

硕士学位论文

融合双目视觉和惯性的定位技术研究

**RESEARCH ON POSITIONING TECHNOLOGY
COMBINING BINOCULAR VISION AND
INERTIAL MEASUREMENT UNIT**

管一凡

哈尔滨工业大学

2020 年 6 月

国内图书分类号：TP242.6
国际图书分类号：681.513

学校代码：10213
密级：公开

工程硕士学位论文

融合双目视觉和惯性的定位技术研究

硕士研究生：管一凡

导师：夏红伟教授

申请学位：工程硕士

学科：控制工程

所在单位：航天学院

答辩日期：2020年6月

授予学位单位：哈尔滨工业大学

Classified Index: TP242.6

U.D.C: 681.513

Dissertation for the Master Degree in Engineering

**RESEARCH ON POSITIONING TECHNOLOGY
COMBINING BINOCULAR VISION AND
INERTIAL MEASUREMENT UNIT**

Candidate :	Guan Yifan
Supervisor :	Professors. Xia Hongwei
Academic Degree Applied for :	Master of Engineering
Speciality :	Control Engineering
Affiliation :	School of Astronautics
Date of Defence :	June, 2020
Degree-Conferring-Institution :	Harbin Institute of Technology

摘 要

随着无人驾驶、人工智能和移动机器人等技术的不断发展,视觉/惯性组合导航定位技术逐步成为这些研究领域不可或缺的研究方向之一,受到了科学人员的广泛关注。视觉/惯性组合导航定位系统将低成本、小型化的惯性测量单元(IMU)和摄像机组合在一起,既可以利用相机提供的丰富场景信息来修正IMU的累积漂移,又能够利用IMU输出的载体自身运动信息为双目视觉系统提供更为精确的初始位姿信息,并通过二者信息融合解算出载体在惯性空间中位置和姿态,该系统充分利用了视觉和惯性的性能互补关系进行载体的导航和定位。

本文通过双目视觉相机和MEMS惯性测量单元组成的导航定位系统进行载体的运动状态估计,采用导航定位精度更高的非线性优化紧耦合方案进行多传感器信息融合的位姿解算。首先,对双目视觉导航系统中涉及到的坐标系进行定义并明确了相机坐标系到世界坐标系的变换关系,分析了摄像机的张氏标定方法,对比分析了SIFT、Harris、ORB和SURF四种主流的图像特征点检测方法,进而验证了其中的ORB特征点检测方法的最优性,并基于RANSAC方法给出了特征点误匹配的去重算法,采用ICP方法估计载体的运动状态,通过特征提取匹配并对图像特征进行跟踪并给出了基于ICP/SFM融合的重构算法;其次,给出了MEMS惯性测量单元坐标系定义及坐标系间的转换方法,建立了IMU的测量模型,通过对IMU数据进行预积分处理的基础上,将视觉和惯性求得的位姿信息进行松耦合算法处理,通过求解线性最小二乘方程获得位姿优化所需要的精确系统初值,包括位姿初始化、陀螺仪的零偏,重力加速度的方向以及载体的运动状态;再次,利用IMU测量误差和视觉重投影误差构造非线性优化的代价函数,基于滑动窗口边缘化的方法来保证视觉导航系统常量的计算消耗和实时性,同时,将边缘化的先验信息加入到代价函数中,避免图像信息的丢失,另外,还引入了基于词袋模型的回环检测方法减少视觉导航系统的累积误差,增加四自由度位姿图优化功能保证组合导航系统运动轨迹的全局一致性,并针对载体纯旋转退化问题给出了一种基于扩展卡尔曼滤波的解决方法;最后,采用Euroc数据集进行了系统仿真验证。

关键词: 双目视觉; 组合导航; 视觉惯性里程计; 回环检测

Abstract

With the continuous development of technologies such as unmanned driving, artificial intelligence and mobile robots, visual inertial integrated navigation and positioning technology has gradually become one of the indispensable research directions in these research fields, and has received widespread attention from scientific researchers. The visual inertial integrated navigation and positioning system combines a low-cost, miniaturized inertial measurement unit (IMU) and a camera, which can not only use the rich scene information provided by the camera to correct the cumulative drift of the IMU, but also use the carrier output by the IMU itself. The motion information provides more accurate initial pose information for the binocular vision system, and the carrier's position and attitude in the inertial space are solved by the fusion of the two information. The system makes full use of the complementary relationship between vision and inertia for the navigation of the carrier And positioning.

In this paper, the motion state of the carrier is estimated by the navigation and positioning system composed of binocular vision camera and MEMS inertial measurement unit, and the nonlinear optimization and tight coupling scheme with higher navigation and positioning accuracy is used to solve the position and pose of multi-sensor information fusion. First, define the coordinate system involved in the binocular visual navigation system and clarify the transformation relationship from the camera coordinate system to the world coordinate system, analyze the camera's Zhang's calibration method, and comparatively analyze the four SIFT, Harris, ORB and SURF A mainstream image feature point detection method, which further verifies the optimality of the ORB feature point detection method, and based on the RANSAC method, a feature point removal algorithm for mismatching is given. The ICP method is used to estimate the motion state of the carrier. Extract the matching and track the image features and give a reconstruction algorithm based on ICP/SFM fusion; secondly, the coordinate system definition of MEMS inertial measurement unit and the conversion method between coordinate systems are given, and the measurement model of IMU is established. On the basis of pre-integration processing of IMU data, the posture information obtained by vision and inertia is processed by a loose coupling algorithm, and the accurate initial system value required for posture optimization is obtained by solving the linear least squares equation, including posture initialization, The gyro's bias, the direction of gravity acceleration and the motion

state of the carrier; again, using the IMU measurement error and visual reprojection error to construct a nonlinear optimized cost function, based on the sliding window marginalization method to ensure that the visual navigation system constant Calculate consumption and real-time, at the same time, add marginal prior information to the cost function to avoid the loss of image information. In addition, a loop detection method based on the bag of words model is introduced to reduce the cumulative error of the visual navigation system and increase The four-degree-of-freedom position and pose optimization function ensures the global consistency of the integrated navigation system's motion trajectory, and provides a solution based on extended Kalman filtering for the problem of pure rotation degradation of the carrier; finally, the Euroc dataset was used for system simulation verification.

Keywords: binocular vision, integrated navigation, visual inertial odometer, loop closure

目 录

摘 要	I
ABSTRACT	II
第 1 章 绪 论	1
1.1 课题背景与研究意义	1
1.2 国内外研究现状	2
1.2.1 微惯性导航技术发展概述	2
1.2.2 视觉里程计发展概述	3
1.2.3 视觉 SLAM 发展概述	4
1.2.4 视觉与 IMU 融合技术发展概述	6
1.3 论文主要研究内容	7
第 2 章 双目视觉导航定位技术	8
2.1 引言	8
2.2 双目视觉定位	8
2.2.1 摄像机标定	8
2.2.2 图像特征提取	10
2.2.3 图像特征匹配	16
2.2.4 运动估计	19
2.3 本章小结	23
第 3 章 双目视觉与 IMU 组合导航定位系统	24
3.1 引言	24
3.2 系统结构	24
3.3 IMU 运动估计	25
3.3.1 导航坐标系	25
3.3.2 坐标系变换关系	26
3.3.3 IMU 测量模型	28
3.4 IMU 预积分	29
3.5 视觉惯性的松耦合	31
3.5.1 陀螺仪零偏的修正	33
3.5.2 系统初始化	34
3.6 本章小结	36
第 4 章 双目视觉与 IMU 的紧耦合优化方法	37

目 录

4.1 引言	37
4.2 视觉-惯性紧融合算法	37
4.2.1 IMU 误差分析	39
4.2.2 滑动窗口边缘化	40
4.2.3 载体位姿优化	41
4.3 回环检测	42
4.3.1 视觉位姿优化	43
4.3.2 纯旋转运动的处理	46
4.4 数据集仿真测试	47
4.5 纯旋转运动姿态估计仿真结果	51
4.6 本章小结	52
结 论	54
参考文献	55
哈尔滨工业大学学位论文原创性声明和使用权限	59
致 谢	60

第1章 绪 论

1.1 课题背景与研究意义

人类很早就对定位技术的重要性有了深刻的认识和理解，随着科技的进步，其技术的发展也突飞猛进，逐步成为在航空、航海、军事、医疗等领域不可或缺的一部分。其中，智能无人设备得到了越来越广泛的应用，世界各国很多科研人员都对其产生了浓厚的兴趣，无人设备想要完成它的任务有一个必要的前提工作，那就是确定自身所处的位置，所以运动过程中确定自身的位置和姿态成了自主导航的关键技术，成为了无人设备领域中的热点研究问题。

实时准确地确定自身的位置和姿态，是无人设备的工作中最重要的一环。目前最常用的定位方式分为两类，第一类是绝对定位方法，绝对定位是指不依靠当前时刻之前的运动过程，根据已知的全局参考信息，依靠传感器获得的数据来确定当前自身位姿的定位方法，包括 GPS、北斗等。以 GPS 为例，它通过测算已知位置的卫星与用户接收器的距离，结合多个卫星的数据解算出用户接收器的位置。第二类是相对定位方式，相对定位的位置是通过累计运算得到相对初始位置信息，该技术通常借助视觉、轮式里程计、MEMS 等传感器进行定位。以视觉定位为例，摄像机根据采集到的图像序列来确定场景与载体的对应位置关系，为载体接下来的路径规划提供有效依据。绝对定位不存在累计误差，但因数据的获取需要依靠传感器，其引入的干扰问题不容忽视，相对定位方式存在累计误差，对长航时保障高精度的要求并不适用。

上述定位方式都存在着优缺点，所以实际的导航系统多是将两种或两种以上的定位方式组合，来提高定位精度。融合视觉与惯性测量单元的定位方法常被采用，该方法中两种传感器的优缺点都很明显。在视觉定位方法中摄像机获取图像后进行特征提取、跟踪，得到估计信息，实现视觉里程计。该方法定位精度较高，而且相机的价格成本低，缺点是容易受到相机特性和环境因素的影响，比如摄像机的抖动、光照变化等因素，抗干扰能力差。另外图像处理过程中，实时性也是一个关键因素。惯性传感器定位技术可以弥补其不足，具有较好的实时性和动态特性，抗干扰能力也较强。但缺点是通过陀螺仪和加速度计解算出的姿态和估计出的位置信息，有较大的误差。

基于前面提到的视觉定位和惯性测量单元的优缺点，本文选取双目视觉系统与微惯性测量单元组合来实现精度高、成本低的组合导航系统。首先在微惯性测量

单元这一方面,融合双目视觉里程计后可以很好的对微惯性测量单元的漂移现象进行有效的修正,反过来在惯性定位的应用下弥补了双目视觉定位中高速运动动态特性不好的缺点。将二者结合可以有效提高导航系统的定位精度,具有体积小,能耗低,成本低廉等优点,非常适用于无人小车、移动机器人、无人机等。

1.2 国内外研究现状

近十几年来,随着计算机视觉的崛起,视觉 SLAM 技术已经深入到了组合导航的实际工程应用领域。美国的火星探测器就是通过视觉 SLAM 技术进行导航的。该技术一个重要的发展方向就是将视觉与传感器相融合,视觉里程计的开发就是该领域的重要突破,对提高定位系统精度和鲁棒性更优起到了很大的作用。

1.2.1 微惯性导航技术发展概述

惯性导航是一种相对定位的导航技术,在已知系统初始信息的前提下,通过惯性定律来求得运载体的运动信息。惯性导航主要由加速度计和陀螺仪两种传感器组成,二者分别用来对运载体具有的加速度和角速度进行测量。惯性导航系统具有极高的自主性、隐蔽性、抗干扰能力强和信息连续等特点。分为旋转式、平台式和捷联式惯性导航系统。旋转式惯导系统利用旋转调制抑制漂移误差,平台式惯导系统顾名思义是把陀螺仪和加速度计固定在平台上,与之相反的是捷联式惯导系统是把陀螺仪和加速度计固定在载体上。近年来,三种方式混合的惯导系统(HINS)被提出,具有很大的发展空间,在未来在高精度导航中的高速运动载体上有广阔的应用前景。

陀螺仪是惯性导航系统的重要器件,世界上第一台陀螺仪在 1908 年被德国的安许茨所创造,当时称作陀螺罗经,凭借着这个新发明,航海导航技术有了质的飞跃,随着陀螺仪生产技术的发展^[1],现如今,陀螺精度有了明显的提高^{[2][3]}。静电陀螺是精度最高的陀螺,但是它价格贵、制作流程复杂,多用于航天、航母等领域。世界上静电陀螺研究水平最高的两个公司是日本的 Tokimec 公司和美国的 Holleywell 公司,其中 Tokimec 公司生产的环形转子静电悬浮陀螺仪^[4]代表着静电陀螺的较高水平,其转速可达到 74000r/min,稳定性有 0.01°/s。

20 世纪 90 年代,随着微电子技术和精密微型机械技术的发展,微电子机械系统(MEMS)诞生,因其有体积小、性能稳定等优点 MEMS 在无人车、无人机等微型应用领域中有很大的比重。国外对于 MEMS 的技术研究比较早,技术水平也优于国内一部分。我国对 MEMS 技术的研究始于 90 年代初期,主要的研究单位有国防科技大学、清华大学等研究院所。目前国际上精度较高的 MEMS 惯性产品主

要由美国的 ADI 公司研制，它的 MEMS 惯性单元 ADIS16497 如图 1-1 所示，具有较高的稳定度。此外还有瑞士 Sensoror 的公司和 CrossBow 公司，二者也分别享有 STIM 系列产品与 NAV 系列产品，Sensoror 公司的惯性单元 STIM300 如图 1-2 所示，它的陀螺仪偏置稳定度可达到 $0.5^{\circ}/h$ ，加速度运动中的偏执稳定度有 $0.05mg$ 。由此可见高精度的 MEMS 惯性单元已经可以满足民用和部分军事的性能要求。



图 1-1 ADIS16497



图 1-2 STIM300

1.2.2 视觉里程计发展概述

视觉里程计（Visual Odometry）初期是由美国 NASA 火星探索项目推动的，当时是用来检测火星探测车的六自由度。在 19 世纪 80 年代，斯坦福大学博士生 Moravec 在研究中对仅依靠视觉输入进行车辆运动估计这一问题进行了研究。他提出的首个视觉定位框架至今仍被使用^[5]，实现了图像特征检测、匹配与追踪等功能。他所提出的 Moravec 角点检测算子为后来的研究者提出 Forstner 算子和 Harris 算子作出了极大的贡献。Moravec 在行星探测车上装备了一个滑块立体相机，该相机可以在滑道上滑动，相机等距离水平移动九个位置分别拍摄照片。正是这一举措，为今后的视觉里程计的发展打下了坚实的基础。

基于 Moravec 的研究基础，Matthies 和 Shafer 选择使用双目相机和沿用 Moravec 的角点检测追踪法^{[6][7]}，但与 Moravec 用数量表示不确定性的方法不一样的是，他们在运动估计时引入三角化特征点的误差协方差矩阵，最后整合到运动估计的步骤中去，优化了行星探测车轨迹恢复的视觉效果。在 5.5m 的路径中相对误差为 2%。2003 年，Olson^{[8][9]}等人引入了一种绝对方位传感器，该使用 Forstner 角点检测，最后发现运算速度比 Moravec 算子快很多。最终实验表明，引入绝对方位传感器后，误差会变为线性增长，这大大优于仅使用相机成像导致的误差累计越来越大。2004 年 Nister 发表了一篇重要论文，在视觉里程计发展史上有里程碑的意义，此举确立了术语 Visual Odometry，Nister 采取了不跨帧追踪特征点，而是在每一帧中独立检测特征点，并只在这些特征点中进行跨帧匹配，这样降低了漂移误差，同时又采取 3D-2D 的 PnP 求解方法来求解位置估计问题，着意味着视觉里程计会

更多的应用于地面车辆的导航定位。2011, Andreas Geiger、Philip Lenz 和 Raquel Urtasun 等人创立了 KITTI 标准视觉测试平台, 提供了常用传感器的各种实验数据^[10]。并公开了这些数据集, 该领域的学者们开动脑筋, 力争设计出最优秀的精确定位算法^{[11][12]}, 最终网站会给出排名算法的排名, 这也大大推广了定位算法的发展。2013 年中国的“玉兔号”月球车等也将视觉里程计应用于系统定位, 完成了探测车在月球表面的导航任务。

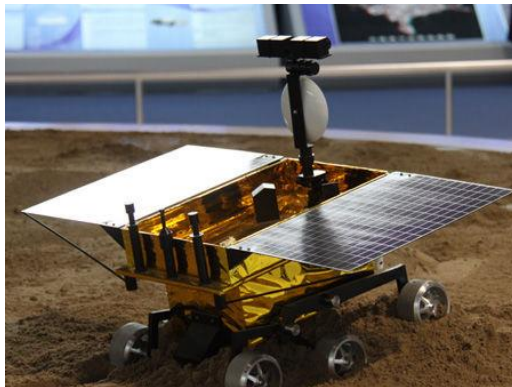


图 1-3 中国“玉兔号”

基于视觉里程计的定位研究虽晚于国外, 但近几年也有着不错的发展。浙江大学在视觉里程计领域的研究最早、最权威, 其对月球车的视觉定位研究取得了丰硕成果。彭勃^[13]通过 Harris 角点检测算法为基础, 为了提高系统实时性引入了线性与非线性最小二乘法。李永佳^[14]通过改进的 Harris 特征点提取与匹配法, 在特征跟踪时加入了惯导数据, 提高了跟踪结果的稳定性, 进而, 他还提出融合 Harris 角点和 SIFT 特征检测来提高旋转、尺度不变性。最终实验表明在 41.25 米的闭环路径, 定位误差为 0.45%。但由于相机是通过离线数据再进行运动估计, 所以并未实现实时定位。2015 年, 卢维^[15]采用局部双目光束平差算法, 硬件系统上采用图形处理单元为处理器, 处理频率达到 40Hz。通过实验算法定位精度可达到 0.39%。从所查文献可发现, 无论是高校或科研所大多选择光照条件好的水平地面进行较短的距离实验, 数据集多采用的是 KITTI 公开数据集^{[16][17][18]}。江燕华在城市环境下通过引入车辆运动约束特性进行了 5 千米实验, 实验结果表明城市环境下引入车辆运动约束的视觉里程计可以满足实时导航定位的要求。

1.2.3 视觉 SLAM 发展概述

在早期, 研究者们都选择通过滤波的手段来对视觉 SLAM 位姿估计问题进行处理。在 2007 年, Davison 提出了一种基于 EKF 方法的 SLAM 系统, 叫作 MonoSLAM^[19], 这个方法通过把当前相机位姿和全部的地图点作为状态量, 通过

更新迭代不断修正它的误差量。提出通过恒速模型对相机的下一帧位姿进行预测以及通过协方差矩阵对接下来的搜索进行预测。MonoSLAM 满足构建小场景特征点地图的实时性,但特征点稠密会对系统的实时性产生较大的影响,出现追踪丢失的问题。

M. Montemerlo 等人通过粒子滤波器 (Particle Filter, PF)代替 EKF, 提出了 FastSLAM^{[20][21]}。这个系统的时间复杂度与状态数量和粒子数量有关,是线性的。而且可以用一定的数据结构对时间复杂度进行优化。整个系统通过迭代不断计算状态量的均值和方差进行优化。

近几年,基于非线性优化的 SLAM 在实际应用中越来越多,基于滤波器的 SLAM 逐渐被取代。当前主流 SLAM 系统都分为前端和后端^[22]。其中前端相当于视觉里程计,负责特征提取、特征匹配与筛选以及相机的运动估计等。即通过图像信息作为输入,通过追踪图像特征点进而对相机轨迹位姿估计。后端进行回环检测和全局优化来消除之前产生的累计误差。

Klein 等人于 2007 年提出一种基于关键帧 SLAM 系统。PTAM (Parallel Tracking and Mapping)^{[23][24]}视觉追踪和地图构建分在两个线程进行计算,也叫作并行处理 SLAM 系统。图像数据能够实时响应,与此同时另外一个线程中优化过程独立进行。在 PTAM 系统被开发之前,没有人意识到非线性优化稀疏性问题,就是指一个位置只能观察到一部分路标点,同时一个路标点也只能在某些位置被观测到,从此之后视觉 SLAM 便开始采取非线性优化来对后端数据进行实时优化。由于 PTAM 的引领作用,后续的大量的视觉 SLAM 系统都是它的扩展。但是 PTAM 缺少回环检测环节,所以系统的累计误差会逐渐增大。

J.Engel^[25]等提出了 LSD-SLAM 系统,它是通过单目直接法来进行大型半稠密建图系统的建立。该系统构建的半稠密地图是无需 GPU 加速的条件,并具有很好地实时性。与稀疏地图相比,该系统可以对机器人的应用场景进行扩展。但实际上 LSD-SLAM 还是与特征点计算有关,因为在进行回环检测时还是需要依赖特征点。LSD-SLAM 直接对像素进行操作,所以诸多因素都会对其造成影响,比如载体运动过快。。

