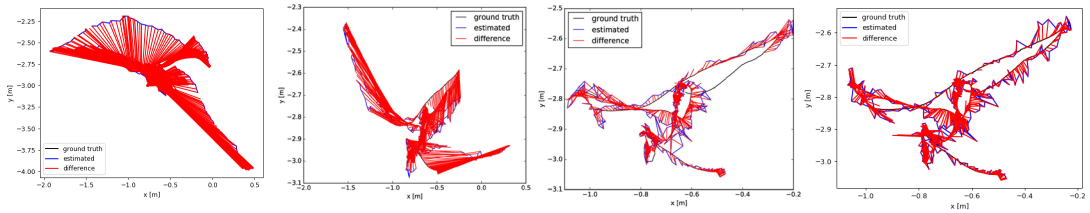
4.3.2 轨迹对比实验

轨迹对比实验选择 walking 序列中的 halfphere、rpy、xyz 和 static 来测试算法精度。ORB-SLAM3、DS-SLAM、Dyna-SLAM 和本章算法构建的轨迹如图 4-7 所示，其中，红色部分为真实轨迹和算法预测轨迹间的误差。从图 4-7 中第一列可以看出，由于 ORB-SLAM3 没有剔除动态物体，动态对象对相机位姿估计产生影响，导致轨迹漂移误差较大。DS-SLAM 和 Dyna-SLAM 都是通过深度学习

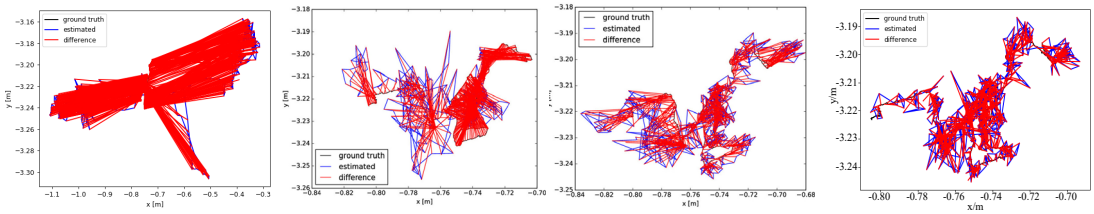
算法对动态物体所在区域进行识别并剔除，减小动态物体对构图的影响，但由于算法本身的缺陷，无法对模糊物体进行准确识别，影响了算法的定位精度。本章算法结合语义信息和特征约束能够准确判断物体的运动状态并剔除动态物体，提升了动态点剔除的合理性，同时最大程度保留了静态点，因此本章算法在动态场景下表现最好，与真实轨迹最为接近。



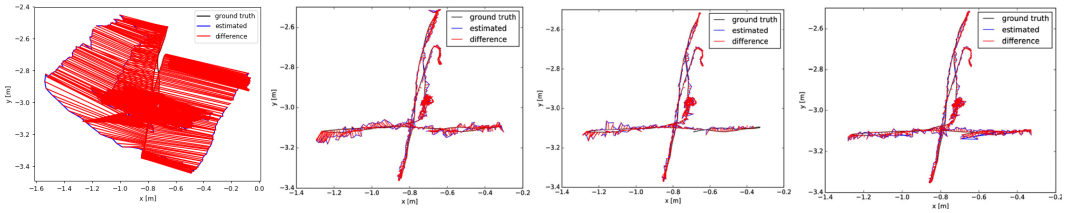
（a）ORBSLAM3 half（b）DS-SLAM half （c）Dyna-SLAM half （d）Our half



（e）ORBSLAM3 rpy （f）DS-SLAM rpy （g）Dyna-SLAM rpy （h）Our rpy



（i）ORBSLAM3 static（j）DS-SLAM static （k）Dyna-SLAM static （l）Our static



（m）ORBSLAM3 xyz （n）DS-SLAM xyz（o）Dyna-SLAM xyz （p）Our xyz

图 4-7 不同算法轨迹图

上述不同算法的绝对轨迹误差的均方根误差（RMSE）和标准差（SD）如表 4-2 所示，数值越低表示系统定位精度越高，从表中可以看出在这 4 种序列中本章算法都取得了较低的误差数值，特别是在由于相机旋转造成运动模糊的 rpy 序列中，相较于 ORB-SLAM3、DS-SLAM 和 Dyna-SLAM，本章算法的绝对轨迹误差 RMSE 平均减少 96.0%、93.0%、16.2%。总的来说，本章算法的平均绝对轨迹误差较于 ORB-SLAM3、DS-SLAM 和 Dyna-SLAM 分别减少了 96.7%、41.7%和

