

学号：19080902492

常州大学  
硕士学位论文

无人机视觉 SLAM 与惯导融合位姿估计算法  
研究

研究生姓名	陈易文
指导教师姓名	储开斌
申请学位类别	工学硕士
学科、专业名称	电路与系统
研究方向	飞行器控制

二〇二二年五月

**An UAV pose estimation algorithm based on vision SLA**

**M and inertial navigation**

A Dissertation Submitted to

**Changzhou University**

**By**

**Chen Yiwen**

**(Circuits And Systems)**

Dissertation Supervisor: Prof. Chu Kai-Bin

May, 2022

## 常州大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下独立进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人已完全意识到本声明的法律结果由本人承担。

作者签名： 陈易文

签字日期：2022 年 6 月 1 日

## 学位论文版权使用授权的说明

本学位论文作者完全了解 常州大学 有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属常州大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。学校可以公布学位论文的全部或部分内容，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。

保密论文注释：本学位论文属于保密范围，在\_\_\_\_年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

学位论文作者签名： 陈易文

签字日期：2022 年 6 月 1 日

导师签名： 傅有明

签字日期：2022 年 6 月 1 日

## 中文摘要

四旋翼无人机以其体积小,操控性强,自由度高等优点,可搭载多种传感器以实现垂直起降、高空悬停等功能,被广泛应用于摄影、救援、军事等各个领域。而无人机如何通过自身传感器获取自身定位,是实现无人机智能化的关键技术。近年来随着计算机视觉的高速发展,视觉 SLAM 已成为无人机获取自身定位的重要技术之一。本文以四旋翼无人机为实验平台,开展基于视觉 SLAM 与惯性导航融合的无人机位姿估计算法的研究。旨在使四旋翼无人机在无 GPS 的情况下,通过融合双目相机获取的位姿信息和惯性测量单元采集的信息,为四旋翼无人机的自主飞行控制提供稳定可靠的姿态、位置信息。主要研究内容如下:

(1) 针对视觉 SLAM 前端 ORB 算法存在图像分布不均匀、匹配程度不高、匹配精度差的问题,本文通过网格划分的方法,计算网格内图像灰度值以提取角点,对于不同金字塔层构建不同深度的四叉树,融合 BRIEF-LBP 特征描述子提升 ORB 算法匹配精度,最后对视觉得到的信息进行位姿估计实验。实验结果表明,对比其他改进算法,本文算法在速度提升了 2%-5%,特征点分布均匀度提升了 8.9%-10.7%,在无人机进行稳定飞行的情况下,有很好的定位效果。

(2) 针对视觉 SLAM 在快速运动时出现模糊、重影,定位精度变差甚至丢失的问题,本文采用基于滑动窗口方法,对 IMU 测量得到的运动数据和视觉信息估计的运动数据进行融合。对 IMU 信息进行预积分处理,根据 IMU 运动模型可以得到下一帧图像的预积分增量预测值,预测值与实际情况下观测到的预积分观测值相比存在一个误差量,再将误差量带回视觉模型中,并对滑动窗口内的数据信息进行边缘化处理,降低系统计算量,提升系统的实时性。实验结果表明,对比视觉惯导融合算法 VINS-mono,本文算法对位姿估计精度有明显的提升,从最大最小轨迹误差至平均轨迹误差上本文算法都要优于 VINS-mono,平均绝对轨迹误差更是比 VINS-mono 提升了 80%左右。

**关键词:** 旋翼无人机, 视觉 SLAM, ORB, 位姿估计, 视觉惯导融合

## ABSTRACT

Four rotor unmanned aerial vehicle (UAV) has characteristics of compact structure, strong maneuverability, flexible operation, there are many kinds of equipment for a vertical take-off and landing, airborne hovering, and other functions, widely used in military, agriculture, film and television and other fields, and unmanned aerial vehicle (UAV) how to use their own positioning sensors, is the key technology to realize intelligent unmanned aerial vehicle (UAV). In recent years, with the rapid development of computer vision, visual SLAM has become an important technology for UAV to obtain its own positioning. In this thesis, the quadrotor UAV is used as the experimental platform to study the UAV pose estimation algorithm based on binocular vision and inertial navigation fusion. The aim is to provide stable and reliable attitude and position information for the autonomous flight control of the quadrotor UAV without GPS by integrating the attitude information obtained by binocular camera and the information collected by inertial measurement unit. The main research contents are as follows:

(1) For visual SLAM the front ORB algorithm, uneven distribution of image matching degree is not high, the problem of poor precision of matching, this paper, by using the method of grid computing grid image grey value to extract angular point, within the pyramid in different layer to build different depth of quadtree, fusion BRIEF - LBP feature descriptor increases the ORB matching precision, Finally, pose estimation experiment is carried out on the visual information. Experimental results show that compared with other improved algorithms, the speed of the proposed algorithm is improved by 2%-5%, and the distribution uniformity of feature points is improved by 8.9%-10.7%, which has a good positioning effect under the condition of UAV stable flight.

(2) In order to solve the problems of blur, double image, poor positioning accuracy and even loss in fast motion of visual positioning, this paper integrates the motion data obtained by IMU measurement with the motion data based on visual information estimation, and adopts the fusion method based on sliding window to pre-integrate IMU information. According to the error between the predicted value of pre-integration increment in the next image frame predicted by the IMU motion model and the observed value of pre-integration in the actual situation, the error is brought back to the visual model, and the data information in the sliding window is marginalized to reduce the calculation amount of the system and improve the real-time performance of the system. Experimental results show that compared with the visual inertial navigation fusion algorithm, the proposed algorithm has a significant

improvement in pose estimation accuracy, from the maximum and minimum trajectory error to the average trajectory error of the proposed algorithm is better than the VINS-mono, and the average absolute trajectory error of the proposed algorithm is improved by about 80%.

**Key words:** rotorcraft UAV, Visual SLAM ,ORB, pose estimation, visual inertial navigation fusion

# 目录

1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 视觉位姿估计算法研究现状.....	2
1.2.2 视觉-惯导融合研究现状.....	3
1.3 研究内容.....	5
1.4 章节安排.....	6
2 视觉惯导 SLAM 相关基础理论.....	7
2.1 视觉 SLAM 框架.....	7
2.2 双目视觉模型.....	7
2.3 相机标定理论.....	9
2.3.1 内参系数估计.....	10
2.3.2 畸变系数估计.....	12
2.3.3 相机标定实验.....	12
2.4 视觉算法介绍.....	13
2.4.1 SIFT 特征点算法.....	14
2.4.2 SURF 特征点算法.....	15
2.4.3 ORB 特征点算法.....	17
2.5 本章小结.....	18
3 基于双目视觉的位姿估计算法.....	19
3.1 基于四叉树的 ORB-LBP 改进算法.....	19
3.1.1 ORB 算法.....	19
3.1.2 FAST 角点检测算法.....	19
3.1.3 BRIEF 描述子.....	20
3.1.4 基于融合描述子的改进 ORB 算法.....	21
3.1.5 构建图像金字塔.....	21
3.1.6 角点提取.....	21
3.1.7 构建四叉树.....	22
3.1.8 BRIEF-LBP 融合描述子.....	23
3.1.9 特征匹配.....	24

3.2 相机位姿解算.....	24
3.3 实验验证.....	26
3.3.1 特征匹配实验.....	26
3.3.2 位姿估计实验.....	30
3.4 本章小结.....	31
4 视觉和 IMU 融合的位姿估计算法.....	33
4.1 视觉-惯导融合整体框架.....	33
4.2 IMU 预积分.....	33
4.3 基于滑动窗口的视觉与 IMU 融合优化模型.....	35
4.4 非线性优化.....	37
4.5 实验验证.....	38
4.6 本章小结.....	42
5 总结与展望.....	43
5.1 总结.....	43
5.2 展望.....	43
参 考 文 献.....	45
攻读学位期间的研究成果.....	49
致 谢.....	50



# 1 绪论

## 1.1 研究背景与意义

无人机 (Unmanned Aerial Vehicle, UAV) 是指由无线电电子计量学遥控系统或由嵌入式电脑程序主动操控的无人驾驶飞行器<sup>[1]</sup>。无人机的基本构造,涵盖了飞行器机身、飞行控制器、推进系统、遥控器等设备。无人机采用嵌入式系统编程,不仅能有效控制其自身状态,还能通过机载相机、各种传感器感知自身的姿态和周围环境,并据此完成飞行任务<sup>[2]</sup>。

无人飞行器,按其配置结构可分为旋翼无人机、固定翼无人机等。其中旋翼无人机具有结构简单,机动性可靠性强,价格低廉等优点,因此被广泛应用到各个领域,在发挥其不可替代的作用时,也面临着其自身因素带来的挑战。在军用领域,无人机具备了容积小、隐藏性高、机动性能好、无人参与战斗的优点,被应用于作战侦查、电子设备干扰、空中打击等方面<sup>[6]</sup>。在民用领域,四旋翼无人机可用于空中航拍、农业植保<sup>[7]</sup>、警力安防、灾害救援、地质勘测<sup>[8]</sup>、野生动物包含<sup>[9]</sup>等方面。但无论在任何应用中,都需要对无人机的自身信息保持良好的掌控,使得无人机能够极好的完成它的任务,避免造成一些不可估计的后果。

当前,无人机可以借助自身所带的传感器,从而能够独立地获取关于自身的所有信息:如速度、位姿、目标信息,并且可以及时处理所得信息并拥有自主决策、独立进行的能力。其重点表现在以主动避障、路线规划、导航监控、目标辨别与追踪为代表的智慧飞行<sup>[11]</sup>,以及以任务共享和位置、地图共享为代表的智能协同,和以自主决定、主动执行为目标的智慧任务。如何进行无人机领域的主动定位和位姿计算则是进行无人机领域智能的核心技术,当前最主要的技术则是即时定位与地图构建 (Simulation Localization and Mapping, SLAM)<sup>[12]</sup>。

根据所采用传感器的不同,SLAM 技术包括视觉 SLAM 和激光雷达 SLAM。相对于传统视觉 SLAM 技术,激光雷达 SLAM 的发展得比较早,技术相对完善,而且定位精度也比较准确,但由于激光雷达重量大、成本高昂、而且没有语义信号,所以激光雷达 SLAM 技术更多的运用于无人汽车以及陆地机器人。而视觉 SLAM 则由照相机作为感应器,按照照相机种类又可分成单目 SLAM<sup>[13]</sup>、多目 SLAM<sup>[14]</sup>和 RGBD SLAM<sup>[15]</sup>。同时,多目 SLAM 系统主要由双目组成。由于双目照相机之间不出现尺寸恢复的问题,所以比单目照相机更加安全,计算结果也更加准确。无人机重量受限且电池续航能力较差,所以通常都只搭载轻便的感应器。因此价格低廉、重量轻、信息丰富的视觉传感器技术应用于无人驾驶飞行器 SLAM 中<sup>[16]</sup>是目前主流方式。

当无人机运动速度过快并且光线变化强烈的情形下,相机会产生运动模糊,又或者两帧之间的交叉区域过小以至于无法实现运动特征匹配时,使得纯视觉 SLAM 技术在高速运动下位姿的估算时准确度并不高,但加入了惯性检测单元(Inertial Measurement Unit,IMU),即便在相机数据失效的时段内,系统还可以在一段时间内维持一个较好的运动位姿估计<sup>[17]</sup>。而且无人驾驶飞行器的运动环境也相对复杂,如在地下停车场、山地等自然环境下,对精确度和鲁棒性都存在着较高需求,但由于 IMU 输出频率较高、抗干扰能力强,因此能够连续获取大量有效数据,在一定程度上也能够克服传统视觉传感器的缺点<sup>[18]</sup>。

本文主要研究了无人机平台上的视觉 SLAM 和惯导融合的位姿估计算法,并以视觉 SLAM 为基础,通过融合惯导数据以获取无人机位姿信息。根据无人机平台上对位姿精确度和鲁棒特性高的需求,以及 SLAM 定位中的难点问题,本文提出了一种融合 BRIEF-LBP 描述子的改进 ORB 算法,提升视觉 SLAM 算法前端特征匹配的精度;基于视觉-惯导耦合算法,融合视觉惯导数据,提升无人机的定位精度,为视觉 SLAM 与惯性导航融合提供一定的理论基础和应用价值。

## 1.2 国内外研究现状

### 1.2.1 视觉位姿估计算法研究现状

目前,国外用视觉对无人机进行位姿估计研究较为成熟<sup>[19]</sup>。通常来说,视觉位姿估计方法可以分为外部视觉定位和机载视觉定位两种。外部视觉定位是指把相机放置于周围环境中,并利用外界摄像机来探测无人机,并估计无人机的实时位置情况;机载视觉定位主要是在机身中放置摄像机,并利用机载摄像机观察环境来获取无人机飞行的位置信号。机载视觉定位又包括纯粹依靠外界标记物(如:人工地标)的位置,和纯粹依靠自然界特点的位置<sup>[20]</sup>。真正根据自然界特点实现位姿估计即视觉 SLAM,是指利用摄像头所获取的自然信号序列形成自身的位姿信号,并形成周围的自然环境地图,这就可以使无人机系统实现在未知条件下真正的自主定位。

VICON 光学运动捕捉系统<sup>[21]</sup>是外部视觉定位的应用,它使用了很多个外部摄像头来观察运动物体的状况,以便于获取运动物体的位姿信息。使用 VICON 系统进行跟踪能够十分准确地获取无人驾驶航空器的位姿信号,定位估算的精确度能够做到微米量级,在高速机动的状况下也很少出现位姿估算错误的现象。运动捕捉系统的位姿估算结果能够给竞赛项目计算提供真实值,无人驾驶航空器可以在 VICON 的感知区域内活动,但是这并没有解决无人驾驶航空器在大范围的自然环境中实现自由航行的要求。

通过机载视觉技术进行位姿估算,是未来无人机主动导航的主要发展趋势。而通过在周围环境中的某些外部标记信息对无人机的位姿进行估算的方式现在使用得比较多<sup>[22]</sup>,技术也较为成熟,通常使用的标记信息有各种特定的图形、地标和二维码等。但这种人工图标信息往往需要提前在周围环境中设定好,之后利用机载相机跟踪对这些已知的地标信息进行定位,还能够通过融合 IMU 信号进行较为精确的位姿估算等。但在现实的工程使用中,这些方式仍然具有一定局限性,也极大影响了无人机的活动范围。因为在许多情形下是无法提取人工图标信号的,因此,在进行搜索和灾后救援时,无人机必须采用视觉 SLAM 技术,根据自然环境的特征进行定位。

通过运用自然环境特点的位姿估计方式不要求任何先验证据,因此,仅利用自然环境的特征就可以独立地确定无人机的高度,这对无人机在未知环境中的长期稳定性工作是有利的,也成为近年来的热门研究课题。对目前的科研状况而言,视觉 SLAM 的方法大致有特征点法和直接法两种。基于特征提取的方法,主要是运用前二帧照片的特性点匹配,来评估照相机当前的位姿。常用的特征检测算法有 SIFT 特征检测法、SURF 特征检测法、ORB 特征检测法<sup>[23]</sup>等,较为典型的特征点法的位姿估计算法有 PTAM 算法<sup>[24]</sup>、ORB-SLAM 算法<sup>[25]</sup>等。直接法的方法用的较少,这种方法不计算关键点和描述子,直接根据图像的像素信息来计算相机运动,且对亮度改变也较为灵敏,较好直接法的算法有 DSO 算法<sup>[26]</sup>、SVO 算法<sup>[27]</sup>等。直接法相对于特征提取的方式来说运算量更大,而且由于必须对周围图片中所有的图像信息进行处理,但是使用直接法就能够直接得到周围环境中的三维场景信息,从而可以得到更稠密的地图。

单纯地通过视觉系统进行的位姿估计,在小区域、短距离内且无人机机动性不高的情形下也是很有效的,所以视觉 SLAM 也被广泛应用在室内机器人的导航位置。由于视觉 SLAM 的运算工作量很大,稳定性不强,对无人机长时间位姿估计存在很大的误差。所以针对较为复杂的环境,通常都必须融合其它传感器的信号才能获得更为精确的位姿估算结果。

### 1.2.2 视觉-惯导融合研究现状

视觉-惯导融合技术,一般可以分为两个方案:一种是利用滤波技术融合惯导和视觉信息;另一种是采用非线性优化技术融合惯导和视觉信息<sup>[28]</sup>。

基于滤波的视觉-惯导组合导航算法,通常又可以为松耦合和紧耦合两种框架。Carrillo 等人<sup>[29]</sup>介绍了一种基于松耦合的视觉-惯导融合的四旋翼位姿估计系统,通过卡尔曼滤波器来融合双目照相机和惯性测量单元输出。Veth<sup>[30]</sup>则提供了一个以视觉辅助低精确度惯性引导的方式,该计算采用了多维随机特性追踪方式,可追踪的特性点

数量维持恒定。Mourikis<sup>[31]</sup>给出了一个多状态变量约束的微惯性-视觉综合搜索导航计算 MSCKF(Multi-State Constraint Kalman Filter), 在观测模式中不要求具有特征点的空间位置, 但在 MSCKF 计算中出现了滤波处理与估计不一致问题: 在不明显的状况下形成了出错的客观性, 例如航线角是不明显的, 而 MSCKF 在经过 EKF 线性化后, 会使航线角形成了出错的客观性。