Forster^[26]等人提出了一种基于半直接法的视觉里程计 (SVO),SVO 需要检测一些关键点且不包含描述子,并不对每一帧图像的特征点进行提取,同时采取块匹配法对光度误差进行最小化处理以此对相机的运动进行估计。SVO 优点是运行速度快,缺点是存在累计误差。其用于无人机等视觉导航时,主要用于俯仰轴相机上,平动轴相机很少使用。。Forster 在视觉 SLAM 领域还有着其它杰出的贡献,2017 年 Forster 等人^[27]为视觉-惯性融合提出了基于流型空间的 IMU 预积分理论。

Labbe M 等人在 2014 年提出的 RTAB-MAP (Real Time Appearance-Based Mapping) [28] 算法, 它被称作是基于 RGB-D 相机的 SLAM 算法中的经典, 它包含了必须的 SLAM 模块, 同时还能在 CPU 上实时构建点云地图和三角网格地图。

2016 年, 文献[29]提出了定位精度较高的视觉 SLAM 系统 ORB-SLAM, 与之前提到的双线程并行不同的是, 它包括了三个线程并发运行: 视觉跟踪、局部建图和回环检测。ORB-SLAM 系统选择视觉里程计和回环检测中都合适的 ORB 特征, 原因是 ORB 特征在计算时速度极快并具有不错的旋转不变性。ORB-SLAM 是目前最精准的实时视觉 SLAM 系统。它的主要流程是首先视觉跟踪线程用来进行检测 ORB 特征, 选取关键帧后大致估计相机位姿; 局部建图线程通过解算 BA 问题, 来对地图点和相机位姿优化; 回环检测线程则是负责检测回环, 去除累计误差, 便于跟踪丢失之后的找回。

1.2.4 视觉与 IMU 融合技术发展概述

近些年, 将低成本、小体积的 IMU 应用与视觉系统成为研究的热点, 视觉惯性导航系统 (VINS) 可以观测俯仰角、横滚角和载体运动的尺度因子, 可以实现绝对导航任务。该系统按融合方式可以分为松耦合和紧耦合两大类。前者通过独立的视觉处理模块获取尺度不固定 (up-to-scale) 的位姿估计, 应用 EKF 对 IMU 数据进行融合, 得到尺度和位姿的估计。后者可选取图优化或滤波器的方法, 把特征点加入状态向量中从而对相机与 IMU 的原始数据进行联合优化, 使它们的噪声模型都被考虑到。

Faessler^[30]等人基于扩展卡尔曼滤波技术, 进行多传感器的 SVO 位姿估计与 IMU 的测量信息融合, 缺点是初始信息的获取需要利用其它的传感器。这也是松耦合这种融合方式的弊端所在。将

Mourikis 等提出了一种新的 MSCKF 方案^[31], 也是将卡尔曼滤波技术应用于视觉导航领域, 对不同时刻的状态进行了约束, 再通过应用了一个滑窗令计算复杂度变低, 加强了实时性。Bloesch 等人也提出一种基于 EKF 的紧耦合方案, 叫作 ROVIO^[32], ROVIO 应用了视觉惯性里程计, 在融合时基于扩展卡尔曼滤波进行融合, 它的优点是计算量小, 缺点是普遍适应性不好, 总需要对参数进行适当的调节。

Leutenegger 等在关键帧方面提出一种视觉惯性融合的方法 OKVIS^[33], 该方法同样是用非线性优化来对滑窗中关键帧位姿进行估计。OKVIS 后端优化的代价函数由三维路标点重投影误差和 IMU 状态误差项构成, 通过开源的 Ceres 库进行求解。多尺度 Harris 角点用于前端特征检测, BRISK 描述子用于相邻帧匹配关系的

检测。OKVIS 未实现回环检测模块和位姿图优化模块,所以一定会存在误差累积。

沈邵劼等在 2017 年提出了 VINS-Mono^[34],它是一种基于滑窗非线性优化的位姿估计器。该方法可以在未知环境下在线标定参数,同样拥有回环检测和重定位功能,具备完整 SLAM 系统条件。

Paul 等人应用预积分理论,与 VINS-Mono 主体思想一致,对 ORB-SLAM 和 IMU 数据进行融合^[35]。前端处理选择 ORB 特征点,能够估计六自由度的位姿。

2018 年,百度 AR 团队与浙江大学联手,提出了一种名为 ICE-BA^[36]的方法用来解决视觉-惯性融合 BA 问题,该方法能够保证效率更高且全局一致。该方法的主要突破点有两个,第一个是通过 SLAM 测量的增量特性引出了解决 BA 问题的新方法,新方法的效率比之前最快的解决方法高十倍以上;第二个是用边缘化算法处理滑窗边缘化偏差与全局回环检测约束的冲突。

1.3 论文主要研究内容

本文通过双目视觉相机和 MEMS 惯性测量单元组成的导航定位系统进行载体的运动状态估计,采用导航定位精度更高的非线性优化紧耦合方案进行多传感器信息融合的位姿解算。本论文主要研究内容及章节安排如下:

第一章主要介绍了视觉惯性导航系统的研究背景及意义,并简要讨论了微惯性导航、视觉里程计、视觉 SLAM 领域以及视觉-惯性融合领域的经典解决方案与框架,最后概述了本论文的主要研究内容和章节安排。

第二章分析了摄像机的张氏标定方法,对比分析了 SIFT、Harris、ORB 和 SURF 四种主流的图像特征点检测方法,进而验证了其中的 ORB 特征点检测方法的最优性,并基于 RANSAC 方法给出了特征点误匹配的去重算法,通过特征提取匹配并对图像特征进行跟踪并给出了基于 ICP/SFM 融合的重构算法。

第三章给出了 MEMS 惯性测量单元坐标系定义及坐标系间的转换方法,建立了 IMU 的测量模型,通过对 IMU 数据进行预积分处理,将视觉和惯性求得的位姿信息进行松耦合算法处理,通过求解线性最小二乘方程获得位姿优化需要的精确系统初值,包括位姿初始化、陀螺仪的零偏,重力加速度的方向和载体的运动状态。

第四章建立了视觉/惯性组合导航定位系统的多种约束关系因子图,以此确定了系统滑动窗口非线性优化的紧耦合方程,并对方程中的误差项进行了分析。另外,还引入了基于词袋模型的回环检测方法以减少视觉导航系统的累积误差,增加四自由度位姿图优化功能保证组合导航系统运动轨迹的全局一致性,并针对载体纯旋转退化问题给出了一种基于扩展卡尔曼滤波的解决方法;并采用 Euroc 数据集进行了系统仿真验证。

第 2 章 双目视觉导航定位技术

2.1 引言

本课题所研究的视觉定位是通过帧间图像特征匹配进行运动估计的双目视觉里程计。双目视觉里程计是以双目相机在运动过程中采集到的图像作为输入，通过特征检测得到的特征点在运动过程中位置的改变从而计算双目摄像机的平移和旋转。双目视觉里程计解决了双目视觉系统尺度不确定性的问题，通过较低的帧率完成高精度的位置姿态测量，是机器人、无人小车等实现导航定位的重要方法。

2.2 双目视觉定位

在双目视觉定位中，最重要的是图像的特征提取和匹配，准确性以及实时性都是最为关键的，这为了最后的位姿估计起到了关键决定性的作用，载体的平移和旋转都离不开它们。

2.2.1 摄像机标定

目前摄像机标定方法大致有两种：一种是传统标定，另一种是自标定。传统标定是需要一个已知的标定参照物，但在实际应用中，往往很多的摄像机无法保证位置不做改变或者摄像机参数不变。自标定变由此而来，自标定技术不需要使用标定参照物，但是必须满足从运动过程中得到摄像机参数，传统标定与自标定相比，自标定技术更繁琐。本文采取介于传统标定和自标定之间的平面图板标定法完成相关标定工作。

平面图板标定法中应用最普遍的是由张正友提出的张氏标定法。这种方法是以前孔模型作为摄像机模型。平面图板相比寻找对应参照物更为简单，棋盘格定板是使用最为广泛的平面图板，常用如图 2-1 所示。

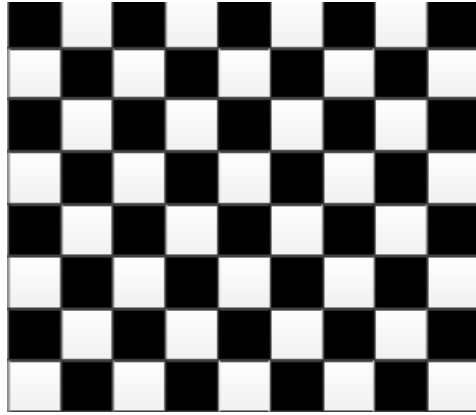


图 2-1 棋盘格标定板

设棋盘格上随机一点的世界坐标为 $P = [X_w \ Y_w \ Z_w \ 1]^T$ ，图像坐标 $q = [u \ v \ 1]^T$ ，由公式(2-1)可得：

$$Z_c q = M_1 \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} P \quad (2-1)$$

空间参考坐标系平面选取棋盘格平面，则 $Z_w = 0$ ，那么 P 可写 $[X_w \ Y_w \ 1]^T$ ，则将式(2-1)进一步简化：

$$\begin{aligned} p &= s M_1 [r_1 \ r_2 \ T] P \\ &= [h_1 \ h_2 \ h_3] P \\ &= HP \end{aligned} \quad (2-2)$$

其中 $s = 1/Z_c$ 叫做比例因子，矩阵 H 被称为单应性矩阵，是由最优化算法解算得到的。旋转矩阵 R 是单位正交矩阵，同时满足 $r_1^T r_2 = 0$ ， $\|r_1\| = \|r_2\| = 1$ ，由式(2-2)中对应关系可得到：

$$\begin{cases} h_1^T M_1^{-T} M_1^{-1} h_2 = 0 \\ h_1^T M_1^{-T} M_1^{-1} h_1 = h_2^T M_1^{-T} M_1^{-1} h_2 \end{cases} \quad (2-3)$$

设 $B = \tilde{M}_1^{-T} \tilde{M}_1^{-1}$ ，矩阵 B 由摄像机内参组成，同时满足对称矩阵关系。将 B 表示成一个六位向量，如(2-4)所示，同时矩阵 H 的第 i 列表示为 $h_i = [h_{1i} \ h_{2i} \ h_{3i}]^T$ ，则(2-3)可表示为式(2-5)

$$B = [B_{11} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}] \quad (2-4)$$

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{12})^T \end{bmatrix} B = 0 \quad (2-5)$$

上式中 $v_{ij} = [h_{1i} h_{1j} \ h_{1i} h_{2j} + h_{2i} h_{1j} \ h_{2i} h_{2j} \ h_{3i} h_{1j} + h_{1i} h_{3j} \ h_{3i} h_{2j} + h_{2i} h_{3j} \ h_{3i} h_{3j}]^T$ 。摄像

机内参数值由拍摄得到的多幅棋盘图像求得。再求取出内参矩阵后,根据(2-2)式所示可得摄像机旋转和平移的参数。通过求得的左相机和右相机参数,最后得到用于校正立体图像的右边摄像机相对左边摄像机的旋转和平移参数。文献^[37]有提及。

最终通过棋盘格标定法,使用一个长宽为 11*8, 每个小方格边长 1.5cm 的棋盘图像。最后得到双目相机标定结果如下所示:

内参矩阵:

$$R_{Left} = \begin{bmatrix} 1401.5306 & 0 & 1004.3107 \\ 0 & 1401.5306 & 536.1793 \\ 0 & 0 & 1 \end{bmatrix}, \quad R_{Right} = \begin{bmatrix} 1401.5306 & 0 & 1014.956 \\ 0 & 1401.5306 & 536.1258 \\ 0 & 0 & 1 \end{bmatrix}$$

畸变系数:

$$D_{Left} = [-0.1845 \quad 0.1284 \quad 0 \quad 0 \quad -0.2069], \quad D_{Right} = [-0.1453 \quad -0.0567 \quad 0 \quad 0 \quad 0.6504]$$

旋转矩阵:

$$R = \begin{bmatrix} 0.9999 & 0.0002 & 0.0058 \\ 0.0002 & 1.0000 & -0.0005 \\ -0.0058 & 0.0005 & 0.9999 \end{bmatrix}$$

平移矩阵:

$$T = [-12.0015 \quad 0.0012 \quad -0.0068]$$

所得到的标定结果 R 阵用作表示右摄像机坐标系过渡到左相机坐标系的旋转阵,而 T 则是代表平移矩阵,由于早就固定了双目相机之间的距离,发现平移矩阵中得到的 12cm 参数也与相机保持一致。

2.2.2 图像特征提取

在计算机中,我们知道色彩和亮度构成的矩阵是组成图像不可或缺的一部分,但是从矩阵来分析运动估计计算量过于庞大。所以要从图像中优先选择具有代表图像特征的特征点,再通过这些特征点进行相机运动估计。此运动估计的前提是在相机视角产生少量变化后特征保持不变。一般情况下,图像中灰度值发生变化较大的像素点一般是特征点,它包括点特征、边缘特征和区块特征。其中边缘特征和区块特征没有图像中间的点辨识度高。就比如说,边缘特征虽然比点特征适用于描述整副图像的特点,但是在两个视角下的同一个边缘特征展示出来的效果却是不同的,所以边缘特征是不如点特征优越的。又因为点特征有更好的抗噪能力,使得点

特征成为最好的特征点选取方式。

特征则是由关键点和描述子组成的。其中关键点包括特征点在整体图像中的位置信息。为了呈现特征点周围像素的信息叫做描述子。规则就是假如两个特征点的描述子相似，那么将这两个特征点视为同一个特征点来匹配。

常用的特征点检测有 SIFT、Harris、ORB 和 SURF。

(a) SIFT 算法:

SIFT 算法^[38]，该算法的大致内容是先是构建一个金字塔，然后在基于高斯差分金字塔条件下，去搜索极值点，其次通过找到的极值点获取其周围区域的梯度直方图，用直方图的结果来确定关键点的方向；最后对关键点和它的邻域内像位点生成向量，获得描述子，详细操作如下：

(1) 建立高斯金字塔；

(2) 关键点检测 DOG 算子，将相邻尺度高斯平滑后的图像相减；

(3) 建立高斯差分金字塔；

(4) 将每一个像素点与周围的 26 个检测点，以及跟它们在同一尺度的相邻点，共计八个，最后包括上下相邻尺度 18 个点共计 26，来进行比较，搜索 DOG 算子空间极值点。

(5) 关键点精确定位

由于关键点的稳定性很重要，所以为提高稳定性，必须对尺度空间 DOG 函数采取曲线拟合处理。以此来去掉不稳定对比度低的极值点。在 Lowe 的试验显示中，得到灰度值低于 10 的极值点可以舍弃；

此外在去除边缘响应时，当 $\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$ ，保留此关键点，其余的舍弃。

其中，矩阵 H 为 Hessian 矩阵。

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2-6)$$

$$\begin{cases} Det(H) = D_{xx} \times D_{yy} - D_{xy} \times D_{xy} \\ Tr(H) = D_{xx} + D_{yy} \end{cases} \quad (2-7)$$

(6) 关键点方向分配

通过求每个极值点的梯度来为极值点赋予方向，通过尺度不变性求极值点，这样会令它享有缩放不变的性质，再通过关键点的邻域像素所具备的梯度方向，通过这些梯度方向的特性为对应关键点指定方向，从而使描述子对图像旋转具有不变性。

关键点主方向指的是极值点附近梯度直方图的主峰值，关键点辅方向称作第

二峰值,第二峰值所含有的能量约等于主峰值的 80%。到这对于图像关键点的检测已经完成,其中每个关键点都包含三个信息:分别为位置、尺度、方向;这也就是为什么方向点具备平移、缩放和旋转不确定性的原因。

(7) 关键点描述

选取一组向量可以把关键点以及它附近的有贡献的像素点描述出来,用作特征匹配的依据,如此一来使关键点不变性质变得越来越多,比如说光照不变性。

对关键点像素块周围的区域进行分块,计算分块后它的梯度直方图,每一个像素块都生成独特的向量。Lowe 最终的实验结果得出采取 $4 \times 4 \times 8 = 128$ 维向量表现出的效果最好^[39]。

(b) SURF 算法:

SURF 算法(Speeded Up Robust Feature),是具有高鲁棒性、运算速度快的特征检测算法^[40]。它通过积分图像、Haar 小波变换以及 Hessian 矩阵运算来增加运算效率,通过 Haar 小波变换可以对鲁棒性产生增强。相比于 SIFT 算法, SURF 的有三方面的区别:第一个是通过不一样的小波滤波模板和积分图像求解 Hessian 矩阵行列式对应的响应图像,接下来对图像上则使用了 3D 非最大值抑制的手段,而不是之前提到的直接求取金字塔图像的方式;第二个是用 Haar 小波响应来得到对应的特征向量;第三个是描述子的维数比 SIFT 算法低。下面是算法的主要步骤:

(1) 求取可运用于不同尺度空间的小波滤波模板。

(2) 求取积分图像。

(3) 使用近似的 Hessian 矩阵,当计算出矩阵的行列式小于零时,该点很有可能是极值点。

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (2-8)$$

其中, $L_{xx}(X, \sigma)$ 表示像素点 (x, y) 与二阶偏导数 $\frac{\partial^2 g(\sigma)}{\partial x^2}$ 的卷积,代表该点在 σ 尺度下在 x 方向上的二阶偏导数, $g(\sigma)$ 是高斯函数。因为高斯函数它的二阶偏导和图像的积分环节计算量很大,所以应该改进为运用盒状小波滤波器来近似代替高斯函数的二阶偏导,以此可以获得更快的运算速度。

(4) 通过变换改变小波滤波器的大小和高斯函数的尺度,建立图像尺度金字塔, SURF 特征具有尺度不变特性。

(5) 把 a 记作极值点所在图像的尺度,通过计算极值点附近邻域内点的 harr 小波响应结果以及权重信息。将极值点定为中心,使其邻域六等分,分别计算出每一个区域 haar 小波响应结果的加权和,最大幅值所对应的俯角即是特征点的方向。

(6) 通过 harr 小波响应值以及得到的高斯权重,计算出一个 64 维向量,用

以表征 SURF 特征描述符。

(c) ORB 算法:

ORB 算法在 FAST 特征检测和 BRIEF 特征描述的基础上对它们做出了优化。这其中 FAST 检测速度快于 SIFT 和 SURF; BRIEF 算法的优点是计算结果为二进制串的特征描述符,这种描述符可以节约更多的存储空间,提高匹配速度,相较于其他算法有更强的优越性^[41]。下面是算法的主要步骤:

(1)先构建图像金字塔。根据金字塔层数 L 和尺度因子进行初始图像的缩放,然后进行高斯平滑处理。

(2) 采取 FAST 特征检测提取角点。在图像中选取一个像素点,记为 p , 设其灰度值为 I_p 。其次设置一个阈值 T (假设为 I_p 的 20%)。再次以像素点 p 为中心,作半径为 3 像素的圆,这个圆的圆周经过了 16 个像素点。接下来如果所选择的圆上,存在连续 N 个点的亮度都大于 $I_p + T$ 或都小于 $I_p - T$, 那么像素 p 则可以被看做特征点 (N 通常取 12, 即为 FAST-12)。最后重复上面四步,对接下来的每一个像素执行同样操作。对于 FAST 角点数量很大且不确定的缺点, ORB 算法在这方面做出优化。把之前所提取到的角点采取非极大值抑制处理,对小空间内集中较多的角点进行剔除。在剩下的角点中挑出前 N 个 Harris 响应值大的,以此来获得最后的特征点集合。

(3) 求取特征点的方向。以特征点为中心,以 15 像素为半径作圆。用灰度质心法来求解圆域的质心坐标以及特征点方向向量。水平方向与方向向量所成的夹角就是特征点的方向。这就是算法 ORB 在原版 FAST 算法上的作出的改进。

(4) 求取描述子,从特征点领域内选择 128、256 或 512 对像素点,比较每个点对的灰度值,对应生成二进制 0 或 1,到最后得到的描述子用一串长度为 128、256 或 512 的二进制数表示。

FAST 角点具备了方向是因为 ORB 算法通过灰度质心法所得到的,这一方式使得改进后的 FAST 角点具备了旋转不变性。与此同时, BRIEF 是没有方向的,一旦图像发生旋转时,同一个特征点如果换成不一致的描述子也会产生变化,这是不足的地方。因此引入 BRIEF 算法,通过旋转矩阵提取出随机点对,如果能保证选取的模式相同,描述子就可以在旋转时保持不变。在匹配机制上, ORB 算法使用了汉明距离,而前面提到的 SIFT 和 SURF 则是使用了欧式距离,当描述子的维度一致的情况下,由二进制数构成的 BRIEF 计算速度明显快于另外两种方法。

(d) Harris 算法:

Harris 角点检测算法的核心思想是让一滑动窗口以每一像素点为中心沿任意方向滑动^[42]。它判断角点的依据是角点是指窗口无论向哪个方向移动都会导致灰度

值发生明显变化。

对于待测像素点 (u, v) ，定义出的灰度变化函数公式如（2-9）所示。

$$E(x, y) = \sum_{u, v} w(u, v) [I(x + u, y + u) - I(u, v)]^2 \quad (2-9)$$

式中， $w(u, v)$ 为窗口函数，可用高斯窗或者是矩形窗都可以。 (x, y) 作为平移量。

当 (x, y) 较小时，通过泰勒展开，可以得到式：

$$E(x, y) \cong [x \quad y] M \begin{bmatrix} x \\ y \end{bmatrix} \quad (2-10)$$

$$M = \sum_{u, v} w(u, v) \begin{bmatrix} I_u^2 & I_u I_v \\ I_u I_v & I_v^2 \end{bmatrix} \quad (2-11)$$