10.2%，验证了本文提出的感知增强和特征约束方法的有效性。

表 4-2 不同算法的绝对轨迹误差对比

测试序列	ORB-SLAM3		DS-SLAM		Dyna-SLAM		Our	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
half	0.476	0.146	0.029	0.018	0.026	0.016	0.024	0.016
rpy	0.784	0.388	0.444	0.235	0.037	0.021	0.031	0.020
xyz	0.721	0.351	0.025	0.015	0.015	0.009	0.014	0.007
static	0.387	0.063	0.008	0.004	0.006	0.004	0.007	0.006

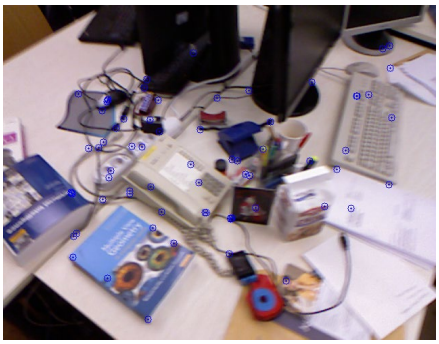
4.3.3 光流跟踪实验

本章算法在 TUM 数据集 freiburg1_desk、freiburg1_xyz 和 fr3_cabinet 序列下进行验证，光流跟踪效果如图 4-8 所示，其中图 4-8（a）为某匹配帧图像进行的特征提取结果，图 4-8（b）是后续某光流帧跟踪结果。在自适应插入匹配帧方案下，本章算法可以很好的筛选出匹配帧和光流帧，并分别进行特征匹配和光流跟踪。在所选的匹配帧图像中均匀地提取 500 个特征点，并在后续光流帧中对这些点进行跟踪，同时展示了光流运动轨迹，不同序列下光流跟踪效果如表 4-3 所示。

表 4-3 不同环境下光流跟踪效果

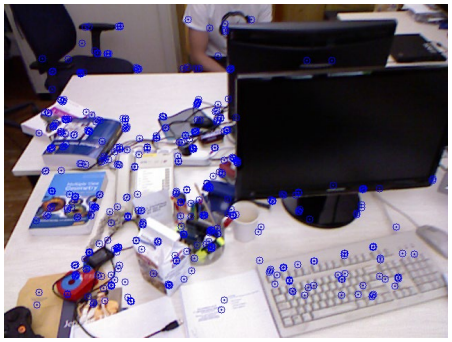
数据集	第一帧特征点提取数	第二帧光流跟踪数	跟踪衰减系数
fr1_desk	500	500	0
fr1_xyz	500	492	0.016
fr3_cabinet	300	287	0.044

（a）特征提取 1



（b）光流跟踪 1

（a）特征提取 2



（b）光流跟踪 2

图 4-8 光流跟踪实验

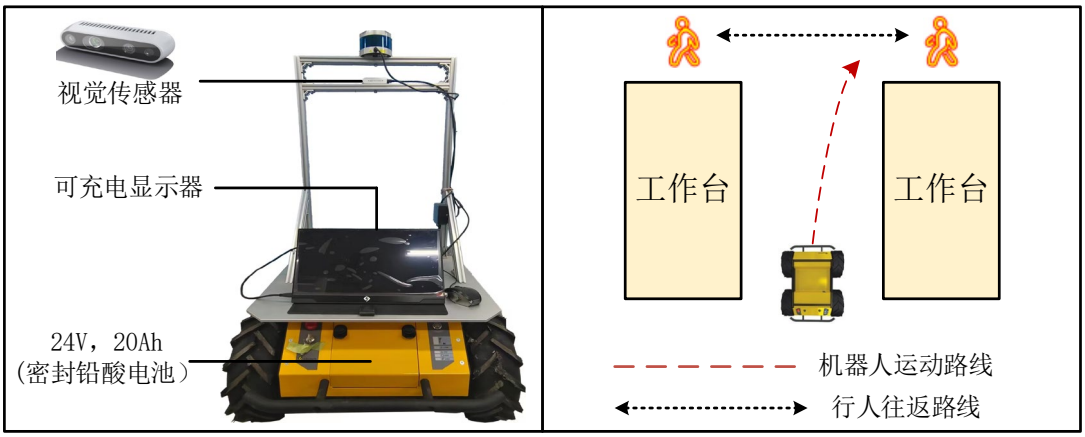
在前端跟踪耗时方面，ORB-SLAM3 和本章算法平均耗时如表 4-4 所示，由表 4-4 可知，ORB-SLAM3 算法在 SLAM 系统运行过程中，由于每帧均执行特征匹配，导致平均每帧计算时间较长，本算法采用了光流与特征融合策略，避免了描述子的计算，在光流跟踪特征点丢失后自适应插入匹配帧，平均处理时间与 ORB-SLAM3 相比平均降低了 20.9%，提升了系统运行效率。