Castellanos 等人<sup>[32]</sup>最开始研究了同步定位的扩展卡尔曼滤波方法, 解决了滤波器一致性的问题; Bloesch 等人<sup>[33]</sup>提出了一种单目视觉-惯性里程计算法(Robust Visual Inertial Odometry, ROVIO), 将多级图像块特征跟踪与底层卡尔曼滤波器紧密结合, 实现了精确地跟踪性能的同时具有非常高的鲁棒性; Indelman 等人<sup>[34]</sup>利用基于三视图几何的视觉辅助导航方法, 旨在不使用先验信息的情况下, 为具有惯性导航系统和单相机的机体提高无 GPS 环境下的位姿估计; Hu 等人<sup>[35]</sup>提出一种单目相机和惯导融合的视觉里程计, 无需重建地图, 能够更好的满足系统实时性需求。

Luoton 等人<sup>[36]</sup>提出一种融合惯性测量单元(IMU)和视觉传感器观测结果的新方法。文中引入了平滑滑动窗口(Sliding Window Forced Independence Smoothing)的技术, 降低了求解状态变量的计算量。预积分思想的确立, 使基于优化的视觉-惯性组合导航计算得到了实施。受此思路影响, Stefan 等人<sup>[37]</sup>还引入了部分边缘化(Partial Marginalization)的技术, 利用强大的“关键帧”概念, 将旧的状态部分边缘化, 以维持一个有限大小的优化窗口, 确保实时操作。该方式并不适用于长航时间的精确引导, 且由于缺乏闭环检测控制功能, 无法校正组合引导系统的累积偏差。Forster 等人<sup>[38]</sup>进一步建立了统计关键帧内部惯性积分增量的基础理论, 将该基础理论推广至旋转群, 并解析了误差传递规律, 但该算子尚未考虑闭环检测问题。ORB-SLAM 的总设计师 Mur-Artal 等人<sup>[39]</sup>运用了预积分理论, 将惯性信号导入了 ORB-SLAM 架构, 并系统设计了具备重定位和闭环监测等功用的视觉-惯性组合导航算法。

国内最为著名且最有代表性的 SLAM 框架, 是由香港科技大学沈邵劼等人<sup>[40]</sup>的单目视觉惯性 SLAM 项目—VINS-mono。VINS-mono 通过紧耦合和非线性优化的方式, 将整套体系包含为五个组成部分: 1、数据分析预处理, 包含图像数据跟踪与 IMU 预积分; 2、视觉效果最初化, 和视觉效果惯性联合最初化; 3、局部优化和重定位技术; 4、全局角度与非线性优化技术; 5、回环测试环节。在 2017 年, VINS-mono 进行了开源, 在即时性、精确度、鲁棒性等各技术领域方面都有着出色的体现, 基本能够媲美业界顶级技术。

Zheng F 等人<sup>[41]</sup>给出了一个运用点线特性的双目视觉惯性 SLAM: Trifo-VIO, Trifo-VIO 是首个运用紧耦合滤波的基于双目的并且运用点线特性的 VIO; 另外, Trifo-VIO 还运用一个创新的轻量级的基于滤波的回环技术, 在不经全局束调节或位姿图

优化的情形下降最低累积偏差, 经过试验显示, Trifo-VIO 和 OKVIS、VINS 等在准确度、低鲁棒特性、实时化等几个主要方面都不相左右。李明阳等人<sup>[42]</sup>给出 FEJ-EKF 滤波算法(the First Estimate Jacobian EKF), Huang 等人<sup>[43]</sup>给出一个可观性约束的 UKF 滤波方法, 都从一定意义上克服了滤波器估计的不统一问题。

面对传统视觉惯性 SLAM 后端耗时较长等情况, Liu 等人<sup>[44]</sup>利用 SLAM 的增量特性, 改良了光束差优化方法(BA), 在不影响 SLAM 工程效能的前提下, 使 SLAM 效能提高近数倍; Zhang 等人<sup>[45]</sup>给出了采用视觉惯性里程计定量评估估计轨道质量问题的原则性办法, 这是对各种计算的正确性提供基准测试的依据。毛家宁等人<sup>[46]</sup>则给出了一个采用滑窗非线性设计的双目视觉 SLAM 计算, 可准确预测相机位置而不出现显著漂移现象, 与 ORB-SLAM 算法比较, 位姿的预测准确性也有了明显提升。

针对基于 SLAM 系统在高速运动和特性缺失时出现精确度低和跟踪易错误的实际问题, 万薇等人<sup>[47]</sup>提议使用双目视觉与惯导数据信息相互融合的位置估算与优化计算, 进一步提高了移动机器人位置估算和移动地图建立的准确度与鲁棒性。孙楠等人<sup>[48]</sup>提供了一个采用紧耦合的非线性优化的立体视觉-惯导 SLAM 算法, 相比视觉 SLAM 算法平均精确度增加了一倍; Heng 等人<sup>[49]</sup>则提供了一个更简单有效的视觉里程估算方法, 设计了一个采用滑动滤波窗口的视觉惯性导航体系, 可以在更复杂场景下进行连续的快速定位; 卢陶然等人<sup>[50]</sup>提供了一个采用点、线等综合特性的视觉 SLAM, 用以无人机在纹理较小、光线差异明显等特殊环境下的自主定向和位姿估计。

纵观国内研发状况, 目前国内针对固定环境下的无人机定位算法, 已经获得了较为丰硕的研究成果。但针对视觉 SLAM 中的误差累积, 以及无人机高速运动或大机动时等在特定情形下的定位还存在精度不高的问题, 本文重点研究了视觉-惯导融合方法, 主要包括在前期特征检测方面对 ORB 特征匹配算法进行改进, 使得特征检测和匹配精度方面有了提升, 再将优化后的匹配结果与惯导信息进行融合。实验结果表明, 本文算法较 VINS-mono 算法在位姿精准度上具有更强的精确性。

### 1.3 研究内容

本文以四旋翼无人机为主要实验平台, 以视觉 SLAM 为基础, 融合惯导信息, 研究了无人机视觉 SLAM 与惯导融合的位姿估计算法。针对无 GPS 信号的环境下, 利用双目相机和惯性测量单元得到的信息进行耦合优化, 使无人机在快速和慢速运动中都能得到准确的位姿信息。本文的主要研究内容有:

(1) 前端特征提取和匹配算法研究。针对 ORB 算法存在图像分布不均匀、匹配程度不高、匹配精度差的问题, 本文通过划分网格计算图像灰度值的方法提取角点。

通过在不同金字塔层上分别构建不同深度的四叉树以提高计算效率，再融合 BRIEF-LBP 特征描述子以提升 ORB 算法匹配精度，最后对视觉所得信息进行位姿估计。

(2) 视觉-惯导融合算法研究。针对纯视觉 SLAM 算法无法满足无人机快速运动的情况，本文使用基于滑动窗口的视觉惯性紧耦合优化方法，对图像的重投影误差以及 IMU 的误差进行优化。该方法首先对 IMU 信息进行预积分处理，然后将预积分处理后的 IMU 信息加入优化模型，与视觉信息共同进行优化处理；最后利用无人机 EuRoC 数据集进行实验验证并与其他视觉-惯导 SLAM 算法进行对比。

## 1.4 章节安排

本文的研究内容包括视觉方面改进算法和视觉惯导融合算法，本文的章节安排如下：

第 1 章，主要介绍了课题研究背景与意义，无人机位姿估计等关键技术的国内研发状况，以及视觉惯导融合的研究现状。最后说明了研究内容与章节安排。

第 2 章，主要讲述了视觉 SLAM 相关理论，内容涉及视觉 SLAM 系统基本架构、双目视觉建模、镜头标定方法，以及一些常用的视觉前端检测算法。

第 3 章，开展对前端特征提取和匹配改进算法研究。针对 ORB 算法进行改进以解决特征点不稳定，匹配效果差的问题，最后再对视觉信息进行位姿解算，并加以实验对比及总结。

第 4 章，研究视觉-惯导融合算法。使用视觉惯性紧耦合优化方法，对 IMU 数据进行处理，并与视觉数据融合，最后进行实验对比及总结。

第 5 章，总结与展望。总结全文科研工作，剖析不足之处，展望未来科研方向。

## 2 视觉惯导 SLAM 相关基础理论

### 2.1 视觉 SLAM 框架

经典视觉 SLAM 是过去数十年的重要研究成果，该框架较为成熟，主要由图 2-1 所示的几大模块组成。该框架首先利用搭载的传感器模块获取信息，对获得的照片信息进行处理，以计算出相应的相机位姿数据，后端接收不同时间测量的相机位姿数据，通过回环检测和进一步优化处理，可以获得无人机完整的运动轨迹。

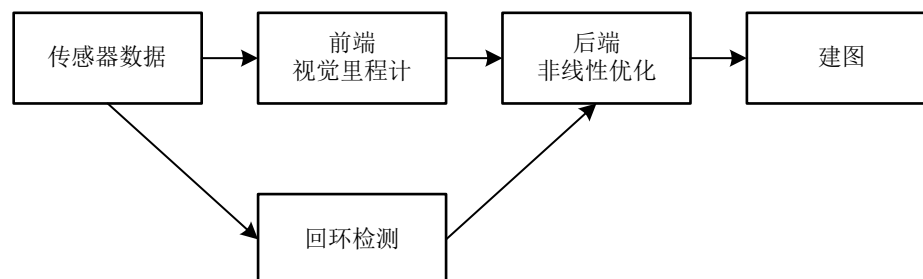


图 2-1 视觉 SLAM 整体框架

视觉 SLAM 流程如下：

(1) 传感器信息读取。在视觉 SLAM 系统中，主要是各类型的摄像头。其次还有惯性传感器信息的读取和同步。

(2) 前端(Front End)。也可以称为视觉里程计(Visual Odometry, VO)。主要工作是利用各种计算方法获取图中像的特征点，并对其进行匹配。相机的位姿信息可以从相应的特征点关系中估算出来，不同时间的位姿信息被储存下来，便形成了相机整个的位姿信息。

(3) 后端优化(Optimization)。前端得到的相机位姿数据传送至后端进行优化，可以获得全局统一的运动轨迹和位置图。

(4) 回环检测(Loop Closing)。判断当前帧所得的位姿信息是否出现过，若检测到有相同位姿信息，将该信息传回后端进行处理。

(5) 建图(Mapping)。按照估计的位置，创建与目标位置相应的地图。

### 2.2 双目视觉模型

双目相机通过同步左右相机拍摄到的同一物体的差异来判断相机距离物体距离。图 2-2 为双目视觉模型，表示了将三维空间中的一个点投影到相机平面上的基本原理。点  $P$  为三维空间中的一个点，图中的两个平行四边形代表了左右摄像头的成像平面，基线  $l$  为两投影中心的距离， $O_L$  和  $O_R$  分别为左右摄像机的投影中心， $P_L$  和  $P_R$  分布代表了点  $P$  在左右成像平面的投影像素点。成像平面和投影平面的距离称为相机的焦

距  $f$ 。

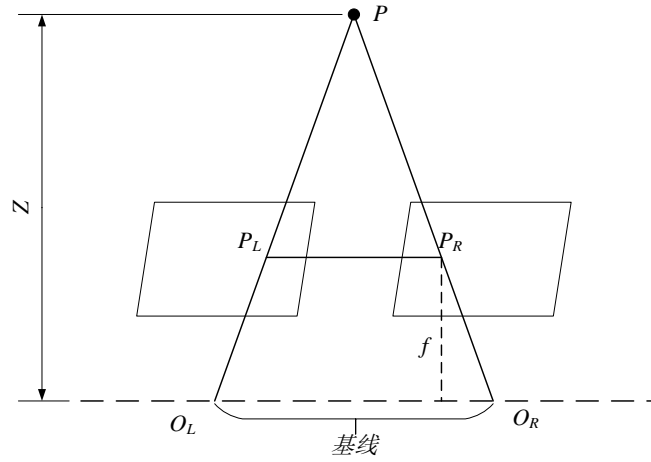


图 2-2 双目视觉模型

设  $P$  点坐标  $(x, y, z)$ ， $P_L$  坐标  $(u_l, v_l)$ ， $P_R$  坐标  $(u_r, v_r)$ ，根据三角形相似原理可得视觉模型比例关系

$$\frac{u_l}{x} = \frac{f}{z}, \frac{u_r}{x-l} = \frac{f}{z}, \frac{v_l}{y} = \frac{f}{z} \quad (2-1)$$

可以计算出成像平面  $P_L$  和  $P_R$  的坐标

$$u_l = x \frac{f}{z}, u_r = (x-l) \frac{f}{z} \quad (2-2)$$

$$v_l = v_r = y \frac{f}{z} \quad (2-3)$$

$$d = u_l - u_r \quad (2-4)$$

其中， $d$  是  $P_L$  和  $P_R$  之间的视差，继而就能算出点  $P$  在相机坐标系下的坐标

$$\begin{aligned} x &= u_l \frac{b}{d} \\ y &= v_l \frac{b}{d} = v_r \frac{b}{d} \\ z &= f \frac{b}{d} \end{aligned} \quad (2-5)$$

根据相机运动过程中的旋转矩阵  $R$  和平移矢量  $t$ ，就能够计算得到空间点  $P$  从照相机坐标系到世界坐标系的转化关系



$$z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_x & 0 \\ 0 & f_y & v_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = K(RP_w + t) = KTP_w \quad (2-6)$$

其中， $u_x$ 和 $v_y$ 是像素坐标系原点相较图像平面原点的平移分量， $P_w$ 为世界坐标系下的齐次坐标， $f_x = \alpha f$ 和 $f_y = \beta f$ 表示焦距 $f$ 的像素度量值， $\alpha$ 和 $\beta$ 表示 $x$ 轴和 $y$ 轴方向的像素个数，共同组成相机的内部参数矩阵 $K$ ，外参矩阵 $T$ 由旋转矩阵 $R$ 与平移向量 $t$ 构成。

基线长度 $f$ 的距离在相机生产是已经固定，根据式(2-6)可知，一旦外部参数矩阵 $T$ 被确定，就可以通过计算整个世界坐标系中空间点 $P_w$ 的位置信息来生成一张地图。如果空间点 $P_w$ 的齐次坐标是已知的，那么通过计算外部参数矩阵 $T$ 就可以得到相机的位置信息。但是，由于相机的镜头轮廓和接合工艺的不同，镜头和成像平面会存在一定的偏差，造成采集的图像不能满足成像的视觉模型，造成严重的变形。因此，通常需要根据对相机标定后的畸变参数对镜头进行修正，并获得镜头内部参数矩阵 $K$ ，用于后续姿态估计。

## 2.3 相机标定理论

相机标定的主要目的是为了得到相机内参误差系数和畸变校正误差系数，而相机标定的准确度会直接影响 SLAM 系统计算量和位姿解算准确度。张正友标准法是现如今应用最为广泛的标准法之一，该方式介于传统的机械标准方式与自动标准方式之间，同时具备了所用仪器简易，标准方法简便，角点变化明显，测量标定准确度较高等优势；缺点是必须经过多次人为转动和平移的平面模板。基于 SLAM 系统精度特点，本文将选用张正友标准算法。图中 2-3 为拍摄的棋盘格，只需要在同一平面内录制了不同部位，不同视角，不同姿势的好几幅标准模板图片，即可计算得到相机的内参外参等参数。

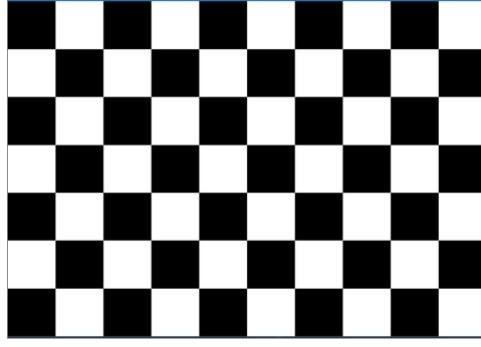


图 2-3 棋盘格标定板

### 2.3.1 内参系数估计

内参的主要功能是要使这相机坐标系与图像坐标系间有线性的变化,以解决单位不统一的问题;外参则是用于描述世界坐标系和相机坐标系之间的变换关系的。因此,将空间点  $P$  从世界坐标系下坐标  $[X_w, Y_w, Z_w]^T$  到相机坐标系下坐标  $[X, Y, Z]^T$  转换可以表示为

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (2-7)$$

将  $Z=0$  平面上点  $P_w [X_w, Y_w, 0, 1]^T$ , 投影在像素坐标系下的坐标  $p_w [u, v, 1]^T$ , 可得世界坐标与像素坐标间映射关系

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \quad (2-8)$$

记单应性矩阵  $H = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}$ , 由于  $\begin{bmatrix} r_1 & r_2 \end{bmatrix}$  是旋转矩阵的分量, 根据旋转矩阵的正交性可知:  $r_1^T r_2 = 0, \|r_1\| = \|r_2\| = 1$ , 可以推出以下的约束条件

$$\begin{cases} h_1^T K^{-T} K^{-1} h_2 = 0 \\ h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \end{cases} \quad (2-9)$$

为求解方便, 记  $B = K^{-T} K^{-1}$ , 将  $B$  展开可得

$$B = K^{-T} K^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{f_x^2} & 0 & \frac{-c_x}{f_x^2} \\ 0 & \frac{1}{f_y^2} & \frac{-c_y}{f_y^2} \\ \frac{-c_x}{f_x^2} & \frac{-c_y}{f_y^2} & \frac{c_x}{f_x^2} + \frac{c_y}{f_y^2} + 1 \end{bmatrix} \quad (2-10)$$

上式为对称矩阵，有效因子仅有 6 个，所以构建新的矩阵  $b$

$$b = [B_{11} \quad B_{12} \quad B_{22} \quad B_{13} \quad B_{23} \quad B_{33}]^T \quad (2-11)$$