设 λ_1, λ_2 为矩阵 M 的两个特征值，当二者都较大时，窗口无论向哪个方向移动都会带来明显的灰度变化，则该点为角点；当二者一个特征值大另一个特征值较小时，窗口沿着某个方向移动时，灰度值变化大，而另一个方向变化却很小，这说明这个点位于图像边缘；当二者特征值都较小时，这个点则是内点，位于图像的平坦区。

通过前面对四种算法的分析，分别对四种算法提取图像特征点的效果进行了实测对比。主要包括四种情况：第一种情况是图片平移后，特征点的检测是否受影响，确定检测效果更好的算法；第二种情况是图片发生旋转，特征点的检测是否受影响，确定检测效果更好的算法；第三种情况是图片发生尺度变化，即缩放，特征点的检测是否受影响，确定检测效果更好的算法；第四种情况是光照强度发生变化，特征点的检测是否受影响，确定检测效果更好的算法。

从上到下依次为 SIFT、SURF、ORB、Harris 的检测效果：

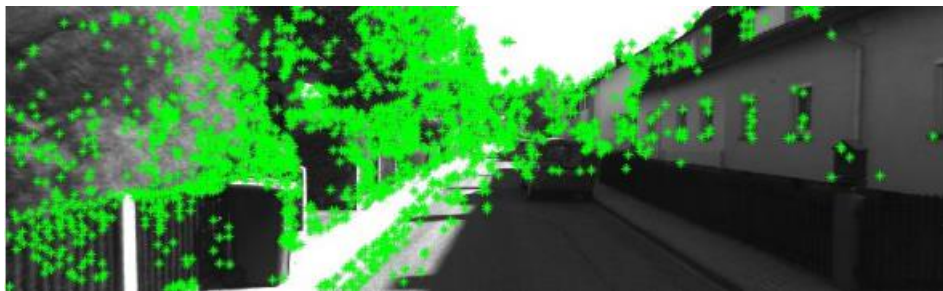


图 2-2 SIFT 特征检测结果

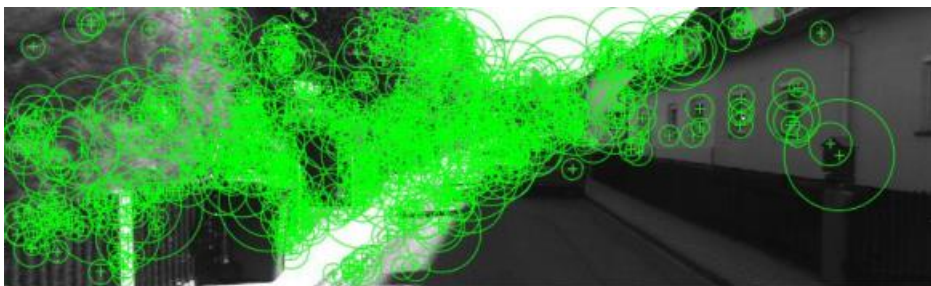


图 2-3 SURF 特征检测结果

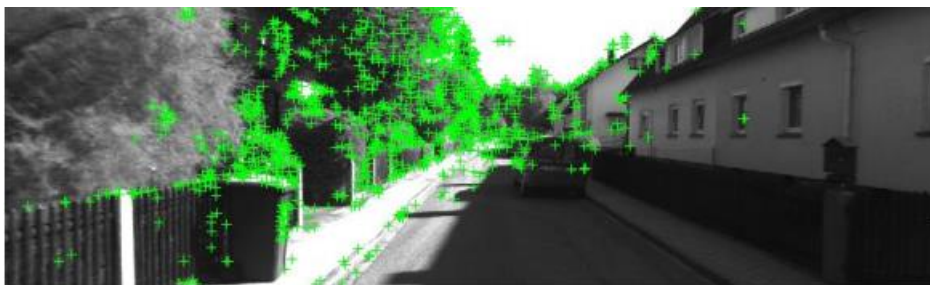


图 2-4 ORB 特征检测结果



图 2-5 Harris 特征检测结果

表 2-1 四种方法对应特征点提取结果

项目	SIFT	SURF	ORB	Harris
特征点个数	2548	886	1251	734
耗时/ms	914.7	201.5	26.2	25.1

通过表 2-1 可看出，SIFT 算法检测到的特征点个数最多，是 SURF 和 Harris 的三倍，是 ORB 的两倍，但是它的耗时也是最长的。综合来看，ORB 算法最这四组算法中最优的，无论是在实时性还是在检测点个数上都有不错的平衡。

然后用同样的方式对尺度、旋转、光照后的图片进行特征点检测，得到的结果如下表 2-2 所示。

表 2-2 实验三性能指标

特征点	原图像	图像缩小为原来的一半	图像顺时针转 20°	低亮度图像
SIFT	706	482	712	149
SURF	1351	1145	1320	589
ORB	1943	1332	2425	401

通过表 2-2 可得出，对 SIFT 算法来说，当尺度信息发生变化后，特征点检测个数下降 1/3 左右；在图像旋转 20°时，特征点数目几乎没有变化；而当图像亮度过低时，所有算法的特征点检测数目都下降 80%以上，影响最大。而相较于 SURF 算法和 ORB 算法而言，检测得到的特征点数量在图像旋转和尺度变化的影响下比较小。

2.2.3 图像特征匹配

在特征点匹配上，比较常用的两个方法分别是暴力匹配法和近似最近邻算法。暴力匹配法使用的是穷举策略，具体为对两幅图像上全部的特征点都进行描述子距离的计算，再对其进行排序，距离最小的两个特征点互为匹配点。由于 SIFT 和 SURF 所使用的描述子是浮点类型，所以 SIFT 和 SURF 的描述子距离使用的是欧式距离；而 ORB 的 BRIEF 的描述子更使用的是二进制描述子，选择采取汉明距离测量它。汉明距离是用来描述两个二进制串所对应的位数不同的个数。由于小车对实时性的要求较高，暴力匹配的计算量大，不宜采用。

近似最近邻算法(approximate nearest neighbor algorithm)，主要包含随机 k-d 树搜索算法，k-means 树优先搜索算法^[43]，局部位置敏感哈希算法等。近似最近邻算法集成在 FLANN 开源库中。

(1) k-d 树上的最邻近查找算法

k-d 树上的最邻近查找算法先是寻找数据集中方差最高的维度，利用方差最高维度所持有的数值把数据分成两个部分，接下来再对每个子集循环之前一样的步骤，形成二叉树，之后再进行搜索。在以目标图像关键点为基准的前提下，搜索的内容是与目标图像的特征点最邻近的原图像特征点和次邻近的原图像特征点。

(2) 局部位置敏感哈希算法

前面提到的基于树的临近查找算法在空间维度较低时有比较可观的匹配速度，但是当空间维度比某一值高了的时候，它的搜索速度便不能满足需求。局部位置敏感哈希算法(Locality Sensitive Hashing)以降低部分精确度为代价提高了搜索的速度，更适合无人车对实时性高的要求。该算法的核心是，用哈希函数对高维向量进行处

理，随之将其转换到低维度空间哈希桶中。对特征点所映射的哈希桶和它的邻近哈希桶进行查找。

下面为 SURF、Harris 和 ORB 的特征点检测和 Flann 匹配效果图。



图 2-6 SURF 特征检测与匹配结果



图 2-7 Harris 特征检测与匹配结果

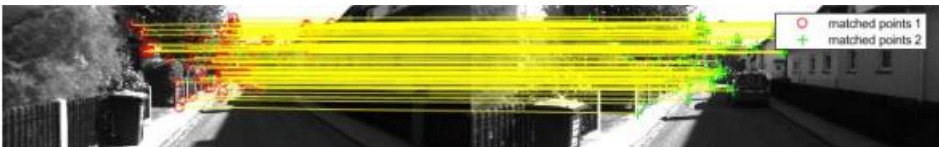


图 2-8 ORB 特征检测与匹配结果

表 2-3 三种特征点匹配结果比较

特征点	SURF	Harris	ORB
A 图特征点	886	734	1251
B 图特征点	1078	973	1588
匹配对个数	270	111	154

在无人小车行进的时候，必然要对相机采集到的图像进行特征点匹配，进而才能得到图与图之间的关联。但经常存在匹配时导致的误匹配情况，生成了一些误匹配对，有一种去除误匹配的方法叫暴力法，但却很可能将正确的对误看作错误的；同时仍有少量的误匹配存在于匹配对中。因此，从上述的两个问题可以看出暴力去除误匹配需要借助另外的约束条件来去除两图像间的误匹配对。

由于不知道不同帧下的相机是怎么运动的，那么就一定先必备早就匹配了的特征点对，只有通过它们才可以计算出是如何变换的。完成特征匹配以后，会存在一部分包括噪声的匹配对，但是如果不知道是如何变换的，就不可以迅速地将把误匹配对清除。基于统计学理论的随机采样一致算法(Random Sample Consensus, RANSAC)，该算法可以用于很多情形下，可以在包含噪声的数据集里面，通过循

环迭代，去寻找最优的变换模型。

RANSAC 算法先是根据具体的问题设定出需求的目标函数，然后为了估计参数的初始值采用循环提取最小点集方法，利用这些初始值把之前得到的所有的数据分为“内点”和“外点”，最终通过所有的内点重新对函数的参数进行计算。采取 RANSAC 算法对变换关系的单应性矩阵 H 及基础矩阵 F ，步骤如下所示：

(1) 在所有数据中随机选取出几组匹配对，初步求得单应矩阵 H_0 ，或者基础矩阵 F_0

(2) 对求得的单应矩阵及基础矩阵设置一个阈值 T_H 和 T_F ，将所有匹配对都进行检测，计算得出内点对的个数。

(3) 将最佳模型的内点对与此时模型的内点对进行对比，如若最佳模型内点对数比当前模型内点对数小，则最优模型就会被赋予给此时的模型。

(4) 把 p 表示为置信度，通过 p 和内点比例求解最大的迭代次数，用 k 表示。公式如下(2-12)所示：

$$\begin{cases} k = \frac{\log(1 - p)}{\log(1 - w^n)} \\ w = \frac{N_{imm}}{N_{um}} \end{cases} \quad (2-12)$$

其中 w 为求得的内点比例。就是内点与总点数的比值。 n 代表最小样本数。当是单应性矩阵， n 最少是 4；当所求的是基础矩阵矩阵， n 最少是 8。通过公式 (2-12) 可以看出， k 是根据当前模型的 w 不断产生变化的，在同样的条件下，哪个模型能够寻找内点的数目越大，它就越离最优估计模型越近。

(5) 继续循环迭代，什么时候迭代数与最大迭代数一致了。那么这个时候得到的 H 以及 F 的求解就是最佳变换模型。

已经通过 RANSAC 法求得到最佳模型以后，接下来通过(2-11)和(2-12)，可以使用它寻找并去除误匹配。



图 2-9 误匹配去除后的效果图

表 2-4 误匹配去除算法性能比较

算法	原始对数	匹配对数	用时/ms
H 矩阵	500	255	4.9
F 矩阵	500	237	20.6

通过观察图 2-9 和 2-4 的表格，可以得知 H 比 F 的去除误匹配更迅速，理由其实就是(2-12)中要选不一样的 n 值，从而导致了 RANSAC 法对 F 求取时更费时。此外，在同样条件下在 H 矩阵的精确匹配对数值更高。综上所述，最终选择 RANSAC 方法所预估出的单应性矩阵 H 来去除帧间匹配中的误匹配点对。

2.2.4 运动估计

在完成左右相机图像特征匹配后，可以得到一系列的匹配点对，把这些求得的匹配点对用在前后帧图像特征匹配上即可得出前后时刻相机的相对运动，从而求出姿态信息和位置信息。

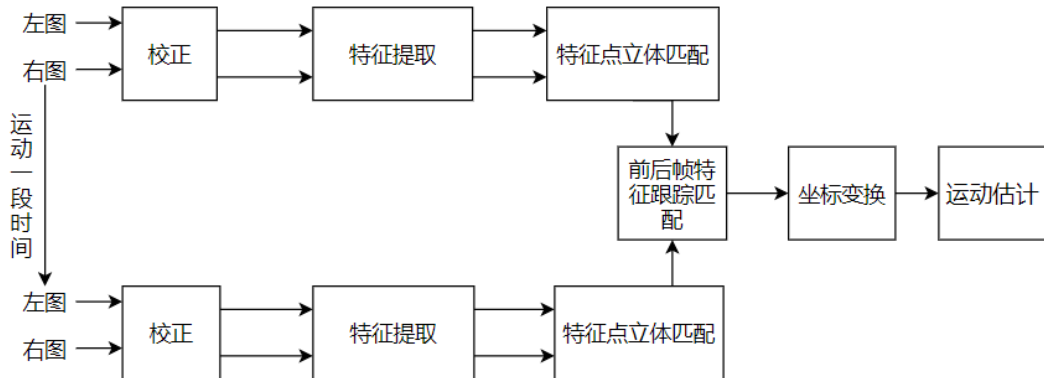


图 2-10 双目视觉定位框图

在小车视觉定位的整个过程中的重点就是位姿估计，因为位姿计算的精确度将直接影响实际的轨迹与求得的轨迹的贴合程度。设空间一点坐标为 $P(X,Y,Z)$ ，它在世界坐标系下的坐标为 $P(X_w,Y_w,Z_w)$ ，假设第 k 帧和 $k+1$ 帧时，这个点被相机所拍到。把它在相机坐标系下的坐标点计为 $P_c^k(X_c^k,Y_c^k,Z_c^k)$ ，那么根据刚体运动约束，存在下式成立：

$$P_w = R_{w_k} P_c^k + t_{w_k} = T_{w_k} P_c^k \quad (2-13)$$

$$P_w = R_{w_k+1} P_c^{k+1} + t_{w_k+1} = T_{w_k+1} P_c^{k+1} \quad (2-14)$$

其中，把第 k 帧从世界到相机坐标系的变换矩阵记为 T_{w_k} ， T_{w_k} 是旋转矩阵

R_{w_k} 和平移向量 t_{w_k} 组成。把式(2-13)及式(2-14)进行联立, 就可以知道某一空间点基于不同帧相机坐标系下是怎样的变换关系, 如下式(2-15)所示:

$$P_c^{k+1} = T_{w_k+1}^{-1} T_{w_k} P_c^k = T_{k+1_k} P_c^k \quad (2-15)$$

式中, T_{k+1_k} 同样由旋转矩阵 R_{k+1_k} 及平移向量 t_{k+1_k} 构成。旋转和平移与点集的三维坐标之间的关系可以用如下式子表示:

$$P_c = R P_p + T \quad (2-16)$$

上式中 P_p 和 P_c 分别表示前一帧和当前帧的匹配点对在相机坐标系下所对应的点的三维坐标。其中 R 为旋转矩阵, T 为平移向量。设 α 、 β 、 γ 分别为绕 x 轴、 y 轴和 z 轴旋转的角度, 它们分别表示偏转、俯仰和侧倾角。

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \quad (2-17)$$

$$R_y = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \quad (2-18)$$

$$R_z = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-19)$$

$$R = R_x R_y R_z \quad (2-20)$$

求解出 α 、 β 、 γ 既可求解出旋转矩阵 R 。根据两副图像特征匹配对的维度, 可以将双目相机运动的估计方法分为两类: 一是 3D-3D, 二是 3D-2D。

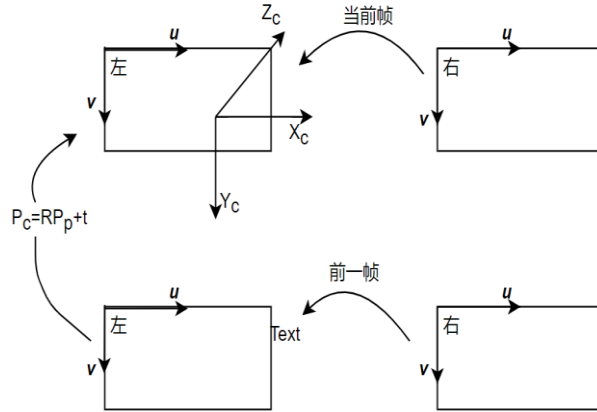


图 2-11 位姿估计

其中 3D-3D 方法是通过已知的特征匹配对的坐标，搜索符合要求的 R 和 t ，使全部匹配对应的相邻帧之间对应 3D 重投影误差求和是最小的。

3D-3D 法经常使用的是 ICP^[44]算法，本文将采取基于 ICP 与 SFM 结合的双目立体视觉三维重构算法。在 SFM 算法中，获取连续图像序列后，基于特征匹配算法在相邻图像对中寻找特征点并完成匹配。接着基于双视角的对极几何，去计算图像特征匹配对解出本征矩阵 E ，并通过对 E 进行奇异值的分解，来求取出左右相机之间的 R 与 t ，再基于三角测量对三维点云进行匹配点的计算。

对于 ICP 算法来说，先是把集合中的所有点依次移到质心位置，然后通过 SVD 分解对载体移动之后的点坐标阵来对最优旋转阵 R 进行求解，最后通过刚才求得的旋转矩阵 R 利用式 (2-21) 解算出平移向量 t ，求解步骤如下：

选取两组点的质心定义如下：

$$p = \frac{1}{n} \sum_{i=1}^n (p_i), \quad p' = \frac{1}{n} \sum_{i=1}^n (p'_i) \quad (2-21)$$

那么优化目标函数可以写成：

$$\min_{R,t} J = \frac{1}{2} \sum_{i=1}^n \left\| p_i - p - R(p'_i - p') \right\|_2^2 + \left\| p - Rp' - t \right\|^2 \quad (2-22)$$

优化目标函数的左项只和旋转矩阵 R 相关，右项既有参数 R 也有参数 t ，但只和质心相关。因此可通过左项为零求出 R ，再通过右项为零求出 t 。

所以，ICP 可分为以下三个阶段求解：

(1) 首先计算两组点的质心位置 p 和 p' ，再求每个点的去质心坐标：

$$q_i = p_i - p, \quad q'_i = p'_i - p' \quad (2-23)$$

(2) 然后下面优化问题求取旋转矩阵:

$$R^* = \arg \min_R \frac{1}{2} \sum_{i=1}^n \|q_i - Rq_i'\| \quad (2-24)$$

(3) 最后根据 R 求

$$t^* = p - Rp' \quad (2-25)$$

先通过SFM法从多角度进行拍摄目标, 再通过双目相机采集的四个角度所采集的图像, 再去求取特征点所对的三维点云即可。整个过程变为前者以及新增图像的特征点进行融合, 再基于上述ICP去求取挨着三维点云之间的 R 和 t 。

3D 到 2D 法与 3D 到 3D 法一样, 同样都是通过帧间对应的 n 对匹配点对去完成运动估计。这里面 3D 指的是匹配点对在前帧相机坐标系下的三维坐标, 2D 指的是后一帧图像相机坐标系下的坐标。配合适当的手段结合通过相机内参, 可以求出相对于帧间的运动, 这被称作 PnP 问题。

n 点透视问题的解法主要有两种: 迭代法和非迭代法。其中前者是通过构建重投影误差代价函数, 用迭代手段减小取值, 直到发现最优相对位姿。迭代法的特点是额初始值选择非常重要, 它直接影响了迭代的效率问题, 且迭代法容易陷入局部最优解的陷阱。

非迭代法可分为专用求解法和通用求解法。经常见到的专用求解法有 P3P、P4P 等, 它们是去选取指定数值的匹配点, 再经过约束方程组, 以此来求出相机的帧间位姿, 专用求解法特点是要求比较高的匹配点对的三维坐标精度, 所以它的鲁棒性不好。通用求解法分为 EPnP 法, 直接线性法。后者通过相机参数模型和匹配点对情况来成立方程组求取帧间的位姿变换。假设特征点个数的是 n , 那么计算复杂度就为 $O(n^3)$, 可以看出效率较低, 因为 EPnP 算法^[45]的计算复杂度仅为 $O(n)$ 。

通过 EPnP 法求解得到的相机帧间位姿可以用光束平差法(BA)进行非线性优化, 其实光束平差法的本质就是迭代。

通过 EPnP 算法计算后, 可以求出用来表示帧间相对位姿的 R 和 t 。则重投影误差为:

$$e = Rp + t - p \quad (2-26)$$

在式(2-27)中:

P 表示在相机坐标系中, 上一帧的匹配点的 3D 坐标。

p 表示相机坐标中, 下一帧的匹配点的 2D 点坐标。

BA 法的最终的追求的是 R 和 t , 它们可以让重投影误差平方和取到最小, 式子表示为:

$$\min_{R,t} e^2 = \sum_{i=1}^n \|RP_i + t - p_i\| \quad (2-27)$$

将式(2-27)表示成为状态变量 X 的函数:

$$\min F(X) = \sum_{i=1}^n \|e(X, P_i, p_i)\|_2^2 = \sum_{i=1}^n e(X, P_i, p_i)^T Q_i^{-1} e(X, P_i, p_i) \quad (2-28)$$

接下来对式(2-28)进行泰勒展开, 并对 ΔX 进行求导。当迭代到一定时候时, ΔX 会低于阈值或迭代上限发生, 被迫停止。这时旋转矩阵和平移向量即为满足相机重投影误差平方和为最小。