表 4-4 本章算法与 ORB-SLAM3 平均耗时对比

数据集	ORB-SLAM3		本章算法		耗时优化%
	Media	Mean	Media	Mean	
fr1_xyz	27.2	27.7	20.1	21.2	26.1
fr1_desk	19.2	19.6	17.8	16.6	15.3
fr2_xyz	24.1	25.2	19.3	19.8	19.9
fr2_flower	36.2	37..8	28.2	29.6	22.1

4.3.4 真实场景实验

此次实验使用加拿大公司 Clearpath Robotics 开发的 Husky 系列机器人，并配备有 D435i 视觉传感器，Husky 机器人及真实实验场景如图 4-9 所示，机器人主要参数如表 4-5 所示，通过此平台在真实动态场景中对本章算法有效性进行验证。真实场景平面布局如图 4-9（b）所示，移动机器人从两工作台间起点开始运动，行人在工作台进行往返走动。



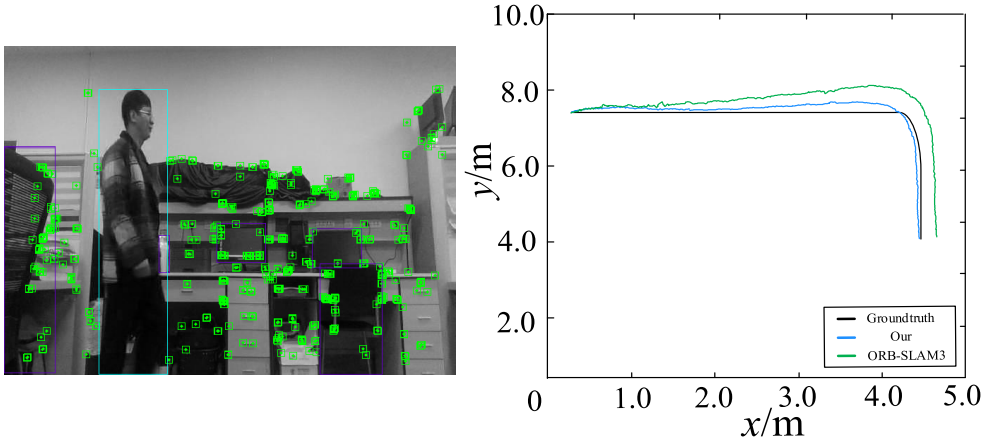
（a）Husky 移动机器人　　（b）真实场景平面布局图

图 4-9 实验机器人及真实实验场景

表 4-5 主要参数设置

变量名称	参数	数值
最大速度	V	1m/s
重量	kg	50kg
最大坡度	Θ	π/4
相机采样频率	H	30fps

本章算法剔除动态点如图 4-10（a）所示，场景中运动的人会对系统位姿估计产生较大影响，本章结合语义信息和特征约束有效地剔除了动态点。本章算法与 ORB-SLAM3 算法在真实场景下构建的轨迹如图 14（b）所示，在运动模糊场景下本章算法轨迹更接近真实轨迹，而 ORB-SLAM3 算法由于场景中存在运动的人从而出现了轨迹漂移最终导致建图误差较大。



