

分类号

学校代码: 10112

密 级:

太原理工大学

硕士学位论文

(专业学位)

基于视觉 GPS 融合二维占据网格地图构建

姓 名: 王永祥

学 号: 2018520256

培养单位: 信息与计算机学院

专业领域: 电子与通信工程

研究方向: 机器视觉与计算智能

导师姓名: 常青 副教授

企业导师: 孙永明 教授级高工

论文提交日期: 2021 年 6 月

A Dissertation Submitted to
Taiyuan University of Technology
In partial fulfillment of the requirement
For the degree of Master

**Construction of 2D Occupied Grid Map Based on
Vision-GPS Fusion**

By
Yongxiang Wang

College of Information and Computer

June 2021

硕士工程类专业学位：

学位论文答辩信息表

论文题目	基于视觉 GPS 融合二维占据网格地图构建		
论文类型 (在[]内打“√”)	[]产品研发 []工程设计 [√]应用研究 []工程/项目管理 []调研报告 []基础研究		
答辩日期	2021 年 5 月 28 日	答辩秘书	张博
学位论文答辩委员会成员			
姓名	职称	工作单位	备注
谢刚	教授	太原科技大学	主席
安临生	高级工程师	中国电信股份有限公司 山西分公司	委员
谢珺	副教授	太原理工大学	委员

摘要

随着智能机器人技术的蓬勃发展,复杂场景下的定位与建图对传感器的要求也越来越高。根据不同传感器的特性,在应对不同场景时可以选择相应的传感器:单纯采用视觉传感器对场景的纹理特征以及运动速度有一定的限制,但是视觉传感器可以提供丰富的场景信息。双目立体相机既可以解决单目相机没有尺度的问题,又可以避免 RGBD 相机在室外被光照干扰的问题。GPS (Global Position System) 可以实时获取自身的绝对位置,但是在高楼树林里 GPS 信号很容易被遮挡,进而导致定位失败。视觉和 GPS 传感器在使用场景上可进行互补。IMU (Inertial measurement unit) 能够在快速运动的状态下获得自身姿态信息,但是随着时间的推移,IMU 会产生积分累积误差。通过 IMU 运动模型来预测自身姿态,利用视觉传感器和 GPS 对其位置进行修正可以应对复杂场景。有了良好的姿态信息以后,可以进行地图构建来判断环境周围的障碍物,但是当前大多数视觉 SLAM (Simultaneous Localization And Mapping) 只提供了姿态信息和稀疏点云地图,而稀疏点云地图对周围环境的描述能力有限,所以可以充分利用视觉信息来对周围环境进行描述。基于此做了以下研究:

(1) 提出了基于视觉里程计的相对状态更新扩展卡尔曼的 VI-rel 方法。

为了给视觉刚体运动提供一个很好的刚体先验约束,通过将 VO (Visual Odometry) 与 IMU 通过紧耦合的方法建立联合优化目标函数,求解姿态并保存。从保存的姿态里面找一个先前的姿态添加到 IMU 预测的状态中,建立增广状态矩阵,并对预测模型的雅各比矩阵进行推导。利用此刻的 VO 测量姿态与之前添加到状态矩阵的姿态构建相对测量模型作为 EKF 的观测模型进行更新,并推导了观测模型的雅各比矩阵。最后在 EuRoC 数据集上实验表明本文算法轨迹与真值比较接近,经 evo 工具分析,在 MH_02_easy 场景下的 RMSE 为 0.153238。在 V1_02_medium 复杂场景下的 RMSE 为 0.240838。

(2) 提出了基于 GPS 绝对状态更新扩展卡尔曼的 IG-abs 方法。

GPS 与 IMU 采用松耦合的 EKF 数据融合方法,IMU 对刚体运动状态预测,提供预测方程,GPS 提供观测方程更新预测状态,但是 GPS 测量会由于受到多径效应等干扰出现异常值,导致观测值误差很大。通过当前卫星数量来判断 GPS 测量值是否可靠,但该方法效果并不明显。故可以通过预测残差平方和与观测残差平方和增量是否异常来判断更新后的状态是否可靠,通过实验与 GPS 测量值对比,平均误差大约在 1.02m 到 2.54m。

(3) 提出了基于 ORB-SLAM2 的二维占据网格地图构建的 VO-Scan 方法。

目前大多数视觉 SLAM 都只输出姿态信息与周围环境的稀疏点云地图,对周围环境的描述效果有限,但是可以利用输出的姿态信息与环境深度信息构建二维网格地图来

对周围环境进行描述。将双目立体相机提供的深度图转换为 ROS 平台下 Scan 格式的数据，同时通过 ORB-SLAM2 提供姿态信息，二者结合来构建二维占据网格地图。并建立与双目视觉相关的逆传感器模型来更新占据概率。最后在 ROS 平台下实验并与真实环境对比表明，可以构建可提供判断周围障碍物的二维占据网格地图。

关键词：双目视觉里程计；IMU；GPS；二维网格地图

ABSTRACT

With the rapid development of intelligent robot technology, the sensor requirements for location and mapping in complex scenes are getting higher and higher. According to the characteristics of different sensors, the corresponding sensors can be selected when dealing with different scenes. The mere use of visual sensors has certain limitations on the texture features and motion speed of the scene, but they can provide rich scene information. Binocular stereo camera can not only solve the problem that monocular camera does not have scale, but also avoid the problem that RGBD camera is interfered by illumination outdoors. GPS (Global Position System) can obtain its own absolute Position in real time, but the GPS signal is easy to be blocked in the high-rise forest, which leads to the failure of positioning. Visual and GPS sensors complement each other in usage scenarios. IMU (Inertial Measurement Unit) can obtain its own attitude information in the state of fast motion, but with the passage of time, IMU will produce integral accumulation error. The IMU motion model is used to predict its own posture, and the visual sensor and GPS are used to correct its position to cope with complex scenes. Have made a very good attitude information, can undertake judging map building environment around obstacles, but the most visual SLAM (Simultaneous Localization And Mapping) provides only attitude information And sparse point cloud maps, sparse point cloud maps on the surrounding environment description ability is limited, so can make full use of visual information on the surrounding environment is described. Based on this, the following research was done:

(1) A method of EKF based on VO relative state update is proposed.

In order to provide a good rigid-body prior constraint for Visual rigid-body motion, the joint optimization objective function was established by tightly coupling VO (Visual Odometry) and IMU, and the attitude was solved and saved. Add a previous attitude from the saved attitude to the predicted state of the IMU, establish the augmented state matrix, and deduct the Jacoby matrix of the prediction model. A relative measurement model was constructed by using the current VO measurement attitude and the previous attitude added to the state matrix to update the observation model of EKF, and the Jacoby matrix of the observation model was deduced. Finally, experiments on Euroc dataset show that the trajectory of the algorithm in this paper is close to the true value. After analysis by EVO tool, RMSE in MH_02_Easy scenario is 0.153238. RMSE of V1_02_Medium complex scene is 0.240838.

(2) A method of updating EKF based on GPS absolute state is proposed.

GPS and IMU adopt the loose-coupling EKF data fusion method. IMU predicts the motion state of rigid body and provides the prediction equation, while GPS provides the observation equation to update the prediction state. However, GPS measurement will have outliers due to interference such as multipath effect, resulting in large error of the observed value. The current number of satellites is used to determine whether GPS measurements are reliable, but this method is not very effective. Therefore, whether the updated state is reliable can be judged by predicting whether the residual sum of squares and the observed residual sum of squares are abnormal. By comparing the experiment with the GPS measured value, the average error is about 1.02m to 2.54m.

(3) A 2D occupied grid map construction method based on ORB-SLAM2 is proposed.

At present, most visual SLAM only outputs sparse point cloud maps of the attitude information and the surrounding environment, and the description effect of the surrounding environment is limited. However, the attitude information and the environmental depth information can be used to build a 2D grid map to describe the surrounding environment. The depth map provided by the binocular stereo camera is converted to Scan format data, and the attitude information is provided by ORB-SLAM2. The combination of the two is used to construct the 2D occupied grid map. An inverse sensor model related to binocular vision was established to update the occupancy probability. Finally, the experiment on ROS platform and the comparison with the real environment show that the 2D occupied grid map can be constructed to judge the surrounding obstacles.

Keywords: binocular vision odometer; IMU; GPS; 2D grid map

目 录

摘 要.....	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 选题来源.....	3
1.3 国内外研究现状	3
1.3.1 视觉 SLAM 研究现状	3
1.3.2 视觉惯性 GPS 融合定位研究现状.....	4
1.3.3 二维网格地图研究现状	6
1.4 主要研究内容以及组织结构.....	7
1.4.1 主要研究内容	7
1.4.2 组织架构.....	7
第 2 章 相关技术研究.....	9
2.1 双目视觉里程计	9
2.1.1 双目相机模型.....	9
2.1.2 ORB 特征提取与匹配	10
2.1.3 视觉里程计位姿表示	11
2.2 IMU 传感器姿态估计.....	13
2.2.1 IMU 传感器基本原理.....	13
2.2.2 IMU 噪声模型	14
2.2.3 IMU 运动模型	14
2.3 GPS 传感器定位技术	15
2.3.1 GPS 定位原理.....	15
2.3.2 GPS 定位误差.....	16
2.4 本章小结.....	16
第 3 章 视觉惯性 GPS 融合定位的 VI-rel 与 IG-abs 算法研究	17
3.1 紧耦合 IMU 的视觉里程计	17
3.1.1 联合优化函数.....	17
3.1.2 重投影误差.....	18

3.1.3	IMU 预积分与测量残差.....	19
3.2	VO/IMU 融合定位的 VI-rel 研究.....	20
3.2.1	IMU 状态传递动力学模型.....	20
3.2.2	基于 VO 相对状态的 EKF.....	21
3.3	IMU/GPS 融合定位的 IG-abs 研究.....	22
3.3.1	基于 GPS 的观测模型.....	22
3.3.2	IMU 与 GPS 信号同步.....	23
3.3.3	基于残差平方和的 EKF.....	23
3.4	实验探究.....	25
3.4.1	VO/IMU 融合性能评估.....	25
3.4.2	IMU/GPS 融合性能评估.....	28
3.5	本章小结.....	30
第 4 章	二维占据网格地图构建的 VO-Scan 算法.....	31
4.1	双目立体视觉逆传感器模型.....	31
4.1.1	理想传感器模型.....	31
4.1.2	实际的逆传感器模型.....	31
4.2	传感器模型的参数选择.....	32
4.2.1	显著因子 k 对 ISM 的影响.....	32
4.2.2	深度相关不确定度 Δp 对 ISM 的影响.....	33
4.2.3	测量距离 z 对 ISM 的影响.....	33
4.3	二维网格地图模型.....	34
4.4	基于视觉定位的二维网格地图构建算法流程.....	36
4.4.1	ORB-SLAM2 概述.....	36
4.4.2	深度图转 Scan 格式的数据原理.....	37
4.4.2	基于 ORB-SLAM2 的二维占据网格地图构建算法流程.....	38
4.5	深度图转 Scan 数据.....	38
4.6	本章小结.....	39
第 5 章	实验与分析.....	41
5.1	引言.....	41
5.2	实验平台.....	41
5.2.1	ROS 操作系统.....	41
5.2.2	MYNTEYE 相机.....	42

5.3 场景测试.....	42
5.3.1 室内实验.....	42
5.3.2 室外实验.....	44
5.4 对比分析实验.....	46
5.5 本章小结.....	47
第6章 结论与展望.....	49
6.1 结论	49
6.2 展望	50
参考文献.....	51
攻读学位期间取得的研究成果	55
致 谢.....	57

第1章 绪论

1.1 研究背景及意义

科学技术发展日新月异，生活中机器人的应用非常广泛，有仓储物流的 AGV 机器人可代替人工搬运快递，有喷洒农药的无人机提高喷洒效率，有消防机器人代替消防人员进入危险区域消防救灾，有军用机器狗搬运弹药减少人员伤亡。这些机器人的应用都需要自主定位，对环境分析来识别周围的障碍物。应用场景复杂程度越高以及机器人智能程度越高，则对机器人的定位能力以及对环境的感知能力要求也越来越苛刻。



图 1-1 智能机器人应用

Fig. 1-1 Application of intelligent robot

目前，主流的定位导航技术有视觉 SLAM（Simultaneous Localization and Mapping）^[1]定位，惯性测量单元 IMU（Inertial Measurement Unit），全球定位系统 GPS（Global Position System）^[2]。

SLAM 同时定位与地图构建是机器人自主导航中的一个非常重要的研究分支，它可用于人类背负系统以及机器人系统的导航应用中。由于 SLAM 算法具有重要的应用价值，因此，很多学者认为它是最终实现全自主移动机器人系统的关键技术^[3]。

惯性导航系统主要采用 IMU 来获取自身载体的加速度与角速度来实时估计自身的姿态，它对环境的依赖程度较低，不需要借助外部信息就可以获得自身姿态，所以应用非常广泛。根据测量精度的差异，IMU 的种类也非常多。

全球定位系统 GPS 为地球上的用户提供定位服务，可以全天候地提供位置以及时间服务，不受外界环境影响。

机器人智能化程度的高低，不仅取决于自身的定位能力，还需要有一定的分析判断决策能力。当处于一个未知的环境中，机器人可以对周围环境进行描述，我们称其为地图，地图的种类有很多，包括拓扑地图，3D 栅格地图，二维网格地图^[4]，稀疏点云地图^[5]等。不同的地图有着不同的应用场景。