通过前后帧图像的平移旋转关系的迭代, 可以获得任意时刻相机相对初始状态的位置关系。设定初始位置信息, 即可换算出任意时刻相机在世界坐标系中的位置。

2.3 本章小结

本章主要对视觉-惯性导航系统中双目视觉进行研究, 首先分析了摄像机的张氏标定方法, 对比 SIFT、Harris、ORB 和 SURF 四种主流的图像特征点检测方法, 进而验证了其中的 ORB 特征点检测方法的最优性, 在特征匹配后基于 RANSAC 方法给出了特征点误匹配的去重算法, 通过特征提取匹配并对图像特征进行跟踪并给出了基于 ICP/SFM 重构算法。

第3章 双目视觉与IMU组合导航定位系统

3.1 引言

VINS之所以是一个高度非线性的系统，主要就是因为存在视觉观测、坐标变换以及IMU积分等因素。本章主要记录了通过松耦合方法对系统初值进行确定，在系统刚开始阶段，通过之前提到的基于ICP/SFM的双目立体视觉三维重构算法对位姿和特征点的三维坐标进行了计算；接着在前端处理方面，对选取出来的两帧图像采取IMU预积分，通过惯性信息求得所选出的两帧图像之间物理关系。通过这些关系构造方程，为后续紧耦合所需的初始数据做准备。

3.2 系统结构

通过实际情况与当今流行视觉惯性导航系统拥有的功能提出了图3-1的框架，该框架包含了IMU测得的惯性信息以及视觉部分通过图像估计出的载体位姿信息和路标点坐标，为了避免位姿估计时产生随时间产生的漂移，整个系统具备了回环检测功能以及全局位姿优化功能。

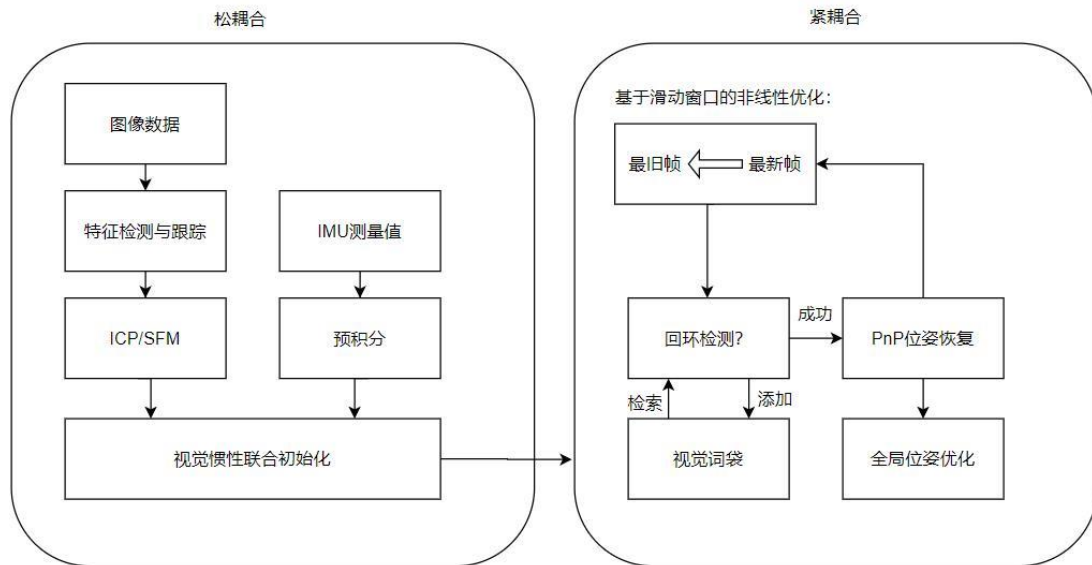


图 3-1 系统整体结构图

系统被分为松耦合端与紧耦合端两部分：左侧是松耦合端，主要包括预处理部分，整个系统的初始化。右侧是紧耦合端，主要包含基于滑窗的非线性优化，基于词袋模型的回环检测部分，本章节主要介绍系统松耦合初始化过程。

3.3 IMU 运动估计

VINS 将视觉部分与惯性部分信息进行融合，先验信息不知道的情况下对载体位姿情况进行估计。所以为了对 VINS 进行非线性优化的研究，就先需要对坐标性变换有深刻的认知，建立出视觉部分传感器的模型。

3.3.1 导航坐标系

(1) 惯性坐标系

惯性坐标系原点 O_i 位于地球的中心地心。共有三个轴向 x_i 、 y_i 、 z_i 。其中 x_i 从原点指向春分点； z_i 所指方位是与自转轴平行的。

(2) 地球坐标系

地球坐标系它的原点是球心，地球自转轴指向哪它的轴就指向哪是一致的。在这里， x_e 是以球心为原点指向子午线； z_e 起点是球心，平行于自转轴； y_e 轴从地心指向赤道与东经 90° 交点，遵循右手坐标系。

(3) 地理坐标系

地理坐标系，它的坐标原点是运载体质心。 x_g 和 y_g 轴处在水平面上。 z_g 从地心指向自己本体，与垂直于地面是一个方向，指向天向。定位中经常提及到的东北天坐标系，就是一种最常应用的地理坐标系，当载体位于北半球时一般采用东北天坐标系，反之南半球用北东地坐标系。本文选择东北天为地理坐标系。

(4) 载体坐标系

载体系最大的特点就是相对于载体本身不动，三个轴向分别用 x_b 、 y_b 、 z_b 表示。坐标原点为物体的质心， x_b 轴向由质心指向物体的前进方向，也叫作横滚轴向； y_b 轴向由质心指向物体右方向，也叫作俯仰轴向； z_b 轴也叫作航向轴向。

(5) 导航坐标系

导航坐标系是计算运载体运动参数用的坐标系。本文用东北天坐标系作导航坐标系。

而我们在视觉惯性导航领域主要会用到世界坐标系、摄像机坐标系以及 IMU 坐标系。其中保持不动的是世界坐标系 W，它的 z 轴和重力方向一致；相机坐标系 C 与相机紧密关联，而相机是固定在载体上的，所以它是随载体的运动而运动的；IMU 坐标系 B 也称为载体系，因为与 IMU 固连，所以它也是一个移动坐标系，故也叫作载体系。

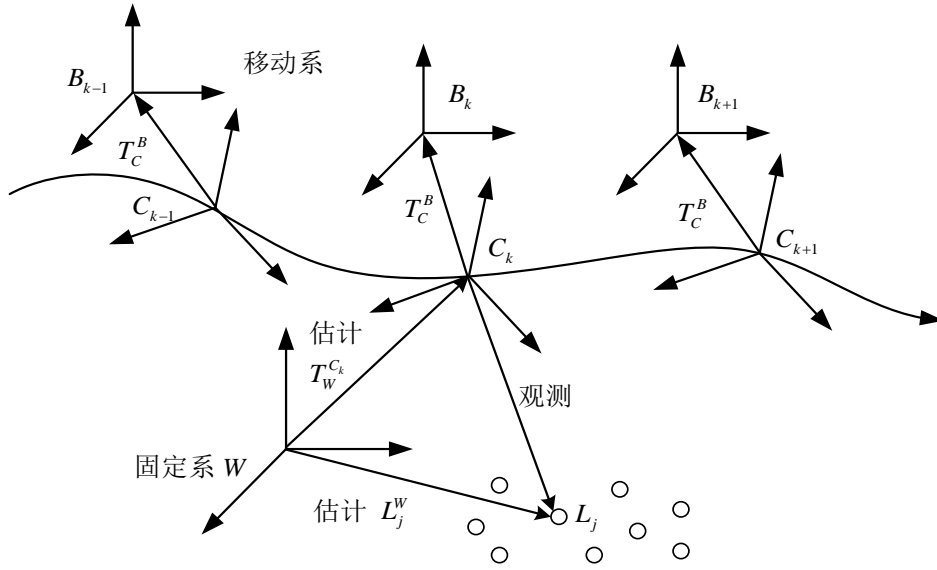


图 3-2 VINS 中坐标系之间的关系

3.3.2 坐标系变换关系

（1）旋转矩阵与欧拉角

旋转运动其实是载体围绕着 X、Y、Z 轴依次进行旋转后，最终得到的结果，三次旋转的角度分别代表着滚转角 ϕ 、俯仰角 θ 和偏航角 ψ 。使用欧拉角这种方式表达坐标系的变换很直观，但是会出现方向锁的问题。还可以使用方向余弦矩阵 R 来表示旋转矩阵的运动，最终 R 可表示为下式：

$$R = \begin{bmatrix} \cos \theta \cos \psi & \cos \phi \sin \psi + \sin \phi \sin \theta \cos \psi & \sin \phi \sin \psi - \cos \phi \sin \theta \sin \psi \\ -\cos \theta \sin \psi & \cos \phi \cos \psi - \sin \phi \sin \theta \sin \psi & \sin \phi \cos \psi + \cos \phi \sin \theta \sin \psi \\ \sin \theta & -\sin \phi \cos \theta & \cos \phi \cos \theta \end{bmatrix} \quad (3-1)$$

载体于世界系经通过相机系转变为载体系表示为 R_{BW} 。因为旋转矩阵是一个单位正交阵，根据矩阵的特性，它具有转置和逆相等的特性，所以从载体系到坐标系的旋转就可以表示为：

$$(R_{WB})^{-1} = (R_{WB})^T = R_{BW} = R_{BC} R_{CW} \quad (3-2)$$

（2）四元数法

1982 年 Hamilton 提出了四元数法，四元数法只需要求解四个未知数，求解过程相对容易，四元数与复数很相像，也是具备实部和虚部，写为下面 4×1 的列向量形式^[46]：

$$q = \begin{bmatrix} \vec{\varepsilon} \\ \eta \end{bmatrix} = q_0 + q_1 \vec{i} + q_2 \vec{j} + q_3 \vec{k}, \quad \eta = q_0, \quad \vec{\varepsilon} = [q_1, q_2, q_3]^T \quad (3-3)$$

上式中, $\vec{\varepsilon}$ 表示三个虚部, η 是实部。四元数也可用三角函数来表达:

$$\vec{\varepsilon} = \begin{bmatrix} k_x \sin(\theta/2) \\ k_y \sin(\theta/2) \\ k_z \sin(\theta/2) \end{bmatrix} = \vec{k} \sin(\theta/2), \quad q_0 = \cos(\theta/2) \quad (3-4)$$

通过三角函数的写法, 发现四元数直白地描述旋转的过程, 其中 θ 为转过角度, \vec{k} 为单位向量。四元数法主要就是四个量描述三维旋转, 所以为了使其能够作用于坐标系的旋转, 对四元数法需要添加一个约束, 即为四元数的单位化:

$$|q| = \sqrt{q^T q} = \sqrt{|\vec{\varepsilon}|^2 + \eta^2} = 1 \quad (3-5)$$

所以, 参数化也是四元数具备的一个性质。两个互为相反数的单位化后的四元数表示沿着相反地去旋转最终依旧到达相同的一个坐标系。互为相反数的两个四元数可以用来表示任意的旋转^[47]。此时载体系中假设有一速度量记为 v^B , 经过旋转之后在世界系下得速度量记为 v^W , 而设四元数 q_B^W 涵盖旋转的所有信息。则从 v^B 经四元数乘法到 v^W 有下式^[48]:

$$v^W = q_B^W \otimes v^B \otimes (q_B^W)^{-1} \quad (3-6)$$

四元数和旋转矩阵二者可以互相转换, 具体为:

$$R_W^B = \begin{bmatrix} 1-2q_2^2-2q_3^2 & 2q_1q_2-2q_0q_3 & 2q_1q_3+2q_0q_2 \\ 2q_1q_2+2q_0q_3 & 1-2q_1^2-2q_3^2 & 2q_2q_3-2q_0q_1 \\ 2q_1q_3-2q_0q_2 & 2q_2q_3+2q_0q_1 & 1-2q_1^2-2q_2^2 \end{bmatrix}^T \quad (3-7)$$

设矩阵 $R_W^B = \{r_{ij}\}$, $i, j \in [1, 2, 3]$, 则根据式 (2-1) 和 (2-7) 可以求出载体姿态角的值:

$$\begin{cases} \phi = \arctan(-r_{32}/r_{33}) \\ \theta = \arcsin(r_{31}) \\ \psi = \arctan(-r_{21}/r_{11}) \end{cases} \quad (3-8)$$

因此俯仰角、偏航角和滚转角的取值范围分别为: $(-90^\circ, 90^\circ]$ 、 $(-180^\circ, 180^\circ]$ 和 $(-180^\circ, 180^\circ]$, 实际计算中应该合理取值。

除此之外, 四元数的微分方程也是它另外一个主要的性质:

$$\dot{q}_B^W = \frac{1}{2} q_B^W \otimes \tilde{\omega}(t) \quad (3-9)$$

也可表示为:

$$\dot{q}_B^W = \lim_{\Delta t \rightarrow 0} \frac{q_B^W(t + \Delta t) - q_B^W(t)}{\Delta t} \quad (3-10)$$

则对于离散时间函数的一阶近似形式为：

$$q_B^W(t + \Delta t) \approx \left(I + \frac{1}{2} R(\tilde{\omega}(t)) \Delta t \right) q_B^W(t) \quad (3-11)$$

其中 $R(\tilde{\omega}(t))$ 表示角速度右乘矩阵。

对于平移变换来说，用 t_C^W 表示世界坐标下的相机坐标系原点，所以可把变换表示成：

$$v^W = R_C^W v^W + t_C^W \quad (3-12)$$

可通过齐次坐标再将上式表示为：

$$\begin{bmatrix} v^W \\ 1 \end{bmatrix} = \begin{bmatrix} R_C^W & t_C^W \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v^C \\ 1 \end{bmatrix} = T_C^W \begin{bmatrix} v^C \\ 1 \end{bmatrix} \quad (3-13)$$

式中 T_C^W 被记作变换矩阵。

3.3.3 IMU 测量模型

IMU 一般是三轴陀螺仪以及三轴加速度计组成，可以称作六轴传感器，但是它的测量值比较容易受到零偏和白噪声的影响。在不考虑地球自转的前提下，可以对 MIMU 的输出作出以下简单建模：

$$\tilde{\omega}^B(t) = \omega^B(t) + b_{\omega_i} + n_{\omega} \quad (3-14)$$

$$\tilde{a}^B(t) = a^B(t) + R_W^B g^W + b_{a_i} + n_a \quad (3-15)$$

上式中上标 B 代表 IMU 的输出都是基于载体系的数据，其中 $\tilde{a}^B(t)$ 和 $\tilde{\omega}^B(t)$ 表示在 t 时刻 IMU 的测量值， $a^B(t)$ 和 $\omega^B(t)$ 是载体运动加速度和角速度的实际值， b_{a_i} 和 b_{ω_i} 代表载体运动加速度和角速度的零偏值， n_a 和 n_{ω} 代表测量时存在的高斯白噪声，它们符合高斯分布， $n_a \sim N(0, \sigma_a^2)$ ， $n_{\omega} \sim N(0, \sigma_{\omega}^2)$ 。考虑测量时加速度计值包括了重力的比力，所以重力加速度被舍弃。其中 IMU 零偏值实际上是随机状态，微分遵循高斯分布，有下式：

$$\dot{b}_{a_i} = n_{b_a} \quad n_{b_a} \sim N(0, \sigma_{b_a}^2) \quad (3-16)$$

$$\dot{b}_{\omega_i} = n_{b_{\omega}} \quad n_{b_{\omega}} \sim N(0, \sigma_{b_{\omega}}^2) \quad (3-17)$$

3.4 IMU 预积分

正常情形下，惯性测量单元的采样频率与双目相机的采样频率相差很大，比如说双目相机采样为两位数，但惯性部分采样却可能为三位数或四位数。在将二者的时间起点保持一致，由于 IMU 更高的频率导致相邻帧间 IMU 数据更多。惯性坐标系中帧间图像 IMU 测量结果通过积分的手段就能够知道它在惯性坐标系的位姿信息，根据 (3-11) 以及 (3-12)，存在^[49]：

$$\begin{aligned} p_{B_{k+1}}^W &= p_{B_k}^W + v_{B_k}^W \Delta t_k + \iint_{t \in [t_k, t_{k+1}]} \left(R_{B_t}^W (\tilde{a}^B(t) - b_{a_t}) - g^W \right) dt^2 \\ v_{B_{k+1}}^W &= v_{B_k}^W + \int_{t \in [t_k, t_{k+1}]} \left(R_{B_t}^W (\tilde{a}^B(t) - b_{a_t}) - g^W \right) dt \\ q_{B_{k+1}}^W &= q_{B_k}^W \otimes \int_{t \in [t_k, t_{k+1}]} \frac{1}{2} q_{B_t}^{B_k} \otimes \begin{bmatrix} 0 \\ \tilde{\omega}^B(t) - b_{\omega_t} \end{bmatrix} dt \end{aligned} \quad (3-18)$$

(3-18) 式中初始状态是不易获得的，如果它的状态变量在后续优化过程中发生了变化，就要对积分进行重新计算。为了避免这一现象，将采取预积分^[50]的手段改成从世界坐标系到上一帧的 IMU 坐标系为参考坐标系的过度，如图 3-3 所示，对 k 和 k+1 帧图像中采取 IMU 积分。对 (3-18) 左右同乘 $R_W^{B_k}$ ，有：

$$\begin{aligned} p_{B_{k+1}}^{B_k} &= p_{B_k}^{B_k} + v_{B_k}^{B_k} \Delta t_k - \frac{1}{2} g^{B_k} \Delta t_k^2 + \alpha_{B_{k+1}}^{B_k} \\ v_{B_{k+1}}^{B_k} &= v_{B_k}^{B_k} - g^{B_k} \Delta t_k + \beta_{B_{k+1}}^{B_k} \\ q_W^{B_k} \otimes q_{B_{k+1}}^W &= \gamma_{B_{k+1}}^{B_k} \end{aligned} \quad (3-19)$$

其中预积分项为，

$$\begin{aligned} \alpha_{B_{k+1}}^{B_k} &= \iint_{t \in [k, k+1]} R_{B_t}^{B_k} (\tilde{a}^B(t) - b_{a_t}) dt^2 \\ \beta_{B_{k+1}}^{B_k} &= \int_{t \in [k, k+1]} R_{B_t}^{B_k} (\tilde{a}^B(t) - b_{a_t}) dt \\ \gamma_{B_{k+1}}^{B_k} &= \int_{t \in [k, k+1]} \frac{1}{2} L(\tilde{\omega}^B(t) - b_{\omega_t}) \gamma_{B_t}^{B_k} dt \end{aligned} \quad (3-20)$$

其中 Δt_k 是时间间隔， $L(\tilde{\omega}^B(t) - b_{\omega_t})$ 是左乘矩阵形式。数值积分方法用来解决离散时间问题，经常用到的有龙格库塔法和欧拉法，对于每两帧图像之间的初始状态，把 $\alpha_{B_k}^{B_k}$ 和 $\beta_{B_k}^{B_k}$ 设为零， $\gamma_{B_k}^{B_k}$ 设为单位四元数，这样 $\alpha_{B_{k+1}}^{B_k}$ 、 $\beta_{B_{k+1}}^{B_k}$ 和 $\gamma_{B_{k+1}}^{B_k}$ 就分别代表相邻帧图像间相对的位移、速度和旋转。

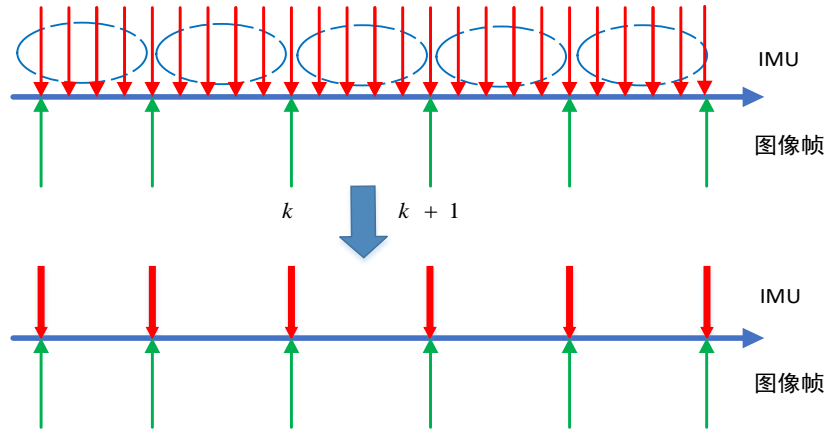


图 3-3 IMU 预积分示意图

预积分与 IMU 以及零偏有关系，与载体状态无关。由于 IMU 的零偏会随着系统的运行产生一些变化，不过因为零偏值较小，将采取一阶近似修正零偏对积分的影响。积分刚开始时是不存在零偏的，会采取视觉数据对零偏进行校准。

本小节也对求解正规方程用到的协方差矩阵得以确定，针对于连续时间系统，得到误差状态方程^[51]如下所示：

$$\begin{aligned}
 \delta \dot{\alpha}_{B_t}^{B_k} &= \delta \beta_{B_t}^{B_k} \\
 \delta \dot{\beta}_{B_t}^{B_k} &= R_{B_t}^{B_k} \left[\tilde{a}_{B_t} - b_{B_{at}} \right]_{\times} \delta \theta_{B_t}^{B_k} - R_{B_t}^{B_k} \delta b_{B_{at}} - R_{B_t}^{B_k} n_a \\
 \delta \dot{\theta}_{B_t}^{B_k} &= - \left[\tilde{\omega}_{B_t} - b_{B_{\omega}} \right]_{\times} \delta \theta_{B_t}^{B_k} - \delta b_{B_{\omega}} - n_{\omega} \\
 \delta \dot{b}_{B_{at}} &= n_{b_a} \\
 \delta \dot{b}_{B_{\omega}} &= n_{b_{\omega}}
 \end{aligned} \tag{3-21}$$