（a）动态点剔除　　（b）算法轨迹图

图 4-10 真实场景实验

4.4 本章小结

针对传统视觉 SLAM 算法在动态场景下，无法有效判断物体的运动状态以及剔除动态物体后特征点数量较少等问题，在感知增强策略基础上提出一种基于特征约束的改进 SLAM 方法，利用重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场的物体运动判断模型；针对传统 SLAM 方法特征提取匹配耗时，全局计算描述符导致系统运行效率不高的问题，本章采取特征和光流相融合的跟踪策略，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算耗时情况，提高了 SLAM 系统的运行效率。在实验部分，采用 TUM 数据集中动态序列对本章算法和主流动态视觉 SLAM 算法进行对比，并在移动机器人平台上进行真实场景测试，实验结果表明，本章提出算法在动态场景下具有更高的精度。

第 5 章 全文总结与展望

5.1 全文总结

本文以高效获取复杂场景下的语义信息，提高系统在实际应用中的鲁棒性，更好地适应各种场景和条件为目的，主要研究内容是面向运动模糊场景下基于感知增强和特征约束的视觉 SLAM 方法，包括单阶段去模糊化识别网络 PENET 和基于特征约束的运动判断模型，有效解决了运动模糊场景下物体识别和运动判断的问题。在公开数据集以及真实场景中验证，并与多个目前主流算法对比，本文方法改善了运动模糊导致的漏检测和轨迹漂移等情况，提升了系统的感知能力和鲁棒性。

论文主要研究内容与创新点概括如下：

（1）针对移动机器人在识别模糊图像中易出现漏检测和错误检测的情况，传统视觉 SLAM 算法在运动模糊场景下特征点提取数量不足的问题。本文提出一种基于感知增强的视觉 SLAM 方法，设计一种单阶段去模糊化识别网络，由基于模糊区域关注模块的去模糊前端和基于增强识别机制的检测后端组成。将相机采集到的图像输入到模糊检测模块，该模块对输入图像进行清晰度计算，判断为模糊的图像输入至去模糊化识别网络，否则直接进行特征点提取，根据语义信息进行数据关联，剔除动态物体所在区域。在公开数据集和真实场景对所提算法进行验证，结果表明该算法能够有效降低运动模糊场景对系统的影响，提升了移动机器人的感知能力。

（2）针对传统动态 SLAM 方法无法有效判断物体的运动状态以及剔除动态物体后特征点数量较少的问题，在感知增强策略基础上提出一种基于特征约束的改进 SLAM 方法。基于重投影误差、极线距离以及共视帧数构造特征约束函数，融合语义信息和特征约束以构建全局条件随机场的物体运动判断模型，将动态点和静态点的筛选转化为特征点标签的二分类问题，根据分类的动态点数量判断物体的实际运动状态，剔除筛选出的动态点集合。在公开数据集和真实场景对所提算法进行验证，本文方法平均绝对轨迹误差较 ORB-SLAM3、DS-SLAM 和 Dyna-SLAM 分别减少了 96.7%、41.7%和 10.2%，提升了 SLAM 系统的定位精度。

（3）针对传统 SLAM 方法特征提取匹配耗时，全局计算描述符导致系统运行效率不高的问题，本文采取特征和光流相融合的跟踪策略，根据是否需要补充特征将输入的图像分为匹配帧和光流帧，仅在匹配帧中提取特征点，后续光流帧中跟踪这些特征点，减少了非必要图像中的描述符计算，改善了跟踪线程中由全局特征匹配引起的计算耗时情况。在公开数据集对所提方法进行验证，实验结果表明该方法平均处理时间与 ORB-SLAM3 相比平均降低了 20.9%，提高了 SLAM 系统的运行效率。

5.2 研究展望

本文围绕“基于感知增强和特征约束的视觉 SLAM 算法”展开研究，通过公开数据集和移动机器人构建的真实场景实验验证本文所提方法，实验表明，本文所提算法在运动模糊场景下有着良好的感知效果以及较高的定位精度，但取得的成果仍有进一步完善的空间，主要分为以下几个方向：

（1）本文重点研究了前端图像处理和特征选取部分，然而并未对视觉 SLAM 后端优化、闭环检测和地图构建环节进行研究，后续工作可针对语义地图方向进一步扩展，充分利用语义信息进行地图构建，设计出更适用于移动机器人定位与导航的语义地图。

（2）本文是基于视觉 SLAM 系统的研究，但当移动机器人遇到纹理稀疏的场景，视觉传感器无法获取到足够的特征信息，可以考虑结合多传感器或者引入线、面等特征，增加数据的类型，避免因特征数量不足导致系统定位失败等问题。

参考文献

[1] 杨傲雷, 金宏宙, 陈灵, 等. 融合深度学习与粒子滤波的移动机器人重定位方法[J]. 仪器仪表学报,2021,42(07):226-233.