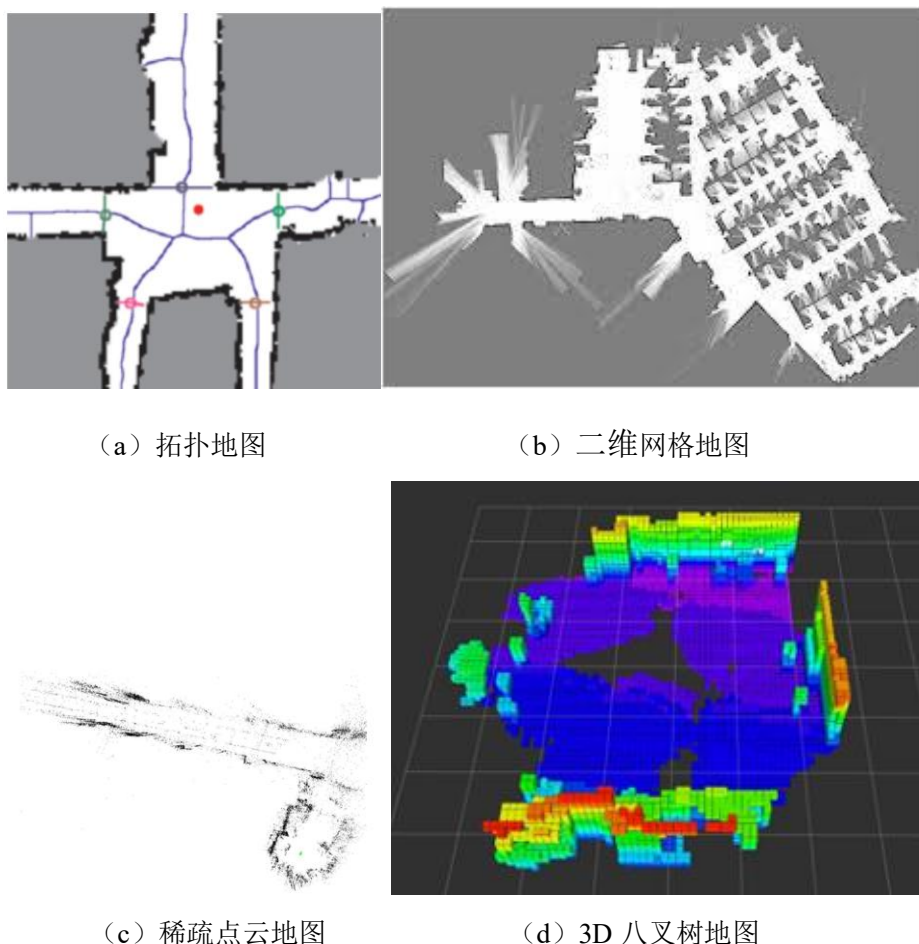


图 1-2 各种类型的地图

Fig. 1-2 All kinds of maps

对于复杂场景下的机器人应用目前还存在以下问题：

纯视觉的 SLAM 对于实际场景的应用受到诸多限制，视觉 SLAM 估计得到的相机位姿缺少绝对定位信息，无法获得场景结构的真实尺度。当图像发生变化的时候，无法判断是相机自身的运动还是外界环境发生了变化导致的相对运动，因此，纯单目视觉 SLAM 存在一定的缺陷，它难以处理动态的障碍物。而且也容易受环境影响导致定位失败。所以多传感器融合也是当下视觉 SLAM 技术研究的热门方向。

在人为干扰、山区丛林、城市高楼林立的情况下，GPS 存在因为信号被遮挡而无法实现定位的问题，但此时环境纹理丰富，特别适合视觉 SLAM 的应用。在空旷的地方，

环境纹理缺乏,不适合视觉 SLAM 定位,而 GPS 在空旷的环境下,可以进行高精度定位,不仅如此,还可以提供绝对定位信息,同时在系统初始化阶段获得场景结构的真实尺度^[6]。目前使用比较多的是 VO/IMU^[7], IMU/GPS^[8]融合的组合方法, GPS 可以为 IMU 提供初始位置与速度信息,但是会很依赖 GPS 来修正 IMU 累积误差。

当前大多数 SLAM 系统无法提供可供机器人判断障碍物的地图,只生成了稀疏的点云地图,所以在机器人寻找自身位置的同时,需要构建可供机器人导航避障的地图。

1.2 选题来源

森林火灾频发,不仅给国家直接带来经济损失,而且还威胁着森林消防人员的生命,2019年3月,四川省凉山州发生森林火灾。四川森林消防总队凉山州支队指战员,和地方扑火队员共689人在原始森林展开扑救。扑火行动中,受风力风向突变的影响,突发林火爆燃,瞬间形成巨大火球,在现场的扑火人员紧急避险,但27名森林消防指战员和3名地方扑火人员失联。为避免人员伤亡,失踪等重大事故的发生,人类或机器人的定位与高分辨的地理环境建图就成为了急需解决的问题。

1.3 国内外研究现状

1.3.1 视觉 SLAM 研究现状

SLAM 技术发展至今,根据传感器类别的不同,大致为激光 SLAM 与视觉 SLAM 两大分支。激光 SLAM 技术虽然相对成熟,但激光雷达成本较高,使其发展受到一定的约束,而视觉 SLAM 技术利用的是相机,相机采集的图像包含的信息丰富,不仅可以定位,还可以通过图像进行物品的识别,是当下 SLAM 技术研究的热点。近些年,视觉 SLAM 技术已经得到了很大的发展。根据使用场景,传感器又能分为 RGBD 相机^[9]、单目相机^[10]、双目相机^[11]。

2008年, Klein 等提出 PTAM (Parallel tracking and mapping)^[12],是一种基于关键帧的单目视觉 SLAM 算法。该项目开创性地将视觉 SLAM 分为前端与后端,所以分为两个线程,一个线程为 Tracking 线程,负责根据每一帧实时计算更新相机位姿并建立局部稀疏地图。另一个线程为 Mapping 线程负责地图优化,由于地图的更新不需要每一帧,所以可以选出关键帧进行非线性优化,这种方法得到的地图具有更高的精度。但是该方法只适用于小场景,并且如果后端优化时间过长没有得到最新的地图很容易导致跟踪丢失,鲁棒性差。

2014年 Engel 等提出 LSD-SLAM (Large-Scale Direct Monocular SLAM)^[13],该算法是一种将直接法应用到半稠密的单目视觉 SLAM 算法。LSD-SLAM 算法分为跟踪模

块、深度地图估计模块和地图优化模块三个部分。通过深度滤波器计算像素点的深度。

2015 年 T Pire 等提出 S-PTAM (Stereo parallel tracking and mapping)^[14] 是一种在 PTAM 的基础上基于 Stereo 视觉的 SLAM 算法。该方法采用了 Stereo 相机, 避免了单目相机产生的自举问题, 在没有任何先验信息的情况下, 可计算出环境的真实尺度。而且最小化了线程间的依赖性。追踪线程采用 BRIEF 描述子, Shi-Tomasi 特征提取, 较 PTAM 提高了系统的鲁棒性。

2016 年 Jakob 等提出了 DSO (Direct Sparse Odometry)^[15], 该方案采用纯直接法计算视觉里程计。DSO 的出现使得采用直接法计算视觉里程计的技术变得成熟。按照经典的 SLAM 框架划分, 由于它缺少回环检测, 地图复用的功能, 所以不是一个完整的 SLAM 系统。所以, 随着时间的推移, 不可避免地会出现累计误差, 而且也没有其他消除累计误差的方案。

2017 年, Raul Mur-Artal 等提出了 ORB-SLAM2 算法^[16], 该系统支持单目、双目、RGBD 等多种相机, 分为跟踪、建图、重定位、闭环检测四个模块, 在大规模、小规模、室内室外的环境都可以运行。对于剧烈运动该算法也体现出很强的鲁棒性, 它支持宽基线的闭环检测和重定位, 而且比较容易初始化, 与 PTAM 算法相比有了较大的提升。但是该算法只能生成稀疏的点云地图, 应用在机器人导航方面还有一定的差距。

1.3.2 视觉惯性 GPS 融合定位研究现状

目前传统的定位导航工具有 GPS 定位和 IMU 惯性传感器定位, 但是 GPS 信号会衰减, 被反射, 在城市, 封闭的环境, 信号接收困难, 会导致定位出错。为了缓解独立使用 GPS 导航出现的缺陷, 经常采用耦合 GPS 和 IMU 传感器的测量方案, 还可以有效地使 IMU 传感器标定和惯性导航的误差得到控制, 实现更好的输出。除了传统的导航定位技术外, 随着计算机视觉科学的进步, 激发了研究人员们对于视觉定位的兴趣。单目相机一般会造成尺度模糊, 所以在复杂的环境中, 通常相机与 IMU 传感器结合一起估计机器人运动。当前传感器融合技术大致分为紧耦合与松耦合两种方案, 紧耦合的精度会更好, 但是在某些情况下容易出现发散的问题, 而松耦合方案具有高扩展性和系统稳定性。

2013 年 Shi Y 等提出了在户外环境中采用全景相机与带有 GPS 支持的严格传感器模型的视觉 SLAM 方案^[17]。为了限制不确定性的传播, 通过 GPS 提供全球地理配准信息来进行全局优化。研究了 GPS 支持的嵌入 6 自由度模型的 BA-SLAM^[18]方法, 其中严格的传感器模型被应用到几何投影模型中, 最后通过几个地面控制手动测量点来检查 GPS 支持的 BA-SLAM 绝对精度, 但该方法特别依赖 GPS 信号, 高层建筑会导致 GPS 信号长时间锁定。

2015 年 Amani Ben Afia 提出了一种基于联合卡尔曼滤波的积分继承系统^[19], 该方案整合了惯性测量单元 IMU, GPS 接收器, 轮速传感器和单目视觉 SLAM 等多个传感

器，通过 IMU 进行状态预测，然后参考其他传感器来修正预测状态的值。

2016 年 Vetrella A R 等提出了一种基于扩展卡尔曼滤波算法。该算法将 GPS 与视觉系统相融合^[20]，是一种多机协作式的导航算法，通过视觉系统与 GPS 来改善惯性传感器的导航性能，并利用飞行器之间的差分 GPS 和基于视觉的跟踪来构建虚拟的附加导航传感器，然后将其信息集成到基于扩展卡尔曼滤波器的传感器融合算法中。

2017 年 Yu Song 等提出了一种多传感器融合的状态估计方案^[21]，它通过双目视觉里程计、GPS、气压计和 IMU 等多种传感器进行状态估计，该系统通过随机克隆 EKF 估计器以松耦合的方式将绝对状态测量信息与相对状态估计信息融合起来，其中绝对状态测量由 GPS 与气压计提供。相对状态估计由双目视觉里程计与 IMU 提供。该方案的 GPS 测量结果的好坏由当前卫星数量来确定。

2017 年香港科技大学提出的 VINS-Mono 是一种实时单目视觉惯性融合的 SLAM 系统^[22]。该方案采用了滑动窗口技术，并通过优化窗口中的变量来提供高定位精度，而且利用单目与 IMU 融合恢复尺度，解决了纯单目视觉没有尺度的缺陷。按照融合类型分类，它是一种紧耦合的数据融合方式。它具有有效的 IMU 预积分，偏差校正，自动估计器初始化，故障检测和恢复，回环检测，在线重定位和 4-DoF 全局姿态图优化，实现了位姿图复用。

2018 年 Xiao Chen 等提出了一种低成本全球导航卫星系统和单目相机融合的方案^[23]，首先系统在两个传感器之间对齐坐标，在每个关键帧选出来之后，将各自的测量结果通过构建目标函数并进行优化求解的方案进行融合，系统可以实时运行获得车辆的绝对位姿。

2018 年 Paul 等提出了一种朝向天空的相机与 GPS 信号融合的方案^[24]，该方案需要通过朝天空的相机拍照对天空进行分类，分为非天空区域（建筑物遮挡）和天空区域，通过卫星信号的反射角度判断哪些卫星处于非天空区域，位于非天空区域的卫星不参与最后位置的解算。

2019 年发布了 vins-fusion 项目^[25]，该项目包含单目与 IMU 融合、纯双目、双目与 IMU 融合、双目与 IMU，GPS 融合等方案。双目 SLAM 与单目 SLAM 相比在初始化方面有一定的优势，静止的情况下就可以初始化。在与 GPS 融合的方案中，初始位置由 GPS 提供，当接受到 GPS 数据时与 VIO（Visual Inertial Odometry）数据，通过 ceres 库构建代价函数进行优化求解。

2019 年 Adam Kalisz 提出了 DSO-GPS 的融合方案，该方案采用 EKF（Extended Kalman Filter）滤波的方式进行融合，首先由 GPS 对初值进行初始化。在预测步骤中采用 IMU 运动模型，更新步骤中采用传感器测量模型来修正，传感器测量采用 GPS 和 DSO 两种方式，所以在更新步骤中，通过依赖传感器测量的可靠程度来选择测量方

式，分别是在更新步骤中仅采用 GPS 测量、仅采用 DSO 测量和 GPS 与 DSO 测量结果加权平均三种方式。但是 DSO 并不是一个完整的 SLAM 系统，没有地图重用的功能。

1.3.3 二维网格地图研究现状

目前大多数视觉 SLAM 系统构建的地图只有稀疏特征地图如 ORB-SLAM、VINS、LSD-SLAM^[26]等，但是稀疏特征地图并不能满足机器人对障碍物的检测以及导航，很多研究人员基于这些方案构建能够满足机器人避障导航的网格地图，3D 栅格地图包含的障碍物信息丰富，但是对于嵌入式系统设备来说，没有足够的内存与计算能力。二维网格地图则对于内存和计算能力没有那么高的要求。

2011 年 Santana 等提出一种使用平面信息即单应矩阵通过单目相机构建视觉二维占用网格地图的方法^[27]。该方法需要进行分割步骤以将图像的各个部分分类为地板或非地板。通过这种分类，可以确定图像的哪些部分是自由的，图像的哪些部分是障碍。但是由于单目视觉没有尺度信息，所以通过此方法建立的二维网格地图在现实中的机器人定位和导航存在缺陷。

2013 年 K Kamarudin 等提出了一种在室内将 Kinect 相机的 3D 深度数据转换为二维地图的方法^[28]，Kinect 相机可以直接得到相机的深度数图，利用每个像素的深度通过几何运算来计算二维地图的 X，Y 坐标。

2016 年 Robert Goeddel 提出了一种快速使用 2D 表示法绘制车辆周围的 3D 结构的方法^[29]，该方案基于 2D 扫描与 3D 点云提取垂直结构的匹配，垂直度限制可确保清晰，易于匹配的扫描，这种 2D 表示形式大大减少了地图存储的需求。

2017 年 Dia R 等提出了一种方法来建立单目标传感器的逆传感器 ISM (Inverse Sensor Model) 概率模型^[30]。评估了传感器精度和网格分辨率变化对逆传感器模型的影响，但并没有通过占据概率构建可用于导航的网格地图。

2017 年 Hull G 等提出了一种基于 LSD-SLAM 的实时占据网格建图^[31]，使用平面信息构建单目 SLAM 地图，利用点云地图投影的方式生成二维网格地图。

2019 年 J Clemens 等提出了一种基于 β 分布的占据网格地图^[32]。传统的网格地图是将每个网格单元的不确定性建模为伯努利分布，但是无法区分由于不同现象引起的不确定性，该文章将占据概率建模为随机变量，假定其服从贝塔分布，分析了造成不确定性的不同原因，并推导了 SLAM 算法与建图所必要的传感器模型，以及具有贝塔分布的随机变量构成的地图。

2019 年 Lichao Xu 提出了通过改进 ORB RGBD SLAM 的方法构建占据网格地图^[33]，而且该方案比较容易部署，除了由 ORB SLAM 自己本身维护的稀疏特征地图之外，还开发了新的二维地图模块来构建和维护二维占据网格地图。二维占据网格地图是由 Visual SLAM 估算的 3D 相机位姿和从 RGBD 相机在当前姿态下观察到的点云信息中提取的 Scan 数据构成的。

2019年 Silva 等基于 ROS (Robot Operating System)^[34]操作系统采用 RGBD 相机进行室内定位与建图实验^[35]。利用 RGBD 传感器分别对 GMapping^[36]和 RTAB-map^[37]两个 SLAM ROS 功能包的定位精度,生成二维网格地图的质量进行了实验分析。