上式 (3-21) 中 $\delta \theta$ 是角度误差，可以写成矩阵形式如下：

$$\begin{aligned}
 \begin{bmatrix} \delta \dot{\alpha}_{B_t}^{B_k} \\ \delta \dot{\beta}_{B_t}^{B_k} \\ \delta \dot{\theta}_{B_t}^{B_k} \\ \delta \dot{b}_{B_{at}} \\ \delta \dot{b}_{B_{\omega}} \end{bmatrix} &= \begin{bmatrix} 0 & I & 0 & 0 & 0 \\ 0 & 0 & R_{B_t}^{B_k} \left[\tilde{a}_{B_t} - b_{B_{at}} \right]_{\times} & -R_{B_t}^{B_k} & 0 \\ 0 & 0 & - \left[\tilde{\omega}_{B_t} - b_{B_{\omega}} \right]_{\times} & 0 & -I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta \alpha_{B_t}^{B_k} \\ \delta \beta_{B_t}^{B_k} \\ \delta \theta_{B_t}^{B_k} \\ \delta b_{B_{at}} \\ \delta b_{B_{\omega}} \end{bmatrix} \\
 &+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ -R_{B_t}^{B_k} & 0 & 0 & 0 \\ 0 & -I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \begin{bmatrix} n_a \\ n_{\omega} \\ n_{b_a} \\ n_{b_{\omega}} \end{bmatrix}
 \end{aligned} \tag{3-22}$$

将上式 (3-22) 简化为:

$$\delta \dot{z}_t^{B_k} = F_t \delta z_t^{B_k} + G_t n_t \quad (3-23)$$

在离散时间系统下也可以写作:

$$\delta z_{t+\Delta t}^{B_k} = (I + F_t \Delta t) \delta z_t^{B_k} + (G_t \Delta t) n_t \quad (3-24)$$

Δt 为连续两个 IMU 测量值的时间间隔。将协方差初始值定为 0, 通过式 (3-23) 进行迭代就可以求得状态协方差矩阵:

$$P_{t+\Delta t}^{B_k} = (I + F_t \Delta t) P_t^{B_k} (I + F_t \Delta t)^T + (G_t \Delta t) Q (G_t \Delta t)^T \quad t \in [k, k+1] \quad (3-25)$$

其中 Q 表示噪声矩阵, $Q = \text{diag}(\sigma_a^2, \sigma_\omega^2, \sigma_{b_a}^2, \sigma_{b_\omega}^2)$ 。

与此同时, $\delta z_{B_{k+1}}^{B_k}$ 相对于 $\delta z_{B_k}^{B_k}$ 的雅克比矩阵可由下式迭代得到, 将雅克比矩阵初值设为 $J_{B_k} = I$:

$$J_{t+\Delta t} = (I + F_t \Delta t) J_t \quad (3-26)$$

矩阵用在 IMU 零偏变化时修正。

3.5 视觉惯性的松耦合

普遍系统静止时, VINS 才开始初始化。无论是陀螺仪零偏还是加速度计零都可用标定获得, 载体的俯仰角和滚转角可以通过加速度计的数据和重力向量计算得到, 而对于三维位置和航向角则可以把它们设置为零。但是这种方式的缺点是系统在开始阶段不能是运动状态, 因为如此一来非线性优化不再准确。本课题将使用在线初始化, 确定初值属于松耦合, 为了保证 IMU 在初始化阶段数据是可观的, 应该要保证载体具有一定的线速度。

IMU、双目、世界坐标系的关系图为:

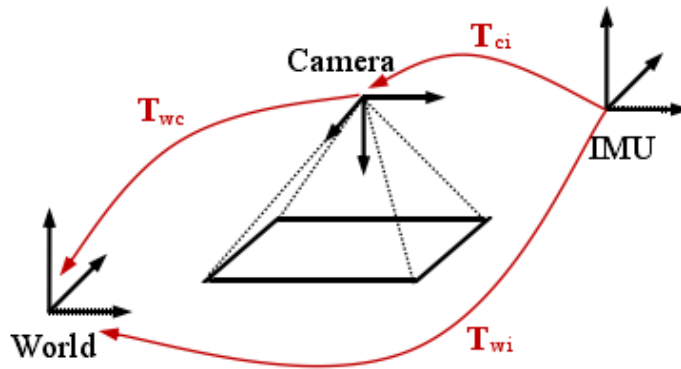


图 3-4 坐标系融合示意图

$T_{wc} = (R_{wc}, t_{wc})$ 表示基坐标系到世界坐标系的变换矩阵。本文采取的数据融合是基于松耦合合法的基础上进行融合的，松耦合把两个系统都视为互相独立的，求出相机位姿后再与 IMU 测量值采取融合策略。它的核心状态为：

$$x_{core}^T = [p_{iw}^T, v_{iw}^T, q_{iw}^T, b_{\omega}^T, b_a^T] \quad (3-27)$$

式 (3-26) 用下标 iw 表示世界坐标系到 IMU 坐标系的转换， p_{iw} 表示世界坐标系原点相对于 IMU 坐标系的位置， v 、 q 表示速度以及姿态， b_{ω} 和 b_a 表示陀螺仪和加速度计偏差。最终运动方程如下：

$$\begin{cases} p_k = p_{k-1} + v_k \Delta \tau + \frac{1}{2} a_k \Delta \tau^2 \\ v_k = v_{k-1} + a_k \Delta \tau \\ q_k = q_{k-1} + q_k \Delta \tau = q_{k-1} + \left(\frac{1}{2} q_k \otimes \omega_k \right) \Delta \tau \end{cases} \quad (3-28)$$

根据式(3-27)，本文把状态误差向量表示下面这种形式为：

$$\tilde{x}_{core}^T = [\Delta p_{iw}^T, \Delta v_{iw}^T, \delta \Theta_{iw}^T, \Delta b_{\omega}^T, \Delta b_a^T] \quad (3-29)$$

式 (3-28) 中 \tilde{x} 表示估计值 \hat{x} 与真实值 x 二者的差值，再定义四元数 $\delta q = q \otimes \hat{q} \approx [\frac{1}{2} \delta \Theta^T \quad 1]^T$ ，融合后系统状态为 IMU 和视觉里程计状态集合，用下面表示：

$$\tilde{x} = [\tilde{x}_{core}, \tilde{x}_{vis}] \quad (3-30)$$

$$\tilde{x}_{vis}^T = [\Delta \lambda, \delta \Theta_{ci}^T, \Delta p_{vw}^T, \delta \Theta_{cw}^T] \quad (3-31)$$

式 (3-30) 表示的是额外状态的误差向量。相机姿态测量模型为：

$$z_{vis} = \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} C_{q_{cw}} (p_{iw} + C_{q_{iw}}^T p_{ci}) \lambda + p_{cw} + n_{p_c} \\ q_{ci} \otimes q_{iw} \otimes q_{cw} + n_{q_c} \end{bmatrix} \quad (3-32)$$

式 (3-32) 中 $C_{q_{iw}}$ 是从世界到 IMU 坐标系的旋转矩阵， $C_{q_{cw}}$ 和 p_{cw} 表示世界到相机系的旋转矩阵和平移向量， n_{p_c} 和 n_{q_c} 为噪声会产生影响。测量误差表示成：

$$\begin{bmatrix} \tilde{z}_p \\ \tilde{z}_q \end{bmatrix} = \begin{bmatrix} C_{q_{cw}}^T (p_{iw} + C_{q_{iw}}^T) \lambda + p_{cw} - (C_{\hat{q}_{wc}}^T (p_{iw} + C_{\hat{q}_{iw}}^T) \hat{\lambda} + \hat{p}_{cw}) \\ \otimes q_{wc} \otimes (\hat{q}_{ci} \otimes \hat{q}_{iw} \otimes \hat{q}_{wc})^{-1} \end{bmatrix} \quad (3-33)$$

将其线性化，得：

$$\tilde{z} = H_{cis} \tilde{x} + \eta \quad (3-34)$$

再通过 EKF 得到更新方程：

$$\tilde{z}_k = H_{k+m} \tilde{x}_{k+m|k} + H_k \tilde{x}_{k|k} + \eta \quad (3-35)$$

式(3-35)中下标 $k+m|k$ 意思是时间点 k 是对于 $k+m$ 的预测，和 \tilde{x} 表示状态量。基于框架能够求取出定值对应的过去状态：

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{k|k}^T & \tilde{x}_{k+m|k}^T \end{bmatrix}^T \quad (3-36)$$

式(3-36)所对应的协方差矩阵如下所示：

$$\tilde{P}_{k+m|k} = \begin{bmatrix} P_{k|k} & P_{k|k} F_{k+m,k}^T \\ F_{k+m,k} P_{k|k} & P_{k+m|k} \end{bmatrix} \quad (3-37)$$

式(3-37)中 $F_{k+m,k} = \prod_{i=1}^m F_{k+i}$ ，它表示具体化的系统动态矩阵。增量差值，协方差阵用下面表示为：

$$\begin{cases} r_{k+m} = z_{k,k+m} - \hat{z}_{k,k+m} \simeq \tilde{H} \tilde{X} \\ S_{k+m} = \tilde{H} \tilde{P}_{k+m|k} \tilde{H}^T + R_r \end{cases} \quad (3-38)$$

式(3-35)中 R_r 代表着视觉 SLAM 位姿的协方差矩阵， $\tilde{H} = \begin{bmatrix} H_{k|k} & H_{k+m|k} \end{bmatrix}$ 包括两个所对应的雅克比矩阵，那么它相对应的卡尔曼增益为：

$$\tilde{K} = \tilde{P}_{k+m|k} \tilde{H}^T \tilde{S}_{k+m}^{-1} = \begin{bmatrix} K_k^T & K_{k+m}^T \end{bmatrix}^T \quad (3-39)$$

系统最优估计方程以及协方差矩阵的更新方程如下式所示：

$$\hat{x}_{k+m|k+m} = \hat{x}_{k+m|k} + K_{k+m} r_{k+m} \quad (3-40)$$

$$P_{k+m|k+m} = P_{k+m|k} - K_{k+m} \tilde{S}_{k+m} K_{k+m}^T \quad (3-41)$$

至此，已经获得了融合后状态估计值，载体位姿状态优化完毕。

3.5.1 陀螺仪零偏的修正

首先进行陀螺仪零偏的计算，连续两帧图像 B_k 和 B_{k+1} 过度的旋转关系可以由 IMU 预积分或者 SFM 获得。正常情况下，无误差的前提下二者值应相等。事实上却有差值，我们将差值最小化来对陀螺仪零偏进行初步校准。根据前面所得到的数据，有：

$$\min_{\delta b_\omega} \sum_{b \in \Gamma} \|q_{B_{k+1}}^{C_0}{}^{-1} \otimes q_{B_k}^{C_0} \otimes \gamma_{B_{k+1}}^{B_k}\|^2 \quad (3-42)$$

其中， Γ 代表在滑动窗口中全部图像帧。 $\gamma_{B_{k+1}}^{B_k}$ 为陀螺仪零偏的一阶泰勒展开：

$$\gamma_{B_{k+1}}^{B_k} \approx \hat{\gamma}_{B_{k+1}}^{B_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} J_{b_w}^\gamma \delta b_\omega \end{bmatrix} \quad (3-43)$$

式 (3-26) 展示的代价函数有最小值, 最小值为单位四元数, 与 (3-27) 联立可得:

$$\hat{\gamma}_{B_{k+1}}^{B_k} \otimes q_{B_k}^{C_0-1} \otimes q_{B_{k+1}}^{C_0} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} J_{b_w}^\gamma \delta b_\omega \end{bmatrix} \quad (3-44)$$

在只关心虚部的情况下, 有:

$$J_{b_w}^{\gamma T} J_{b_w}^\gamma \delta b_\omega = 2 J_{b_w}^{\gamma T} \left(\hat{\gamma}_{B_{k+1}}^{B_k}{}^{-1} \otimes q_{B_k}^{C_0-1} \otimes q_{B_{k+1}}^{C_0} \right)_{xyz} \quad (3-45)$$

通过滑窗中相邻帧求解到超定方程, 对这个方程求解最小二乘解。在以后的非线性优化中也采取针对于的平移和速度量的泰勒展开:

$$\alpha_{B_{k+1}}^{B_k} \approx \hat{\alpha}_{B_{k+1}}^{B_k} + J_{b_a}^\alpha \delta b_{ak} + J_{b_\omega}^\alpha \delta b_{\omega k} \quad (3-46)$$

$$\beta_{B_{k+1}}^{B_k} \approx \hat{\beta}_{B_{k+1}}^{B_k} + J_{b_a}^\beta \delta b_{ak} + J_{b_\omega}^\beta \delta b_{\omega k} \quad (3-47)$$

以后都将用这个方式对获得的了 IMU 新的零偏进行修正。

3.5.2 系统初始化

不光是零偏要确定, 初始化过程也需要确定载体的速度、绝对尺度和重力加速度。与前面一样, 依旧对连续的两个图像帧进行计算处理, 二者对应的速度、平移能从 IMU 预积分获得。式 (3-20) 中 $\alpha_{B_{k+1}}^{B_k}$ 由 IMU 预积分计算得到, 而 $p_{B_{k+1}}^{B_k}$ 和 $p_{B_k}^{B_k}$ 的具体值是未知的, 与 (3-16) 联立有:

$$\alpha_{B_{k+1}}^{B_k} - R_{C_0}^{B_k} \left[s \left(\bar{p}_{B_{k+1}}^{C_0} - \bar{p}_{B_k}^{C_0} \right) + R_{B_k}^{C_0} v_{B_k}^{B_k} \Delta t_k - \frac{1}{2} g^{C_0} \Delta t_k^2 \right] \approx 0 \quad (3-48)$$

同理, 联立式 (3-15) 有:

$$\beta_{B_{k+1}}^{B_k} - R_{C_0}^{B_k} \left(R_{B_{k+1}}^{C_0} v_{B_{k+1}}^{B_k} \Delta t_k + R_{B_k}^{C_0} v_{B_k}^{B_k} - g^{C_0} \Delta t_k \right) \approx 0 \quad (3-49)$$

将初值写成列向量的形式, 则可以将上述两式写为矩阵形式:

$$\begin{bmatrix} \beta_{B_{k+1}}^{B_k} \\ \alpha_{B_{k+1}}^{B_k} \end{bmatrix} - \begin{bmatrix} -I_{3 \times 1} & R_{C_0}^{B_k} R_{B_{k+1}}^{C_0} & R_{C_0}^{B_k} \Delta t_k & 0 \\ -I_{3 \times 1} \Delta t_k & 0 & \frac{1}{2} R_{C_0}^{B_k} \Delta t_k^2 & R_{C_0}^{B_k} \left(\bar{p}_{B_{k+1}}^{C_0} - \bar{p}_{B_k}^{C_0} \right) \end{bmatrix} \begin{bmatrix} v_{B_k}^{B_k} \\ v_{B_{k+1}}^{B_k} \\ g^{C_0} \\ s \end{bmatrix} \approx 0 \quad (3-50)$$

把 (3-50) 简写为: $Y_{B_{k+1}}^{B_k} - A_{B_{k+1}}^{B_k} X \approx 0$, 考虑多帧图像之间, 可将问题转化成解线性最小二乘方程:

$$X = \arg \min_X \sum_{k \in \Gamma} \|Y_{B_{k+1}}^{B_k} - A_{B_{k+1}}^{B_k} X\|^2 \quad (3-51)$$

基于这个方法在载体系下滑窗中每一图像帧表示的运动速度 $v_{B_k}^{B_k}$ 都可以求得，而在双目视觉参考系下重力向量 g^{C_0} 以和尺度因子 s 也都可以求得。

如下图 3-5 所示，求取到重力后便把视觉系转为惯性系。具体如下：世界坐标系的重力方向竖直向下，但是在参考系中重力的方向 $\bar{g}^{C_0} = g^{C_0} / \|g^{C_0}\|$ ，为了更好的修正重力，引入世界坐标系下重力大小 $|g| \approx 9.8m/s^2$ 的约束。确定 \bar{g}^W 和 \bar{g}^{C_0} 夹角便得到旋转矩阵 $R_{C_0}^W$ 。 $R_{C_0}^W$ 把物理量转变到惯性坐标系下，速度转换到惯性坐标系下。航向角的变化不影响重力向量三个投影的，所以重力使俯仰角和滚转角具有可观性。

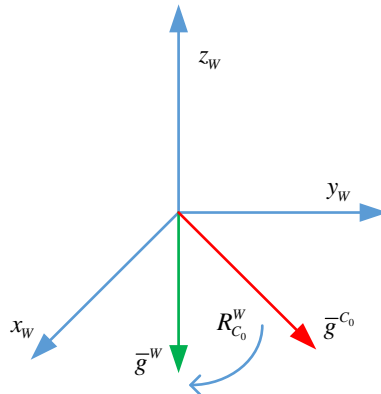


图 3-5 视觉坐标系到世界坐标系的转换

到此为止，初始化的部分已经完毕，所得到的物理量为后续的紧耦合优化的做了铺垫。图 3-6 中，五角星表示路标点，线段表示镜头到路标点的观察路线，红色虚线表示预积分过程。对 IMU 提供的绝对尺度以及重力方向和视觉部分测得的无漂移的位子估计进行融合，最终得到 3-6 惯性视觉对齐后的状态。

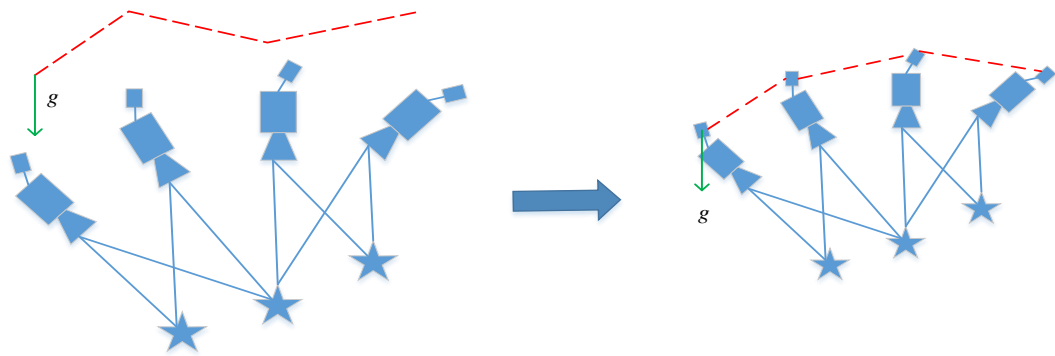


图 3-6 视觉-惯性导航系统初始化

3.6 本章小结

本章建立了整体系统框架模型，给出了双目视觉导航系统中相机坐标系到世界坐标系的转换方程，给出并对 IMU 的测量数据进行了预积分处理，获得了图像帧之间的位姿和非线性优化迭代的初值，确定特征点的三维坐标，并将双目视觉里程计与 IMU 预积分处理后的数据求得载体位姿后进行松耦合，构建线性最小二乘方程组，确定非线性优化所需要的初始速度、陀螺仪的零偏、以及重力方向。

第 4 章 双目视觉与 IMU 的紧耦合优化方法

4.1 引言

本章建立视觉/惯性组合导航定位系统的多种约束关系因子图，采取紧耦合方法对双目视觉与 IMU 进行优化。系统的状态约束最终通过因子图^[52]的手法给出，对需要优化的状态变量进行非线性优化方程的创建。采用基于词袋模型的回环检测去消除状态量随着时间积累的累积误差，从而完成重定位，采取全局位姿优化保证了全局一致性。最后提出纯旋转的情况，提供在纯旋转情况中如何处理视觉惯性里程计不能正常工作的解决办法。

4.2 视觉-惯性紧融合算法

由于约束条件不同，导致视觉惯性里程计和视觉实时构图定位不一样，它由于 IMU 的引入使得除了最开始的图像帧与路标点的约束外，仍有两个约束，它们为：IMU 预积分的速度和零偏带来的随机游走过程约束。

基于上述约束关系可以构建因子图，因子图如（4-1）所示，它由因子和变量组成。其中的圆节点是未估计状态量；方节点表示约束关系，因子图中因子就是方框节点。

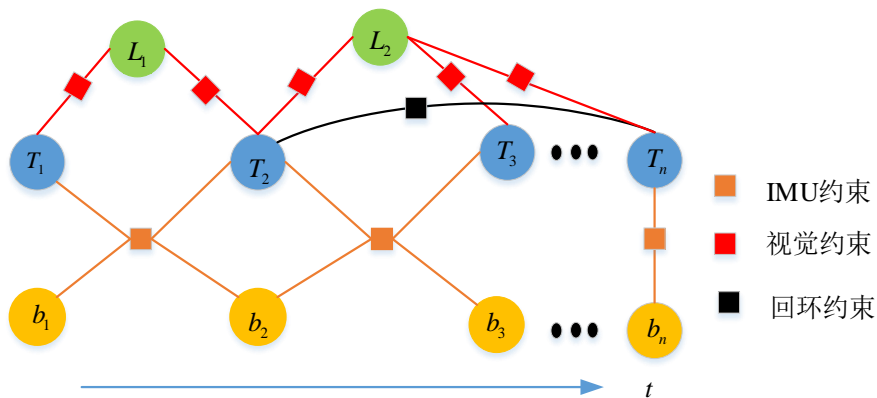


图 4-1 视觉-惯性导航系统因子图

把它融汇到视觉惯性导航系统中，蓝色节点、黄色节点，绿色节点分别代表相机位姿、IMU 测量值和路标点；红色方框表示图像帧与路标点的约束外，黄色方框表示 IMU 约束，黑色方框回环检测系统的回环约束。由于因子对应非线性最小二乘的误差项，可以把问题等同于非线性优化问题，每个约束都是一个目标函数，