带入式(2-9)的上式可得

$$h_i^T B h_j = v_{ij}^T b \quad (2-12)$$

其中  $v_{ij} = [h_{i1}h_{j1} \quad h_{i1}h_{j2} + h_{i2}h_{j1} \quad h_{i2}h_{j2} \quad h_{i3}h_{j1} + h_{i1}h_{j3} \quad h_{i3}h_{j2} + h_{i2}h_{j3} \quad h_{i3}h_{j3}]$ ，由约束条件(2-9)可得

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} b = 0 \quad (2-13)$$

相机内参矩阵中的未知数有 5 个，由于每拍摄一幅图像都能够获取两个约束方程，因此只需要拍摄 3 张定标模板图像就可求解内参系数。将式(2-13)简化为  $Vb = 0$ ，然后再考虑所有最小特征值的特征向量  $V^T V$ ，就能够得到  $V$  并估算  $b$ ，接着再用 cholsky 分解计算相机内参矩阵  $K$ ，内参矩阵中元素如下

$$\begin{cases} f_x = \sqrt{\frac{1}{sB_{11}}} \\ f_y = \sqrt{\frac{B_{11}}{s(B_{11} + B_{12} - B_{12}^2)}} \\ c_x = -sB_{13}f_x^2 \\ c_y = \left( \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2} \right) \\ s = \frac{B_{11}}{B_{33} - [B_{13}^2 + c_y(B_{12}B_{13} - B_{11}B_{23})]} \end{cases} \quad (2-14)$$

将所得内参系数带入单应性矩阵中，得到外参系数

$$\begin{cases} r_1 = \frac{K^{-1}h_1}{s} \\ r_2 = \frac{K^{-1}h_2}{s} \\ t = \frac{K^{-1}h_3}{s} \end{cases} \quad (2-15)$$

### 2.3.2 畸变系数估计

目前装配的工艺，已能较好地使镜片和图像平面的平行，从而可以把横向畸变确定为零，只需求解镜头的径向畸变系数。径向畸变的齐次表达式可描述为

$$\begin{bmatrix} xr^2 & xr^4 \\ yr^2 & yr^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} x_c - x \\ y_c - y \end{bmatrix} \quad (2-16)$$

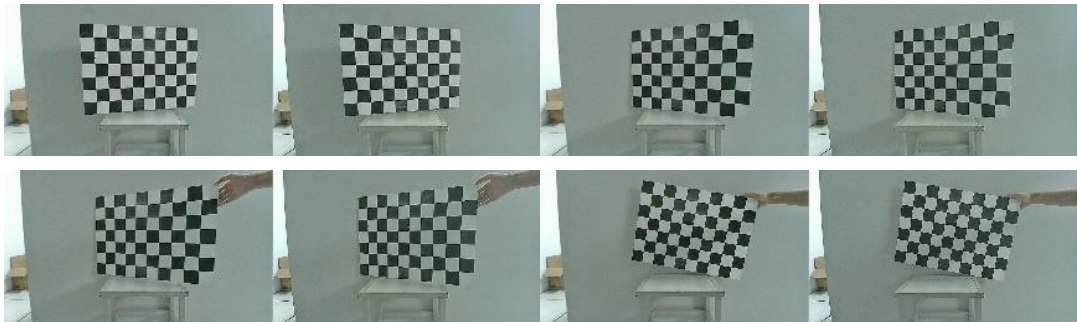
其中， $k_1, k_2$  为畸变参数， $r$  为外参矩阵中的向量。将上式简化为  $Dk = d$ ，通过最小二乘求解可得  $k = (D^T D)^{-1} D^T d$ 。

制作了  $n$  张平面模板图片，而每张图片上都有  $m$  个可提取的角点。第  $i$  张图片上第  $j$  角点  $P_{ij}$  的理想投影为  $m_{ij}$ 。采用非线性优化计算联合优化了畸变系数和内参系数，得到系数的优化。

$$\arg \min_{K, k_1, k_2, R_i, t_i, P} \sum_{i=1}^n \sum_{j=1}^m \left\| \hat{m}(K, k_1, k_2, R_i, t_i, P_{ij}) - m_{ij} \right\| \quad (2-17)$$

### 2.3.3 相机标定实验

本文所用标定板为自制的  $9 \times 6$  棋盘格标定板，将标定板中每个小格子的尺寸设定为  $30\text{cm} \times 30\text{cm}$ ，通过固定所采用的照相机，并将棋盘格进行移动、翻转，得到了十二幅各种姿态的棋盘格图片，实验步骤如图 2-4 所示。



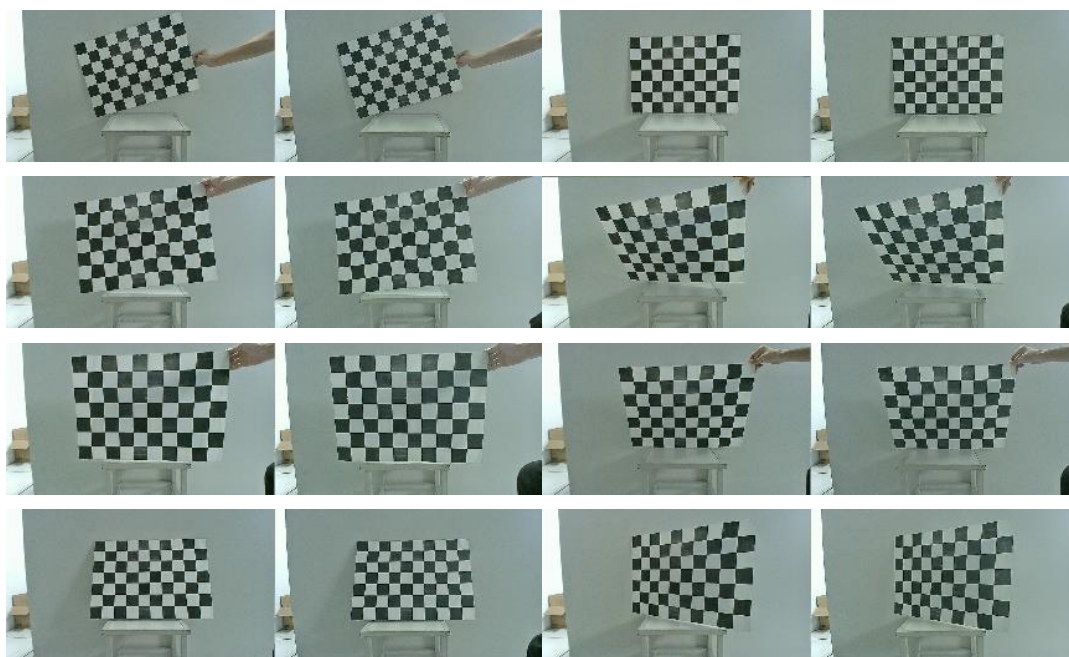


图 2-4 不同方位的棋盘格图像

从每幅图像中获得角点后，使用张氏标定算法确定相机左右目的参数，可以得到如表 2-1 所示标定结果。

表 2-1 相机标定参数

相机参数	左目	右目
$f_x$	716.1983	722.3156
$f_y$	719.8814	726.0727
$c_x$	626.1806	657.1226
$c_y$	385.5039	385.4363
$k_1$	0.1165	0.1060
$k_2$	-0.1546	-0.0987

表 2-1 给出了相机的参数。在视觉 SLAM 中，相机标定的基本要求是可以精确获取相机的内参数，而张氏标定法相对简单、成本低廉，可以获得相机精确的内参信息，因此更加适合于应用。

## 2.4 视觉算法介绍

视觉前端算法是指通过相邻帧的视觉图像数据，粗略地估算相机运动范围，并把这些粗略的估计值作为初值提交到后端。精确的初始值有助于减少后端迭代次数，从而缩短执行时间。通常，可以使用特征点法、直接法或者光流法来对相机运动进行估计。直接法根据灰度不变的设定，通过最小化光度误差预测相机运动，这个方案适用于对照度标定较好的相机。光流法基于对场景光线亮度恒定、相机短距离移动以及在

空间上的统一性的假定，分为密集光流追踪和稀疏光流追踪。特征点法包括 SIFT、SURF 和 ORB 等特征点，由于要求大量运算关键点和描述子，因此特性点法在低动态和光线的环境条件下应用得更为良好。

### 2.4.1 SIFT 特征点算法

SIFT 算法是由 Lowe 提出的，具体步骤如下：

#### (1) 特征点选取

为能在大尺度空间结构中测量到稳定性的主要特征点，Lowe 引入了高斯差分函数与图像进行卷积，以实现大尺度空间结构的构建，并测量出稳定性主要特征点。

对图像进行了平滑处理，然后再将经过平滑后的图像进行下采样就能够形成高斯金字塔，高斯金字塔可以有很多组，每组又分为很多层，一组的多个层之间的尺度是不一样的，也就是使用的高斯参数不同，图像金字塔中的组数，在尺度空间  $\sigma$  中，有如下关系

$$\sigma(o,s)=\sigma_0 2^{\frac{o+s}{S}} \quad (2-18)$$

其中， $\sigma_0$  表示基准尺度空间， $o$  表示所在的组， $s$  表示所在的层， $S$  表示每组的层数。

为了找出尺度空间的极值点，每个像素点必须与图像和尺度空间中的每个邻近点比较，当其比所有邻近点都大或都小时，则该点便可确定为极值点。如图 2-5 所示，中间的就是要检测的点，只有当其比所在层周围以及上下两层 26 个邻近点都大或都小时，则认为其是极值点。

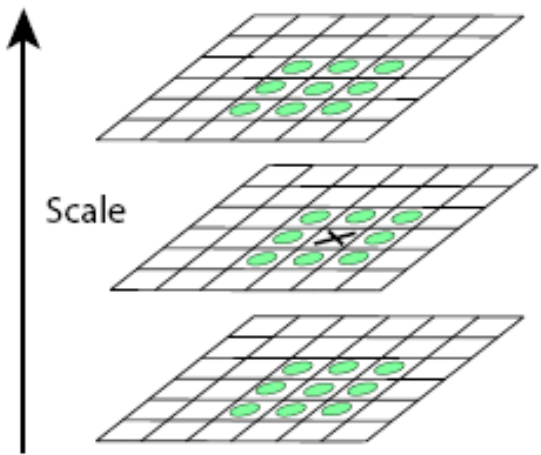


图 2-5 DOG 空间极值检测

## (2) 特征点剔除

选取特征点后，所得特征点中会含有对比度较低或者处于边缘区域内的误差点，因此必须对这类误差点加以消除。过拟合三维二次函数去除明暗度较低的误差点，再用 Hessian 矩阵方法去除边缘点。

## (3) 特征点方向选择

使用特征点邻域图像的梯度信号就能确定其方向，同时通过图像梯度直方图就可获得特征点的稳定方向。每个特征点  $L(x,y)$ ，其梯度的模  $m(x,y)$  和方向  $\theta(x,y)$  分别由式 (2-19) 和式 (2-20) 确定

$$m(x,y)=\sqrt{[L(x+1,y)-L(x-1,y)]^2+[L(x,y+1)-L(x,y-1)]^2} \quad (2-19)$$

$$\theta(x,y)=\tan^{-1}[(L(x+1,y)-L(x-1,y))/(L(x,y+1)-L(x,y-1))] \quad (2-20)$$

得到阶梯走向之后，可以把特性点邻域内像素的阶梯走向和大小用直方图进行计算，到直方图的峰值即是特性点的主走向。如图 2-6 所示

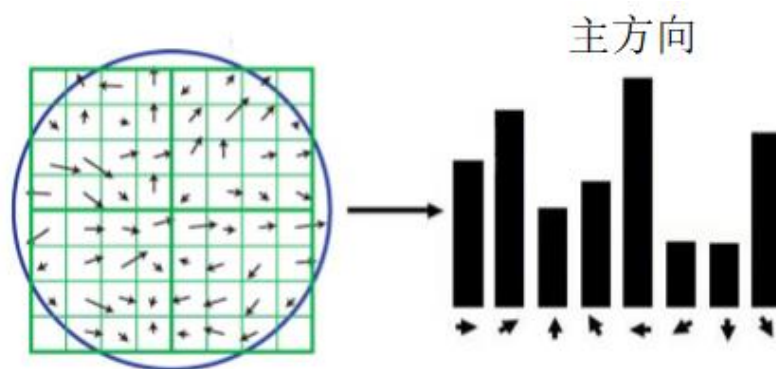


图 2-6 关键点方向直方图

## (4) 特征描述生成

首先把坐标轴转换为特征点的方向，以特征点为中心的  $16 \times 16$  的窗口图像的梯度幅值和方向，将窗内的图像分为 16 块，每块都是其中心图像内 8 个方向的直方图统计，共可以形成 128 维的特征向量。

## (5) 特征点匹配

SIFT 描述子可以在当前帧与上一帧建立特征点集，对比特征点集中的特征点，通过欧式距离决定特征点的匹配度，欧式距离越小，对应特征点的匹配度越高。

## 2.4.2 SURF 特征点算法

加速稳健特征 (SURF) 是由 Bay 对 SIFT 算法的改进提升的算法的执行效率，步骤如下：



### (1) 构建 Hessian 矩阵

Hessian 矩阵用于得到图像稳定的边缘点，完成对 SURF 特征点的提取，Hessian 矩阵如式 (2-21) 所示

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (2-21)$$

由于 Hessian 矩阵求解的执行工作量大，因而可通过采用盒式滤波器代替高斯二阶偏导数来降低工作量，盒式滤波器如图 2-7 所示。

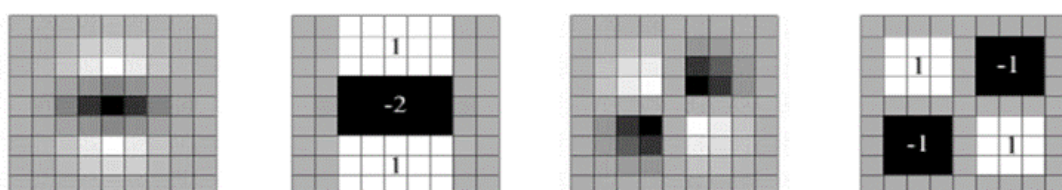


图 2-7 盒式滤波器

SURF 算子利用增加盒式滤波器的宽度进行空间尺度的变化。

### (2) 计算特征点主方向

特征点主方向选择用的是统计特征在圆形邻域中使用了 harr 特征。即在统计了  $60^\circ$  扇形区域内所有点的垂直和水平 harr 小波特征之和后，以  $0.2$  弧度的幅度为间隔进行旋转，重新统计该区域内 harr 小波的数值，将特征点主方向定为数值最大的扇形区域所在的方向。

### (3) 特征描述生成

沿着特征点主方向取一个  $4 \times 4$  的矩形区域块，在每个子区域块内统计 harr 响应值，然后对响应值进行统计形成特征向量，然后以特征点为中心，用高斯核函数进行加权计算，最后分别统计对每个子块的加权响应值，最终形成  $64$  维向量的特征描述子。SURF 特征点描述如图 2-8 所示



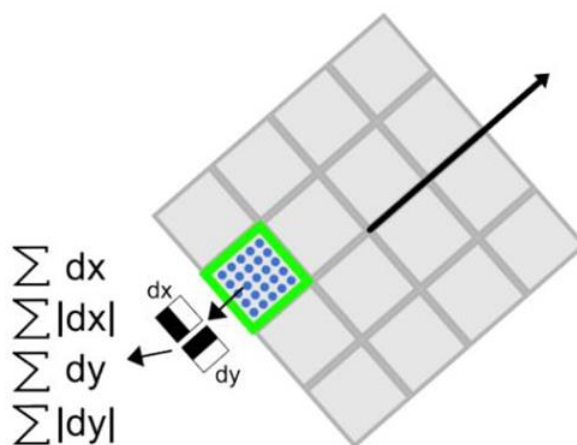


图 2-8 SURF 特征点描述

### (5) 特征点匹配

SURF 通过计算两个特征点间的欧式距离判断特征点是否匹配，欧氏距离越短，就表示两个特征点的匹配度越好。此外还加入了 Hessian 矩阵迹作为判断，只有当两个迹符号相同时，这两个特征的明暗变化方向相同，否则，两个特征点的明暗变化方向是相反的，这时候便不通过欧式距离长短作为判断条件，直接将这两特征点排除。

## 2.4.3 ORB 特征点算法

ORB 特征点由一种改进的 FAST 角点“Oriented FAST”，和一个“Rotated BRIEF”描述子两部分构成，因此，ORB 特征提取可以分为如下两个步骤。

### (1) FAST 角点提取

如图 2-9 所示，先从图像中选择任一像素点  $p$ ，并将它的亮度值设为  $I_p$ ；人为设定一个合适的阈值  $T$ ；考虑像素点  $p$  周围的 16 个像素点；假如这 16 个像素点有一半以上像素点的亮度值大于  $I_p+T$  或小于  $I_p-T$ ，则像素点  $p$  可以被看作角点。