1.4 主要研究内容以及组织结构

1.4.1 主要研究内容

本文主要研究在不同场景下机器人定位的传感器信息融合方法,以及针对现有 SLAM 系统构建的稀疏点云地图无法进行机器人导航以及对障碍物的判断的情况,进行构建可用于机器人避障导航的二维网格地图的研究。具体描述如下:

(1) 纯视觉定位并不满足任何场景,在空旷的地方,特征纹理不明显的地方,纯视觉定位很容易跟丢,IMU 测量会随着时间累积,由于积分产生的误差越来越大,GPS 传感器在高山,高楼,隧道等场景下。测量结果误差会很大,但是可以提供一个绝对定位。而且不会存在累积误差。本文研究了多种传感器的融合方法,对 VO 重投影误差,IMU 测量误差,噪声模型,以及 GPS 的测量模型,进行了推导,本文基于 EKF,利用 IMU 确定动力学方程作为系统状态方程,采用视觉里程计相对测量的观测模型来更新系统状态方程。当视觉里程计跟踪效果欠佳的情况下也可以利用 GPS 绝对测量来更新系统状态。

(2) 卫星信号的好坏通常由当前 GPS 接收机接收到的卫星个数来确定,卫星数量越少,协方差越大,定位误差就会很高,这样的判断效率并不高。因此可以通过判断当前状态的残差平方和增量是否异常增大来判断当前值是否异常。

(3) 目前大多数视觉 SLAM 系统只能输出相机轨迹与稀疏点云地图,无法适用于机器人进行导航、避障任务。通过单目相机虽然可以构建网格地图,但没有尺度信息,RGBD 相机虽然具有尺度信息,但是在室外受环境光影响比较大。本文在 ORB-SLAM2 的基础上,采用双目视觉计算里程计计算相机位姿,DepthToScan 通过立体视觉的深度图像得到深度信息,在此基础上采用逆传感器模型计算占据概率,并在 ROS 通用平台下与 Occupancy Grid Map 结合重新构建了一种可用于机器人路径规划、导航的二维占据网格地图算法。

1.4.2 组织架构

论文组织结构如下:

第一章绪论:阐述了论文的研究背景与意义,分析了当前视觉 SLAM 的发展现状,梳理了 VO、IMU、GPS 传感器融合方法的研究现状,以及二维网格地图的研究状况。并对本文的主要研究内容以及组织结构进行阐述。

第二章相关技术研究：介绍了双目相机模型、深度计算原理，基于 ORB 特征的视觉里程计，同时还研究了 IMU 传感器的基本原理，噪声模型，GPS 的定位原理等相关技术。

第三章视觉惯性 GPS 融合定位的 VI-rel 与 IG-abs 算法研究：针对纯视觉定位的局限性，采取 VO IMU GPS 等多个传感器进行信息融合。分析了视觉重投影误差与 IMU 残差模型，并建立联合优化函数。此外，提出了 VO 相对定位与 GPS 绝对定位的方法，利用先前视觉定位或者 GPS 的状态来作为观测模型来更新 EKF 系统。同时在配准方面采用了残差平方和是否异常增大的方法来判断结果是否异常。并通过实验进行数据分析。

第四章二维占据网格地图构建的 VO-Scan 算法：推导了 ISM 逆传感器模型与二维网格地图模型，设计了双目立体视觉构建二维网格地图的实现方案。

第五章实验分析：通过搭建实验设备，在真实场景进行室内与室外的实验，在 ROS 系统的基础上实现二维网格地图的构建，同时能够为机器人进行避障和导航。

第六章总结与展望：对所做的工作进行总结。

第2章 相关技术研究

2.1 双目视觉里程计

视觉里程计由于采用不同的视觉传感器有其不同的应用场景。常见的视觉传感器有，单目相机，RGB-D 相机和双目相机。单目相机结构简单成本较低，方便标定和识别，但是没有尺度信息，无法确定一个物体的真实大小，真实环境的轨迹与测量值存在一个尺度因子，具有尺度不确定性。RGB-D 相机通过结构光或者 ToF（Time of Fly）的物理方法测量物体的深度信息，但是测量范围窄，噪声大，视野小，容易受日光干扰，主要用在室内，室外很难应用。双目相机类似人眼，两个镜头之间存在一个基线，基线距离越大，测量的距离就越远，而且能够得到一个真实的尺度，可以运用到室内与室外的环境，但是配置与标定较为复杂，需要消耗大量的资源来计算误差^[38]。为了满足在复杂环境中的使用，决定采用双目相机作为视觉传感器。

2.1.1 双目相机模型

双目相机由两个单目相机组成，一般分为左眼相机和右眼相机，两个相机的光圈中心都水平放置在 x 轴上，两个光圈中心的距离为双目相机的基线，相机出厂后这个值就是固定的，是双目相机的重要参数，真实尺度可根据基线参数进行计算。

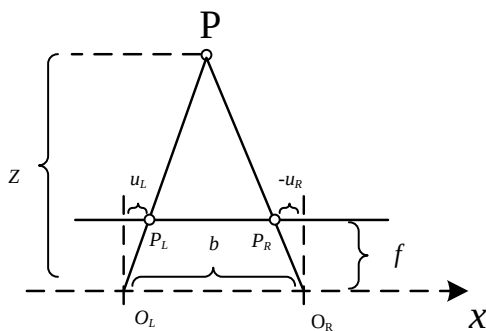


图 2-1 双目立体视觉模型

Fig. 2-1 Stereo vision model

$$\begin{aligned} d &= u_L - u_R \\ z &= \frac{fb}{d} \end{aligned} \quad (2-1)$$

图 2-1 中 f 为焦距， b 为基线， u_L 为左像素点的横坐标， u_R 为右像素点坐标， d 为左右像素点的横坐标之差，称为视差， z 即为通过视差值计算出得测量距离。

2.1.2 ORB 特征提取与匹配

ORB^[39] (Oriented Fast and Rotated Brief) 可以用来快速地对图像中的关键点建立可用于识别图像中对象的特征向量,如图 2-2 所示。其中,FAST^[40](Features from Accelerated Segments Test) 和 BRIEF^[41] (Binary Robust Independent Elementary Features) 分别是特征检测算法和向量创建算法,是一种使用二进制字符串描述特征点的方法。它利用图像金字塔实现,在一定范围内可以保证尺度的不变性。利用了灰度质心法计算出特征的方向,然后计算旋转后的 BRIEF 描述子,保证了旋转不变性。计算 BRIEF 的时候并不是使用一个点的灰度,而是使用了点周围区域的灰度,提高了抗噪能力。不仅如此,它还有最重要的一个优点就是计算速度也很快,速度是 SIFT 的 100 倍, SURF 的 10 倍,是实现实时系统的关键。



图 2-2 ORB 特征点

Fig. 2-2 ORB feature points

ORB 特征检测算法首先会寻找关键点,也就是图像中比较特殊的区域。通常选择像素值从浅色急剧变化为深色的区域,例如角点。找到关键点后,ORB 算法会计算出每个关键点的特征向量。由于创建的这些特征向量中只包含了 1 和 0 两种情况,所以称之为二元特征向量。如果关键点和其周围的像素区域发生变化,则该关键点的特征向量中 1 和 0 的顺序也会随之变化。关键点的特征向量表示它周围的像素的强度模式,如果要识别更大的区域或则图像中的特定对象,可以使用多个特征向量。ORB 算法除了计算速度超快之外,在一定程度上,旋转和缩放变换也不会受到影响。

ORB 特征检测算法的第一步是利用 FAST 算法查找图像中的关键点,它可以快速选择关键点,算法步骤如下:

在图中给定一个像素点 p 像素值为 I_p , FAST 算法来比较像素点 p 周围的 16 个像素,每个像素按照高于 I_p , 小于 I_p , 与 I_p 相似进行分类,一共分为三类。通常并不是直接按 p 点像素值划分,而是给 p 点像素值引入一个阈值 h , 将亮度高于 I_p+h 像素值的分为一类,经对比并对像素分类后,如果像素点 p 周围的 16 个像素点上存在 8 个以上的像素相连,则将像素点 p 选作关键点,如图 2-3 (a) 所示。而 FAST 如此高效的原因是,仅将像素点 p 与周围区域中的 4 个等距像素相比,如图 2-3 (b) 所示,这种方

法已经证明和比较 16 个周围像素的效果相同。如果这四个等距像素中至少存在一对连续像素的亮度高于或低于 p 点像素值, 则将 p 选作关键点, 通过优化使得在整张图像中搜索关键点的时间缩短了四倍。

ORB 特征检测算法的第二步是为第一步中发现的关键点计算特征向量, 这些特征向量可以共同表示一个对象。ORB 算法利用 BRIEF 算法为关键点创建特征向量。它的作用是根据一组关键点创建二元特征向量。二元特征向量又称为二元描述符, 是仅包含 1 和 0 的特征向量, 1 和 0 代表关键点附近两个像素的大小关系。由于内存中存的都是 1 和 0, 所以计算速度也很快。在 BRIEF 中每个关键点由一个二元特征向量描述, 该向量一般为 128-512 位的字符串, 代表关键点周围 128-512 个像素, 关键点周围的像素点一般是随机选取。描述子距离表示了两个特征之间的相似程度。对于二进制描述子通常采用汉明距离作为度量。

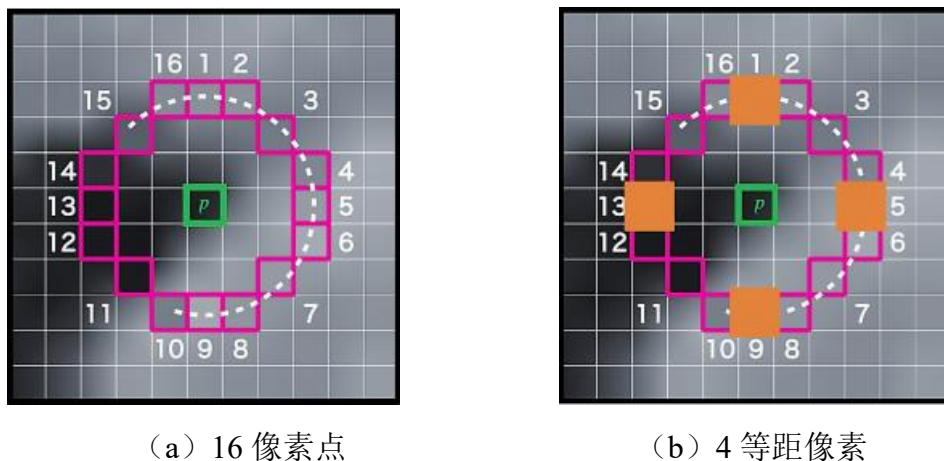


图 2-3 FAST 特征提取表示

Fig. 2-3 FAST feature detection illustration

2.1.3 视觉里程计位姿表示

本节根据每帧的特征点来求解相机的位姿, 一般分为以下几种情况。如果是单目相机, 空间点位置信息未知, 属于 2D-2D 位姿求解类型, 采用对极几何的方法, 如果采用的是 RGBD 相机或者是双目相机, 可以测量出空间的深度信息, 那么空间 3D 点坐标与其对应的特征点像素位置是已知的, 属于 3D-2D 位姿求解类型, 采用 PnP^[42]算法求解, 如果已知的是两帧图像所对应同一空间点坐标, 属于 3D-3D 求解类型, 采用 ICP^[43]算法求解。

特征点匹配后, 建立了 3D 到 2D 的对应关系后就可以通过 PnP 算法求解当前相机位姿, 即相机的旋转矩阵 R 与平移 t 。

已知空间点 P 的齐次坐标为 $(X, Y, Z, 1)$, 对应图像中特征点的齐次坐标为 $(u, v, 1)$ 。当前帧下在世界坐标系下的旋转矩阵为 R , 平移为 t , 由于 R, t 都是待求解的变量, 所以定义一个增广矩阵来同时表示旋转与平移。

定义增广矩阵为：

$$T = (R \parallel t) = \begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_5 & t_6 & t_7 & t_8 \\ t_9 & t_{10} & t_{11} & t_{12} \end{pmatrix} \quad (2-2)$$

空间点 P 与像素之间存在如下关系：

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_5 & t_6 & t_7 & t_8 \\ t_9 & t_{10} & t_{11} & t_{12} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2-3)$$

将矩阵 T 的行向量重新表示如下：

$$\begin{cases} t_1 = (t_1 & t_2 & t_3 & t_4)^T \\ t_2 = (t_5 & t_6 & t_7 & t_8)^T \\ t_3 = (t_9 & t_{10} & t_{11} & t_{12})^T \end{cases} \quad (2-4)$$

通过简化表示可以将式 (2-4) 简写为：

$$\begin{cases} t_1^T P - t_3^T P u = 0 \\ t_2^T P - t_3^T P v = 0 \end{cases} \quad (2-5)$$

此时根据图像中的一个特征点可以列出两个方程，但是代求变量 T 是有 12 个未知变量，那就至少需要 6 对特征点来进行求解，由于采用的是相机，图像中特征点数量多，所以很容易找到其他特征点，该方法被称为直接线性变换法。由上面可列出方程组如下所示：

$$\begin{pmatrix} P_1^T & 0 & -u_1 P_1^T \\ 0 & P_1^T & -v_1 P_1^T \\ \vdots & \vdots & \vdots \\ P_N^T & 0 & -u_N P_N^T \\ 0 & P_N^T & -v_N P_N^T \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 0 \quad (2-6)$$

此刻求解出的 T 并不一定满足要求，因为旋转矩阵 $R \in (SO_3)$ ，求解出的 T 矩阵里面的每一个元素之间没有任何约束，它只是一个一般矩阵，对于 T 中的旋转矩阵 R 必须通过直接线性变换来估计 T 左边的 3×3 的矩阵块，可以通过 QR 分解寻找一个最好的旋转矩阵对它进行近似。

上述方法为直接线性法，求解 3D-2D 问题除以上方法外还可以采用 P3P 算法，它是利用三对 3D-2D 点通过空间几何的方法来求解旋转矩阵 R 与平移 t 的。记 3D 点为 A, B, C ，2D 点为 a, b, c 其中小写字母代表的点为对应大写字母代表的点在相机成像平面上的投影，如图 2-4 所示。相机光心为 O ， A, B, C 是在世界坐标系下的坐标，不是相机坐标系下的坐标。