从而开始最大后验估计。

系统需要优化的状态量有：

$$\begin{aligned}\chi &= [x_{B_0}^W, x_{B_1}^W, \dots, x_{B_n}^W, x_C^B, L] \\ x_{B_k}^W &= [p_{B_k}^W, v_{B_k}^W, q_{B_k}^W, b_a, b_\omega], k \in [0, n] \\ x_C^B &= [p_C^B, q_C^B] \\ L &= \left[\frac{1}{\lambda_0}, \frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_m} \right]\end{aligned}\quad (4-1)$$

其中，在 k 时刻，除了 IMU 的零偏是在载体系外表示的， $x_{B_k}^W$ 剩余状态都在 W 中表示；其中 x_C^B 是进行估计的外参，选择的原因是当载体运动时存在微惯性器件和相机摇晃的影响； L 表示逆深度集合。

整个系统中误差被滑窗中别的图像帧约束成三种，分别是 IMU 预积分误差、视觉重投影误差以及先验信息。为了获取系统状态的最优估计可以采取将各测量误差马氏距离的平方和进行最小化：

$$\chi^* = \arg \min_{\chi} \left\{ (b_p - \Lambda_p \chi) + \sum_{k \in \Gamma} \left\| e_{\Gamma}(\hat{z}_{B_{k+1}}^{B_k}, \chi) \right\|_{P_{B_{k+1}}^{B_k}}^2 + \sum_{(l,j) \in C} \left\| e_C(\hat{z}_l^{C_j}, \chi) \right\|_{P_l^{C_j}}^2 \right\} \quad (4-2)$$

这里面 Λ_p 是先验信息， Γ 代表全部的惯性数据，因为只有多次观测才能进行重投影，所以 C 代表了在窗口中多次被观测的路标点的集合，它们所对应的残差定义为 $e_{\Gamma}(\hat{z}_{B_{k+1}}^{B_k}, \chi)$ 和 $e_C(\hat{z}_l^{C_j}, \chi)$ ， P 则表示协方差矩阵，它对两个传感器的置信度起着决定性的作用， $(\hat{\cdot})$ 代表估计值。

用 IMU 误差来推出其正规迭代方程，如下所示：

$$\min_{\Delta \chi} \sum_{k \in \Gamma} \left\| e_{\Gamma}(\hat{z}_{B_{k+1}}^{B_k}, \chi + \Delta \chi) \right\|_{P_{B_{k+1}}^{B_k}}^2 = \min_{\Delta \chi} \sum_{k \in \Gamma} \left\| e_{\Gamma}(\hat{z}_{B_{k+1}}^{B_k}, \chi) + J_{B_{k+1}}^{B_k} \Delta \chi \right\|_{P_{B_{k+1}}^{B_k}}^2 \quad (4-3)$$

上式中其实就是 (4-2) 中间一项的迭代写法。其中 $J_{B_{k+1}}^{B_k}$ 代表 e 对待优化变量的雅克比矩阵，展开后使增量 $\Delta \chi$ 的一阶导数等于 0，就有：

$$J_{B_{k+1}}^{B_k} {}^T P_{B_{k+1}}^{B_k} {}^{-1} J_{B_{k+1}}^{B_k} \Delta \chi = -J_{B_{k+1}}^{B_k} {}^T P_{B_{k+1}}^{B_k} {}^{-1} e_{\Gamma} \quad (4-4)$$

那么可以把式 (4-2) 重写成：

$$(\Lambda_p + \Lambda_{\Gamma} + \Lambda_C) \Delta \chi = (b_p + b_{\Gamma} + b_C) \quad (4-5)$$

把 (4-6) 记为状态量更新：

$$\chi = \tilde{\chi} \oplus \Delta \chi \quad (4-6)$$

这里面，可以把速度、位置以及逆深度近似写成：

$$\begin{aligned} p &= \check{p} + \Delta p \\ v &= \check{v} + \Delta v \\ \lambda &= \check{\lambda} + \Delta \lambda \end{aligned} \quad (4-7)$$

因为旋转不是加法计算过程，所以更繁琐，最后通过旋转流型切空间的扰动 Δq 实现更新：

$$q = \check{q} \otimes \Delta q \quad \Delta q \approx \begin{bmatrix} 1 \\ \frac{1}{2} \Delta \theta \end{bmatrix} \quad (4-8)$$

$\Delta \theta$ 是李代数形式。旋转阵也写作：

$$R = \check{R} \cdot (I + \Delta \theta^\wedge) \quad (4-9)$$

接下来分别介绍这三项误差的表示形式，以及对于状态变量的雅克比矩阵计算。

4.2.1 IMU 误差分析

在滑动窗口中针对所相邻的两帧图像 B_k 、 B_{k+1} 中间关于 IMU 数值的预积分分存在的误差由下式 (4-10) 表示：

$$e_\Gamma(\hat{z}_{B_{k+1}}^{B_k}, \chi) = \begin{bmatrix} \delta \alpha_{B_{k+1}}^{B_k} \\ \delta \beta_{B_{k+1}}^{B_k} \\ \delta \theta_{B_{k+1}}^{B_k} \\ \delta b_a \\ \delta b_\omega \end{bmatrix} = \begin{bmatrix} R_W^{B_k} \left(p_{B_{k+1}}^W - p_{B_k}^W + \frac{1}{2} g^W \Delta t_k^2 - v_{B_k}^W \Delta t_k \right) - \hat{\alpha}_{B_{k+1}}^{B_k} \\ R_W^{B_k} \left(v_{B_{k+1}}^W + g^W \Delta t_k - v_{B_k}^W \right) - \hat{\beta}_{B_{k+1}}^{B_k} \\ 2 \left[q_{B_k}^{W-1} \otimes q_{B_k}^W \otimes \hat{\gamma}_{B_{k+1}}^{B_k-1} \right]_{xyz} \\ b_{aB_{k+1}} - b_{aB_k} \\ b_{\omega B_{k+1}} - b_{\omega B_k} \end{bmatrix} \quad (4-10)$$

其中 $\hat{\alpha}_{B_{k+1}}^{B_k}$ 、 $\hat{\beta}_{B_{k+1}}^{B_k}$ 和 $\hat{\gamma}_{B_{k+1}}^{B_k}$ 代表含有噪声的陀螺仪和加速度计测量值经过预积分处理所得， $\delta \theta_{B_{k+1}}^{B_k}$ 代表四元数虚部数值，也就是角度误差。

通过下式可以得到 IMU 误差对于状态变量所表示的 J ：

$$J_{imu} = \begin{bmatrix} \frac{\partial e_\Gamma}{\partial \chi_k} & \frac{\partial e_\Gamma}{\partial \chi_{k+1}} \end{bmatrix}^T \quad (4-11)$$

其中， $\frac{\partial e_\Gamma}{\partial \chi_k} = \begin{bmatrix} \frac{\partial e_\Gamma}{\partial p_{B_k}^W} & \frac{\partial e_\Gamma}{\partial v_{B_k}^W} & \frac{\partial e_\Gamma}{\partial R_{B_k}^W} & \frac{\partial e_\Gamma}{\partial b_{aB_k}} & \frac{\partial e_\Gamma}{\partial b_{\omega B_k}} \end{bmatrix}^T$ ， $\frac{\partial e_\Gamma}{\partial \chi_{k+1}}$ 它的格式一致。接下来分

别把这两个部分的雅克比矩阵列出：

若有 $\frac{\partial e_\Gamma}{\partial \chi_k} = [A_k \quad B_k]$ ，则：

$$A_k = \begin{bmatrix} -R_W^{B_k} & -R_W^{B_k} \Delta t_k & \left(R_W^{B_k} \left(p_{B_{k+1}}^W - p_{B_k}^W + g^W \frac{\Delta t_k^2}{2} \right) \right)^\wedge \\ 0 & -R_W^{B_k} & \left(R_W^{B_k} \left(v_{B_{k+1}}^W - v_{B_k}^W + g^W \Delta t_k \right) \right)^\wedge \\ 0 & 0 & -L \left(q_{B_{k+1}}^{W-1} \otimes q_{B_k}^W \right) R \left(\hat{\gamma}_{B_{k+1}}^{B_k} \right) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4-12)$$

$$B_k = \begin{bmatrix} -J_{b_{aB_k}}^\alpha & -J_{b_{\omega B_k}}^\alpha \\ 0 & -L \left(q_{B_k}^{W-1} \otimes q_{B_k}^W \otimes \hat{\gamma}_{B_{k+1}}^{B_k-1} \right) J_{b_{\omega B_k}}^\gamma \\ -J_{b_{aB_k}}^\beta & J_{b_{\omega B_k}}^\beta \\ -I & 0 \\ 0 & -I \end{bmatrix} \quad (4-13)$$

若 $\frac{\partial e_\Gamma}{\partial \chi_{k+1}} = [A_{k+1} \ B_{k+1}]$, 则:

$$A_{k+1} = \begin{bmatrix} R_W^{B_k} & 0 & 0 \\ 0 & 0 & L \left(q_{B_{k+1}}^W \hat{\gamma}_{B_{k+1}}^{B_k-1} \otimes q_{B_k}^{W-1} \otimes q_{B_{k+1}}^W \right) \\ 0 & R_W^{B_k} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4-14)$$

$$B_{k+1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \quad (4-15)$$

4.2.2 滑动窗口边缘化

因为系统对于实时性要求比较高,且载体处理器计算资源有限,想对所有时刻的相机位姿和路标点进行优化是很困难的,所以要选择一定的相机状态和特征点进行滑窗边缘化,此举可作为计算量优化。时间逐步累计,去除古老帧就变得理所当然,并逐步加入当前帧进行非线性优化。考虑到 B_o 在经历预积分处理后,在路标点对比后与 B_n 生成约束,若抛弃原始帧可能会引起这一部分量测得信息缺失,会对位姿估计结果产生较大影响。所以选择采取约束信息变成状态先验信息参与

优化。如果要对 x_o 边缘化处理，在滑窗中应去除状态量 x_o ；设定待求变量是 x_n ，它对应窗口全部状态量^[53]。窗口没滑动前，使用高斯-牛顿法基于 δx 更新 x ，这样会更精确。但此时需要丢掉 x_o ，只估计 x_n ，此时的正规方程为：

$$\begin{bmatrix} \Lambda_o & \Lambda_{on} \\ \Lambda_{on}^T & \Lambda_n \end{bmatrix} \begin{bmatrix} \delta x_o \\ \delta x_n \end{bmatrix} = \begin{bmatrix} b_o \\ b_n \end{bmatrix} \quad (4-16)$$

进行高斯消元，把上式变为：

$$\begin{bmatrix} \Lambda_o & \Lambda_{on} \\ 0 & \Lambda_n - \Lambda_{on}^T \Lambda_o^{-1} \Lambda_{on} \end{bmatrix} \begin{bmatrix} \delta x_o \\ \delta x_n \end{bmatrix} = \begin{bmatrix} b_o \\ b_n - \Lambda_{on}^T \Lambda_o^{-1} b_o \end{bmatrix} \quad (4-17)$$

其中 $\Lambda_{on}^T \Lambda_o^{-1} \Lambda_{on}$ 被称作 Λ_o 在 Λ_{on} 中的舒尔补，则滑动窗口中的状态量 x_n 则由 (4-18) 计算求得：

$$(\Lambda_n - \Lambda_{on}^T \Lambda_o^{-1} \Lambda_{on}) \delta x_n = b_n - \Lambda_{on}^T \Lambda_o^{-1} b_o \quad (4-18)$$

上式由式 (4-17) 转化而来，其中 $\Lambda_n - \Lambda_{on}^T \Lambda_o^{-1} \Lambda_{on}$ 被称作先验信息，尽管它去掉了某一帧图像但它的约束仍被保留下，所以用此方法求取的 x_n 不会存在缺少约束的情况。

本文采取的滑窗边缘化方法是视觉 SLAM 当前一种主流处理手段，用最新图像帧的上一帧是否为关键帧对滑动窗口中的某个图像帧进行取舍，基本可以理解成倘若次新帧为关键帧则保留，就把最老帧和它的 IMU 值移除；反之就移除次新帧视觉数据，但惯性部分 IMU 测量值会被保留，为后面的预积分过程做准备。采取这个滑窗策略对滑窗中图像帧之间存在足够视差来对特征点的三角化，对系统初始化起到了保障作用，此外最新的特征点对于特征跟踪是非常重要的。

4.2.3 载体位姿优化

在 VINS 里，因为 IMU 加速度计测量了重力，导致俯仰角、滚转角以及尺度都是在惯性系下展示，而三维位置和航向角是在参考系展示，如此就会导致航向角并不是绝对的，如此一来就会在四自由度带来累积误差。本文引入回环检测部分为了提高系统精度和鲁棒性并且减小位姿估计的漂移。回环检测是系统检测到历过的场景，就会用回环约束去修改位姿优化结果，也可以当遭遇视线遮挡时系统轨迹丢失，去依靠回环检测完成重定位。回环检测构建全局一致性轨迹、重新定位时尤为关键，一个完备的 SLAM 系统必须有着回环检测和全局位姿图优化功能。回环检测的核心技术是对两帧图像之间的特征描述和匹配，准确检测图像回环是至关重要的，主要方法包括词袋模型、暴力匹配法和机器学习法。回环检测独立存在于 VINS 中，不依赖于 IMU 的测量值，只与视觉信息有关联。

4.3 回环检测

在视觉 SLAM 领域，很多研究比如 VINS-Mono、ORB-SLAM 等都采取了用词袋模型对系统进行回环检测，特征匹配采取 Kmean++ 算法，这样一来，视觉单词就会产生，接下来视觉单词从待检索图像中的某些特征点进行选取，再在词袋中依次求相似度，最终比较出相似度最高的作为回环帧。本文采用的是 DBoW2^[54] 作为词袋模型，来保证视觉惯性导航系统能够快速检测到回环帧。

在对图像进行比较时，提取图像特征点耗时比较大，为了提高系统的实时性，并保证效率，运算速度快、占空间小的 BRIEF 描述子和 FAST 关键点就是视觉字典的选取标准。前端部分应用 ORB 算法，实现回环检测的要求，可以求出足够的 FAST 关键点并计算其周围的 BRIEF 描述子。随着系统运行时间的加长，有越来越多的图像被采集，全部图像帧都加入字典是大可不必的，回环检测也不现实。原因有以下两个方面，一是过近的图像帧之间会有不正确的回环约束；另外就是存在多帧图像与之前同一个历史帧图像存在回环约束，实际上利用其中的一帧即可，所以关键帧如何挑选就变得尤为关键。本文设定初始 50 个关键帧不具备回环信息，接下来按照图 4-2 所示的流程建立视觉字典。

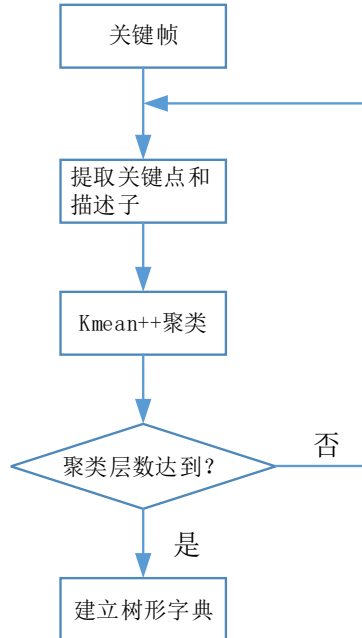


图 4-2 视觉字典创建

在后面的过程中，若新图像是关键帧就会进行回环检测，检测到回环就进行重定位以及位姿优化，没检测到就添加至字典里。首先会把新图像转成词袋向量，再根据下图步骤开始检索，通过建立 K 叉树存储模型，里面每一个底层节点都是一

个视觉单词，检索方式分为正向和反向。这里面反向检索表示某个单词所在的全部关键帧信息和它所在的图像的权重，这样就使得在检测过程中便于更快的寻找到存

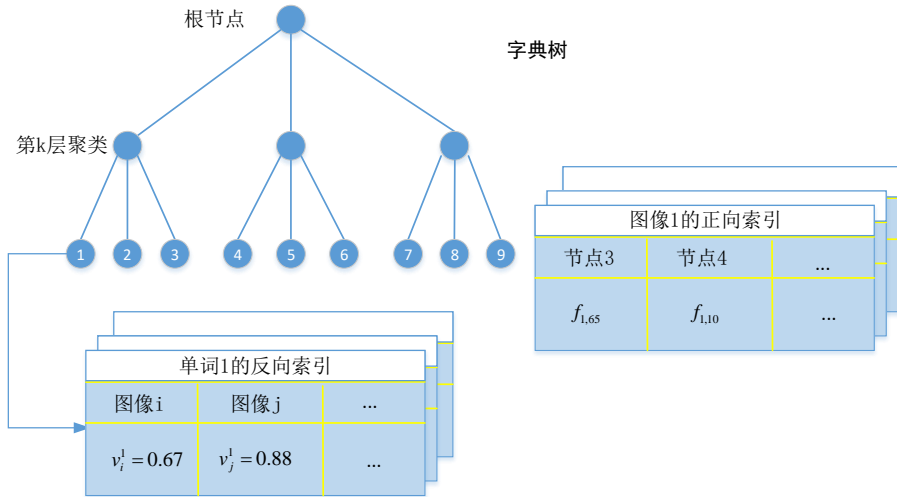


图 4-3 视觉字典检索

在共同特征的图像帧，再进行相似度的计算来选出后选帧图像，反向索引在数据库搜索图像时会被访问；当字典中添加了新的图像帧的时候，它所具备的每个单词的反向索引就会变化。正向索引包含了每一个图像的特征信息，还有它的相关联节点的特征，它有着加速回环检测中匹配图像对的几何一致性检测的过程。当数据库中出现了被新添加的图像后，会更新直接索引，候选匹配帧被检验代表访问^[55]直接索引。

回环被检测到后应该对前后图像帧的关系进行建立，查看是否匹配，这个时候就不能用 KLT 光流法对特征追踪进行操作了。关于前后帧特征点对应关系考虑采用 BRIEF 算子计算二者的汉明距离，但是缺点是采取这个办法会得到一些误匹配情况，通过回环检测后相对位姿大致被复原，为了接下来的一致性测验做准备。如果在外点已经被剔除情况下选取的两帧图像依旧存留满足数量的特征匹配点则说明几何检验已经通过，那么就此时的图像候选帧认定为回环帧接着下一节的操作。

4.3.1 视觉位姿优化

回环检测可以快速地将系统运动的丢失恢复到正确的位姿优化过程中，这就是所谓的重定位，而且回环检测可以修正累积误差导致的漂移位姿轨迹。图 4-4 描述了当遇到连续跟踪失败情况下，此时的重定位流程，其中蓝线表示未丢失轨迹，紫线表示当前轨迹，绿线表示回环约束，橙线代表 IMU 预积分与视觉观察的约束，

虚线表示暂时丢失。由此可见丢失会使得此时位姿估计产生较大的误差，为了使系统回到准确位置，采用回环检测并结合后文的优化方法对当前位姿进行优化，进而接着进行定位建图。

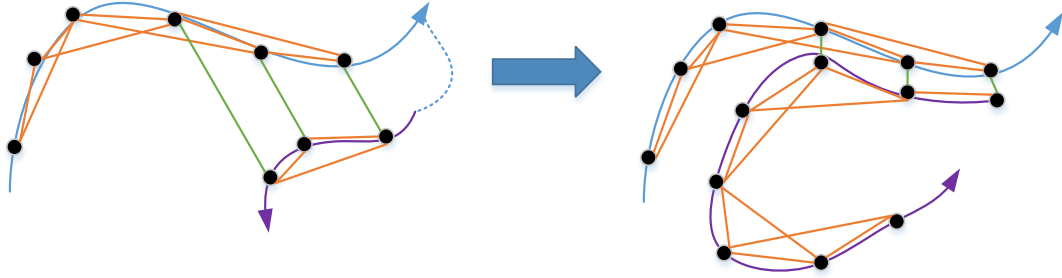


图 4-4 系统跟踪失败后的重定位

重定位不只是在系统丢失之后的才进行的操作，实际上为了减少滑窗中相机状态累积误差，每次检测回环约束，都会进行重定位。假设当前关键帧 m 和之前检测到回环关键帧 v 之间存在回环约束，回环帧位姿是存在记录的，可直接记作 $(\hat{q}_v^w, \hat{p}_v^w)$ 。通过式 (4-10) 一直到 (4-12) 可构建出关键帧 m 和回环关键帧 v 二者的视觉测量残差，在滑窗里对 IMU 值、视觉测量值以及通过回环检测测出的对应关系采取紧耦合式 (4-19) 的非线性优化。

$$\chi^* = \arg \min_{\chi} \left\{ (b_p - \Lambda_p \chi) + \sum_{k \in \Gamma} \left\| e_{\Gamma}(\hat{z}_{B_{k+1}}^{B_k}, \chi) \right\|_{P_{B_{k+1}}^{B_k}}^2 + \sum_{(l,j) \in C} \left\| e_C(\hat{z}_l^{C_j}, \chi) \right\|_{P_l^{C_j}}^2 + \sum_{(l,v) \in \kappa} \left\| e_C(\hat{z}_l^{C_v}, \chi, \hat{q}_v^w, \hat{p}_v^w) \right\|_{P_l^{C_v}}^2 \right\} \quad (4-19)$$