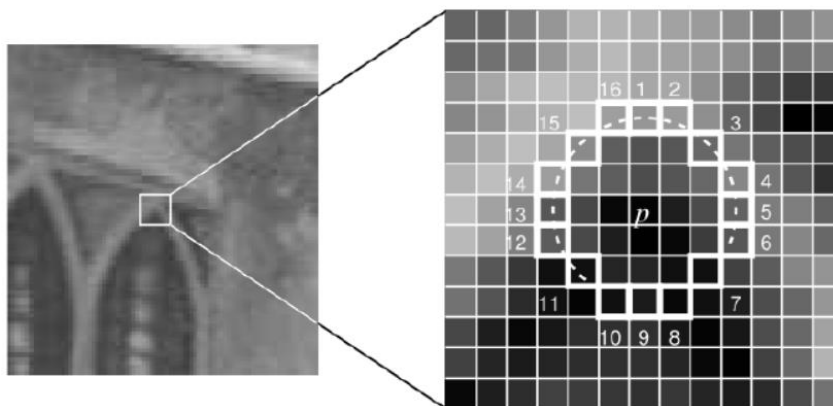


图 2-9 像素点检测示意图

计算如下

$$N = \sum_{x \in P} |I(x) - I(p)| \geq T \quad (2-22)$$

其中  $I(x)$  为周围任意一点的像素值， $I(p)$  是所选点的像素灰度值。

计算特征点附近的图像灰度质心来选取特征点的方向，操作步骤如下：

1) 在一个小的图像块中，定义图像块的矩为

$$m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y), p, q = \{0,1\} \quad (2-23)$$

2) 通过矩可以找到图像块的质心

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (2-24)$$

3) 通过连接图像块的几何中心  $O$  与质心  $C$ ，得到一个方向向量  $\overrightarrow{OC}$ ，特征点的方向定义为

$$\theta = \arctan(m_{01}/m_{10}) \quad (2-25)$$

(2) BRIEF 描述子

BFIEF 是一种由 0 和 1 组成二进制描述子，在以特征点为中心的邻域内任选两个点  $x$  和  $y$ ，其灰度值分别为  $p(x)$ 、 $p(y)$ ，对比其像素灰度值大小，最后得到 128 维向量。计算如式(2-26)所示

$$\tau(p; x, y) = \begin{cases} 1: p(x) < p(y) \\ 0: p(x) \geq p(y) \end{cases} \quad (2-26)$$

## 2.5 本章小结

本章主要研究了视觉 SLAM 相关理论。首先介绍了目前视觉 SLAM 的经典框架以及各组成部分，接着分析了双目视觉模型和相机标定内外参矩阵参数，同时对所用相机进行了标定实验，并获得了正确的相机内外参数，最后介绍是目前视觉 SLAM 前端常用的特征检测方法，为后续文章的深入研究和实验提供了重要的理论基础。

### 3 基于双目视觉的位姿估计算法

前端的特征提取和匹配是视觉 SLAM 的关键一环，拥有精确的特征提取及匹配对后续位姿估计的准确性有着巨大的影响。因此，本章基于双目相机的位姿估计算法，从特征检测、特征匹配以及位姿估计方法等方面进行研究。在 SLAM 前端部分，本章提出了一种基于四叉树的融合 ORB-LBP 改进算法，提升了特征检测及匹配的精度，并使用 EPNP 的方法对相机进行位姿解算，得到了较为精准的相机位姿信息。

#### 3.1 基于四叉树的 ORB-LBP 改进算法

传统 ORB 特征提取算法已经得到了广泛的应用，是一种运行速度快，能很好满足无人机视觉 SLAM 系统实时性的算法，但是快速运动或明暗度变化不明显的场景，会出现特征点聚集或者漏检的情况，从而导致特征匹配精度差，定位精度也达不到理想要求。近年来出现了很多对 ORB 算法的改进方法。董永峰等<sup>[51]</sup>在 SURF 算法基础上融合 BRIEF 描述子，在提升匹配效率和准确度的同时，可以满足实时性的要求。李硕等<sup>[52]</sup>提出一种融合 BRIEF 和 FREAK 描述子的改进 ORB 算法，在尺度变化、明暗度变化时，特征匹配率和鲁棒性都有了很大的提升。

上述改进在一定程度上提升了 ORB 匹配精度。然而，在图像变化不明显的区域，不能检测到特征点，而在一些图像变化剧烈的地方容易出现特征点聚集现象，对后续特征匹配与位姿估计的精度会有巨大的影响。本文针对特征点分布不均匀，匹配精度差的问题，对 ORB 算法进行了改进，通过建立四叉树及融合 BRIEF-LBP 的方法，以期达到在少量增加计算时间的前提下，特征点分布更为均匀、匹配精度更高。

##### 3.1.1 ORB 算法

ORB 是一种局部特征提取方法。利用加速分割测试特征 (FAST) 关键点检测器提取图像特征关键点；利用二值稳定独立基本特征描述 (BRIEF) 描述子提取图像特征描述子。

##### 3.1.2 FAST 角点检测算法

以待检测像素为圆心，三个像素的值为半径做一个圆，该圆交点的周围共有 16 个像素，当检测这 16 个像素中有大多数连续的与中心点的像素差大于设定的阈值时，则把这个像素点认作为角点，然后再循环这个步骤检测每一个像素。

### 3.1.3 BRIEF 描述子

在检测到的角点附近框选一个方形窗口，在这个窗口内随机挑选出 128 组点对。再对每一个点对比较它们的灰度值，重复操作，直至所有 128 组点对都进行比较，生成了一个 128 维的二进制字符串，就是 BRIEF 描述子。

图 3-1 为传统 ORB 的特征点检测结果，可以看出所检测到的特征点都容易集中于部分区域特征点分布极度不均匀，只有在一部分特征明显的区域存在密集的特征点，而在大部分区域都未能检测到特征点，从而会导致后续匹配精度差。



图 3-1 传统 ORB 特征点检测

图 3-2 为使用 RANSAC 匹配算法对 ORB 特征点进行匹配得到的结果，可以看出图中匹配精度低，效果差，出现了很多误匹配现象。

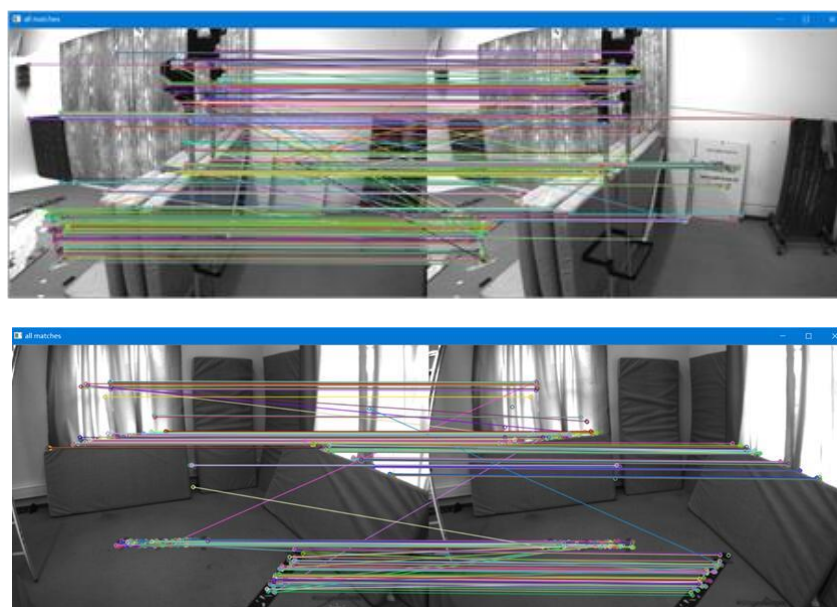


图 3-2 传统 ORB 特征点匹配

### 3.1.4 基于融合描述子的改进 ORB 算法

针对 ORB 算法存在的特征点分布不均匀、匹配程度低、精度差等问题, 本文通过构建四叉树、融合 BRIEF-LBP 特征描述子的方法来改进 ORB 算法。

图 3-3 为算法整体流程图: 首先构建图像金字塔, 通过划分网格计算角点提取阈值来提取 FAST 角点; 其次, 通过构建四叉树使角点分布均匀; 然后, 通过融合 BRIEF-LBP 的方法对提取的角点进行特征描述; 最后, 利用 RANSAC 的方法进行特征匹配。

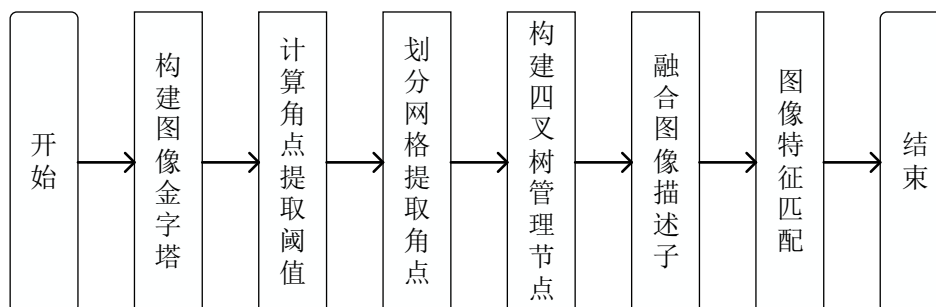
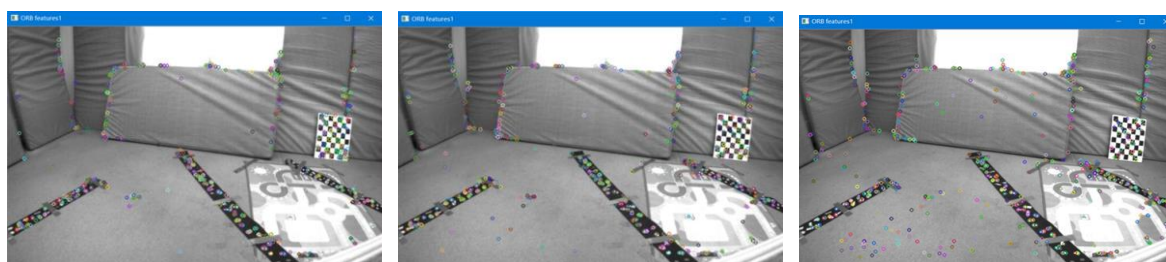


图 3-3 本文算法流程图

### 3.1.5 构建图像金字塔

本文通过建立图像金字塔, 使 ORB 特征具有尺度不变性。为获得不同分辨率的图像, 其中最底层金字塔为初始图像, 进行一定比例的缩放, 对不同金字塔层的特征点进行匹配, 实现了尺度不变性。

图 3-4 为不同金字塔层所检测的特征点结果, 分别建立了七层、五层、三层金字塔, 可以看出在不同层数金字塔下图像所提取到的角点都有所区别。通过实验, 最终建立 3 层金字塔层, 达到效果最优。



(a)七层金字塔

(b)五层金字塔

(c)三层金字塔

图 3-4 不同金字塔层特征点

### 3.1.6 角点提取

将图层划分为  $M \times N$  的网格, 记录每个网格中任意一个像素点的灰度值为  $I_p$ ,

网格平均值灰度值为  $I_q$ ，并定义阈值  $T$

$$T = \alpha \times \frac{1}{m} \sqrt{\sum_{i=1}^m (I_p - I_q)^2} \quad (3-1)$$

其中  $\alpha$  是一个自适应变量，本文  $\alpha$  取 1.2； $m$  为网格内像素点数量。在每个网格中对各像素与阈值  $T$  进行比较，进行角点提取，若在该网格中未提取到角点，则将阈值  $T$  降低为原有的 1/2 再进行角点提取。遍历所有网格，直至提取到的角点数量满足设定的角点数量。

图 3-5 为自适应阈值下的角点提取，可以看出角点的分布情况较图 3-4(c) 三层金字塔提取到的特征点分布相比已经有了改善，但提取到的角点在一定程度上仍会在部分网格内密集分布。由于这部分角点对阈值的响应较大，使得这部分角点占据极大一部分所要求的角点数量，导致一些网格内还未检测到角点就结束角点提取操作。



图 3-5 角点提取

### 3.1.7 构建四叉树

在 3.1.6 节提取到的角点在一定程度上仍会在部分网格内密集分布，会对算法的计算效率有很大的影响，因此构建四叉树限制角点分布。四叉树示意图如 3-6 所示

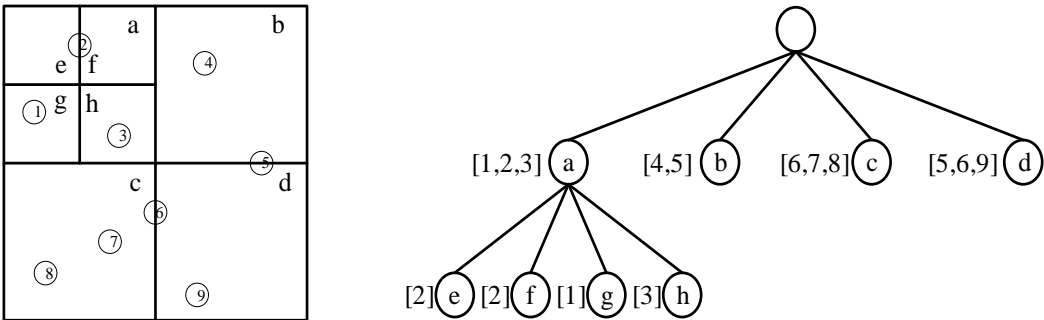


图 3-6 四叉树示意图

首先通过 3.1.6 节的方法对整张图像进行角点检测，然后将整个图像空间划分为四个相同的子空间，若在该子空间内检测到的角点数量大于 1，则继续将该子空间分裂为四个新的子空间继续检测，但若是特征点数量太多，会导致四叉树的深度过深，在

进行空间检索时效率会比较低下，因此本文限制了四叉树的深度为 4 层，若深度达到 4 层后子空间内的角点数量仍大于 1，则留取子空间内响应值最大的角点，舍弃其他角点。

### 3.1.8 BRIEF-LBP 融合描述子

ORB 使用改进的 BRIEF 进行特征描述，减少了特征匹配所需的时间，但是其描述符的相关性仍然很大，使得图像信息的利用不完整，容易导致特征点的不匹配。针对上述问题本文将 BRIEF 描述子与 LBP 描述子融合的方法，其融合流程如图 3-7 所示。

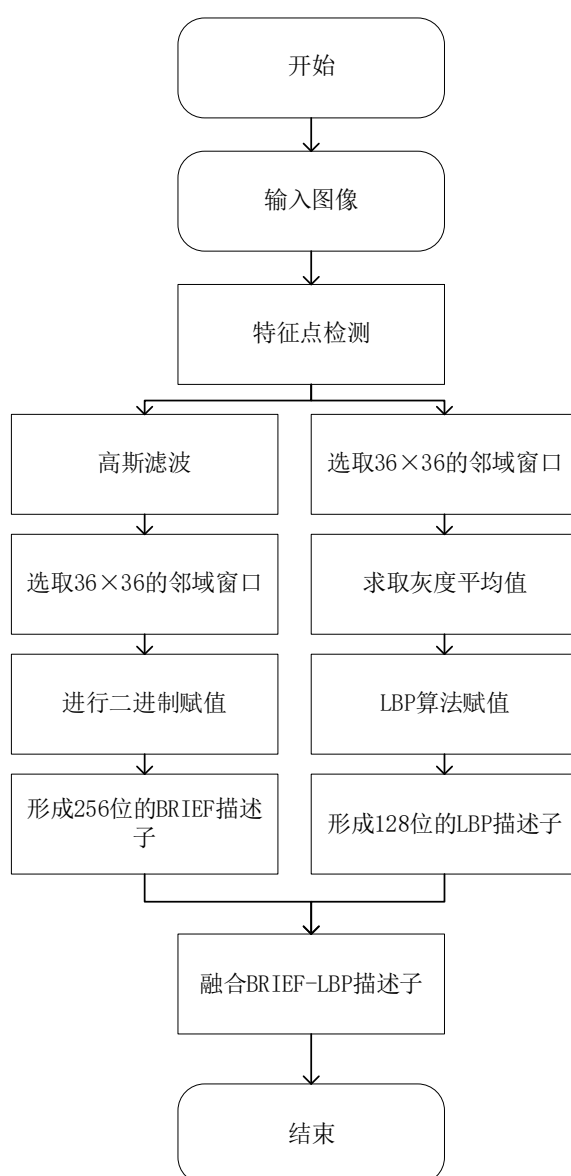


图 3-7 BRIEF-LBP 描述子融合流程图

首先,通过 FAST 角点检测算法进行特征点的提取。以特征点周围选取邻域窗口,在窗口内随机选取像素点块,比较平均值大小,进行二进制赋值,如式(3-2)所示。

$$\lambda(p; r_1, c_1, r_2, c_2) = \begin{cases} 1: I(r_1, c_1) < I(r_2, c_2) \\ 0: I(r_1, c_1) \geq I(r_2, c_2) \end{cases} \quad (3-2)$$

其中  $I(r_1, c_1)$  和  $I(r_2, c_2)$  分别为  $(r_1, c_1)$  和  $(r_2, c_2)$  处像素的强度。如果  $I(r_1, c_1)$  小于  $I(r_2, c_2)$ , 则  $\lambda = 1$ ; 否则,  $\lambda = 0$ 。重复操作, 最终形成一个 256 位的二进制编码。

其次以特征点为中心, 半径为  $R$  做圆形区域。一个中心像素点  $(x, y)$  周围的其他像素点  $g_k$  的坐标  $(x_k, y_k)$  可以表示为

$$(x_k, y_k) = \left( x_r + R \cos\left(2\pi \frac{r}{k}\right), y_r - R \sin\left(2\pi \frac{r}{k}\right) \right) \quad (3-3)$$

其中,  $R$  为圆形 LBP 半径,  $r$  为第  $r$  个采样点,  $k$  为样本点数量。对计算得到的坐标采用双线性内插值取整, 以取整后的坐标作为选取像素点的坐标, 将中心像素点的灰度值与选取像素点的灰度值进行比较, 若中心点的灰度值大, 则记为 0, 否则为 1。重复操作, 最终获得 128 位二进制值, 并将此二进制值循环排列直至得到最小的一组二进制值便作为该像素点的 LBP 值。