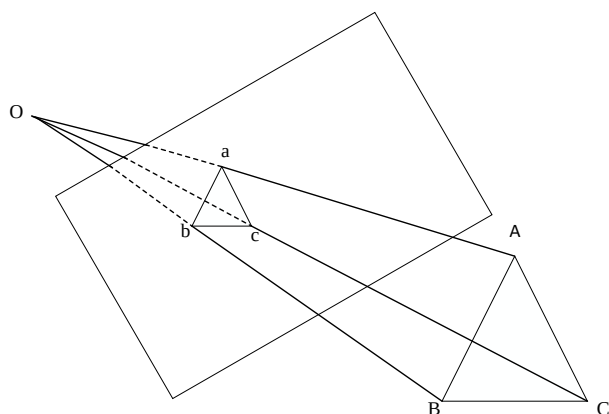


图 2-4 P3P 算法求解示意图

Fig. 2-4 P3P algorithm to solve the schematic diagram

2.2 IMU 传感器姿态估计

IMU 是惯性测量单元，它由加速度计与陀螺仪两部分组成。在惯性坐标系下，通过加速度计可以测得载体的三个轴向的加速度信息，对加速度积分运算可得载体的速度信息，对速度积分可得载体的位移信息。也可以通过陀螺仪测得载体旋转角度，通过矩阵运算，可以算出载体的横滚角、俯仰角以及偏航角等姿态信息，灵敏度高。

2.2.1 IMU 传感器基本原理

IMU 加速度计测量模型可以简单的看作质量块、弹簧、与指示计三个部分来表示，加速度计的测量值 a_m 为弹簧拉力对应的加速度，如图 2-5 所示：

$$a_m = \frac{f}{m} = a - g \quad (2-7)$$

f 表示弹簧拉力， a 表示物体在惯性系下的加速度， g 表示重力加速度

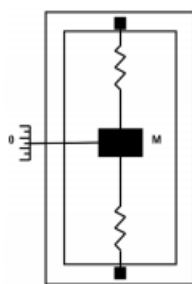


图 2-5 加速度测量

Fig. 2-5 Acceleration measurement

陀螺仪主要用来测量物体的旋转角速度，它的原理就是陀螺在转动的过程中会产生角动量，通过角动量信息就能计算出载体的运动方向的信息。以 MEMS 陀螺为例，它

包含一个主动轴和一个敏感轴，在旋转的坐标系中，运动的物体会受到科氏力的作用。采用震动陀螺原理通过测量科氏力间接获得角速度。如图 2-6 所示：

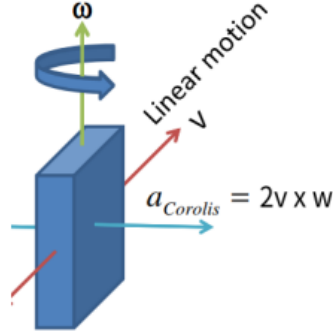


图 2-6 陀螺仪测量

Fig. 2-6 Gyroscope measurement

2.2.2 IMU 噪声模型

IMU 通过加速度计和陀螺仪测量的加速度与角速度并非 IMU 真值，测量结果会包括测量噪声和偏置等误差，加速度计和陀螺仪的误差可以分为确定性误差和随机误差。确定性误差可以事先标定来确定，例如器件出厂时由于制作工艺造成的误差，随机误差通常假设噪声服从高斯分布，包括高斯白噪声，随机游走偏差等。理想情况下，当 IMU 传感器没有受到外部作用时，输出的值应该为 0，但实际数据并不为 0，因此，存在一个偏置，该偏置就是一个确定性的误差，可以通过六面标定法得到一个尺度因子来消除偏置。由于尺度因子可以人为消除，在处理随机误差时，先忽略其影响。只考虑白噪声和随机游走偏差时的 IMU 误差模型如下：

$$\begin{cases} \tilde{\omega}^b = \omega^b + b^g + n^g \\ \tilde{a}^b = q_{bw}(a^w + g^w) + b^a + n^a \end{cases} \quad (2-8)$$

上标 g 表示陀螺仪数据，上标 a 表示加速度计测量数据， w 表示在世界坐标系 world， b 表示 IMU 坐标系 body。 ω 和 a 表示 IMU 的真实值， $\tilde{\omega}$ 和 $\tilde{a}$ 表示测量值。

2.2.3 IMU 运动模型

IMU 的位移 Pose，速度 Velocity，旋转 Quaternion 分别用 p 、 v 、 q 表示，它们对时间求导可得：

$$\begin{aligned} \dot{p}_{wbt} &= v_t^w \\ \dot{v}_t^w &= a_t^w \\ \dot{q}_{wbt} &= q_{wbt} \otimes \begin{bmatrix} 0 \\ \frac{1}{2} \omega^{bt} \end{bmatrix} \end{aligned} \quad (2-9)$$

从 i 时刻的 p 、 v 、 q ，通过 IMU 的测量值进行积分，得到第 j 时刻的 p 、 v 、 q

$$\begin{aligned} p_{wb_j} &= p_{wb_i} + v_i^w \Delta t + \iint_{t \in [i, j]} (q_{wb_t} a^{b_t} - g^w) \delta t^2 \\ v_j^w &= v_i^w + \int_{t \in [i, j]} (q_{wb_t} a^{b_t} - g^w) \delta t \\ q_{wb_j} &= \int_{t \in [i, j]} q_{wb_t} \otimes \begin{bmatrix} 0 \\ \frac{1}{2} \omega^{bt} \end{bmatrix} \delta t \end{aligned} \quad (2-10)$$

2.3 GPS 传感器定位技术

对于许多实际应用需要精确地定位到地球上的位置，美国在 1973 年建设了 GPS 全球定位系统，它是通过接收卫星的无线电信号来工作的，不幸的是，GPS 的使用仍然存在一定的挑战，当处于高山，高楼林立的环境中，卫星信号很容易被遮挡，或者卫星信号被反射，这样就会导致 GPS 无法计算出目标的位置，或者计算出一个误差很大的位置。尽管如此，在空旷的环境中 GPS 的应用还是很广泛的。

2.3.1 GPS 定位原理

GPS 定位系统由地面站、接收机以及 24 颗卫星均匀分布在空中组成，信号覆盖全球，GPS 卫星时时刻刻都在向全球广播自己的位置，所以在地球上的任何位置，都会观测到 GPS 定位卫星。

为了计算其在三维空间中的位置，GPS 定位需要至少四颗卫星来计算接收机在地球上的位置，而且，接收机必须在每个卫星的信号上保持锁定一段时间，足以接收编码信息的时间。实际上，通过计算出卫星到接收机的伪距，伪距测量表示 GPS 卫星和接收器天线之间的距离，再通过数学模型即可计算出接收机的经纬度以及高度。接收机生成 PRN 码副本与从卫星接收到的副本对齐所需要的时移量，可从 PRN 码中获得，GPS 接收器使用此信息来确定卫星信号从卫星传播到接收机的时间按光速提供伪距测量，因为卫星和接收器的时钟并不完美与 GPS 时间同步，测量结果包含时钟偏差。因此，只通过伪距计算，会有一定的误差，为了减少误差，实际上，需要考虑一个时钟漂移 b ，所以在计算公式中需要将其考虑在内。假设接收机的位置为 $[x, y, z]^T$ ，每一个卫星到接收机的伪距为 S_i ，卫星的位置为 $[x_i, y_i, z_i]^T$ ，那么计算公式为：

$$S_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} + b \quad (2-11)$$

上式中接收机的位置有三个未知数，时钟差也是一个未知数，因此，一共四个未知数，求解时需要四组方程，这也是至少需要四颗卫星来计算位置的原因。如图 2-7 所示：

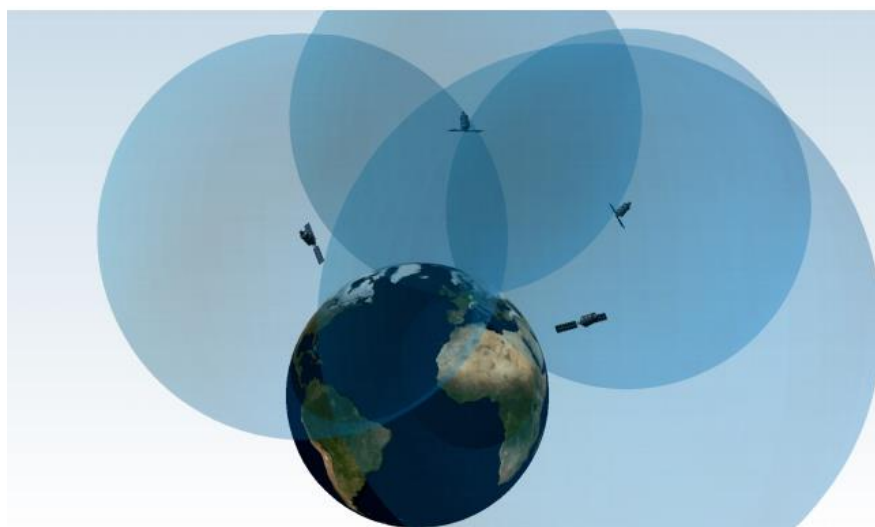


图 2-7 GPS 定位

Fig. 2-7 GPS positioning

2.3.2 GPS 定位误差

GPS 的误差可以分为电离层误差、对流层误差、接收机钟差和卫星星历差异性以及多路径效应^[44]等。电离层和对流层的误差可以通过相应的模型进行消除或者减少，接收机钟差作为一个待估参数，可以求解出钟差漂移。

GPS 定位会受同步时钟误差，天线相位中心偏差，天线信号噪声，以及接收机所在环境的影响，而且电磁波的传播也会受到干扰，所以定位会存在误差，测得距离称为伪距。

2.4 本章小结

本章主要介绍了三种传感器的基本原理，分为三个部分。第一部分介绍双目视觉里程计，首先介绍了双目相机模型，以及通过相机内参、基线、视差图计算得到深度图。为了满足系统实时性要求，本文选择了可以快速提取特征和匹配的 ORB 特征子，通过匹配到的图像，利用 PnP 算法来估计相机位姿。第二部分介绍了 IMU 传感器基本原理，噪声模型，以及运动模型，并推导了利用加速度与角速度，求得 IMU 姿态的具体积分过程，方便后面进行传感器数据融合。第三部分则是论述了 GPS 的定位方式，以及 GPS 定位的解算过程。GPS 定位最少需要接收到四颗以上的卫星才可以进行解算位置，还论述了其定位环境，在高楼大山等环境中 GPS 信号会受到影响，定位精度会大大降低，甚至定位失败。

第3章 视觉惯性 GPS 融合定位的 VI-rel 与 IG-abs 算法研究

视觉里程计成本低, 功耗低, 而且还可以提供令人满意的精度, 双目相机可以通过基线计算出环境的真实尺度。在测量快速运动的物体的姿态时 IMU 被广泛的应用, 但随着时间的推移, 会存在累积误差, 使得测量结果越来越不精确。视觉里程计与 IMU 都是测量的相对位置, 要想测量出在地球上的绝对位置, 可以采用 GPS 定位, 而且也可以通过 GPS 去修正 IMU 产生的累积误差, 但是 GPS 采样速率比较慢, 容易受环境影响导致定位不精确, 所以可以将这些传感器的信息融合起来提供一个精确的定位^[45]。

传感器信息融合一般分为紧耦合与松耦合, 本文将通过 IMU 提供状态预测, GPS 提供初始位置, 当 GPS 信号受到干扰的时候, 我们采用紧耦合的方式将视觉里程计与 IMU 融合建立联合优化目标函数求解。当视觉里程计所处环境纹理特征差的时候, 此刻一般处于空旷的场景, 采用 GPS 提供观测信息来修正 IMU。如图 3-1 所示:

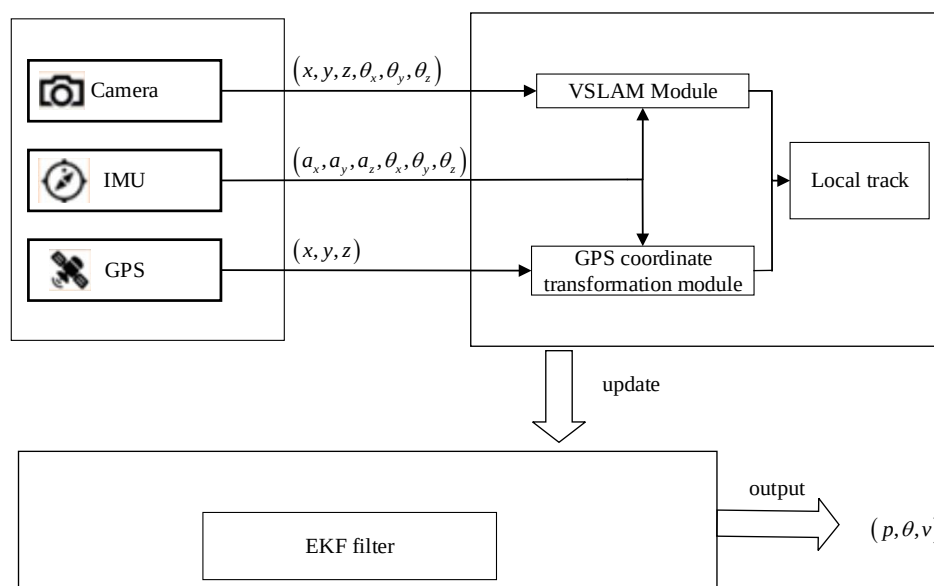


图 3-1 视觉惯性 GPS 融合定位

Fig. 3-1 Visual inertial GPS fusion positioning

3.1 紧耦合 IMU 的视觉里程计

3.1.1 联合优化函数

视觉里程计中通过紧耦合的方式集成 IMU 测量信息有一个好处, 就是为视觉里程计的运动提供了一个很好的刚体运动先验约束。

$$\varepsilon = \min \left\{ \underbrace{\|r_p - H_p X\|^2}_{\text{先验信息}} + \underbrace{\sum_{k \in B} \|r_B(\tilde{Z}_{b_j}^{b_i}, X)\|^2}_{\text{IMU 残差}} + \underbrace{\sum_{(l,j \in C)} \|r_C(Z_l^{c_j}, X)\|^2}_{\text{视觉重投影误差}} \right\} \quad (3-1)$$

该联合优化函数分为三部分，第一部分为先验信息，第二部分为 IMU 残差和，第三部分为重投影误差。视觉里程计与 IMU 联合建立优化目标函数，通过误差模型对状态变量一阶泰勒展开可得雅可比矩阵，得到雅可比矩阵即可采用 LM 算法^[46]进行迭代求解当前姿态：

$$(J^T J + \lambda I) \Delta X = -J \varepsilon \quad (3-2)$$

J 表示雅各比矩阵， ε 表示残差

3.1.2 重投影误差

当前相机姿态 (R, t) 一般通过优化非线性重投影残差函数得到，它是由像素坐标与空间中 3D 点按照当前估计的位置进行投影得到的像素坐标位置之间的误差。如下图所示：