[2] 孙辉辉, 胡春鹤, 张军国. 移动机器人运动规划中的深度强化学习方法[J]. 控制与决策,2021,36(06):1281-1292.

[3] 陈孟元, 田德红. 基于多尺度网格细胞到位置细胞的仿生 SLAM 算法[J]. 计算机辅助设计与图形学学报,2021,33(05):712-723.

[4] 贾晓辉, 徐文枫, 刘今越, 等. 基于惯性测量单元辅助的激光里程计求解方法[J].仪器仪表学报,2021,42(01):39-48.

[5] 宁凯, 张东波, 印峰, 等. 基于视觉感知的智能扫地机器人的垃圾检测与分类[J]. 中国图象图形学报,2019,24(08):1358-1368.

[6] 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.

[7] 徐韬, 陈孟元, 刘晓晓, 等. 动态场景下基于注意力机制与几何约束的 VSLAM 算法[J]. 传感技术学报,2023,36(09):1395-1406.

[8] Farahi F, Yazdi H S. Probabilistic Kalman filter for moving object tracking[J]. Signal Processing: Image Communication, 2020, 82: 115751.

[9] Binch A, Das G P, Fentanes J P, et al. Context dependant iterative parameter optimisation for robust robot navigation[C].2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 3937-3943.

[10] Hodge V J, Hawkins R, Alexander R. Deep reinforcement learning for drone navigation using sensor data[J]. Neural Computing and Applications, 2021, 33(6): 2015-2033.

[11] Zhang K, Ren W, Luo W, et al. Deep image deblurring: A survey[J]. International Journal of Computer Vision, 2022, 130(9): 2103-2130.

[12] Liu Y, Pu H, Sun D W. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices[J]. Trends in Food Science & Technology, 2021, 113: 193-204.

[13] 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.

[14]陈孟元, 钱润邦, 郭行荣, 等.动态场景下基于实例分割与运动一致性约束的VSLAM 算法[J]. 中国惯性技术学报,2023,31(10):986-995.

[15]Ali W, Liu P, Ying R, et al. A feature based laser SLAM using rasterized images of 3D point cloud[J]. IEEE Sensors Journal, 2021, 21(21): 24422-24430.

[16]田野, 陈宏巍, 王法胜, 等. 室内移动机器人的SLAM算法综述[J]. 计算机科学,2021,48(09):223-234.

[17]权美香, 朴松昊, 李国. 视觉 SLAM 综述[J]. 智能系统学报,2016,11(06):768-776.

[18]Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C].2011 international conference on computer vision. IEEE, 2011: 2320-2327.

[19]Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C].European conference on computer vision. Cham: Springer International Publishing, 2014: 834-849.

[20]Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C].2014 IEEE international conference on robotics and automation (ICRA). IEEE, 2014: 15-22.

[21]Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6): 1052-1067.

[22]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. IEEE, 2007: 225-234.

[23]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.

[24]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.

[25]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.

[26]杨世强, 范国豪, 白乐乐, 等. 基于几何约束的室内动态环境视觉 SLAM[J]. 计算机工程与应用,2021,57(16):203-212.

[27]赵洋, 刘国良, 田国会, 等. 基于深度学习的视觉 SLAM 综述[J]. 机器人,2017,39(06):889-896.

[28]Wang R, Wan W, Wang Y, et al. A new RGB-D SLAM method with moving object detection for dynamic indoor scenes[J]. Remote Sensing, 2019, 11(10): 1143.

[29]Dai W, Zhang Y, Li P, et al. Rgb-d slam in dynamic environments using point correlations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 373-389.

[30]Gao C, Zhang Y, Wang X, et al. Semi-direct RGB-D SLAM algorithm for dynamic indoor environments[J]. Robot, 2019, 41(3): 372-383.

[31]Yang S, Fan G H, Bai L L. Geometric constraint-based visual SLAM under dynamic indoor environment[J]. Computer Engineering and Applications, 2021, 57(16): 203-212.

[32]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.

[33]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). IEEE, 2018: 1168-1174.

[34]Fan Y, Zhang Q, Tang Y, et al. Blitz-SLAM: A semantic SLAM in dynamic environments[J]. Pattern Recognition, 2022, 121: 108225.