本文将得到的 256 位 BRIEF 描述子与 128 位 LBP 描述子级联得到新的 384 位的二进制描述符, 将该新的描述子作为特征点的描述子。

### 3.1.9 特征匹配

本文采用暴力匹配方法对特征点的进行匹配, 对所有待匹配的特征点进行汉明距离测试, 舍弃汉明距离较大的特征点, 对于剩下的匹配点利用随机抽样一致性 (RANSAC) 方法进行进一步的剔除。

## 3.2 相机位姿解算

完成了特征点匹配之后, 就能够使用匹配结果对相机位姿进行估计。根据所使用的传感器类型不同, 解算的方式有 2D-2D, 3D-3D, 3D-2D 三种。本章实验使用的双目相机可以得到特征点的 3D 定位信息, 并且对特征点在镜头中的 2D 投影情况也能够判断, 即可通过 PnP(Perspective-n-Point) 的方法来解算相机位姿。

EPnP 方法是求解相机位姿最为高效的算法之一, 该方法指出在三维空间中任何一个点都分别为四个不共面的点加权总和表示, 这四个不共面的点称为控制点。

世界坐标系下任意选取  $n$  个三维点  $P_i^w = [x_i \ y_i \ z_i]^T, i = 1, 2, \dots, n$ , 在像素坐标



系下的投影点为  $u_i = [u_i \quad v_i]^T, i=1,2,\dots,n$ ，设四个控制点在世界坐标系下的坐标为  $Q_j^w, j=1,2,3,4$ ，在相机坐标系下的坐标为  $Q_j^c, j=1,2,3,4$ 。

则任意三维点可以表示为

$$\begin{aligned} P_i^w &= \sum_{j=1}^4 \alpha_{ij} Q_j^w, \sum_{j=1}^4 \alpha_{ij} = 1 \\ P_i^c &= \sum_{j=1}^4 \alpha_{ij} Q_j^c, \sum_{j=1}^4 \alpha_{ij} = 1 \end{aligned} \quad (3-4)$$

根据第二章的双目视觉模型，可得

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K P_i^c = K \sum_{j=1}^4 \alpha_{ij} Q_j^c = K \sum_{j=1}^4 \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \quad (3-5)$$

式中， $s_i$  表示空间点的深度

$$s_i = \sum_{j=1}^4 \alpha_{ij} z_j^c \quad (3-6)$$

将深度信息带回方程(3-5)，可得

$$\begin{aligned} \sum_{j=1}^4 \alpha_{ij} f_x x_j^c + \alpha_{ij} (u_o - u_i) z_j^c &= 0 \\ \sum_{j=1}^4 \alpha_{ij} f_y y_j^c + \alpha_{ij} (v_o - v_i) z_j^c &= 0 \end{aligned} \quad (3-7)$$

其中

$$f_x = \frac{f}{Q_x}, f_y = \frac{f}{Q_y} \quad (3-8)$$

令  $X = [Q_1^{cT} \quad Q_2^{cT} \quad Q_3^{cT} \quad Q_4^{cT}]^T$ ，是一个  $12 \times 1$  的矩阵，可以得到线性方程组

$$MX = 0 \quad (3-9)$$

其中  $M$  为一个  $2n \times 12$  的矩阵。

对上式进行分解求解，在求解  $M^T M$  的特征值与特征向量时，用  $K_i$  表示  $M^T M$  零特征值所对应的特征向量，则可求得  $X$ ：

$$X = \sum_{i=0}^N \beta_i K_i \quad (3-10)$$

其中  $N$  表示核空间的维数，通常  $N$  取 1, 2, 3, 4。

由于控制点距离是恒定的，所以不管相机坐标系或者世界坐标系，每对控制点之

间的距离是恒定的，即

$$\|q_i^w - q_j^w\|^2 = \|q_i^c - q_j^c\|^2, \{i, j\} \in \{1, 2, 3, 4\} \quad (3-11)$$

由于  $N$  的数值不是固定的，该约束公式会随着  $N$  的变化而相应变化。

当  $N=1$  时，约束可写成

$$\|\beta K^i - \beta K^j\|^2 = \|q_i^w - q_j^w\|^2 \quad (3-12)$$

其中  $K^i K^j$  是与控制点  $q_i^w, q_j^w$  对应的特征向量中的子向量。可求得

$$\beta = \frac{\sum_{\{i, j\} \in \{1, 2, 3, 4\}} \|K^i - K^j\| \cdot \|q_i^w - q_j^w\|}{\sum_{\{i, j\} \in \{1, 2, 3, 4\}} \|K^i - K^j\|^2} \quad (3-13)$$

当  $N=2, 3, 4$  时，约束可写成

$$\|\beta_1 K_1^i + \beta_2 K_2^i - (\beta_1 K_1^j + \beta_2 K_2^j)\|^2 = \|q_i^w - q_j^w\|^2 \quad (3-14)$$

此时以  $\beta$  为优化目标，可以建立以下优化函数

$$Error(\beta) = \sum_{(i, j)} \left( \|q_i^w - q_j^w\|^2 - \|q_i^c - q_j^c\|^2 \right), \{i, j\} \in \{1, 2, 3, 4\} \quad (3-15)$$

使用高斯-牛顿法进行优化，找到误差最小时的  $\beta$  值，并进一步找到相机坐标系中控制点的坐标，再根据质心坐标系数得到相机坐标系下三维点的坐标。最后，从世界坐标系和相机坐标系的变换关系中，求解得到相机位姿。

### 3.3 实验验证

为验证本文算法的有效性，对本文算法进行实验验证。所用的实验平台配置为 Intel(R) Core(TM) i5-8265U，内存 1.60GHz。采用了 EuRoC 公开数据集中的图片，该数据集包括了视觉(双相机)测量得到的数据。

#### 3.3.1 特征匹配实验

将本文算法与 ORB 算法、SIFT 算法以及论文[51]、论文[52]改进的 ORB 算法进行对比实验验证。分别做了特征点分布均匀度实验，算法耗时实验和算法匹配召回率实验。

为了验证特征点的分布均匀度。本文采用了基于区域统计信息计算分布均匀度的方法。

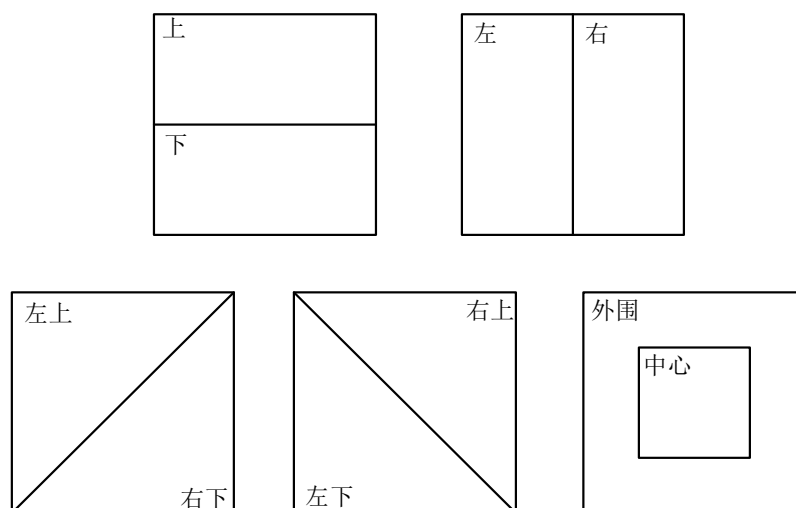


图 3-8 图像分割示意图

通过把图像划分为上、下、左、右、左上、右下、左下、右上、中心和外围 10 个区域。统计每个区域所检测到的特征点数量。然后统计每个区域分布向量的方差  $V$ 。然后计算所得特征点的分布均匀度  $u$

$$u = 101\log(V) \quad (3-16)$$

公式(3-16)是朱<sup>[53]</sup>给出的特征点分布均匀性的客观评价标准,利用公式(3-16)就可以计算得到整幅图像内特征点的分布均匀度,该数值越小,就表示不同区域内的特征点数量差别越小,分布均匀度越好。

对实验结果进行均匀化可视化计算,本文选取了 EuRoc 数据集中的部分图像进行特征点均匀化实验以及耗时实验,实验结果分布记录于表 3-1 和表 3-2。由表 3-1 分析可知,对比传统的 ORB 算法和论文[51]、论文[52]改进算法来说,本文算法对特征的均匀度有较大的提高。

表 3-1 不同算法分布均匀度比较

图像序列	分布均匀度			
	传统 ORB	董 <sup>[51]</sup> 改进算法	李 <sup>[52]</sup> 改进算法	本文算法
1	210.03	184.75	185.54	164.56
2	201.62	175.44	182.53	174.15
3	208.01	203.88	191.59	178.10
4	208.61	180.17	187.62	171.59
5	206.00	181.31	207.78	160.62
6	207.83	181.03	183.10	166.97

表 3-2 为不同算法所耗时间比较。可以看出本文算法比 ORB 算法所耗时间稍增加了一些，但比起论文[51]、论文[52]所改进算法仍稍快一筹。

表 3-2 不同算法时间比较

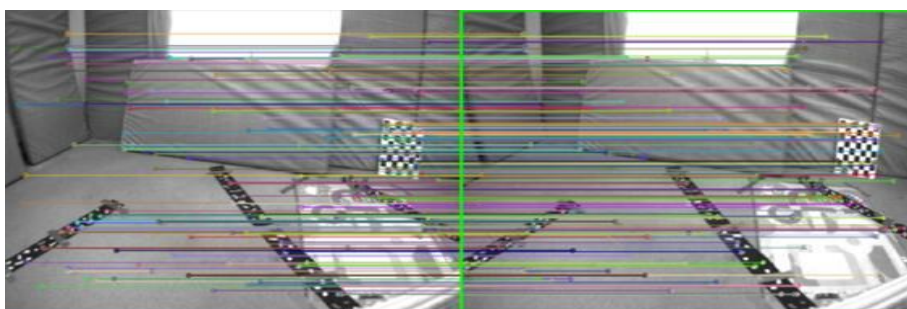
图像序列	时间/s			
	传统 ORB	董 <sup>[51]</sup> 改进算法	李 <sup>[52]</sup> 改进算法	本文算法
1	0.1741	0.2640	0.2550	0.2332
2	0.1768	0.2501	0.3281	0.2283
3	0.1623	0.2743	0.2454	0.2300
4	0.1611	0.2536	0.2553	0.2197
5	0.1651	0.2627	0.2509	0.2229
6	0.1635	0.2733	0.2594	0.1889

图 3-9 为传统 ORB 算法、SURF 算法和本文算法特征点提取结果，从图中可以看出传统 ORB 算法的提取结果极为不均匀，大部分都集中于一部分地区，无法有效的得到整幅图像的信息；SURF 算法提取结果比与 ORB 算法相差不大；而本文算法提取到的特征点遍布整幅图像。



图 3-9 特征点提取结果

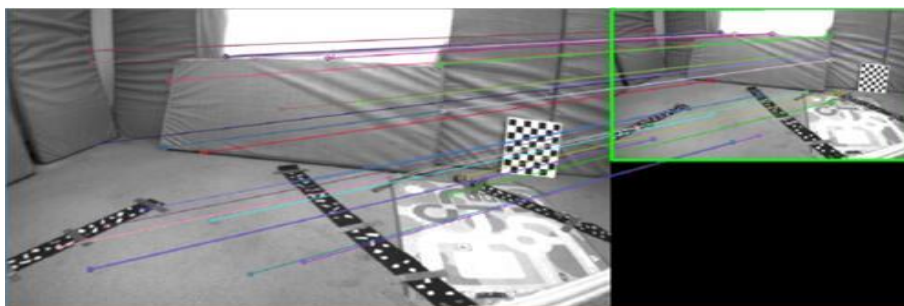
选取 EuRoc 数据集，并将数据集中的图像做明暗度变化、缩放、旋转操作，通过本文算法提取 400 个特征点进行匹配实验，实验结果如图 3-10 所示。图 3-10(a)为原始图像匹配结果，3-10(b)为改变图像明暗度后匹配结果，3-10(c)为图像缩小 50%后的匹配结果，3-10(d)为旋转 90° 后的匹配结果。匹配实验结果可以看出本文算法对于特征匹配结果质量高，匹配结果都未出现误匹配，结果表明本文算法对于图像明暗度变化、图像缩放、旋转等操作都具有较强的鲁棒性。



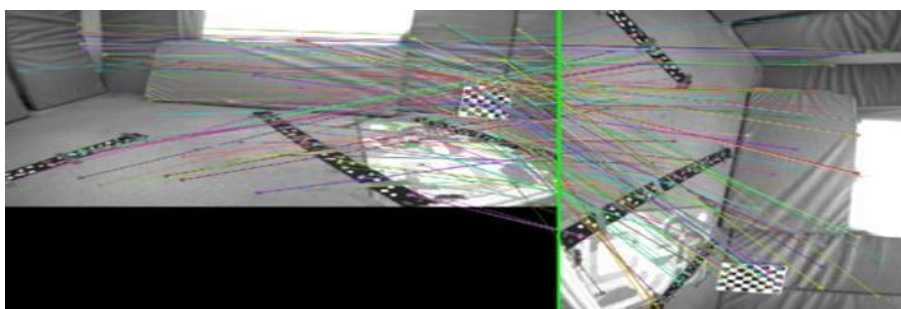
(a)本文算法匹配结果



(b)本文算法改变明暗度匹配结果



(c)本文算法尺寸缩放匹配结果



(d)本文算法旋转 90° 匹配结果

图 3-10 本文算法匹配结果

本文通过大量实验比较了 ORB 算法、SURF 算法、论文[51]、论文[52]改进算法与本文算法的平均所用时长，实验结果记录于表 3-3。实验结果表明，SURF 算法所用时间最长，ORB 算法所用时间最短，本文算法略高于 ORB 算法，仍远低于 SURF 算法，对于论文[51]、[52]改进 ORB 算法来说，本文算法速度更快。

表 3-3 不同算法平均所用时间/s

算法	正常	亮度减少 50%	缩放	旋转 90°
ORB	0.174	0.199	0.183	0.197
SURF	1.516	1.446	0.919	1.514
董 <sup>[51]</sup> 改进算法	0.264	0.273	0.239	0.277
李 <sup>[52]</sup> 改进算法	0.255	0.300	0.275	0.361
本文算法	0.233	0.244	0.235	0.266

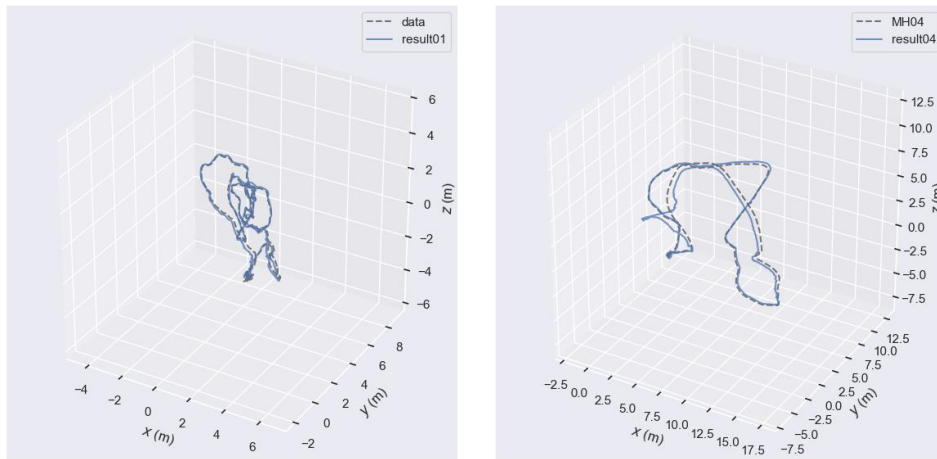
召回率可以反映匹配算法的查全率，即正确匹配占实际全部匹配的比例。本文经过了多次实验，得到的结果记录与表 3-4 中，对比 ORB 算法，本文算法的召回率有明显的提升，比起 SURF 算法仍有不足之处，这是因为 SURF 是构造尺度空间，通过高斯微分函数来识别兴趣点的而不是直接提取角点，鲁棒性非常好。论文[52]改进 ORB 算法对于正常的图像来说具有较高的召回率，但其对于旋转的鲁棒性较差。而本文算法在改变明暗度、对图像进行缩放、旋转操作时，召回率变化不大，可以看出本文算法对图像明暗变化，缩放旋转等操作时都有较强的鲁棒性。

表 3-4 不同算法平均召回率/%

算法	正常	亮度减少 50%	缩放	旋转 90°
ORB	34.00	26.50	25.25	38.25
SURF	80.00	55.25	49.75	78.75
董 <sup>[51]</sup> 改进算法	7.02	3.08	1.12	5.33
李 <sup>[52]</sup> 改进算法	55.00	50.60	31.09	7.25
本文算法	49.05	36.11	37.08	40.86

### 3.3.2 位姿估计实验

选取 EuRoC 数据集中的 MH01\_easy 和 MH04\_difficult 两个序列进行实验，实验结果如图 3-11 所示，图 3-11(a)为 MH01 实验轨迹对比图，图 3-11(b)为 MH04 实验轨迹对比图。其中虚线部分为参考轨迹，实线部分为本文实验结果轨迹。从实验结果可以看出，本章使用改进 ORB 算法进行的位姿估计在简单的地图中位姿比较精确，在较难的地图下获取的位姿信息误差较大，说明了纯视觉位姿估计对于环境复杂，运动速度变化大的情况所呈现的效果不好。