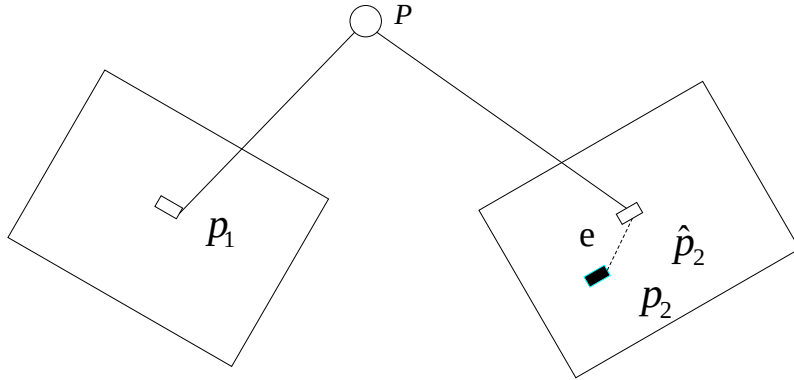


图 3-2 重投影误差

Fig. 3-2 Reprojection error

p_1 与 p_2 是同一个空间点 P 的观测值，点 P 的投影 $\hat{p}_2$ 与观测值 p_2 之间存在一定的误差。为了优化求解，则需要考虑多个空间点 $P_i = [X_i, Y_i, Z_i]^T$ ，其对应的像素坐标为 $U_i = [u_i, v_i]^T$ ，空间点坐标与像素坐标之间的关系如式 3-3 所示：

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K \exp(\xi^\wedge) \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (3-3)$$

s_i 代表像素深度， K 代表相机内参， $\exp(\xi^\wedge)$ 代表相机外参即 (R, t) ，由于观测点存在误差，而且相机外参未知，所以可以构建最小二乘问题，通过 LM 方法求解，使得

误差最小，得到相机位姿，如下所示：

$$\xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) P_i \right\|_2^2 \quad (3-4)$$

3.1.3 IMU 预积分与测量残差

(1) 预积分

IMU 采集数据的速率比较高，而相机的帧率相对较低，则在相邻两张图片之间存在多个 IMU 采集的数据，所以需要 IMU 数据进行预积分^[47]处理，这样有一个好处就是在每次状态更新后不需要马上进行积分运算，减少计算量。

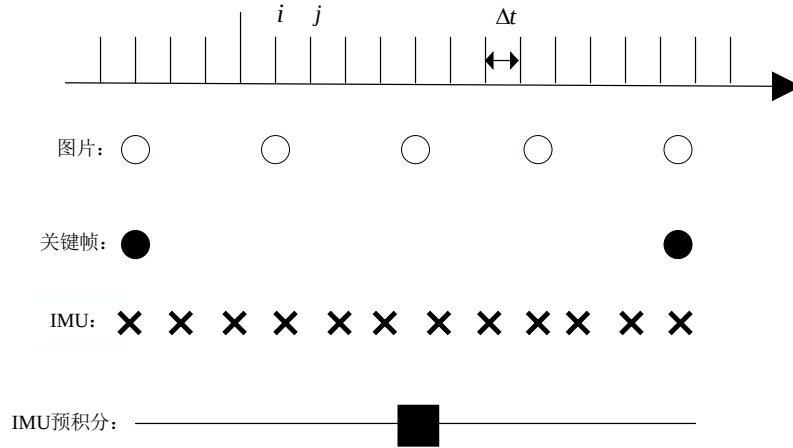


图 3-3 IMU 与相机测量数据关系

Fig. 3-3 Relationship between IMU and Camera measurement data

通过 IMU 噪声模型，需要将相邻两帧之间所有的 IMU 数据进行积分。根据第 i 时刻的位置、速度以及旋转状态求解第 j 时刻的位置、速度以及旋转状态如式 (3-5) 所示：

$$\begin{aligned} p_{b_j}^w &= p_{b_i}^w + v_{b_i}^w \Delta t + \iint_{t \in [i, j]} [R_b^w (\hat{a} - b^a) - g^w] dt^2 \\ v_{b_j}^w &= v_{b_i}^w + \int_{t \in [i, j]} [R_b^w (\hat{a} - b^a) - g^w] dt \\ q_{b_j}^w &= q_{b_i}^w \otimes \int_{t \in [i, j]} \frac{1}{2} \Omega(\hat{\omega} - b^g) q_b^w dt \end{aligned} \quad (3-5)$$

世界坐标系下的位置，速度，旋转状态由 p, v, q 表示 R_b^w 表示将机体坐标系下的加速度转到世界坐标系下， q_b^w 表示将载体坐标系下的角加速度转到世界坐标系下， g^w 代表重力加速度， $\hat{a}$ 表示 IMU 测得的加速度， $\hat{\omega}$ 表示 IMU 测得的角加速度， b^a 表示加速度偏差， b^g 表示陀螺仪偏差。

后端优化过程中，状态更新后，为了避免重复积分第 j 时刻的姿态，所以需要进行预积分，将世界坐标系转变到第 i 时刻的机体坐标系下计算：

$$\begin{aligned}
 R_{b_j}^{b_i} p_{b_j}^w &= R_{b_i}^{b_j} (p_{b_i}^w + v_{b_i}^w \Delta t - \frac{1}{2} g^w \Delta t^2) + \alpha_{b_j}^{b_i} \\
 R_{b_j}^{b_i} v_{b_j}^w &= R_{b_i}^{b_j} (v_{b_i}^w - g^w \Delta t) + \beta_{b_j}^{b_i} \\
 q_w^{b_i} \otimes q_{b_j}^w &= \gamma_{b_j}^{b_i}
 \end{aligned} \tag{3-6}$$

$$\begin{aligned}
 \alpha_{b_j}^{b_i} &= \iint_{t \in [i, j]} (R^{b_i} (\hat{a} - b^a) dt^2 \\
 \beta_{b_j}^{b_i} &= \int_{t \in [i, j]} R^{b_i} (\hat{a} - b^a) dt \\
 \gamma_{b_j}^{b_i} &= \int_{t \in [i, j]} \frac{1}{2} \Omega (\hat{\omega} - b^g) q^{b_i} dt
 \end{aligned} \tag{3-7}$$

第 i 时刻的位姿发生变化的时候, 第 j 时刻的位姿也会受到影响, 但不会影响加速度与角加速度, 所以当位姿发生变化就不需要重新积分。IMU 预积分只与加速度和角加速度有关系。

(2) 测量残差

将 IMU 噪声模型与预积分公式结合, 在两帧之间图像之间的 IMU 数据建立残差模型为:

$$\begin{bmatrix} \Delta \alpha_{b_j}^{b_i} \\ \Delta \gamma_{b_j}^{b_i} \\ \Delta \beta_{b_j}^{b_i} \\ \delta b_a \\ \delta b_g \end{bmatrix} = \begin{bmatrix} R_{b_j}^{b_i} (p_{b_j}^w - p_{b_i}^w - v_{b_i}^w \Delta t + \frac{1}{2} g^w \Delta t^2) - \alpha_{b_j}^{b_i} \\ 2[(\gamma_{b_j}^{b_i})^{-1} \otimes (q_{b_i}^w)^{-1} \otimes q_{b_j}^w]_{xyz} \\ R_{b_j}^{b_i} (v_{b_j}^w - v_{b_i}^w + g^w \Delta t) - \beta_{b_j}^{b_i} \\ b_{b_j}^a - b_{b_i}^a \\ b_{b_j}^g - b_{b_i}^g \end{bmatrix} \tag{3-8}$$

上面式子中 $[\bullet]_{xyz}$ 代表只取四元数中的虚部。该模型在两关键帧之间利用预测值减掉测量值, 从而产生误差, 通过迭代优化可求解未知变量位置 p 、速度 v 、姿态 q 、与 IMU 偏置量。

3.2 VO/IMU 融合定位的 VI-rel 研究

3.2.1 IMU 状态传递动力学模型

基于 IMU 积分方程和偏差模型, EKF 系统的状态 S 定义如下:

$$S = (p, \delta \theta, v_b, \delta b_g, \delta b_a)^T \in R^{15} \tag{3-9}$$

$p \in R^3$ 表示的是世界坐标系下的位置, $\delta \theta \in R^3$ 是 IMU 坐标系下的旋转误差向量,

$v_b \in R^3$ 表示的是 IMU 坐标系下的速度, $\delta b_g \in R^3$, $\delta b_a \in R^3$ 表示当前陀螺仪与加速度计的偏差项。旋转矩阵的估计值被定义为 $\hat{R} \in SO(3)$, 那么旋转矩阵的真实值被定义为 $R \in SO(3)$, 通过右乘矩阵来补偿旋转误差, 如下所示:

$$R = \hat{R} \exp([\delta\theta]_{\times}) \quad (3-10)$$

$[\delta\theta]_{\times}$ 是旋转误差向量 $\delta\theta$ 的矩阵形式。

基于以上系统状态定义, 系统状态的动力学模型预测值为 $\dot{S}$ 被定义为

$$\begin{aligned} \dot{p} &= \hat{R} \exp([\delta\theta]_{\times}) v_b \\ \delta\dot{\theta} &= \exp([\delta\theta]_{\times}) (\hat{\omega} - \delta b_g - n_g) \\ \dot{v}_b &= -[\hat{\omega} - \delta b_g - n_g]_{\times} v_b + (\hat{R} \exp([\delta\theta]_{\times}))^T g + \hat{a} - \delta b_a - n_a \\ \delta\dot{b}_g &= r_g \\ \delta\dot{b}_a &= r_a \end{aligned} \quad (3-11)$$

因此, 预测值对于系统状态的雅各比矩阵 $\frac{d\dot{S}}{dS} \in R^{15}$ 可以计算得出如下结果:

$$\frac{\partial \dot{S}}{\partial S} = \begin{pmatrix} 0_{3 \times 3} & -\hat{R}[v_b]_{\times} & -\hat{R}[\delta\theta]_{\times} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & -[\hat{\omega} - \delta b_g - n_g]_{\times} & 0_{3 \times 3} & -\exp([\delta\theta]_{\times}) & 0_{3 \times 3} \\ 0_{3 \times 3} & [\hat{R}^T g]_{\times} & -[\hat{\omega} - \delta b_g - n_g]_{\times} & -[v_b]_{\times} & -I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \end{pmatrix} \quad (3-12)$$

其中 $I_{3 \times 3}$ 表示的是 3×3 的单位矩阵, $0_{3 \times 3}$ 表示的是 3×3 的零矩阵。包含 IMU 测量噪声和系统噪声 IMU 偏差, 系统输入噪声如下所示:

$$W = (n_g, n_a, r_g, r_a)^T \in R^{12} \quad (3-13)$$

所以动力学预测值 $\dot{S}$ 对于系统输入噪声的雅各比矩阵为:

$$\frac{\partial \dot{S}}{\partial W} = \begin{pmatrix} 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ -I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ -[v_b]_{\times} & -I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \end{pmatrix} \quad (3-14)$$

3.2.2 基于 VO 相对状态的 EKF

状态估计系统采用 VO 相对测量值来更新, 此时系统状态需要找一个先前的 VO

预先测量的姿态 p_{vo} , $\delta\theta_{vo}$ 。此时的系统状态增广为:

$$\tilde{S} = (S^T, p_{vo}, \delta\theta_{vo})^T \in R^{21} \quad (3-15)$$

系统状态传递过程中, 新添加进来的状态 p_{vo} , $\delta\theta_{vo}$ 由于是找的先前的 VO 的测量值, 所以始终为常量, 那就意味着它们的微分都为 0, 则对于新状态 $\tilde{S}$ 的雅各比矩阵为:

$$\begin{aligned} \tilde{J}_s &= \begin{pmatrix} Js & 0_{15 \times 6} \\ 0_{6 \times 15} & I_{6 \times 6} \end{pmatrix} \in R^{21 \times 21} \\ \tilde{J}_w &= \begin{pmatrix} J_w \\ 0_{6 \times 12} \end{pmatrix} \in R^{21 \times 21} \end{aligned} \quad (3-16)$$

增广状态的协方差为 $\tilde{P}(k) \in R^{21 \times 21}$, 则增广系统的协方差传递方程为:

$$\tilde{P}(k+1|k) = \tilde{J}_s \tilde{P}(k) \tilde{J}_s^T + \tilde{J}_w Q(k) \tilde{J}_w^T \quad (3-17)$$

基于 VO 的相对测量模型被定义为:

$$\begin{aligned} \Delta p &= \exp(-[\delta\theta_{vo}]_x) R_{vo}^T (p - p_{vo}) \\ \Delta\theta &= \log(\exp(-[\delta\theta_{vo}]_x) R_{vo}^T \hat{R} \exp([\delta\theta]_x)) \end{aligned} \quad (3-18)$$

式 (3-18) 中 $\Delta p \in R^3$ 是位移 p_{vo} 到当前位移 p 的增量, R_{vo} 是先前状态的旋转矩阵, $\hat{R}$ 是当前状态的旋转矩阵, $\delta\theta_{vo}$ 表示先前旋转角度误差, $\delta\theta$ 是当前旋转角度误差。

那么 VO 相对测量模型的雅各比矩阵表示如下:

$$H_{vo} = \begin{pmatrix} R_{vo}^T & 0_{3 \times 12} & -R_{vo}^T & [R_{vo}^T (p - p_{vo})]_x \\ 0_{3 \times 3} & R_{vo}^T \hat{R} & 0_{3 \times 12} & -I_{3 \times 6} \end{pmatrix} \quad (3-19)$$

EKF 的状态更新如下:

$$\begin{aligned} K &= \tilde{P}(k+1|k) H^T (\tilde{P}(k+1|k) H^T + R)^{-1} \\ \tilde{S}(k+1) &= \tilde{S}(k) + K\tilde{r} \end{aligned} \quad (3-20)$$

协方差更新方程如下所示:

$$\tilde{P}(k+1) = (I - KH) \tilde{P}(k+1|k) (I - KH)^T + KRK^T \quad (3-21)$$

3.3 IMU/GPS 融合定位的 IG-abs 研究

3.3.1 基于 GPS 的观测模型

IMU 与 GPS 融合采用松耦合的方式进行数据融合, 在东北天坐标系下, GPS 可以提供绝对位置和速度, GPS 测量模型为:

$$Z_{gps} = \begin{bmatrix} p + \hat{R} \exp([\delta\theta]_{\times}) t_{ig} \\ R \exp([\delta\theta]_{\times}) (v_b + [\hat{\omega} - \delta b_g]_{\times} t_{ig}) \end{bmatrix} \quad (3-22)$$

其中 $t_{ig} \in R^3$ 表示 GPS 接收机的坐标到 IMU 坐标系的变换矩阵, 所以 GPS 测量模型的雅各比矩阵如下所示:

$$H_{gps} = \begin{pmatrix} I_{3 \times 3} & -\hat{R}[t_{ig}]_{\times} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & -R[v_b]_{\times} & \hat{R} & [t_{ig}]_{\times} & 0_{3 \times 3} \end{pmatrix} \quad (3-23)$$

本文并未考虑高度值, 因为 GPS 在测量高度方面有很大得不确定性, 所以 GPS 的观测模型并未利用高度方向的位移与速度两个状态来更新 EKF 的状态, 而只考虑了东和北两个方向上的状态。