κ 表示匹配的特征点的集合，集合的元素都来自回环图像与当前帧中。最后面记录的是重投影误差项，记录着特征点 l 在回环帧中的检测。考虑到回环帧位姿状态为常数，所以此时的优化对系统状态的维度无影响。倘若在检测时好几个图像帧都测到回环，这样重定位变得繁琐，所以需要采取非线性优化，使系统精度提高，在重定位后，为了保障轨迹平滑还需要对全局位姿优化。

因为系统是一个长期运行的状态，尽管重定位能使得回环过程中消除滑窗状态量的累积误差，但是累积误差会导致某一时刻轨迹远离真实轨迹无法回到闭合回路。如图 4-5 所示，当前帧是黄色的，回环约束用绿色表示，五角星表示路标点。为了使前面提到的累积误差得到纠正，经常采用的是让误差均匀地分配给所有时刻的相机位姿，如此一来控制了误差的问题，使得轨迹的一致性、平滑性性能变好。每一帧都有着上百个路标点会被观测，所以全局 BA 会导致要考虑的路标点过多，

优化所有的相机位姿和坐标点会导致时间多久, 计算量过大。本文中最终舍去路标点的优化, 用下面的位姿优化演示修正轨迹漂移的效果。

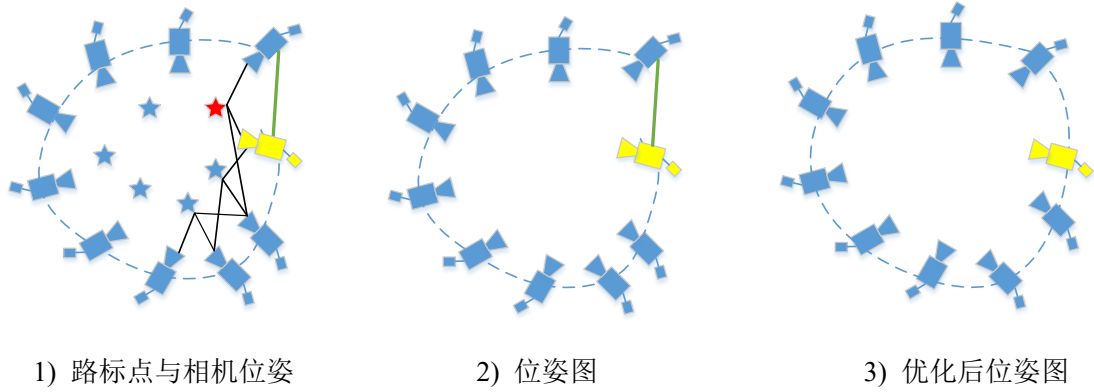


图 4-5 闭合回环

本文只在四自由度上进行全局位姿优化。每次通过边缘化出来的关键帧都被改为待优化状态。如下图 4-6 所示, 位姿图约束关系与因子图很类似, 图中包含两种约束关系, 分别是顺序约束和回环约束。

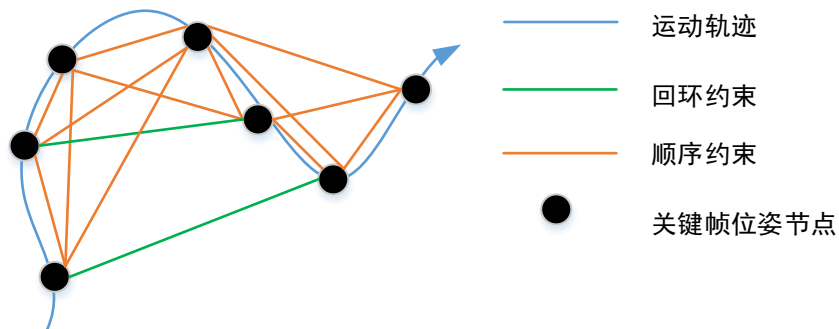


图 4-6 位姿图约束关系

其实无论是哪种约束, 它们的实质都是基于特征匹配得到的关系。以顺序约束举例, 为了计算当前关键帧可采用之前提到的视觉惯性里程计估计结果, 用 i 表示当前关键帧, 历史对应约束关键帧用 j 表示, 它们之间的相对位移 \hat{p}_{ij}^w 和偏航角 $\hat{\psi}_{ij}$ 如下所示:

$$\begin{aligned}\hat{p}_{ij}^i &= R_i^{w-1} (\hat{p}_j^w - \hat{p}_i^w) \\ \hat{\psi}_{ij} &= \hat{\psi}_j - \hat{\psi}_i\end{aligned}\quad (4-20)$$

回环约束相对变换形式是 (4-20) 进行重定位后获得的。

通过测量结果与估计值构造误差方程, 因为横滚角和俯仰角不存在累计误差, 不参与位姿图优化。那么涉及到的两条边的残差如下式所示:

$$e_{i,j}(p_i^w, \psi_i, p_j^w, \psi_j) = \begin{bmatrix} R(\hat{\phi}_i, \hat{\theta}_i, \psi_i)^{-1} (p_j^w - p_i^w) - \hat{p}_{ij}^i \\ \psi_j - \psi_i - \hat{\psi}_{ij} \end{bmatrix} \quad (4-21)$$

上式(4-21)中 $\hat{\phi}_i, \hat{\theta}_i$ 表示通过紧耦合非线性优化估计并固定横滚角和俯仰角, $p_i^w, \psi_i, p_j^w, \psi_j$ 分别表示由两组观测的位置和偏航角, \hat{p}_{ij}^i 和 $\hat{\psi}_{ij}$ 表示式(4-20)的结果。则可以通过最小化下式的代价函数来得到两个约束条件的整个图优化:

$$\arg \min_{p, \psi} \left\{ \sum_{(i,j) \in S} \|e_{i,j}\|^2 + \sum_{(i,j) \in U} \|e_{i,j}\|^2 \right\} \quad (4-22)$$

上面 S 表示顺序约束集合, U 表示回环约束集合。

考虑到系统的实时性,本文选择把重定位和位姿图优化分线程运行,二者分别在独立的线程运行。到此为止,视觉部分与IMU融合的位姿估计结果已全部完成,但是实际运行时会遇到旋转退化的问题造成数据融合困难,如何解决也是一个问题。

4.3.2 纯旋转运动的处理

无人机、无人车等载体有时会存在做纯旋转的动作,这个时候会面临视觉部分深度信息丢失的情况,只靠着IMU预积分去测量姿态角且在MIMU系统存在噪声的干预下,会导致积分存在明显的累积误差,导致姿态估计结果发散,所以采取了基于扩展卡尔曼滤波的方法对二者数据进行融合^[56]。来对载体准确的姿态估计进行求取。对于判断是否做了纯旋转运动采取了对图像进行了旋转补偿后检测视差的手法,若已经产生运动退化,系统就会变为姿态解算状态,接下来当系统正常了,再对视觉惯性紧耦合模块初始化。在偏航角漂移时,必须通过磁强计解决问题,仅依靠初始系统是不能解决漂移问题的。利用四元数作为状态变量,得到状态方程:

$$q_B^w(k+1) = \underbrace{\left(I + \frac{1}{2} R(\tilde{\omega}(t)) \Delta t \right)}_{A_k} q_B^w(k) + w(k) \quad (4-23)$$

上式中, A 为状态转移矩阵, $w(k)$ 表示系统的过程噪声。

根据陀螺仪的角速度读数来进行状态预测、计算协方差矩阵,以进行下一步预测:

$$\hat{q}_k^- = A_k \hat{q}_{k-1} \quad (4-24)$$

$$P_k^- = A_k P_{k-1} A_k^T + Q_{k-1} \quad (4-25)$$

式中 Q_{k-1} 表示过程噪声的协方差矩阵。

前文中知道姿态角的表示可利用四元数法:

$$h(\hat{q}_k^-) = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \arctan\left(-\frac{2q_2q_3 - 2q_0q_1}{1 - 2q_1^2 - 2q_2^2}\right) \\ \arcsin(2q_1q_3 + 2q_0q_2) \\ \arctan\left(-\frac{2q_1q_2 - 2q_0q_3}{1 - 2q_2^2 - 2q_3^2}\right) \end{bmatrix} \quad (4-26)$$

因为 (4-26) 式是非线性方程, 根据加速度测量值 $z_{k1} = [\tilde{a}_x, \tilde{a}_y, \tilde{a}_z]$, 根据加速度计对重力向量的观测得到滚转角和俯仰角的值:

$$\begin{bmatrix} \tilde{a}_x \\ \tilde{a}_y \\ \tilde{a}_z \end{bmatrix} = R_W^B \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} = \begin{bmatrix} g \sin \theta \\ -g \cos \theta \sin \phi \\ g \cos \theta \cos \phi \end{bmatrix} \quad (4-27)$$

故可得:

$$\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \arcsin(\tilde{a}_x/g) \\ \arctan(-\tilde{a}_y/\tilde{a}_z) \end{bmatrix} \quad (4-28)$$

根据测得的磁强计的测量值 $z_{k2} = [\tilde{m}_x, \tilde{m}_y, \tilde{m}_z]$, 根据 (4-28) 所示的载体倾角, 获得水平坐标系下的磁强计的测量值:

$$\begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \phi \sin \theta & \cos \phi \sin \theta \\ 0 & \cos \phi & -\sin \phi \\ -\sin \theta & \sin \phi \cos \theta & \cos \phi \cos \theta \end{bmatrix} \begin{bmatrix} \tilde{m}_x \\ \tilde{m}_y \\ \tilde{m}_z \end{bmatrix} \quad (4-29)$$

再加上磁偏角 D 即为载体的偏航角如式 (4-30) 所示:

$$\psi = \arctan(h_x/h_y) + D \quad (4-30)$$

由于加速度计会测量除了重力加速度以外的加速度, 这就会导致测量误差, 所以采取加权处理, 最后获得更新过程:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (4-31)$$

$$q_k = \hat{q}_k^- + K_k (z_k - h(\hat{q}_k^-)) \quad (4-32)$$

$$P_k = (I - K_k H_k) P_k^- \quad (4-33)$$

其中 R_k 表示测量噪声协方差矩阵, z_k 表示姿态角观测值。

4.4 数据集仿真测试

为了对算法的定位精度进行研究, 本文选择 euroc 数据集中的十一组不同数据来对算法进行检验。选取 MH_05 数据序列作为展示, 满足在 ROS 机器人操作系

统下离线回放，通过数据集获得视觉和 IMU 初始数据，接下来采取前端预处理和联合初始化。无人机在系统启动回环检测模块后轨迹如下，真实的是红颜色线，估计轨迹是绿颜色。

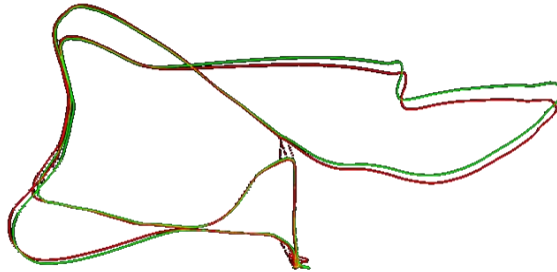


图 4-7 无人机运动轨迹和回环约束

不难发现，二者轨迹几乎重合，产生漂移后，也可以很快地回到正轨轨迹，说明回环检测与重定位修正轨迹效果不错。图 4-8 展示了第 500 帧图像和第 35 帧图像的回环约束及特征点匹配，此时在它们之间检测到了回环关系，再利用之前提到的 RANSC 法去除错误匹配。

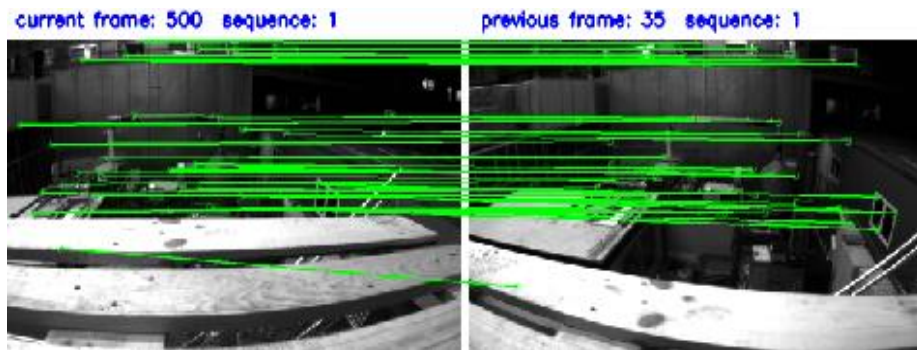


图 4-8 回环约束检测

接下来把估计轨迹对齐到真实轨迹求取定位误差便成了视觉惯性里程计定位的一个关键步骤。通过文献所述，当实际与估计坐标系不一致时，因为 VIO 绝对尺度可得，仅仅依靠有平移向量 t 和旋转矩阵 R_z 的刚体变换即可完成。对下式进行求解可获得该变换，展示如下：

$$\{t, R_z\} = \arg \min_{\{t, R_z\}} \sum_{i=0}^{N-1} \|p_i - R_z \hat{p}_i - t\|^2 \quad (4-34)$$

上式里 p_i 、 \hat{p}_i 分别为真实轨迹、估计轨迹， R_z 只有航向角变化。位姿估计时许多图像帧都被舍弃，会导致实际的位姿数大于结果。如图 4-9 所示，展示了从左侧到右侧的轨迹形式变化图。

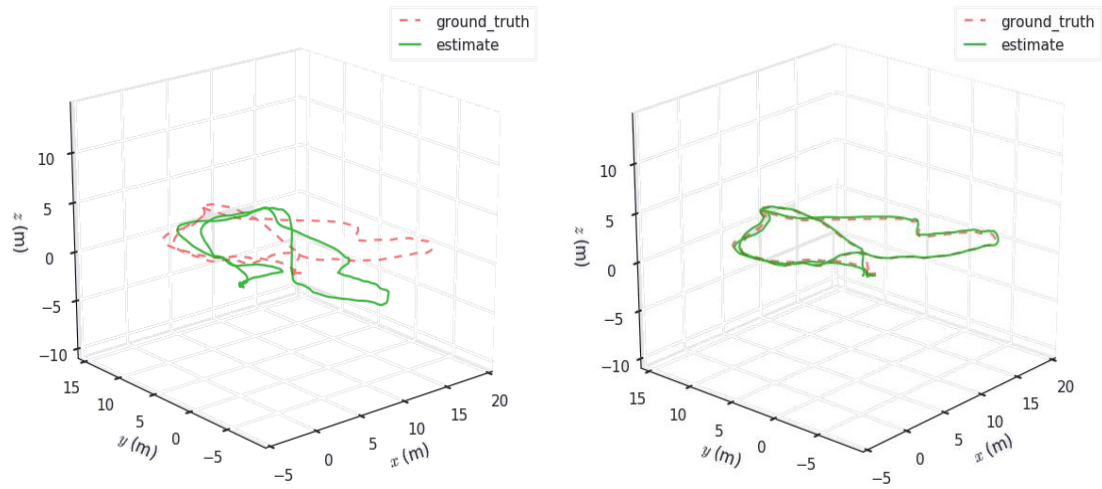


图 4-9 轨迹对齐

这样一来在对齐结束后，选择绝对轨迹误差计算结果，因为 MH_01 序列和 MH_05 序列在平移、旋转和光照的变化大，故选取它们对算法的尺度误差、姿态误差和位置误差进行剖析，记录在图 4-10 以及 4-11。

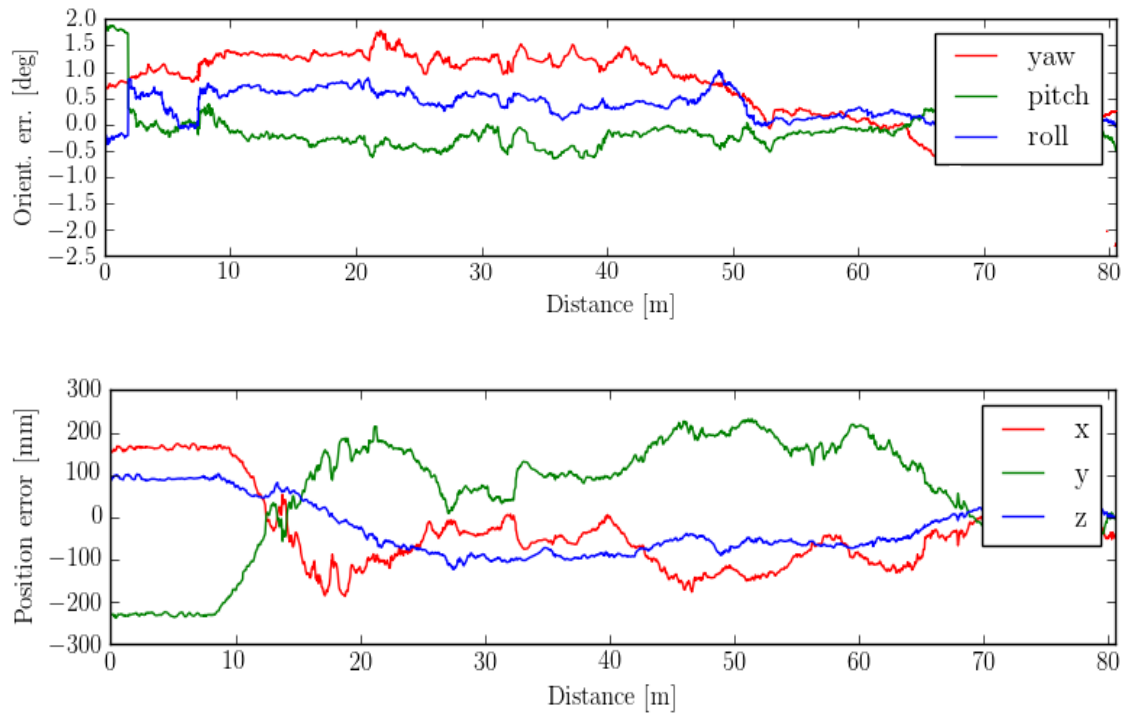


图 4-10 MH_01 序列的状态误差

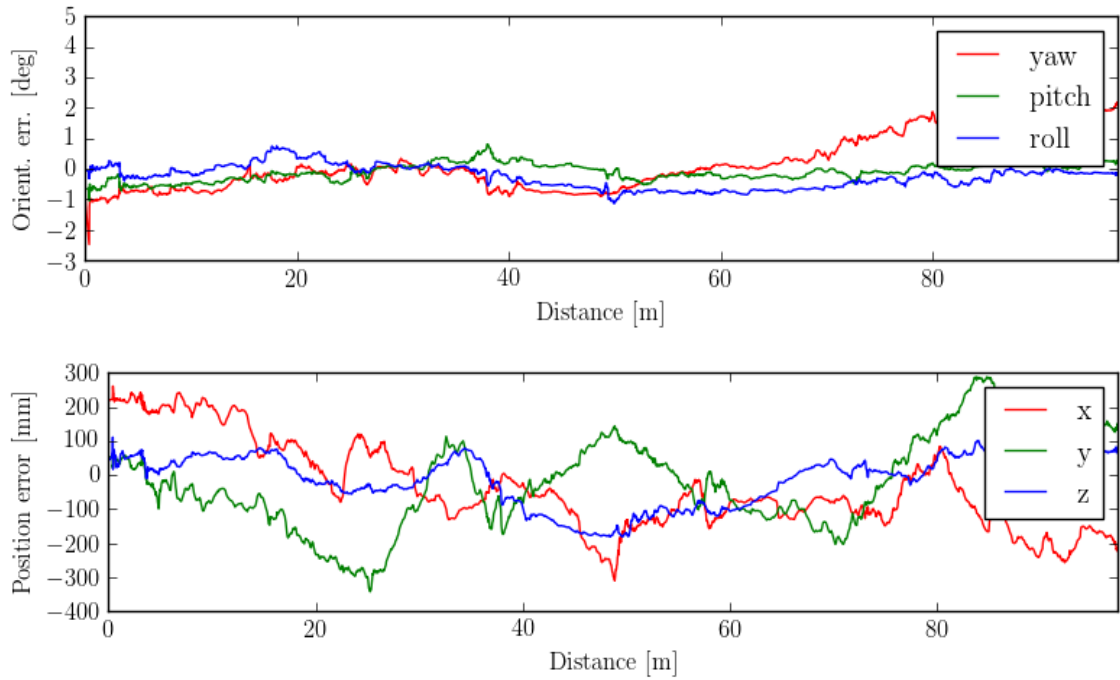


图 4-11 MH_05 序列的状态误差

从图中不难发现，x、y、z 三轴误差都保持在 30cm 内，基本满足定位需求。航向角的误差大于俯仰角和横滚角误差，造成它的原因是后面两者没有存在角度漂移，回环检测和位姿优化虽可以降低误差，但还是远远不够的。最终航向角的误差控制在 $\pm 4^\circ$ 左右，俯仰角和横滚角误差控制在 $\pm 2^\circ$ 以内，累积误差得到控制，可以得出回环检测与全图位姿优化起到了很关键的作用。最后根据均方根误差判定回环检测效果，也与别的流行算法作比较：

$$ATE_{rot} = \left(\frac{1}{N} \sum_{i=0}^{N-1} \|\angle(\Delta R_i)\|^2 \right)^{\frac{1}{2}} \quad (4-35)$$

$$ATE_{pos} = \left(\frac{1}{N} \sum_{i=0}^{N-1} \|\Delta p_i\|^2 \right)^{\frac{1}{2}} \quad (4-36)$$