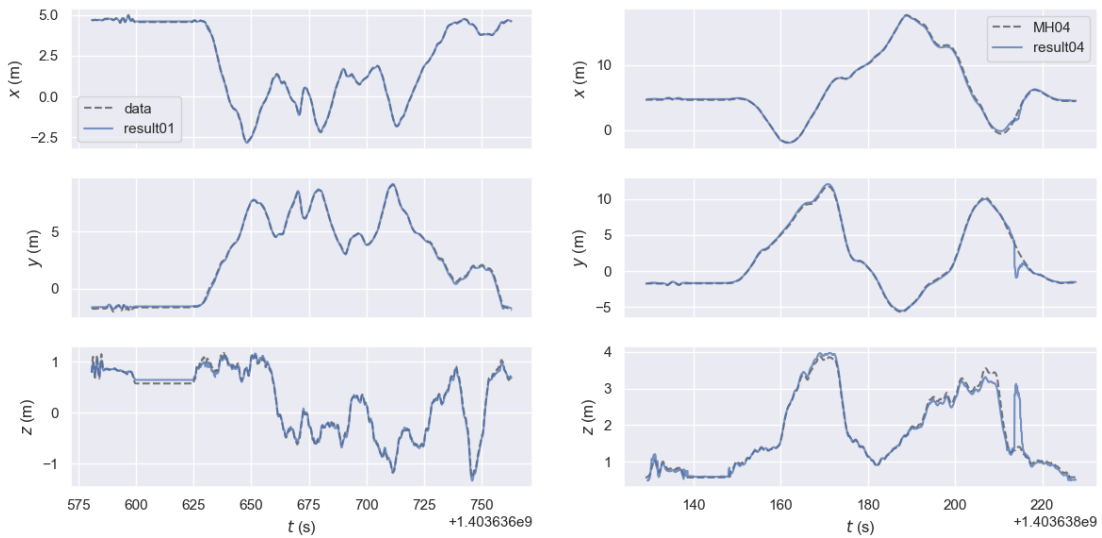


(a)MH01 实验结果图

(b)MH04 实验结果图

图 3-11 不同难度数据集下位姿估计结果图

继续分析在序列 MH01\_easy 和 MH04\_difficult 实验结果在  $x,y,z$  三轴的位置误差, 实验结果如图 3-12 所示。可以清晰的看出在运动稳定, 幅度小的 MH01 序列下, 仅在起飞阶段本章实验结果稍有误差; 而在运动速度快、幅度大的 MH04 序列下, 本文实验结果在大多数时间内仍具有较高的鲁棒性, 但也出现了部分时间出现跟踪丢失的情况, 因此本文下一步融合惯导信息进行进一步的优化。



(a)MH01 实验结果图

(b)MH04 实验结果图

图 3-12 不同难度数据集下位姿估计结果三轴对比图

### 3.4 本章小结

本文算法通过构建四叉树和融合 BRIEF-LBP 描述子的方法来实现 ORB 算法中特征点的均匀分布以及特征匹配的准确性。首先, 对输入图像构建金字塔尺度空间,

在每一图层上进行 FAST 关键点的检测；其次，通过构建四叉树来限制特征点分布；最后，在生成描述子的过程中用 128 位 LBP 描述算子融合 BRIEF 中 256 位描述子形成新的 384 位的特征描述子，新的描述子可以更充分的表达图像信息，使得匹配结果更为精确。

实验结果表明：对比传统 ORB，本文算法虽然在速度上比 ORB 算法降低了 5%，但均匀度提升了 66cm 左右，召回率也提升了 10%；与论文[51]、[52]相比，本文算法在速度是提升了 2%至 5%的同时，特征点分布均匀度提升了 48cm 和 49cm，召回率也提升了 36.63%和 4.925%。实现了在少量增加计算量的同时，特征点均匀度和匹配精度方面效果有较大提升。同时对改进算法进行了位姿解算，从结果看出本文算法在简单运动的情况下具有良好的定位结果，然后面对复杂运动的情况本文实验效果出现了较大的误差。



## 4 视觉和 IMU 融合的位姿估计算法

纯视觉算法的特征完全依靠摄像头获取的数据，易受光照变化干扰，其鲁棒性较低，极易使得所估计的位姿不准确形成累积误差。且当无人机的应用场景为室内、山区、灾难现场等光照不足的地方，可以通过自身惯导信息来弥补纯视觉的不足，因此本文研究视觉-惯导融合算法提升 SLAM 在无人机平台的鲁棒性与实时性。

### 4.1 视觉-惯导融合整体框架

一般情况下，视觉-惯导融合算法可以分为两类：紧耦合和松耦合。松耦合的基本思路，是将惯性测量单元和视觉传感器分开处理，由视觉传感器收集的图像信号处理，惯性测量单元采集角速度、加速度等信息进行处理。通过计算视觉处理图像得到位姿信息，而惯导信息进行积分处理之后也可以得到位姿信息。再把两种信息加以融合，就得到了融合过的综合位姿信息。

紧耦合的基本思路，是把图像信息和惯导信息相结合，再通过融合后的信息计算出位姿信息。紧耦合的方法可以有效利用各传感器的信息，提高测量精度，对比松耦合的方法，紧耦合的方法具有更强的鲁棒性。本文采用了紧耦合的方法来实现视觉与惯导信息融合。整个算法的框架如图 4-1 所示：

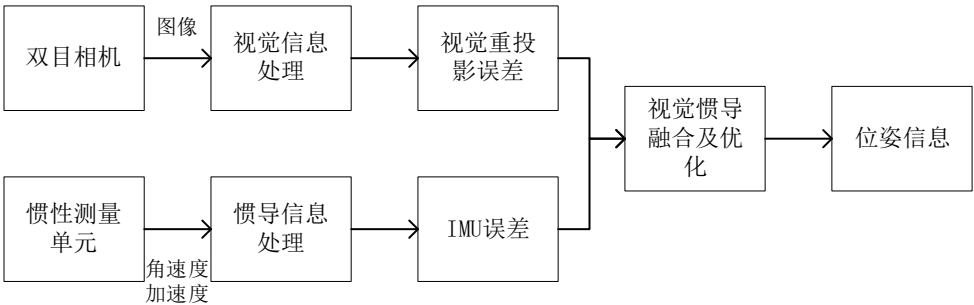


图 4-1 视觉惯导融合框图

首先，用惯性测量单元读取角速度、加速度的信息，并对惯导数据进行预积分处理以解决惯导频率与图像帧频率不一致问题，并根据两帧图像间的惯导信息计算以获取自身位姿信息。对于图像数据，使用第三章的改进 ORB 算法，对图像信息进行特征提取和匹配，这样就可以得到摄像机的姿态信息。在此基础上，将惯导信息与图像信息进行了融合，并进行了非线性优化以整合二者信息，将惯导信息的偏差和图像中的重投影误差结合在一起通过解算优化方程，就可以得到融合后的精确位姿信息。

### 4.2 IMU 预积分

由于摄像机拍摄的频率要比 IMU 的信号频率低得多，因此在将 IMU 与图像信息进行融合之前，必须要考虑到时间的同步或测量频率。利用预积分的方法，克服世界坐标系统中的误差积累，解决了摄像机和 IMU 频率不统一的问题。

如图 4-2 所示， $i$  和  $j$  为两帧相邻的图像帧，由于 IMU 频率远高于图像频率，IMU 在相邻两帧图像  $(i, j)$  间每隔时间  $t$  便会产生一次信息，将  $i$  与  $j$  两帧图像之间所有 IMU 信息进行预积分处理，得到 IMU 一段时间内的相对位姿，以便匹配图像和 IMU 的频率。由于调整了图像和 IMU 的频率，这些 IMU 信息就变成滑动窗口里的状态量，以用于后端联合优化。

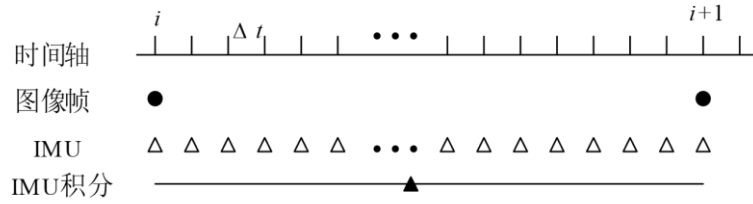


图 4-2 IMU 预积分示意图

假定当前 IMU 的测量结果是两幅相邻的图像帧之间的测量差值，在世界坐标系下，利用 IMU 的预积分值求出平移信息  $p$ 、速度  $v$  以及旋转四元数  $q$  的公式为：

$$P_{b_{k+1}}^W = P_{b_k}^W + v_{b_k}^W V t_k + \iint_k^{k+1} \left( q_t^W \left( a_t^{\wedge} - b_{a_t} - n_a \right) - g^W \right) dt^2 \quad (4-1)$$

$$v_{b_{k+1}}^W = v_{b_k}^W + \int_k^{k+1} \left( q_t^W \left( a_t^{\wedge} - b_{a_t} - n_a \right) - g^W \right) dt \quad (4-2)$$

$$q_{b_{k+1}}^W = q_{b_k}^W \otimes \int_k^{k+1} q_t^W \otimes \left[ \begin{matrix} 0 \\ \frac{1}{2} \left( w_t^{\wedge} - b_{w_t} - n_w \right) \end{matrix} \right] dt \quad (4-3)$$

式中， $w$  代表世界坐标系， $b_k$  表示第  $k$  帧图像。 $q_t^W$  为在  $b_k$  时的旋转信息。 $p_{b_{k+1}}^W$  是世界坐标系  $k+1$  帧图像时的位移信息，同样， $v_{b_{k+1}}^W$ 、 $q_{b_{k+1}}^W$  是世界坐标系下  $k+1$  帧图像时的速度信息和旋转信息。 $a_t^{\wedge}$ 、 $b_{a_t}$ 、 $n_a$  表示加速度计读数、偏差和噪声误差， $w_t^{\wedge}$ 、 $b_{w_t}$ 、 $n_w$  表示陀螺仪的读数、偏差和噪声误差。 $g^W$  表示重力加速度。

通过上述分析判断，有必要对两幅图像帧间的惯性测量值加以换算，以减少原始状态对位姿的影响，而使其转换为只与两帧间的惯性测量值有有关的相对约束。经过转换后的状态公式为

$$\begin{aligned} q_W^{b_k} P_{b_{k+1}}^W &= q_W^{b_k} \left( P_{b_k}^W + v_{b_k}^W V t_k - \frac{1}{2} g^W V t_k^2 \right) + \alpha_{b_{k+1}}^{b_k} \\ q_W^{b_k} v_{b_{k+1}}^W &= q_W^{b_k} \left( v_{b_k}^W g^W V t_k \right) + \beta_{b_{k+1}}^{b_k} \\ q_W^{b_k} \otimes q_{b_{k+1}}^W &= \gamma_{b_{k+1}}^{b_k} \end{aligned} \quad (4-4)$$

式中,  $W$  表示世界坐标系,  $P_{b_{k+1}}^w$  表示下  $b_{k+1}$  帧的位置信息,  $q_w^{b_k}$  表示  $b_k$  时刻的旋转四元数,  $g^w$  表示重力加速度,  $\alpha_{b_{k+1}}^{b_k}$  为平移变化量,  $\beta_{b_{k+1}}^{b_k}$  为速度的变化量,  $\gamma_{b_{k+1}}^{b_k}$  为旋转四元数的变化量。  $\Delta t_k$  为两图像帧之间的时间差, 其中  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$  分别为

$$\begin{aligned}\alpha_{b_{k+1}}^{b_k} &= \iint_k^{k+1} \gamma_{b_i}^{b_k} (a_t^\wedge - b_{a_i} - n_a) dt^2 \\ \beta_{b_{k+1}}^{b_k} &= \int_k^{k+1} \gamma_{b_i}^{b_k} (a_t^\wedge - b_{a_i} - n_a) dt \\ \gamma_{b_{k+1}}^{b_k} &= \int_k^{k+1} \gamma_{b_i}^{b_k} \otimes \begin{bmatrix} 0 \\ \frac{1}{2} (w_t^\wedge - b_{w_i} - n_w) \end{bmatrix} dt\end{aligned}\quad (4-5)$$

其中,  $a_t^\wedge$ 、 $w_t^\wedge$  表示加速度计和陀螺仪在  $t$  时刻的读数,  $b$  表示测量偏差,  $n$  表示噪声误差。从公式(4-4)可知, 只需要计算出  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$  三个量便可得到 IMU 预积分计算结果, 这三种变化量均可以由与  $b_k$  时刻旋转无关的变化值表示。这也减少了当原始状态信息改变后必须再次进行运算的视觉, 大大降低了计算的复杂性。根据公式(4-5)用欧拉积分法展开, 便可最终可以得到这三个变量的值。这样就完成了 IMU 值预积分过程。

### 4.3 基于滑动窗口的视觉与 IMU 融合优化模型

在视觉-惯导融合的算法中, 基于滤波器的算法和基于优化的算法各自有相应的优点, 基于滤波器的算法计算量比较低, 但是在精度方面不如优化的方法; 而由于对优化方案的进一步研究, 并且计算机性能也更为强大, 基于优化的视觉-惯导融合的优化目前已被广泛应用。因此本文采用了基于优化模型的视觉-惯导融合模型, 系统模型如图 4-3 所示:

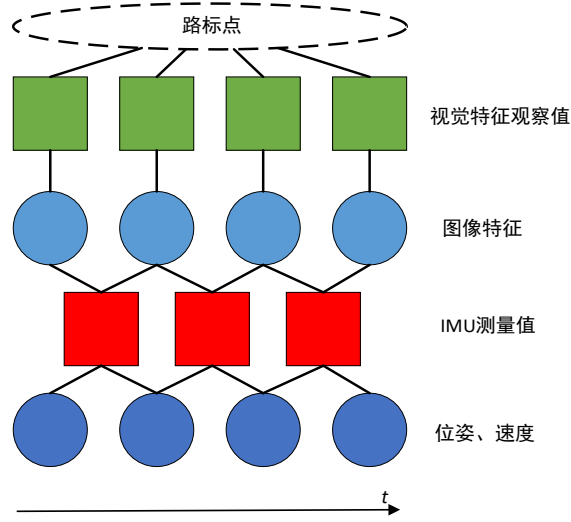


图 4-3 视觉惯导紧耦合优化模型

由上图可以看出，经由视觉传感器可以获得视觉特征测量值，进而通过相应算法得到对应图像特征信息，再融合同一时间段 IMU 测量值后，得到更为精确的位姿、速度信息。

视觉-惯导具体融合步骤如下：

1) 对两个相邻的图像帧之间的 IMU 数据进行预积分处理，从而得到系统状态增量  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$ 。

2) 根据 IMU 运动模型，能够在下一图像帧出现对 IMU 进行积分，从而获得 IMU 增量的预测值。

3) 将预测的 IMU 预积分预测值与实际情况下由 IMU 测量观测得到下一个图像帧时的观测值作差，可以得到惯导数据的残差。

4) 在视觉 SLAM 的最优代价函数中引入惯导数据的残差，从而使最优代价函数变成了视觉重投影误差与惯导数据残差之和。

公式(4-6)表示了运动方程和观测方程描述机体运动的过程。 $k+1$  时刻的运动状态  $X_{k+1}$  可以根据  $k$  时刻的状态和  $k$  时刻到  $k+1$  时刻之间的 IMU 测量值来预测。观测方程是用于描述相机在  $k+1$  时刻的视觉检测特征点的过程。

$$\begin{aligned} X_{k+1} &= f(X_k, u_{k+1}) + \omega_{k+1} \\ p_{k+1} &= h(p_{k+1}, X_{k+1}) + v_{k+1} \end{aligned} \quad (4-6)$$

$X_{k+1} = (R_{k+1}, p_{k+1}, v_{k+1}, b_{k+1}^g, b_{k+1}^a)^T$  表示了系统在  $k+1$  时刻的状态，包含了在  $k$  时刻的旋转信息、速度信息、平移信息、陀螺仪和加速度计的零偏， $u_{k+1}$  为 IMU 提供的运动测量。

机体的运动学过程是一个循环迭代的过程，根据运动学方程和输入来预测下一时刻的状态，并重新使用传感器的测量方程来矫正下一时刻状态。在视觉-惯导融合的

算法中, IMU 数据也起到了输入的作用, 从而可以利用运动学方程对系统状态做出更精准的预测。然而, 加入的 IMU 数据同样存在误差, 因此, 在非线性优化时除了需优化视觉重投影误差外, 还需优化 IMU 误差。

#### 4.4 非线性优化

本文优化方程可以分为视觉重投影误差和 IMU 误差两部分组成。视觉的重投影误差信息, 指的是一个三维空间点在图像平面上的重投影, 与计算中值得到的虚拟的像素点的误差。IMU 误差, 指的是 IMU 运动学模型对下一帧图像帧时的预积分预测值和真实测量值的偏差。将视觉-惯导信息进行融合, 即将两者的误差进行求和便得到组成的优化方程, 然后根据优化库中解决优化问题的方法即可对其进行求解。