3.3.2 IMU 与 GPS 信号同步

通常 IMU 的采样频率要比 GPS 的采样频率高很多, 他们得到测量结果的时间并不同步。通过线性外推法^[48]使得 GPS 到达时来获得 IMU 的预测位置, 公式如下:

$$P_{IMU}(t_{GPS}) = P_{IMU}(t_{k-1}) + \frac{P_{IMU}(t_{k-1}) - P_{IMU}(t_{k-2})}{t_{k-1} - t_{k-2}} (t_{GPS} - t_{k-1}) \quad (3-24)$$

P_{IMU} 表示 IMU 的预测位置, 当 GPS 得到测量数据后, 系统对 $P_{IMU}(t_{GPS})$ 进行更新。

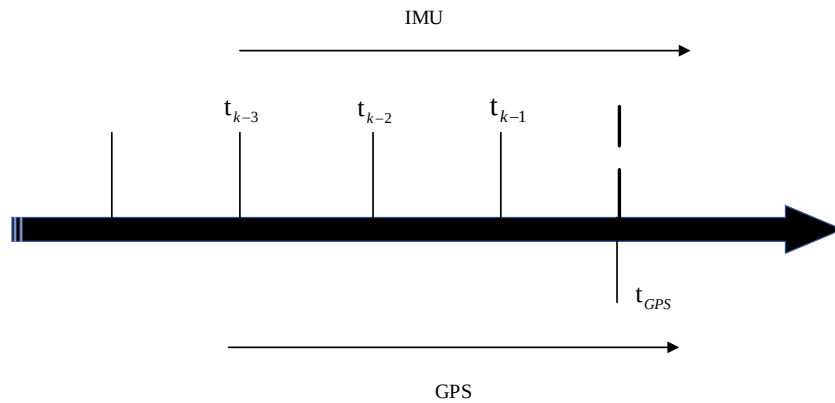


图 3-4 GPS 与 IMU 数据同步

Fig. 3-4 GPS IMU data synchronization

3.3.3 基于残差平方和的 EKF

GPS 接收机可以知道当前被多少颗卫星观测到, 根据观测卫星的数量, 可以知道当前卫星测量的协方差, 如果当前观测卫星数比较少, 可能会得到一个很大的协方差, 这就可以知道观测值准确性, 但是根据卫星数量来判断观测值效率并不高, 所以本文采用基于残差平方和的卡尔曼滤波对系统误差进行最优估计

IMU 与 GPS 数据融合的状态预测方程与观测方程为:

$$\begin{aligned} X_k &= AX_{k-1} + Q \\ Z_k &= HX_k + R \end{aligned} \quad (3-25)$$

状态方程中 A 为状态转移矩阵, H 为测量矩阵, Q 为系统噪声, R 为观测噪声。

根据 $k-1$ 时刻的状态可以预测 k 时刻的状态为:

$$\hat{X}_{k,k-1} = A\hat{X}_{k-1} \quad (3-26)$$

滤波增益为:

$$K = P_{k,k-1}H_k^T(H_kP_{k,k-1}H_k^T + R_k)^{-1} \quad (3-27)$$

根据观测值可以更新预测值如下所示:

$$\hat{X}_k = \hat{X}_{k,k-1} + K(Z_k - H\hat{X}_{k,k-1}) \quad (3-28)$$

则残差平方和增量在第 k 时刻的残差平方和增量为:

$$\Delta x_k^2 = \Delta x_k^T P_{k,k-1} \Delta x_k + (Z_k - H_k \hat{X}_k)^T R_k^{-1} (Z_k - H_k \hat{X}_k) \quad (3-29)$$

上式等号右边为预测值残差平方和增量与观测值残差平方和增量之和。

$\Delta x_k = X_k - X_{k,k-1}$ 为预测值残差, $P_{k,k-1}$ 是根据 $k-1$ 时刻的协方差预测 k 时刻的协方差, $Z_k - H_k \hat{X}_k$ 为观测值残差。

残差平方和增量是利用当前 k 时刻的估计值 $\hat{X}_k$, 说明系统已经完成了更新, 即使发现问题, 也已经晚了。所以需要利用预测值进行计算残差平方和增量。

预测值与估计值之间的差为预测值残差为:

$$\Delta x_k = X_k - X_{k,k-1} \quad (3-30)$$

将公式 (3-27) 和 (3-28) 代入公式 (3-30) 得:

$$\Delta x_k = P_{k,k-1}H_k^T(H_kP_{k,k-1}H_k^T + R_k)^{-1}(Z_k - H_k\hat{X}_{k,k-1}) \quad (3-31)$$

观测值与估计值之差为:

$$Z_k - H_k \hat{X}_k \quad (3-32)$$

将公式 (3-27) 和 (3-28) 代入 (3-32) 化简得:

$$Z_k - H_k \hat{X}_k = R_k(H_kP_{k,k-1}H_k^T + R_k)^{-1}(Z_k - H_k\hat{X}_{k,k-1}) \quad (3-33)$$

那么预测值残差平方和为:

$$\begin{aligned} \Delta x_k^T P_{k,k-1} \Delta x_k &= (Z_k - H_k X_{k,k-1})^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \\ &[I - R_k (H_k P_{k,k-1} H_k^T + R_k)^{-1}] (Z_k - H_k X_{k,k-1}) \end{aligned} \quad (3-34)$$

观测值残差平方和为:

$$\begin{aligned} (Z_k - H_k \hat{X}_k)^T R_k^{-1} (Z_k - H_k \hat{X}_k) &= (Z_k - H_k X_{k,k-1})^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \\ &R_k (H_k P_{k,k-1} H_k^T + R_k)^{-1} (Z_k - H_k \hat{X}_{k,k-1}) \end{aligned} \quad (3-35)$$

由上面的式子可以看出残差平方和增量只与预测值有关,当发现残差平方和增量异常增大时,那就说明该时刻的值存在问题。

3.4 实验探究

3.4.1 VO/IMU 融合性能评估

本节将采用 EuRoC 数据集^[49]对 VO/IMU 融合算法进行评估。它包含两个场景,一个是苏黎世联邦理工学院的 EHT 机房,另一个为 Vicon 会议室。不同的场景中分别包含了容易、中等、复杂三个难度的数据。本文采用了 MH_02_easy 与 V1_02_medium 两个不同场景且不同难度的数据集来验证,并与主流双目视觉惯性 VIO 系统 VINS-fusion 通过 evo 工具箱^[50]做了比较分析。



(a) EHT 机房

(b) Vicon 会议室

图 3-5 EuRoC 数据集场景

Fig. 3-5 EuRoC dataset scenario

该数据集通过飞行器来收集双目视觉与 IMU 的加速度与角速度数据,而且采用了 Leica MS50 激光跟踪扫描仪来提供 groundtruth 的数据采集,定位精度可以达到毫米级,它与视觉传感器之间通过外部校准使得时间戳同步。双目相机为全局快门,单色,相机频率为 20Hz,IMU 测量频率为 200Hz。IMU 数据以视觉图像的时间戳为基准进行对齐。数据集里面提供了 yaml 格式的配置文件,记录了相对于 Body 坐标系的坐标变换,以及传感器自身的参数信息。

图 3-6 展示了本文算法与 VINS-fusion 在 MH_02_easy 数据集下的三维轨迹图,MH_02_easy 是 EuRoC 数据集中比较容易运行的数据集,整体运动平滑,没有剧烈运动。其中图 3-6(a)为 VINS-fusion 运行轨迹,3-6(b)为本文算法运行轨迹,两者都与 rereference 进行对比,图中 reference 为 MH_02_easy 提供的 groundtruth,本文算法与 VINS-fusion 相比,整体上都比较接近 groundtruth。

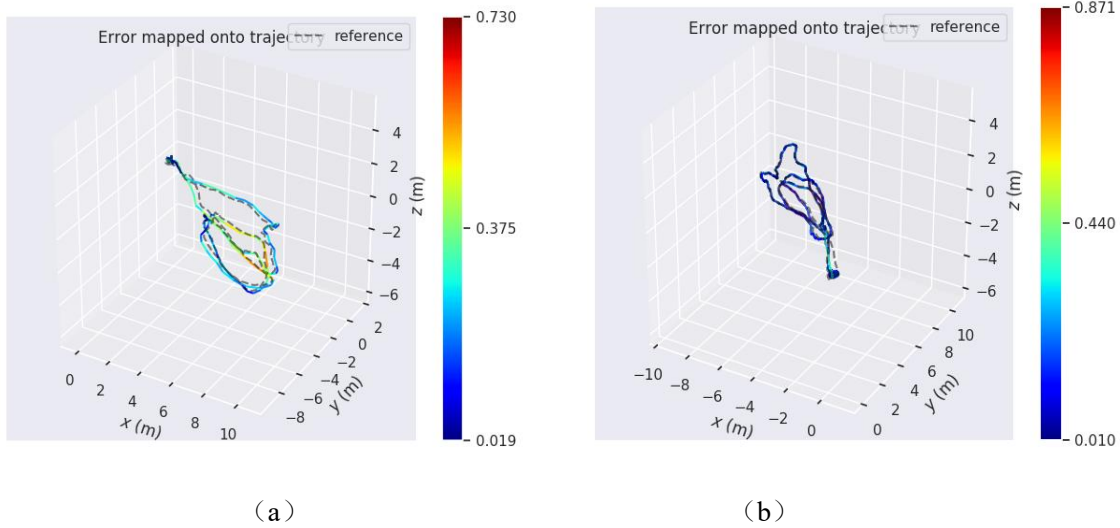


图 3-6 MH_02_easy 轨迹图

Fig. 3-6 MH_02_easy trajectory diagram

图 3-7 展示了本文算法和 VINS-fusion 分别与 MH_02_easy 中 groundtruth 的绝对误差，图 3-7 (a) 是 VINS-fusion 运行结果与 groundtruth 之间的绝对误差，轨迹的最大误差为 0.730125m，最小误差为 0.019285m，平均误差为 0.216125m，RMSE 为 0.240838。图 3-7 (b) 为本文算法运行轨迹与 groundtruth 之间的绝对误差，轨迹的最大误差为 0.87055m，最小误差为 0.009595m，平均误差为 0.114342m，RMSE 为 0.153238。本文算法较 VINS-fusion 相比，平均误差和 RMES 较小，说明系统在没有剧烈运动的场景下性能较好。下面是 V1_02_medium 数据集下的实验。

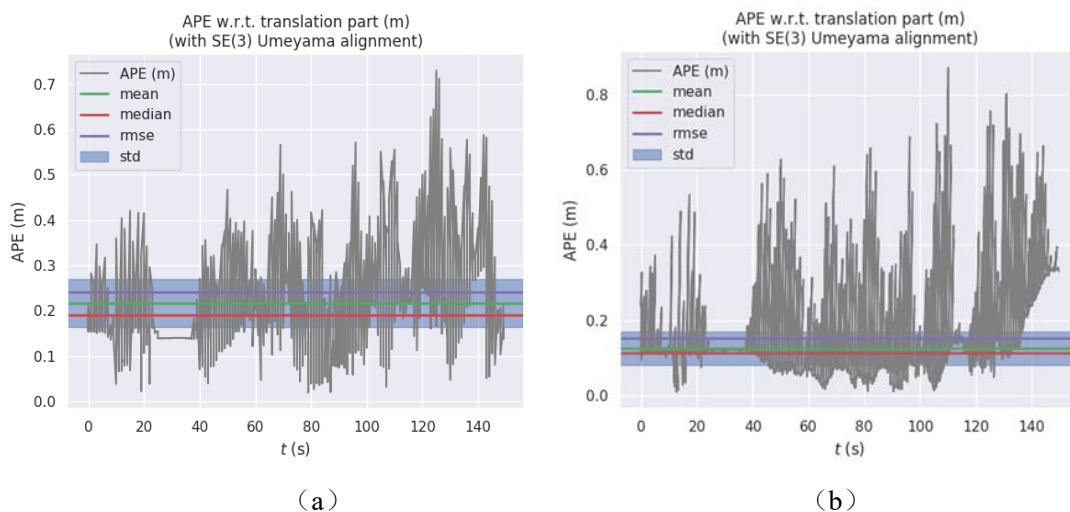


图 3-7 MH_02_easy 绝对误差

Fig. 3-7 MH_02_easy absolute error

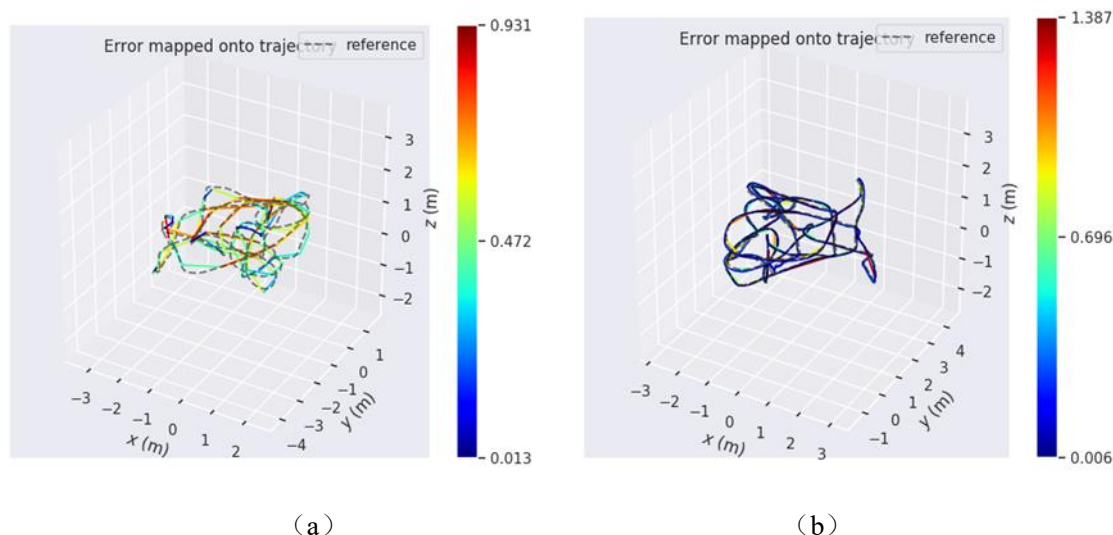


图 3-8 V1_02_medium 轨迹图

Fig. 3-8 V1_02_medium trajectory diagram

图 3-8 展示了本文算法与 VINS-fusion 在 V1_02_medium 数据集下的三维轨迹图，V1_02_medium 是 EuRoC 数据集中较运行难度偏中等的数据集，运动轨迹比较复杂，存在剧烈运动，具有一定的挑战性，像 VIO RB 这样的系统会存在丢失特征点而导致系统运行中断的情况。与 MH_02_easy 相比，误差相对较大。其中图 3-8 (a) 为 VINS-fusion 运行轨迹，3-8 (b) 为本文算法运行轨迹，图中 reference 为 V1_02_medium 提供的 groundtruth。两者都与 rereference 进行对比，本文算法与 VINS-fusion 相比，由轨迹图可知，本文算法性能整体较好。下面进行对轨迹误差进行明确定量分析。