[35]Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments[J]. Procedia computer science, 2022, 199: 1066-1073.

[36]Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[37]Girshick R. Fast r-cnn[C].Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[38]Yang S, Fan G, Bai L, et al. SGC-VSLAM: A semantic and geometric constraints VSLAM for dynamic indoor environments[J]. Sensors, 2020, 20(8): 2432.

[39]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). IEEE, 2018: 1001-1010.

[40] Soares J C V, Gattass M, Meggiolaro M A. Crowd-SLAM: visual SLAM towards crowded environments using object detection[J]. Journal of Intelligent & Robotic Systems, 2021, 102(2): 50.

[41] Mortensen E N, Deng H, Shapiro L. A SIFT descriptor with global context[C].2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, 2005, 1: 184-190.

[42] Rosten E, Drummond T. Machine learning for high-speed corner detection[C].Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer Berlin Heidelberg, 2006: 430-443.

[43] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C].2011 International conference on computer vision. Ieee, 2011: 2564-2571.

[44] 詹煜欣, 董文永. 基于对极几何约束的动态背景下运动目标检测[J]. 计算机应用研究,2018,35(11):3462-3465.

[45] Li P, Wang R, Wang Y, et al. Evaluation of the ICP algorithm in 3D point cloud registration[J]. IEEE access, 2020, 8: 68030-68048.

[46] Song J, Yang K, Zhang Z, et al. Iterative PnP and its application in 3D-2D vascular image registration for robot navigation[J]. arXiv preprint arXiv:2310.12551, 2023.

[47] Khodarahmi M, Maihami V. A review on Kalman filter models[J]. Archives of Computational Methods in Engineering, 2023, 30(1): 727-747.

[48] 梁明杰, 闵华清, 罗荣华. 基于图优化的同时定位与地图创建综述[J]. 机器人,2013,35(04):500-512.

[49] 陈佛计, 朱枫, 吴清潇, 等. 生成对抗网络及其在图像生成中的应用研究综述[J]. 计算机学报,2021,44(02):347-369.

[50] 赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述[J]. 中国图象图形学报,2020,25(04):629-654.

攻读学位期间发表的学术论文与取得的科研成果

一、学术论文

【1】 陈孟元， 程浩， 李鹏飞， 郭行荣， 运动模糊场景下结合感知增强与特征约束的 SLAM 算法，中国惯性技术学报（EI 收录）， 录用.（导师第一，本人第二）

【2】 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. EI 检索（本人第一）

二、发明专利

【1】 陈孟元， 程浩， 一种基于深度学习的感知增强与运动判断算法、存储介质及设备，授权公告号：CN116468940B，专利号：ZL 2023 1 0390675.0，发明专利已授权（导师第一，本人第二）

三、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

行文至此，落笔为终。在毕业论文完成之际，我的学生生涯也即将进入尾声，纵观二十年求学之路，特别是在硕士研究生三年的学习，有许多人给了我无私的关爱和帮助，为我的学生时代画上了完美的句号。在此，我要对曾经帮助我的人致以最诚挚的谢意：

首先，我要感谢我的导师陈孟元教授，在我三年的研究生学习阶段，陈老师三年来的悉心教导，给我提供了许多有价值的意见和建议，让我少走了很多弯路，。孜孜不倦学者路，谆谆教诲师者心。作为研究生导师，陈老师始终坚守着师者的初心，怀揣着对科研的热情，展现了新时代研究生导师的风范，他的才华和品格是我追求的楷模。

其次，我要感谢我的师兄刘金辉、韩朋朋、陈何宝、李明好、贡鹏昊等人，在我刚进入研究生阶段课题学习的时候提供了许多宝贵的学术经验，引领我更快入门 SLAM，留下了很多美好的回忆，在此，祝愿他们前程似锦。我还要感谢我的同门龚光强、郭行荣、钱润邦、姚满宇等人，我们一起讨论学习，相互交流，希望以后工作的都能一帆风顺，读博的多发论文。我也要感谢我的室友杨威和胡俊，虽然我们不是一个导师，但你们的陪伴和支持为我带来了许多愉快和难忘的时光，这些回忆我会永远铭记。

最后，我要感谢我的家人，特别是姜海英女士，如果不是你这么多年的支持，我一定走不到这一步，正是你们对我的支持，让我全身心投入学习中，感谢你们的付出。

尚德教学 唯实惟新