通常在 IMU 预积分后可以获得  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$  三个系统增量, 对下一个图像帧到来时的 IMU 预积分的增量值进行预估, 从而可以得出系统增量预测值。当下一帧图像到达时, 能够得出实际的 IMU 测量值, 在进行预积分运算之后就能够得出实际的增量值。利用以上所得的增量预测值与实际增量值可以求得到一个差值, 此差值即为 IMU 误差  $r_B(z_{b_{k+1}}^{b_k}, X)$ 。

$$r_B(z_{b_{k+1}}^{b_k}, X) = \begin{bmatrix} \delta\alpha_{b_{k+1}}^{b_k} \\ \delta\beta_{b_{k+1}}^{b_k} \\ \delta\gamma_{b_{k+1}}^{b_k} \\ \delta b_a \\ \delta b_g \end{bmatrix} = \begin{bmatrix} q_W^{b_k} \left( P_{b_{k+1}}^W - P_{b_k}^W - v_{b_k}^W \Delta t_k + \frac{1}{2} g^W \Delta t_k^2 \right) - a_{b_{k+1}}^{b_k} \\ q_W^{b_k} \left( v_{b_{k+1}}^W - v_{b_k}^W + g^W \Delta t_k \right) - \beta_{b_{k+1}}^{b_k} \\ q_W^{b_k} \otimes q_{b_{k+1}}^W \otimes (\gamma_{b_{k+1}}^{b_k})^{-1} \\ b_{ab_{k+1}} - b_{ab_k} \\ b_{wb_{k+1}} - b_{wb_k} \end{bmatrix} \quad (4-7)$$

式中,  $\delta\alpha_{b_{k+1}}^{b_k}$ 、 $\delta\beta_{b_{k+1}}^{b_k}$ 、 $\delta\gamma_{b_{k+1}}^{b_k}$  为  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$  三个系统增量在两帧图像之间的误差值,  $\alpha_{b_{k+1}}^{b_k}$ 、 $\beta_{b_{k+1}}^{b_k}$ 、 $\gamma_{b_{k+1}}^{b_k}$  为通过 4.2 节中的 IMU 预积分求得的三个测量值。 $\delta b_a$ 、 $\delta b_g$  为加速度计和陀螺仪偏差的误差。

优化的代价函数如式(4-8)所示

$$\min \left\{ \sum_{(i,k) \in C} \left\| r_C(z_k^{C_i}, X) \right\|_{P_k^{C_i}}^2 + \sum_{k \in B} \left\| r_B(z_{b_{k+1}}^{b_k}, X) \right\|_{P_{b_{k+1}}^{b_k}}^2 \right\} \quad (4-8)$$

其中, 第一项为视觉的重投影误差, 第二项为 IMU 误差。 $C$  表示全部图像帧;  $B$  表示两帧图像帧之间所有 IMU 测量信息;  $i$  表示第  $i$  个特征点,  $k$  表示第  $k$  帧图像,  $z_k^{C_i}$  表

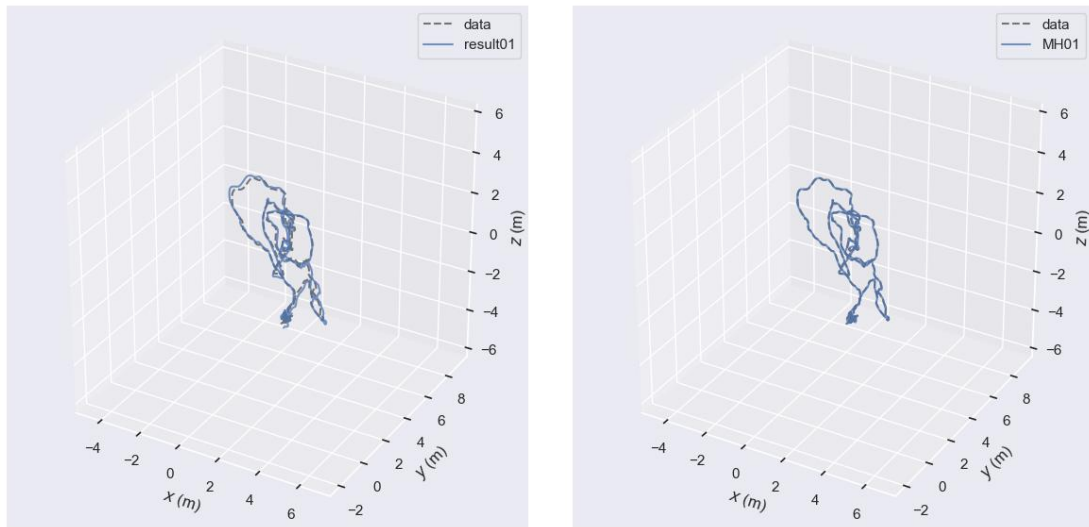
示第  $i$  个特征点在第  $k$  帧图像中的观测值,  $p_k^{C_i}$  表示第  $i$  个特征点在第  $k$  帧图像中对于观测的投影向量;  $z_{b_{k+1}}^{b_k}$  为由  $\alpha_{b_{k+1}}^{b_k}$  和  $\beta_{b_{k+1}}^{b_k}$  构成的向量,  $X$  为当前的状态值,  $p_{b_{k+1}}^{b_k}$  表示在  $b_k$  帧图像坐标系下  $b_{k+1}$  帧图像的位置信息。

最后, 再对上式利用高斯牛顿法进行求解, 如此便可获得更为精确的状态信息。

## 4.5 实验验证

为检验本文算法的有效性, 对本文算法进行实验验证。所用的实验平台配置为 Intel(R) Core(TM) i5-8265U, 内存 1.60GHz。采用了 EuRoC 公开数据集中的数据, 该数据集包括了视觉(双相机)和 IMU 测量得到的数据。

选取 EuRoC 数据集的 MH01\_easy 序列进行实验, 实验结果如图 4-4 所示, 图 4-4(a)为 VINS-mono 实验轨迹对比图, 图 4-4(b)为本文算法实验轨迹对比图。其中虚线部分为参考轨迹, 实线部分为本文实验结果轨迹。实验结果表明, VINS-mono 算法与参考轨迹有明显的误差之处, 而本文实验结果与参考轨迹具有很强的一致性, 最大误差仅有 0.09m, 平均误差也只有 0.02m 左右。

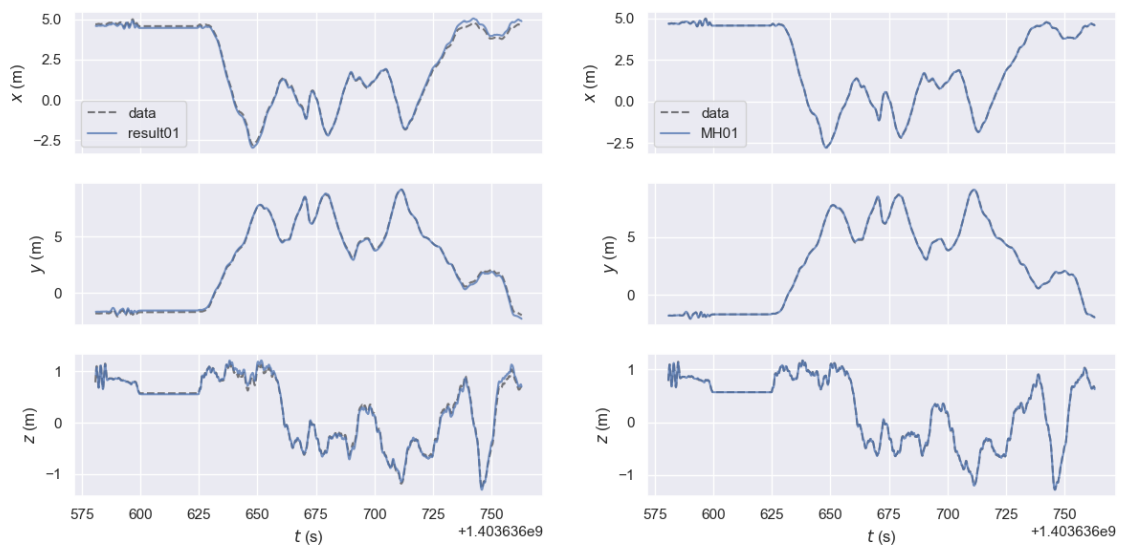


(a)VINS-mono 轨迹对比图

(b)本文轨迹对比图

图 4-4 本文实验结果对比图

图 4-5 为 VINS-mono 和本文实验结果在  $x, y, z$  轴方向的位置误差图, 对比两种算法的位置误差图, 整体上看, 本文算法的表现优于 VINS-mono, 三个方向上的定位精度都有所提升。从细节上看, VINS-mono 轨迹大致上与参考轨迹一致, 但仍有很多地方与参考轨迹具有一定的误差, 而本文算法所得结果与参考轨迹几乎完全一致, 效果比 VINS-mono 的效果更好。

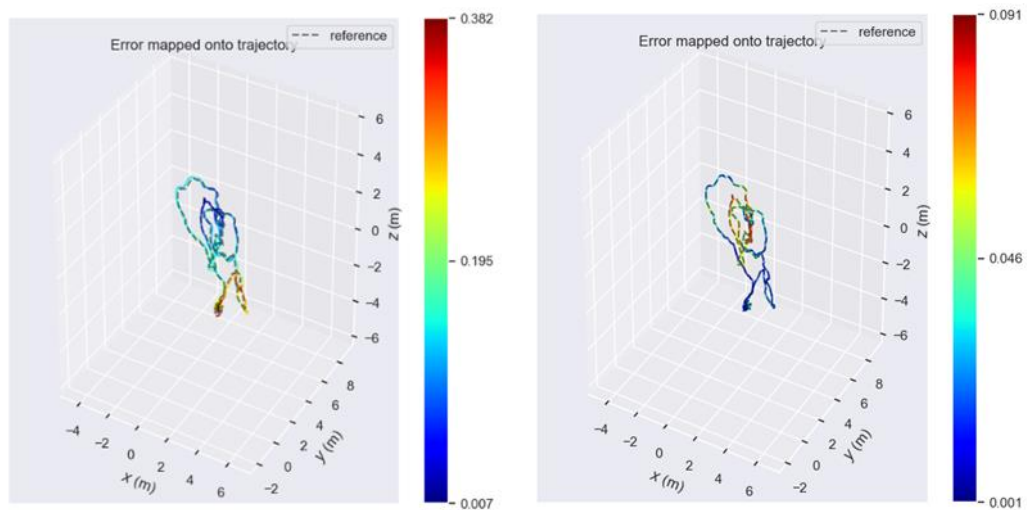


(a)VINs-mono 轨迹对比图

(b)本文轨迹对比图

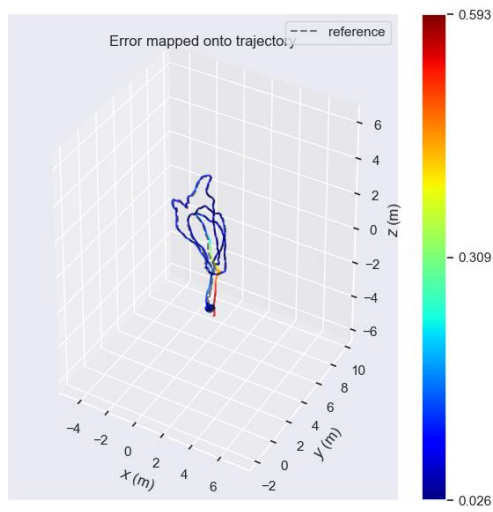
图 4-5 本文实验结果 xyz 轴对比图

本文选取了不同定位难度的序列进行实验，分别在 MH01\_easy, MH02\_easy, MH03\_medium, MH04\_difficult, MH05\_difficult 序列将本文算法与 VINS-Mono 进行对比。图(a1)、(a2)、(a3)、(a4)、(a5)为 VINS-Mono 实验图，(b1)、(b2)、(b3)、(b4)、(b5)是本文实验结果图，虚线为参考轨迹，实线为实验结果图，两者偏差越接近颜色越蓝，两者偏差越大颜色越红。

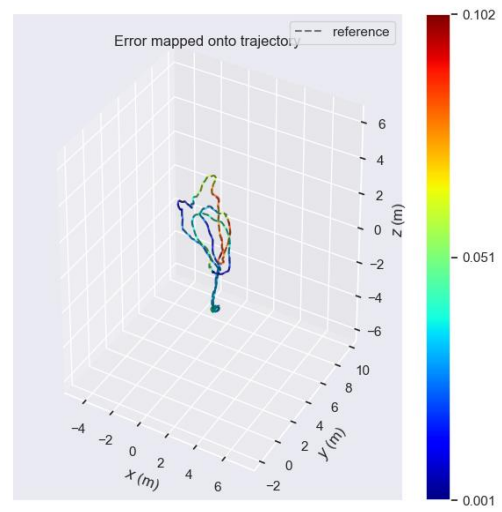


(a1) VINs-mono 轨迹对比图

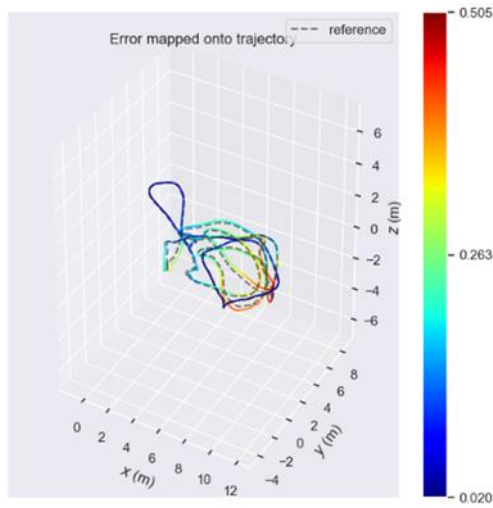
(b1) 本文轨迹对比图



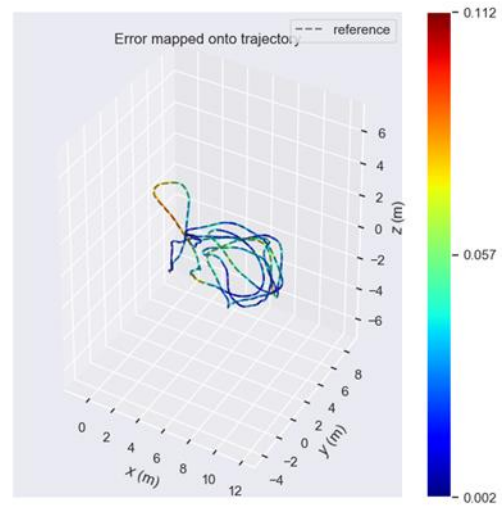
(a2) VINs-mono 轨迹对比图



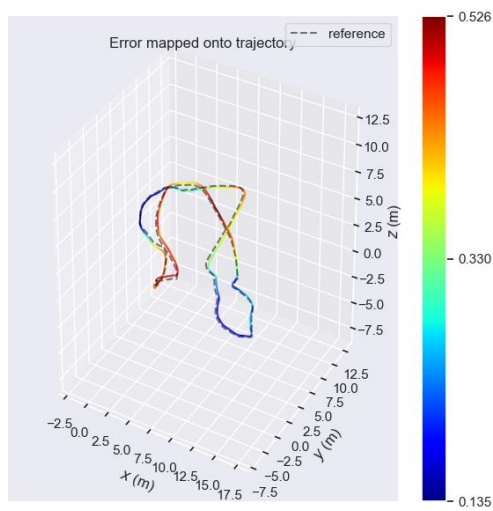
(b2) 本文轨迹对比图



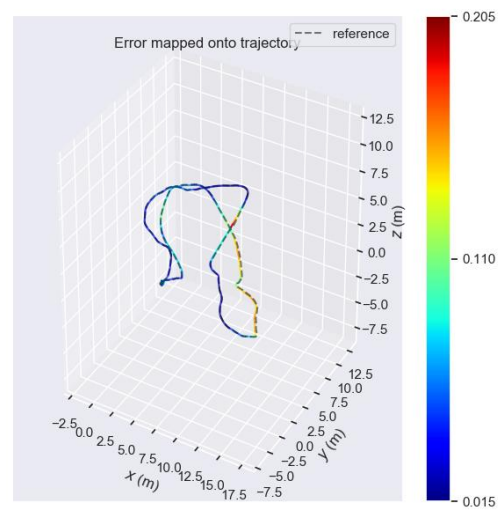
(a3) VINs-mono 轨迹对比图



(b3) 本文轨迹对比图

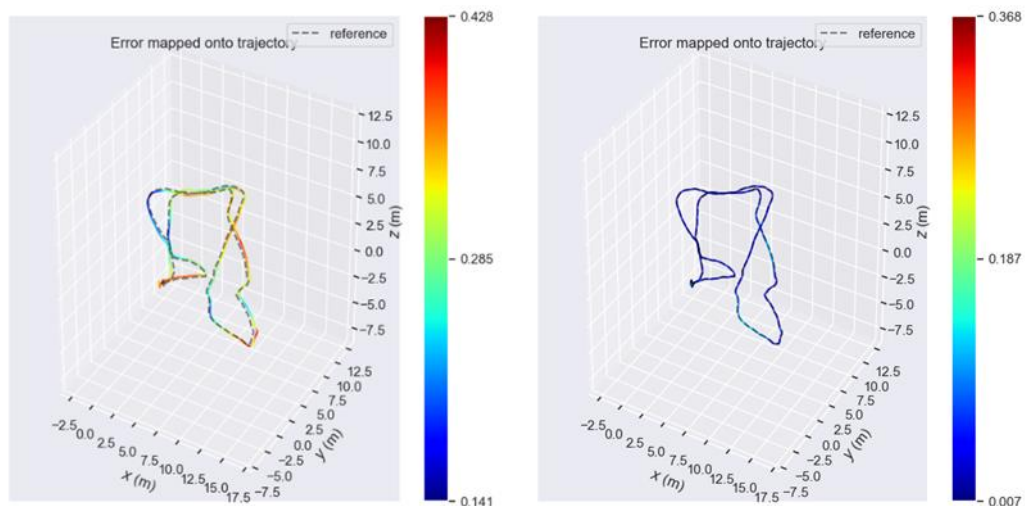


(a4) VINs-mono 轨迹对比图



(b4) 本文轨迹对比图





(a5) VINS-mono 轨迹对比图 (b5) 本文轨迹对比图

图 4-7 本文实验结果与 VINS 实验结果对比图

由图 4-7 可知，本文实验结果较 VINS-mono 算法具有更高的鲁棒性，VINS-mono 的平均误差在 27cm 左右，而本文的平均误差为 10cm 左右，本文算法的误差极值也比 VINS-mono 的更小，因此本文的算法在该序列定位精度更高。