其中 $\angle(\Delta R_i)$ 代表 ΔR_i 的角度形式， Δp 表示位置误差。最终结果记录在表 4-1 中，显而易见回环检测可以较好地减小误差，但对于采集时间较短的数据集来说，回环检测的效果并不足够好，这就表明回环检测其实是需要长时间才可以对误差累积的消除起到比较好的作用，短时间效果并不明显，所以应保证系统不会短时间停止。算法开启回环检测与在不开启回环检测模块时，结果如下表，表明不使用回环检测定位精度也比 OKVIS 更好。

表 4-1 有/无回环 RMSE 对比

数据集	VIO_no_loop (m)	VIO_loop (m)	OKVIS (m)
MH 01	0.24	0.08	0.33
MH 02	0.17	0.08	0.25
MH 03	0.23	0.09	0.31
MH 04	0.14	0.07	0.31
MH 05	0.32	0.08	0.44
V1 01	0.09	0.06	0.12
V1 02	0.13	0.07	0.19
V1 03	0.14	0.12	0.25
V2 01	0.07	0.05	0.12
V2 02	0.08	0.05	0.11
V2 03	0.19	0.10	0.30

4.5 纯旋转运动姿态估计仿真结果

下面展示的是纯旋转运动姿态估计，实验采用动态实验与姿态解算模块输出作比较，比较情况从 4-12 到 4-14 所示，

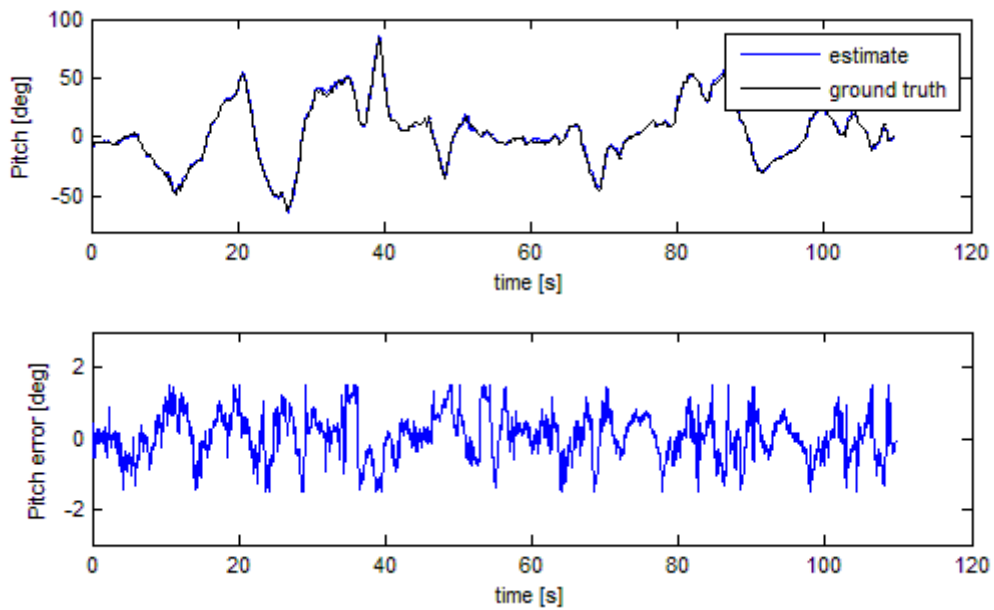


图 4-12 俯仰角和误差曲线

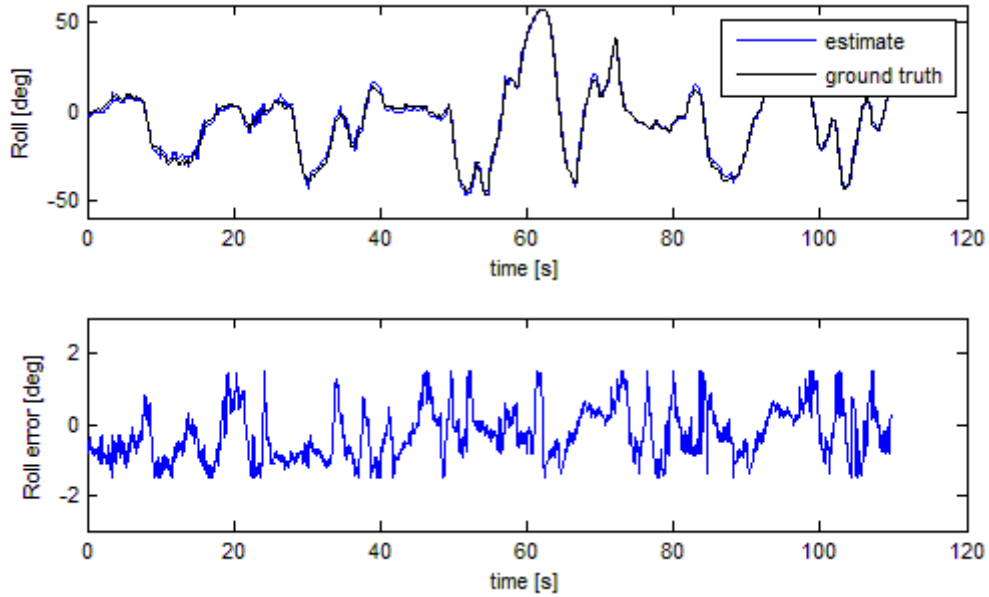


图 4-13 横滚角和误差曲线

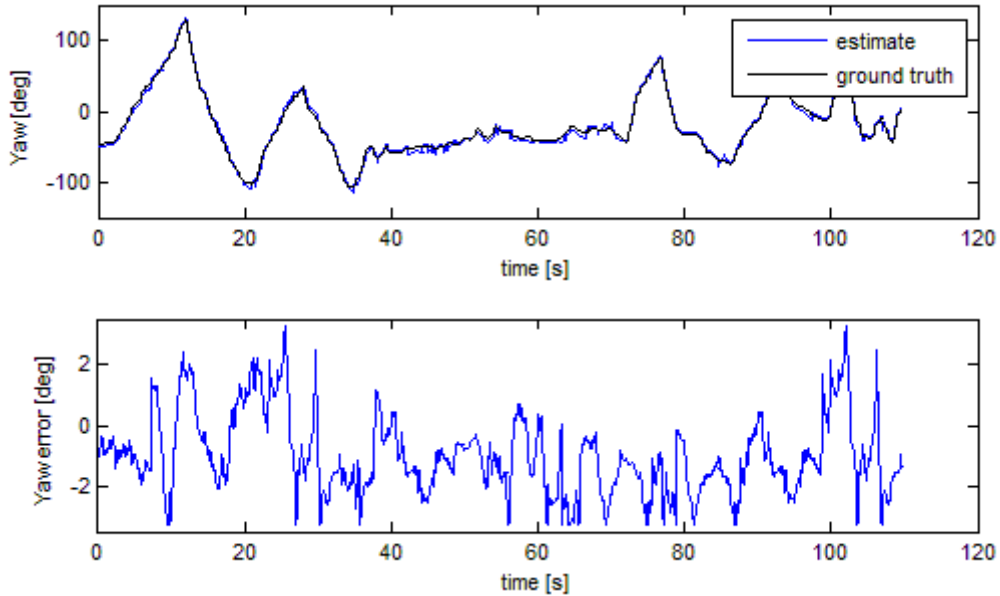


图 4-14 偏航角和误差曲线

通过实验结果可得到当运载体进行纯旋转运动时得到的姿态角误差控制在 $\pm 2^\circ$ 之内，确实可以在载体运动退化的情况下进行位姿估计，且精度及稳定性都得到满足。

4.6 本章小结

本章建立了视觉/惯性组合导航定位系统的多种约束关系因子图，以此确定了系统滑动窗口非线性优化的紧耦合方程，并对方程中的误差项进行了分析，引入滑

窗边缘化方法。通过引入基于词袋模型的回环检测和四自由度位姿图优化的方法，减小了导航系统长时间运行的累积轨迹漂移，为接下来的全局位姿优化减少漂移做了极大的贡献。同时，针对载体纯旋转退化问题，基于扩展卡尔曼滤波方法给出了融合 IMU 信息的姿态估计算法。最后，给出了采用 Euroc 开源数据集的仿真验证方法，对视觉/惯性融合算法进行了定量分析。

结 论

本文通过对比分析视觉/惯性组合导航定位技术的国内外现状,采用松耦合算法确定系统初值和基于滑动窗口的紧耦合非线性优化相结合的系统结构,针对融合双目视觉相机和 MEMS 惯性测量单元的运动位姿估计算法进行了研究,达到满足无人移动机器人导航定位的应用需求。本文主要研究工作总结如下:

(1) 给出了双目视觉导航系统中相机坐标系到世界坐标系的转换方程,对比分析了 SIFT、Harris、ORB 和 SURF 四种特征点检测方法,验证了 ORB 特征点检测方法的最优性,给出了基于 RANSAC 方法的特征点误匹配去除算法,通过特征提取匹配并对图像特征进行跟踪,并给出了基于 ICP/SFM 融合的重构算法。

(2) 建立了 IMU 的测量模型,对 IMU 数据进行了预积分处理,获得了图像帧间的位姿关系和非线性优化迭代的初值,确定特征点的三维坐标,并将 IMU 与视觉得到的位姿进行松耦合处理,通过确立的最小二乘方程组,将非线性优化所需要的陀螺仪的零偏、系统速度以及重力加速度。

(3) 建立了视觉/惯性组合导航定位系统的多种约束关系因子图,以此确定了系统滑动窗口非线性优化的紧耦合方程,并对方程中的误差项进行了分析。通过引入基于词袋模型的回环检测和四自由度位姿图优化的方法,减小了导航系统长时间运行的累积轨迹漂移。同时,针对载体纯旋转退化问题,基于扩展卡尔曼滤波方法给出了融合 IMU 信息的姿态估计算法。

(4) 给出了采用 Euroc 开源数据集的仿真验证方法,对视觉/惯性融合算法进行了定量分析,通过仿真计算系统 ATE 及均方根误差分析了回环检测精度,实验结果表明横滚角和俯仰角误差在 $\pm 2^\circ$ 左右,航向角误差在 $\pm 4^\circ$ 左右,本文导航定位方法的 x、y、z 三轴误差都保持在 30cm 左右。在载体做纯旋转运动时,姿态角的精度在 $\pm 2^\circ$ 左右。

本文通过回环检测方法提高了系统的导航定位精度,但依然有不足之处,后续可以从以下几个方面进行研究:

(1) 本文研究的算法均处在静态场景中,若处于动态场景中定位精度会被严重影响,后期应该朝着针对动态特征点的跟踪继续研究。

(2) 加入后续的路径规划、避障等功能,通过三维坐标构建八叉树、TSDF 等算法完成一个更多功能完善的系统。

(3) 本文研究的方法融合方法仅限于松紧耦合,对于深一层的组合模式还未有所研究,可选择从这一方面进行研究,进一步提高系统精度。

参考文献

- [1]关政军.航海仪器[M].大连:大连海事大学出版社,2009,53-57.
- [2]薛连莉,陈少春,陈效真.2017 年国外惯性技术发展与回顾[J].导航与控制,2018,17(02):1-9+40.
- [3]薛连莉,戴敏,葛悦涛,陈少春.2018 年国外惯性技术发展与回顾[J].飞航导弹,2019(04):16-21.
- [4]S. Nakamura,MEMS inertial sensor toward higher accuracy & multi-axissensing[C]. SENSORS, 2005 IEEE, Irvine, CA, 2005,939-942.
- [5]Moravec H P. Obstacle avoidance and navigation in the real world by a seeing robot rover[M]. Stanford University, 1980.
- [6]Lourakis M I. Design and implementation of a sparse bundle adjustment software library based on the Levenberg-Marquardt algorithm[J]. 2004.
- [7]L. Matthies, “Dynamic stereo vision,” Ph.D. dissertation, Carnegie Mellon Univ.Pittsburgh, PA, 1989.
- [8]C. Olson, L. Matthies, M. Schoppers, and M. W. Maimone, “Robust stereo egomotion for long distance navigation,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2000, pp. 453–458.
- [9]C. Olson, L. Matthies, M. Schoppers, and M. Maimone, “Rover navigation using stereo ego-motion,” Robot. Autonom. Syst., vol. 43, no. 4,pp. 215–229, 2003.
- [10]Skog I, Händel P. Calibration of a MEMS inertial measurement unit[J]. Electrical Engineering Electronic Engineering Information Engineering, 2016.
- [11]Geiger A. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:3354-3361.
- [12]Geiger A, Ziegler J, Stiller C. StereoScan: Dense 3d reconstruction in real-time[J]. IEEE Intelligent Vehicles Symposium, 2011, 32(14):963-968.
- [13]彭勃. 立体视觉里程计关键技术与应用研究[D]. 浙江大学, 2008.
- [14]李永佳. 基于视觉的机器人定位研究[D]. 浙江大学, 2011.
- [15]卢维. 高精度实时视觉定位的关键技术研究[D]. 浙江大学, 2015.
- [16]江燕华. 车辆运动特性约束的智能车辆视觉里程计系统研究[D]. 北京理工大学, 2014.
- [17]厉茂海, 洪炳镨, 罗荣华. 基于单目视觉的移动机器人全局定位[J]. 机器人, 2007, 29(2):140-144.

- [18]高云峰, 李伟超, 李建辉. 室内移动机器人视觉里程计研究[J]. 传感器与微系统, 2012(2):26-29.
- [19]Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-Time Single Camera SLAM[J]. IEEE Trans Pattern Anal Mach Intell, 2007, 29(6):1052-1067.
- [20]MONTEMERLO M, THRUN S, KOLLER D, et al. FastSLAM: a factored solution to the simultaneous localization and mapping problem[C]. Proceedings of the 18th International Conference on Artificial Intelligence. Pittsburgh, PA, USA, 2002: 2004.
- [21]MONTEMERLO M, THRUN S. Simultaneous localization and mapping with unknown data association using FastSLAM[C]. Proceedings of 2003 IEEE International Conference on Robotics and Automation. Taipei, Taiwan, China, 2003: 1985-1991.
- [22]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.
- [23]Klein G. Parallel tracking and mapping for small AR workspaces[C]. Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007), Nara, Japan, Nov. IEEE, 2007.
- [24]Klein G, Murray D W. Improving the Agility of Keyframe-Based SLAM[C]. Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II. Springer-Verlag, 2008.
- [25]Engel J, Schöps T, Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM[M]. Computer Vision – ECCV 2014. 2014.
- [26]BayH, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3):404-417.
- [27]Forster C, Carlone L, Dellaert F, et al. On-Manifold Preintegration for Real-Time Visual-Inertial Odometry[J]. IEEE Transactions on Robotics, 2015, 33(1):1-21.
- [28]Labbé M, Michaud F. Online global loop closure detection for large-scale multi-session graph-based SLAM[C]. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014.
- [29]Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2017, 31(5):1147-1163.
- [30]Faessler M, Fontana F, Forster C, et al. Autonomous, Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle[J]. J. Field Robotics, 2016, 33(4): 431-450.
- [31]Mourikis A, et.al. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation[C]. IEEE International Conference on Robotics & Automation. IEEE, 2007.

- [32]Bloesch M, Omari S, Hutter M, et al. Robust visual inertial odometry using a direct EKF-based approach[C] IEEE/RSJ International Conference on Intelligent Robots & Systems. 2015.
- [33]Leutenegger S , Lynen S , Bosse M , et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3):314-334.
- [34]Tong Q, Li P, Shen S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2017, PP(99):1-17.
- [35] 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):1-8.
- [36]H. Liu, M. Chen, G. Zhang. Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam[J]. IEEE Conference on Computer Vision and Pattern Recognition, June 2018, 34(3):314–334.
- [37](美)布拉德斯基(Bradski, G.), (美)克勒(Kaehler,等).学习 OpenCV[M].清华大学出版社, 2009.
- [38]Lowe D G.Distinctive Image Feature from Scale-Invariant Keypoints[J].International Journal of Computer Vision, 2004, 60(2):91-110.
- [39]基于上下文相关性的视觉特征表示及其应用[D]. 娄天禹.北京交通大学 2012.
- [40]Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J].Computer Vision & Image Understanding, 2008, 110(3):404-417.
- [41]Rublee E , Rabaud V , Konolige K , et al. ORB: An efficient alternative to SIFT or SURF[C]// 2011 International Conference on Computer Vision. IEEE, 2012:2564-2571.
- [42]Chris Harris, Mike Stephens, A Combined Corner and Edge Detector, 4th Alvey Vision Conference, 1988, pp147-151
- [43]Sirait K, Tulus, Nababan E B. K-Means Algorithm Performance Analysis With Determining The Value Of Starting Centroid With Random And KD-Tree Method[C]// 2017:012016.
- [44]Arun K S, Huang T S, Blostein S D. Least-Squares Fitting of Two 3-D Point Sets[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1987, 9(5):698-700.
- [45]Morenonoguer F, Lepetit V, Fua P. Accurate Non-Iterative O(n) Solution to the PnP Problem[J]. Proc.int.conf.on Computer Vision, 2007:1-8.
- [46]Morenonoguer F, Lepetit V, Fua P. Accurate Non-Iterative O(n) Solution to the PnP Problem[J]. Proc.int.conf.on Computer Vision, 2007:1-8.
- [47] 基于 IMU 与单目视觉融合的位姿估计方法研究[D]. 王晨曦.哈尔滨工业大学

2012.

- [48]Morenonoguer F, Lepetit V, Fua P. Accurate Non-Iterative $O(n)$ Solution to the PnP Problem[J]. Proc.int.conf.on Computer Vision, 2007:1-8.
- [49] 高翔,张涛,刘毅,等.视觉 SLAM 十四讲:从理论到实践. 电子工业出版社, ISBN: 97878121311048, 2017.
- [50] 秦永元. 惯性导航(第二版)[M], 科学出版社, ISBN 978-7-03-039465-1: 244-259
- [51] Tong Q, Li P, Shen S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2017, PP(99):1-17.
- [52] Forster C, Carlone L, Dellaert F, et al. On-Manifold Preintegration for Real-Time Visual--Inertial Odometry[J]. IEEE Transactions on Robotics, 2015, 33(1):1-21.
- [53] Solà, Joan. Quaternion kinematics for the error-state Kalman filter[J]. 2017.
- [54] Frank D , Michael K . Factor Graphs for Robot Perception[C] 2017:1-139.
- [55] Sibley G , Matthies L , Sukhatme G . Sliding window filter with application to planetary landing[J]. Journal of Field Robotics, 2010, 27(5):587-608.
- [56] Gálvez-López D , Tardos J D . Bags of Binary Words for Fast Place Recognition in Image Sequences[J]. IEEE Transactions on Robotics, 2012, 28(5):1188-1197.
- [57] Yang S, Scherer S A, Zell A. Robust Onboard Visual SLAM for Autonomous MAVs[M] Intelligent Autonomous Systems 13. 2016.
- [58] Sabatelli S, Galgani M, Fanucci L, et al. A Double-Stage Kalman Filter for Orientation Tracking With an Integrated Processor in 9-D IMU[J]. IEEE Transactions on Instrumentation & Measurement, 2013, 62(3):590-598.
- [59]Morenonoguer F, Lepetit V, Fua P. Accurate Non-Iterative $O(n)$ Solution to the PnP Problem[J]. Proc.int.conf.on Computer Vision, 2007:1-8.

致 谢

时光荏苒，岁月如梭，研究生生活一转眼就要结束了，这期间我学习到了更多的专业知识，收获了珍贵的友谊，找到了一份满意的工作。我遇到了太多优秀的人，他们给予了我很多帮助，教会了做人、做事的道理。在这里我要向他们致以最诚挚的感谢。

首先我要感谢我的研究生导师夏红伟教授，夏老师对我的论文提供了大量的指导，帮助我修正了很多细节上的错误，使我的论文更加完善。

同时我要感谢闻帆老师、奚伯齐老师和屈桢深老师，老师们在科研和论文上给予了我悉心的指导，为我点明了研究的方向；也让我学会了很多为人处世的道理。

其次我要感谢研究生期间遇到的每一位同学和朋友。我们相互督促努力进步，友谊是世间最宝贵的财富，我会铭记一生。

感谢我的父母，在疫情的关键时期，感谢他们支持我的每个决定，从未给我带来任何压力。在今后的日子里，我将会继续成为你们的骄傲。

感谢伟大的祖国，疫情期间防控得当，现在已经是世界上最安全的国家，祖国母亲的爱我感恩于心。

天高水长，我们后会有期。