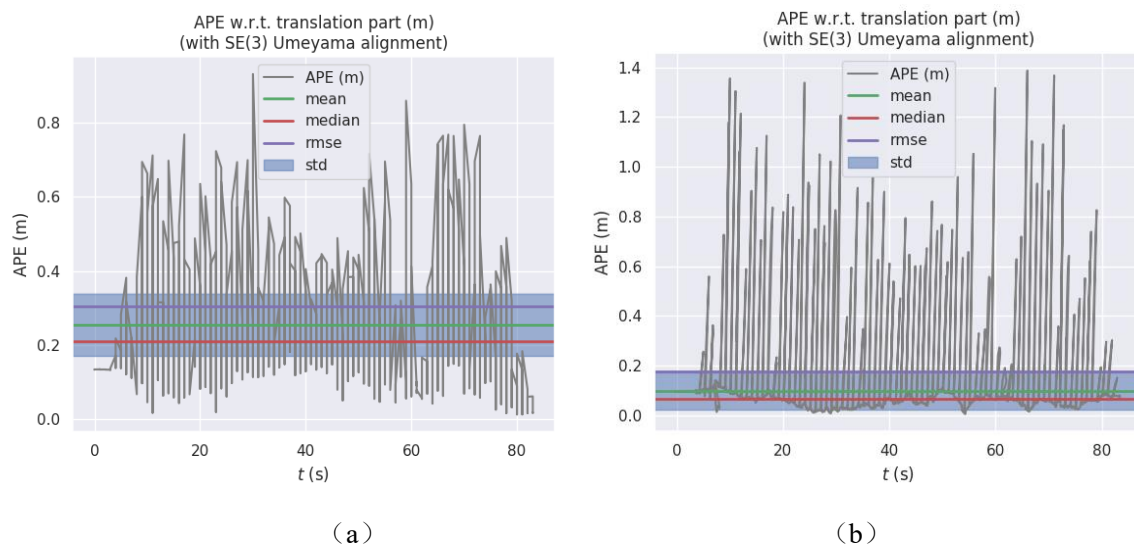


图 3-9 V1_02_medium 轨迹图

Fig. 3-9 V1_02_medium trajectory diagram

图 3-9 展示了本文算法和 VINS-fusion 分别与 V1_02_medium 中 groundtruth 的绝对误差，图 3-9 (a) 是 VINS-fusion 运行结果与 groundtruth 之间的绝对误差，轨迹的最大

误差为 0.931331m, 最小误差为 0.012645m, 平均误差为 0.253826m, RMSE 为 0.304272。图 3-9 (b) 为本文算法运行轨迹与 groundtruth 之间的绝对误差, 轨迹的最大误差为 1.386630m, 最小误差为 0.005989m, 平均误差为 0.097429m, RMSE 为 0.176800。本文算法较 VINS-fusion 相比, 虽然最大误差比起 VINS-fusion 较大 但是平均误差与 RMES 都相对较小。说明本文算法整体性能较好。

3.4.2 IMU/GPS 融合性能评估

本文采用 utbm_robotcar_dataset 数据集^[51]和 WTGPS-200 模块来采集 GPS 与 IMU 的加速度与角速度数据进行实验验证。utbm_robotcar_dataset 数据集包含带有 RTK 基站 Magellan ProFlex 500 GNSS 接收器和 Xsens MTi-28A53G25 IMU 采集的 GPS 数据、加速度和角速度数据, WTGPS-200 支持北斗和 GPS 的卫星接收机芯片并集成了三轴陀螺仪, 三轴加速度计。

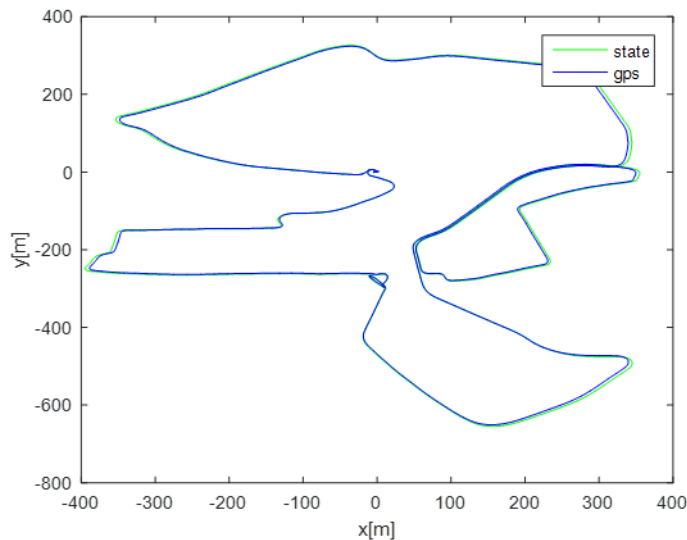


图 3-10 Utbm_robotcar_dataset 轨迹图

Fig. 3-10 Utbm_robotcar_dataset trajectory

图 3-10 展示了 GPS 和数据融合后的轨迹图, 其中蓝色部分为 GPS 测量轨迹, 绿色部分为 GPS 与 IMU 融合后的轨迹。两条轨迹大致相当。图 3-11 展示了 GPS 轨迹与融合后的轨迹之间在 x y 方向上的的轨迹误差, x 方向上的最大误差为 5.8795m, 绝对平均误差为 2.5392m, y 方向上的最大误差为 3.8237m, 绝对平均误差为 1.0237m。

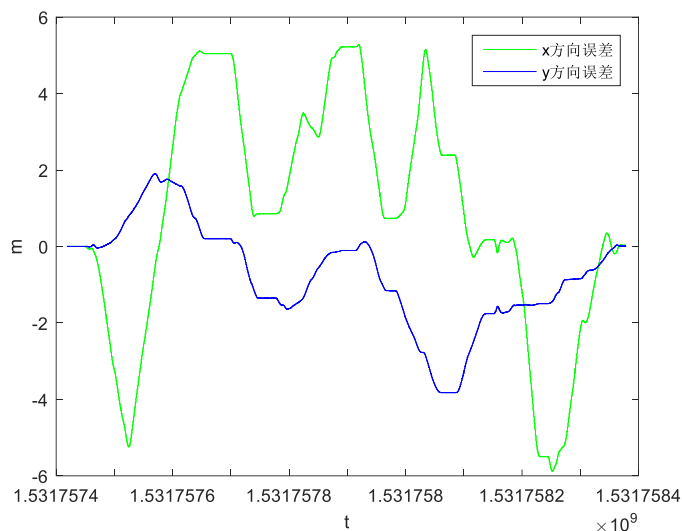


图 3-11 轨迹误差

Fig. 3-11 Trajectory error



图 3-12 WTGPS-200 传感器

Fig. 3-12 WTGPS-200 sensor

实际场景实验采用低成本的 WTGPS-200 传感器，能够同时采集到原始的加速度与角速度数据，GPS 采样频率为 5Hz，IMU 采样频率为 100Hz。图 3-9 中绿色部分为 IMU 原始数据积分得到的轨迹，由于积分累积误差，轨迹越来越差，蓝色部分为 GPS 绘制轨迹，红色部分为融合轨迹。

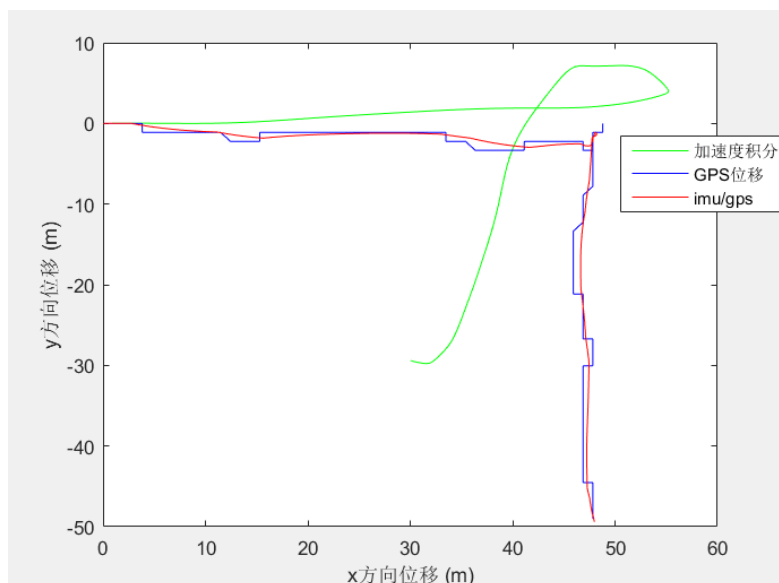


图 3-13 融合轨迹图

Fig. 3-13 WTGPS-200 sensor

3.5 本章小结

本章分别提出了了 VO/IMU 的数据融合方法以及 IMU/ GPS 的融合方法，并通过 EuRoC, utbm_robotcar_dataset 数据集进行实验验证。首先介绍了 IMU 与 VO 建立联合优化函数，通过 LM 算法迭代求解位姿，可以为刚体运动建立一个先验约束。将先前的状态加入到状态矩阵中建立增广状态矩阵，并通过先前的状态与此刻的状态构建 VO 相对测量观测模型，进行对 EKF 的状态更新，最后在 EuRoC 数据集下进行实验验证。其次，通过基于残差平方和的 EKF 将 GPS 与 IMU 进行数据融合，GPS 建立观测模型并通过观测值与预测值的残差平方和来判断当前更新是否可靠，提高系统的鲁棒性，并通过 utbm_robotcar_dataset 数据集进行验证。

第4章 二维占据网格地图构建的 VO-Scan 算法

目前主流的机器人自主导航避障基本都采用网格地图，基于激光 SLAM 构建的网格地图精度已经很高，但是由于激光传感器的成本较高，而激光雷达利用激光束来测量深度，所以为了节省成本，利用视觉传感器来获取深度图像也可以构建网格地图。三维地图虽然可以将空间信息表示出来，但是便携式嵌入式设备资源有限，所以三维地图并不适用，因此本章将研究基于双目视觉传感器的二维网格地图算法。

4.1 双目立体视觉逆传感器模型

逆传感器模型描述了传感器数据到地图的转换，传感器到障碍物之间所有的网格单元都需要填充一个概率值，通过这个概率得知这些单元网格是占据还是空闲状态的可能性，而没有被传感器探测到的单元网格则被描述为未知状态。单元网格的概率状态分布有助于在地图中引入噪声和不确定度^[52]。传感器测量的噪声与不确定度通常会破坏测量结果，所以在设计逆传感器模型的时候，需要将噪声与不确定度考虑进来。

4.1.1 理想传感器模型

深度图像的每个像素到物体的距离都被认为是逆传感器模型的测量距离。假设空间一点 P 在相机坐标系下的坐标为 $P_c=(x_c, y_c, z_c)^T$ 测量距离 l_p 定义如下：

$$l_p = \sqrt{x_c^2 + y_c^2 + z_c^2} \quad (4-1)$$

由于给出的图像坐标不精确，会存在一个不确定度 Δd ，则由视差函数算出的测量距离也存在一个不确定度为：

$$\Delta z_c = \Delta d \cdot \frac{z_c^2}{bf} \quad (4-2)$$

深度不确定度为：

$$\Delta l_p = \Delta z_c \cdot \frac{l_p}{z_c} \quad (4-3)$$

理想状态下的传感器模型为：

$$P_{occ}(l) = \begin{cases} p_{\min} & \text{if } 0 < l \leq l_p \\ 0.5 & \text{if } l > l_p \end{cases} \quad (4-4)$$

4.1.2 实际的逆传感器模型

$P_{\min}$ 是在单元网格中初始的概率，为了减少不确定度的影响，在此引入了一个二次指数函数来估计测量距离。指数函数如下：

$$e^{-\frac{1}{2}(\frac{r-l_p}{\Delta l_p})^2} \quad (4-5)$$

二次指数函数导致了曲线的概率在测量范围附近对称增加并且其宽度与测量的不确定度有关，由于双目传感器测量的距离随着距离的增加而增加，所以需要引入一个尺度因子如下所示：

$$\frac{k}{\Delta l_p \sqrt{2\pi}} + 0.5 - P_{occ}(l) \quad (4-6)$$

显著性因子 k 表明了单次测量的权重，它确保远处的测量结果的影响小，近处的测量结果影响大。

因此双目视觉系统单次测量单元网格地图的占据概率为：

$$P(s(l)) = P_{occ}(l) + (\frac{k}{\Delta l_p \sqrt{2\pi}} + 0.5 - P_{occ}(l)) e^{-\frac{1}{2}(\frac{r-l_p}{\Delta l_p})^2} \quad (4-7)$$

4.2 传感器模型的参数选择

4.2.1 显著因子 k 对 ISM 的影响

在讨论 k 值影响的时候采用控制变量的方法，将测量距离设定为 $z=3$ ，深度相关不确定度设定为 $\Delta l_p=0.1$ 。通过改变 k 值来分析 ISM 的变化，如图 4-1 所示：

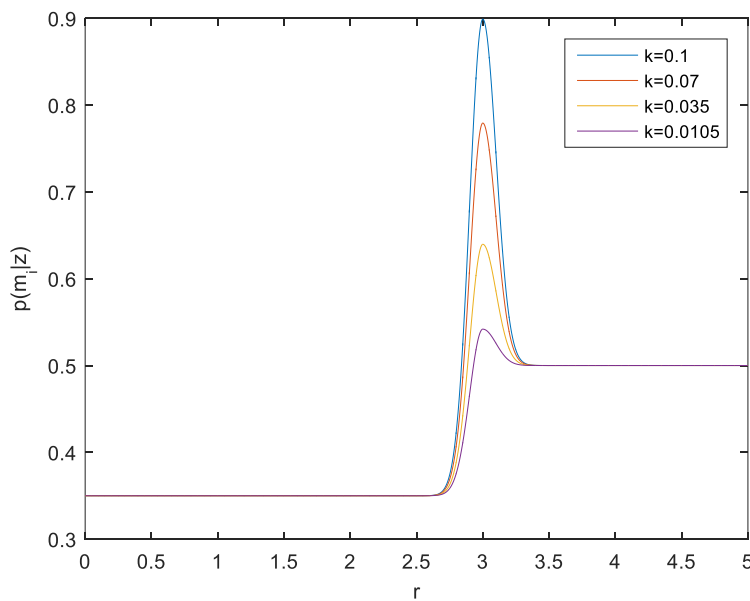


图 4-1 显著性因子 k 对 ISM 的影响

Fig. 4-1 The influence of significance factor k on ISM

由图(4-1)可知显著性因子 k 对基于深度测量结果概率的影响。 k 值越大, 测量结果的概率范围就越大; k 值越小, 测量结果的概率就越小。 k 是一个作为度量测量结果的尺度因子, 描述了单次测量的重要性, 所以选择一个合适的 k 值对于传感器模型的好坏至关重要。经过实验分析, 当 $k=0.1$ 时, 逆传感器模型的效果较好。