选取以上五个序列中的运行难度不同的三个序列对定位误差进行量化分析，主要从均方根误差(Rmse)、误差平方和(std)、误差极值(Max、Min)和误差中值(Median)五个角度进行分析，进一步说明了本文算法具有更好的鲁棒性和定位精度，误差统计特性表如表 4-1 所示。

表 4-1 误差量化指标对比表

序列	MH01_easy		MH03_medium		MH05_difficult	
	VINS-mono	本文算法	VINS-mono	本文算法	VINS-mono	本文算法
Rmse	0.2014	<b>0.0384</b>	0.2310	<b>0.0386</b>	0.3153	<b>0.0505</b>
Std	0.0766	<b>0.0201</b>	0.1052	<b>0.0191</b>	0.0557	<b>0.0271</b>
Max	0.3822	<b>0.0906</b>	0.5051	<b>0.1119</b>	0.4276	<b>0.3679</b>
Median	0.1635	<b>0.0249</b>	0.2016	<b>0.0292</b>	0.3179	<b>0.0379</b>
Min	0.0072	<b>0.0014</b>	0.0198	<b>0.0021</b>	0.1414	<b>0.0067</b>
mean	0.1862	<b>0.0327</b>	0.2057	<b>0.0335</b>	0.3104	<b>0.0271</b>

从表 4-1 的数据可以看出，与 VINS-mono 相比，3 个序列的误差均值平均降低 82%，误差中值平均降低 84%，误差极大值极小值的表现也优于 VINS-mono。对绝对轨迹误差的均方根进行更深一步实验，实验结果如表 4-2 所示。

表 4-2 绝对轨迹误差 RMSE 对比

序列	VINS-mono(m)	本文算法(m)
MH01_easy	0.2014	<b>0.0384</b>
MH02_easy	0.1876	<b>0.0466</b>
MH03_medium	0.2310	<b>0.0386</b>
MH04_difficult	0.3872	<b>0.0761</b>
MH05_difficult	0.3153	<b>0.0505</b>

从表中可以看出，VINS-mono 平均绝对轨迹误差在 26cm 左右，而本文算法平均绝对轨迹误差仅有 5cm 左右，较 VINS-mono 具有较大的提升。

#### 4.6 本章小结

本章节主要研究视觉-惯导的松耦合方法，阐述了视觉与惯导融合定位模型。采用预积分的方式对惯导信息进行处理以解决惯导信息与相机信息频率不匹配问题；采用了基于滑动窗口的视觉与 IMU 的紧耦合优化模型，并利用非线性优化算法实现了对系统的整体优化，满足了无人机平台所需实时性、鲁棒性的需求；通过 RuRoC 无人机数据集进行实验验证，结果表明，本文算法的较 VINS-mono 具有较大的提升，误差均值较 VINS-mono 平均降低 82%，误差中值平均降低 84%，误差极大值极小值的表现也优于 VINS-mono，在平均绝对轨迹误差上更是比 VINS-mono 提升了 80%左右。

## 5 总结与展望

### 5.1 总结

本文以四旋翼无人机为实验平台，开展基于视觉 SLAM 与惯性导航融合的无人机位姿估计算法的研究。旨在使四旋翼无人机在无 GPS 的情况下，通过融合双目相机获取的位姿信息和惯性测量单元信息，为四旋翼无人机的自主飞行控制提供稳定可靠的姿态、位置信息。主要研究内容如下：

研究了视觉 SLAM 相关基础理论。首先介绍了视觉 SLAM 经典框架及各组成部分，然后介绍了双目视觉模型，同时对所用相机进行了标定实验，得到了准确的相机内外参数，最后介绍是视觉 SLAM 前端常用的特征检测方法。

针对视觉 SLAM 前端 ORB 算法存在图像分布不均匀、匹配程度不高、匹配精度差的问题，本文通过划分网格计算图像灰度值的方法进行角点提取。通过在金字塔层上构建四叉树的方法，在不同金字塔层分别构建不同深度的四叉树以提高计算效率，最后融合 BRIEF-LBP 特征描述子以提升 ORB 算法匹配精度。融合 BRIEF-LBP 特征描述子以提升 ORB 算法匹配精度。最后对视觉得到的信息进行位姿估计实验，实验结果表明，本文算法在特征点均匀度和匹配精度效果有较大提升，在无人机进行稳定飞行的情况下具有很好的定位效果。

此外针对视觉定位在快速运动时出现模糊、重影，定位精度变差甚至丢失的问题，本文融合了 IMU 测量得到的运动数据和基于视觉信息估计的运动数据，采用基于优化模型的视觉与 IMU 组合定位框架模型。针对无人机平台的实时性、鲁棒性要求，本文采用了基于滑动窗口的双目视觉与 IMU 紧耦合优化模型，并用非线性优化的方法对误差进行优化；最后利用无人机 EuRoC 数据集进行实验，结果表明，本文算法的较 VINS-mono 具有较大的提升，从最大最小轨迹误差至平均轨迹误差上本文算法都要优于 VINS-mono，在平均绝对轨迹误差上更是比 VINS-mono 提升了 80%左右。

### 5.2 展望

本文围绕视觉 SLAM 前端特征检测部分和视觉惯导融合部分两方面进行了深入研究，其中所提出的改进 ORB 算法和视觉惯导融合算法实现了一定的研究目的，但由于作者的技术水平、能力有限，研究工作仍有诸多缺陷有待于进一步的完善，主要有：

(1) 本文实现的视觉 SLAM 前端特征提取和匹配方法采集的主要是图片的点特征，在处理单一纹理图像或缺失纹理图像时，容易失去点特征，造成不匹配。而线形

特征则是根据图形中的构造信息来实现的,对单个或丢失的纹理图像具有很好的代表性,能够很好地解决点状特征的缺陷。因此,在未来的几年里,将会致力于在多个复杂的环境中,利用点与线结合的特征抽取与匹配来进一步增强其稳定性与鲁棒性。

(2) 随着人工智能的兴起,视觉 SLAM 算法涌现出深度学习等智能算法,需要对数据进行大量训练,对计算机性能具有较高的要求,无法满足视觉 SLAM 的实时性需求。然而,随着计算机硬件和深度学习的快速发展,预计深度学习方法的结合将成为未来的主流,因为它有望提高计算能力,解决神经网络训练耗时长等问题。本论文将会继续运用深度学习来进行图象处理,并以此为基础,对基于深度学习的 SLAM 实时视觉系统进行研究。

## 参 考 文 献

- [1] 全权.多旋翼飞行器设计与控制[M].电子工业出版社,2018.
- [2] 赵爽. 四旋翼无人机姿态解算与自抗扰控制研究[D].常州大学,2021.DOI: 10.27739/d.cnki.gjsgy. 2021.000116.
- [3] 张石磊. 固定翼无人机的设计运用研究[J]. 数码设计（上）,2020,9(11):39.
- [4] 刘富春,高焕丽. 小型无人直升机的模型预测控制算法研究[J]. 控制理论与应用,2018,35(10):1538-1545. DOI:10.7641/CTA.2018.80088.
- [5] 代君,管宇峰,任淑红. 多旋翼无人机研究现状与发展趋势探讨[J]. 赤峰学院学报（自然科学版）,2016,32(16):22-24. DOI:10.3969/j.issn.1673-260X.2016.16.008.
- [6] 陈浩. 面向无人机的双目视觉与惯导融合的位姿估计方法[D]. 江苏:东南大学,2019.
- [7] Mellinger D, Shomin M, Michael N. Cooperative Grasping and Transport Using Multiple Quadrotors[M] Distributed Autonomous Robotic Systems. 2013.
- [8] 范彦铭.无人机的自主与智能控制[J].中国科学:技术科学, 2017,(03):5-13.
- [9] 周彦,旷鸿章,牟金震,王冬丽,刘宗明.面向半稠密三维重建的改进单目 ORB-SLAM[J/OL].计算机工程与应用:1-7[2020-10-12].
- [10] 林利蒙,王梅.改进点线特征的双目视觉 SLAM 算法[J].计算机测量与控制,2019,27(09):156-162.
- [11] 马树军,王英蕾,金铁铮,白昕晖.一种融合单目信息的 RGB-D SLAM 优化方法[J]. 东北大学学报(自然科学版),2020,41(06):841-845+851.
- [12] 陈含欣.无人机视觉特征提取及同步定位与地图构建方法研究[D].南京航空航天大学, 2016.
- [13] 王亭亭,蔡志浩,王英勋.无人机室内视觉 / 惯导组合导航方法 [J].北京航空航天大学学报,2018,13(2):34-38.
- [14] 徐晓苏,代维.室内环境下基于图优化的视觉惯性 SLAM 方法[J].中国惯性技术学报,2017,25(03):313-319.
- [15] Klein G., Murray D. W.. Parallel Tracking and Mapping for Small AR Workspaces[C]. International Symposium on Mixed and Augmented Reality, 2007: 1-10.
- [16] 张毅,郑潇峰,罗元,庞冬雪.基于高斯分布重采样的 Rao-Blackwellized 粒子滤波 SLAM 算法[J].控制与决策,2016,31(12):2299-2304.

- [17] Strasdat H, Montiel J M M, Davison A J. Visual SLAM: Why filter? [J]. Image & Vision Computing, 2012, 30(2):65-77.
- [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, 2007, 29(6):1052-1067.
- [19] Engel J, Sturm J, Cremers D. Camera-based navigation of a low-cost quadrocopter[C]// 2012:2815-2821.
- [20] 熊敏君. 基于视觉和惯导信息融合的小型无人机位姿估计研究[D]. 国防科技大学, 2017.
- [21] How J P, Bethke B, Frank A, et al. Real-time indoor autonomous vehicle test environment: a testbed for the rapid prototyping of unmanned vehicle technologies[J]. Control Systems IEEE, 2008, 28(2):51-64.
- [22] Sharp C S, Shakernia O, Sastry S S. A vision system for landing an unmanned aerial vehicle[C]//Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on. IEEE, 2001, 2: 1720-1727.
- [23] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// IEEE International Conference on Computer Vision. IEEE, 2012:2564-2571.
- [24] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]// IEEE and ACM International Symposium on Mixed and Augmented Reality. 2007:1-10.
- [25] Mur-Artal R, Tardos J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras[J]. IEEE Transactions on Robotics, 2016, PP(99):1-8.
- [26] Engel J, Koltun V, Cremers D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-17.
- [27] 张礼廉,屈豪,毛军,胡小平.视觉/惯性组合导航技术发展综述[J].导航定位与授时, 2020, 7(04):50-63. DOI:10.19306/j.cnki.2095-8110.2020.04.007.
- [28] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast Semi-Direct Monocular Visual Odometry[C]// IEEE International Conference on Robotics and Automation. 2014:15-22.

- [29] Carrillo L R, Lopez A E, Lozano R, et al. Combining Stereo Vision and Inertial Navigation System for a Quad-Rotor UAV[J]. Journal of Intelligent and Robotic Systems, 2012, 65(1): 373-387.
- [30] Veth M, Raquet J F. Fusing Low-Cost Image and Inertial Sensors for Passive Navigation[J]. Annual of Navigation, 2007, 54(1): 11-20.
- [31] Mourikis A I, Roumeliotis S I. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation[C]. ICRA, 2007: 3565-3572.
- [32] Castellanos J A, Martinez-Cantin R, Tardós J D, et al. Robocentric Map Joining: Improving the Consistency of EKF-SLAM[J]. Robotics and Autonomous Systems, 2007, 55(1): 21-29.
- [33] Bloesch M, Omari S, Hutter M, et al. Robust Visual Inertial Odometry Using a Direct EKF-based Approach[C]. Intelligent Robots and Systems, 2015:298-304.
- [34] Indelman V, Gurfil P, Rivlin E, et al. Real-Time Vision-Aided Localization and Navigation Based on Three-View Geometry[J]. IEEE Transactions on Aerospace and Electronic Systems, 2012, 48(3):2239-2259.
- [35] Hu J-S, Chen M-Y. A Sliding-window Visual-IMU Odemeter Based on Tri-focal Tensor Geometry[C]. ICRA, 2014.
- [36] Lupton T, Sukkarieh S. Visual-Inertial-Aided Navigation for High-Dynamic Motion in Built Environments Without Initial Conditions[J]. IEEE Transactions on Robotics, 2012, 28(1): 61-76.
- [37] Stefan Leutenegger, et al., Keyframe-Based Visual-Inertial SLAM Using Nonlinear Optimization[J]. International Journal of Robotics Research, 2015,34(3):314-334.
- [38] Strasdat H, et al. Real-time Monocular SLAM: Why Filter?[C]. ICRA, 2010.
- [39] Mur-Artal R, et al. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J], IEEE Transactions on Robotics, 2015, 31(5):1147-1163.
- [40] Tong, Qin, Peiliang, et al. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2018.
- [41] Zheng F, Tsai G, Zhang Z, et al. Trifo-VIO: Robust and Efficient Stereo Visual Inertial Odometry using Points and Lines[C] European Conference on Computer Vision(ECCV). IEEE Computer Society, 2018:478-482.
- [42] Li M, Mourikis A I. High-precision, Consistent EKF-based Visual-inertial Odometry[J]. The International Journal of Robotics Research, 2013, 32(6): 690-711.

- [43] Huang G, Mourikis A I, and Roumeliotis S I. A Quadratic-complexity Observability-constrained Unscented Kalman filter for SLAM[J]. IEEE Transactions on Robotics, 2013, 29(5):1226–1243.
- [44] HM Liu, MY Chen, GF Zhang, HJ Bao. ICE-BA: Incremental, Consistent and Efficient Bundle Adjustment for Visual-Inertial SLAM[C] IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2018:3234-3243.
- [45] ZC Zhang, Scaramuzza D. A Tutorial on Quantitative Trajectory Evaluation for Visual-Inertial Odometry[C] International Conference on Intelligent Robots and Systems(IROS). IEEE Computer Society, 2018:1156-1162.
- [46] 毛家宁,赵龙.基于滑窗非线性优化的双目视觉 SLAM 算法[J].应用科技,2020,47(01):55-60.
- [47] 万薇.基于双目视觉与惯导信息融合的 SLAM 系统实现[D].浙江:浙江工业大学,2019.
- [48] 孙楠,裴信彪,王春军,李继辉,彭程,白越.基于立体视觉——惯导 SLAM 的四旋翼无人机导航算法[J].微电子学与计算机,2020,37(05):33-38+42.
- [49] Huai Z , Huang G . Robocentric Visual-Inertial Odometry[C]2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018:2671-2680.
- [50] 卢陶然. 面向无人机的视觉-惯性里程计算法研究[D].电子科技大学,2020.
- [51] 董永峰,雷晓辉,董瑶等. 改进 ORB 特征提取与匹配算法研究[J]. 传感器与微系统,2020,39(4):59-62.
- [52] Shou L, Zhiqiang W,et al. A Research of ORB Feature Matching Algorithm Based on Fusion Descriptor[C]. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), 2020:417-420.
- [53] 朱海峰,赵春晖. 图像特征点分布均匀性的评价方法[J]. 大庆师范学院学报,2010,30(3):9-12.



## 攻读学位期间的研究成果

发表论文:

陈易文, 储开斌, 张继, 冯成涛. 基于四叉树的 ORB-LBP 改进算法[J]. 传感器与微系统, 已录用

## 致 谢

时节如流，岁月不居，犹记初入校园时激动与兴奋的面庞，不知不觉都到了与常大说分别的时刻了。告别生活了近3年的大学，即将步入社会的我，要和那些熟悉的老师同学们说再见的时候了。此时的我心情复杂，既有对过去的不舍，又有对未来的期待，但最多的还是感激。

首先感谢我的导师储开斌老师，感谢他从未嫌弃我的愚笨，永远都在用耐心和包容对待我，正是因为有他严厉的教导，亲切的关怀，耐心的指导，我才能取得如今的成绩。在研一时根据我的兴趣和基础，给我确定研究方向；在研二指导我做实验，帮助我开题和撰写小论文；在研三指导我撰写毕业论文，关心我的就业。无论是学业水平，还是专业素质，储老师都是行业内的佼佼者，能够得到储老师的教诲令我倍感荣幸。新的未来或许并不容易，但我会永远铭记您的叮嘱，我将努力前行，勇攀高峰。

感谢张继老师，冯成涛老师等在我在校期间有着巨大帮助的老师，向你们道一声：老师，辛苦了，谢谢你们！

感谢我们团队实验室的各位师弟师妹们，感谢你们给我营造了一个的学习氛围，开组会时一起讨论学术问题，一起交流，让我成长了许多。谢谢！

感谢我的室友们，能和你们朝夕相处真是幸运，你们让我的生活习惯更加健康，自身的生活方式更加精致。谢谢！

感谢我的父母，感谢你们能够全力的支持我，无论我做什么决定都能够支持鼓励我，让我拥有坚实的后盾。

最要感谢的莫过于我的母校，“责任”将永记我心，无论未来在何方，心中永远有我的母校。

谢谢！