4.2.2 深度相关不确定度 Δl_p 对 ISM 的影响

在讨论深度相关不确定度对 ISM 模型的影响时, 同样采取控制变量的方法, 显著性因子 $k=0.1$, 测量距离 $z=3\text{m}$, 通过选取不同的 Δl_p 分析 ISM 的性能, 如图 3 所示。

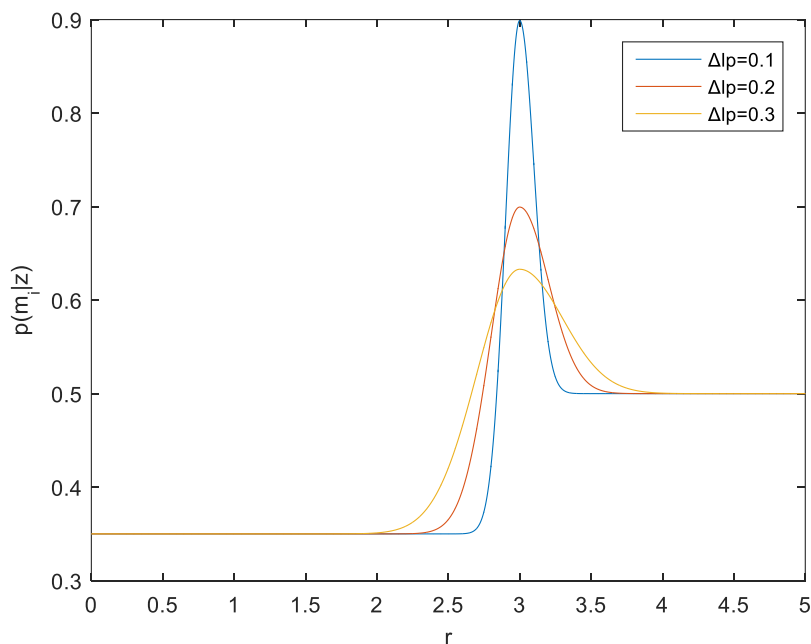


图 4-2 深度相关不确定度 Δl_p 对 ISM 的影响

Fig. 4-2 The influence of significance factor K on ISM

由图(4-2)可知深度相关不确定度 Δl_p 为不同值时对 ISM 的影响, 不确定度越高, 波峰越低, 同时波峰也越宽; 而不确定度越低, 波峰越高, 同时波峰也越窄。

4.2.3 测量距离 z 对 ISM 的影响

由于测量的距离不同, 就会导致不同的测量不确定度, 那么测量结果也就不同。由于 z 的变化导致了 Δl_p 的变化, 选取 $k=0.1$, 通过 z 与 Δl_p 之间的变换关系计算出 Δl_p , 如图 4-3 所示。

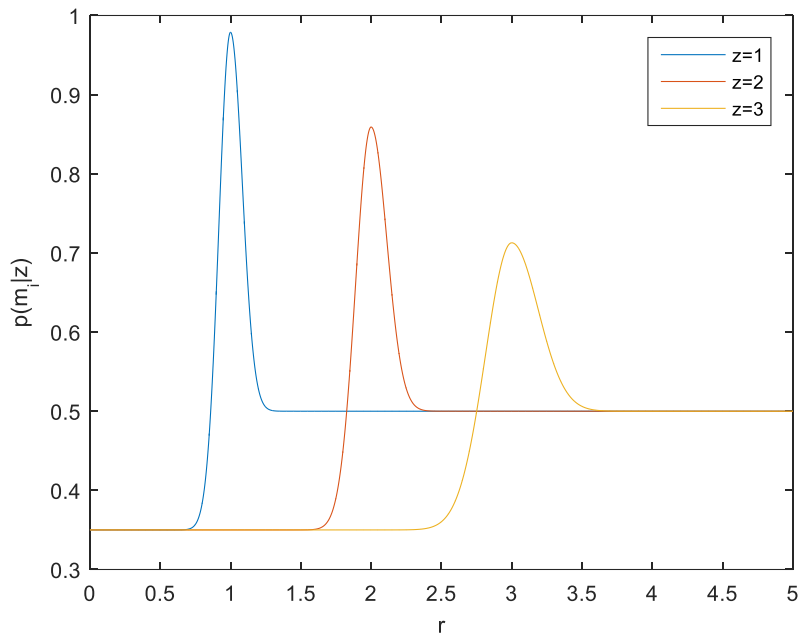


图 4-3 测量距离 z 对 ISM 的影响

Fig. 4-3 The influence of significance factor K on ISM

由图（4-3）可知测量距离对测量结果概率的影响，越近的测量值被占用的可能性就越高，并且每次测量的概率都会随着距离的增加而减小，直到达到最大的距离为止，此时单元格变为未知状态。

4.3 二维网格地图模型

占据网格地图的目的是将环境分成尺寸大小均匀的单元，这些单元存储着环境的相关信息，通过这些信息来判断当前环境状况，有占据、空闲、未知三种状态。占据状态代表该单元网格内有障碍物，空闲状态代表该单元网格无障碍物，机器人可自由通行，未知状态代表未探索区域。每一个单元网格都可以通过一个概率值表示它是否被占据，概率从 0 到 1 的变化表示从空闲逐渐变为占据状态，0.5 表示该单元网格状态未知。

通过传感器的测量值 z 以及从开始到 t 时刻的姿态 x ，就可以计算地图 m 的后验概率，可表示为 $p(m|z_{1:t}, x_{1:t})$ 。

假设当前处于世界坐标系下，同时传感器测量值与机器人姿态 x 相关，所以将姿态合并到测量值中去，则地图 m 的后验概率又可以表示为 $p(m|z_{1:t})$ 。

如果地图中的网格单元数量很多，则地图就会有很多种可能，会导致地图很复杂，所以需要研究每一个单网格的后验分布，地图 m 被分成相同尺寸的网格单元，将第 i 个网格单元的占据变量记为 m_i ， $p(m_i=1)$ 表示第 i 个网格单元被占据的概率， $p(m_i=0)$ 表示第 i 个单元网格空闲的，第 i 个网格单元的后验分布表示为 $p(m_i|z_{1:t})$ 。

假设地图网格单元之间是相互独立的,就可以将地图 m 表示为每一个网格单元后验分布的乘积:

$$p(m_1, m_2, \dots, m_n | z_{1:t}) = \prod_i^N p(m_i | z_{1:t}) \quad (4-8)$$

当传感器有了新的测量结果,地图中一部分单元网格的占据概率发生变化,就需要对地图进行更新。由于第 i 个网格单元的传感器测量值 z 随着时间推移条件独立,采用二进制贝叶斯滤波器来解决二进制状态表示的静态环境的占据概率,通过贝叶斯法则可得:

$$\begin{aligned} p(m_i | z_{1:t}) &= \frac{p(z_t | m_i, z_{1:t-1}) p(m_i | z_{1:t-1})}{p(z_t | z_{1:t-1})} \\ &= \frac{p(z_t | m_i) p(m_i | z_{1:t-1})}{p(z_t | z_{1:t-1})} \end{aligned} \quad (4-9)$$

使用贝叶斯法则将 $p(z_t | m_i)$ 展开可得:

$$p(z_t | m_i) = \frac{p(m_i | z_t) p(z_t)}{p(m_i)} \quad (4-10)$$

将 (4-10) 式代入 (4-9) 式可得第 i 个单元网格被占据的概率为:

$$p(m_i | z_{1:t}) = \frac{p(m_i | z_t) p(z_t) p(m_i | z_{1:t-1})}{p(m_i) p(z_t | z_{1:t-1})} \quad (4-11)$$

同理可得第 i 个单元网格空闲得概率为:

$$p(\bar{m}_i | z_{1:t}) = \frac{p(\bar{m}_i | z_t) p(z_t) p(\bar{m}_i | z_{1:t-1})}{p(\bar{m}_i) p(z_t | z_{1:t-1})} \quad (4-12)$$

利用公式 (4-11) 与公式 (4-12) 做商可得:

$$\frac{p(m_i | z_{1:t})}{p(\bar{m}_i | z_{1:t})} = \frac{p(m_i | z_t)}{p(\bar{m}_i | z_t)} \frac{p(m_i | z_{1:t-1})}{p(\bar{m}_i | z_{1:t-1})} \frac{p(\bar{m}_i)}{p(m_i)} \quad (4-13)$$

对公式 (4-13) 等号两端同时取对数可得:

$$\log \left(\frac{p(m_i | z_{1:t})}{p(\bar{m}_i | z_{1:t})} \right) = \log \left(\frac{p(m_i | z_t)}{p(\bar{m}_i | z_t)} \right) + \log \left(\frac{p(m_i | z_{1:t-1})}{p(\bar{m}_i | z_{1:t-1})} \right) - \log \left(\frac{p(m_i)}{p(\bar{m}_i)} \right) \quad (4-14)$$

为了避免数值上的不稳定将占据概率的值稳定地限制在 0-1 之间,采用 log-odd 率来表示占据概率,如下所示:

$$L(m_i | z_{1:t}) = \log \left(\frac{p(m_i | z_{1:t})}{p(\bar{m}_i | z_{1:t})} \right) \quad (4-15)$$

则公式 (4-14) 可写为:

$$L(m_i | z_{1:t}) = L(m_i | z_t) + L(m_i | z_{1:t-1}) - L(m_i) \quad (4-16)$$

分析公式 (4-16) 可知 $L(m_i | z_{1:t})$ 是代表的是地图更新后的第 i 个网格单元的占据概率, $L(m_i | z_t)$ 是通过传感器测量值得到的 t 时刻的概率, 该概率可以通过 ISM 传感器模型确定 $L(m_i | z_{1:t-1})$ 为前一时刻的占据概率, $L(m_i)$ 是地图的先验信息。

4.4 基于视觉定位的二维网格地图构建算法流程

4.4.1 ORB-SLAM2 概述

ORB-SLAM2 是一个基于特征点的实时的 SLAM 系统, 支持单目, 双目以及 RGB-D 相机, 该系统包含了跟踪 (Tracking)、建图 (Mapping)、重定位 (Relocalization)、闭环检测 (Loop closing)。该系统能够实时计算出相机的轨迹, 同时生成稀疏的点云地图。

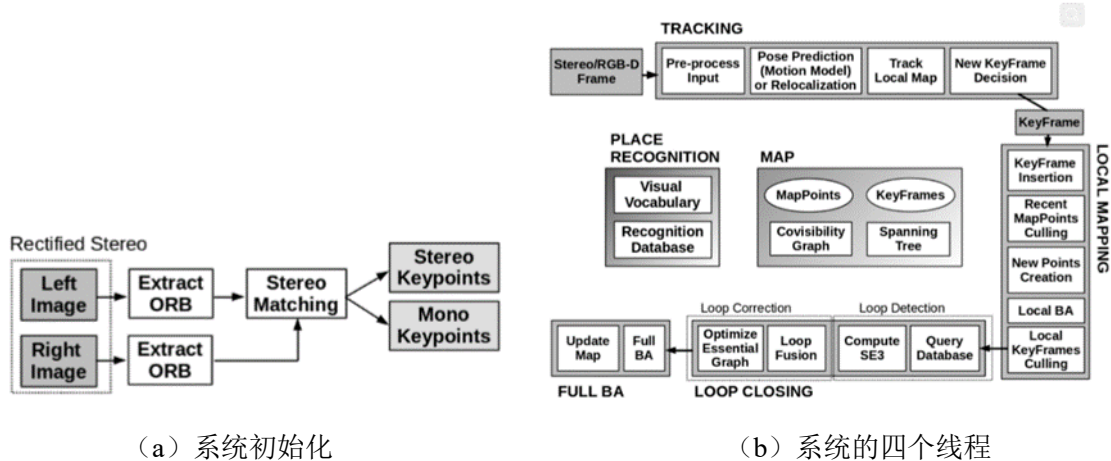


图 4-4 ORB-SLAM2 算法流程图

Fig. 4-4 ORB-SLAM2 algorithm

跟踪线程通过特征提取匹配并将这些特征与前一帧最小化重投影误差, 从而计算相机每一帧下的定位, 局部建图线程采用局部 BA 来优化关键点与关键帧, 回环线程负责采用姿态图优化来检测重新访问的区域并校正累积的漂移。姿态图优化过后启动第四个线程优化关键帧间的整个地图轨迹。ORB-SLAM2 能够使得累积误差最小化, 并且有很高的定位精度。它从输入图像中提取了 ORB 特征以稀疏点云的形式表示对环境的编码。这些功能在旋转和缩放方面很稳定, 同时又可以快速提取和匹配, 满足了实时性的要求。

4.4.2 深度图转 Scan 格式的数据原理

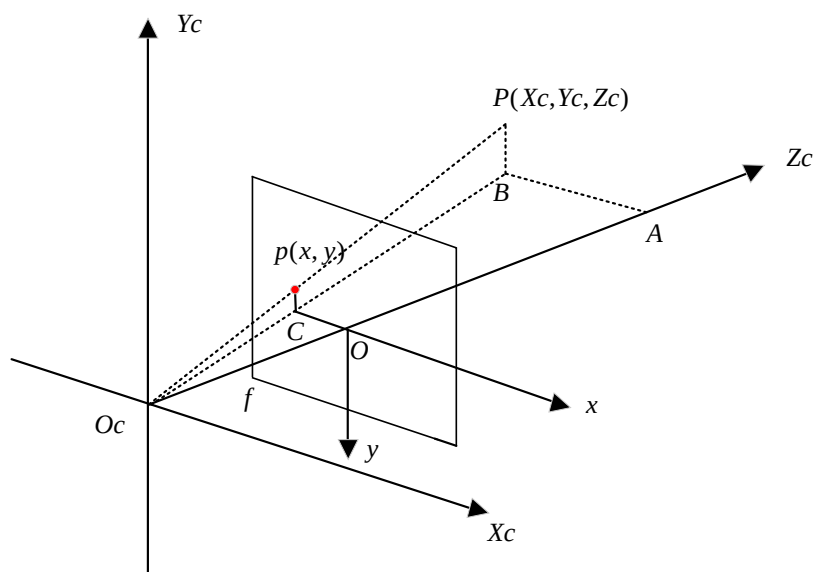


图 4-5 空间点在相机平面上的投影

Fig. 4-5 The projection of a space point onto the camera plane

Scan 数据是对激光数据格式的一种定义，它是在通用平台下对深度信息的一种描述。图中 $p(x, y)$ 为深度图中一点，相机原点 O_c 到物体所在平面的距离为 Z_c ，相机的焦距为 f ， P 点的相机坐标为 $P(X_c, Y_c, Z_c)$ 。

深度图转换为 scan 格式数据的主要思想就是将一定高度范围内的数据进行投影，那么投影角度就是如图 4-5 中所示直线 AO_c 和 BO_c 的夹角 $AOCB$ ，计算公式如下所示：

$$\theta = \arctan\left(\frac{x}{z}\right) \quad (4-17)$$

得到投影角度之后，将角 $AOCB$ 映射到深度图中相对应的 Scan 数据的扫描范围，由深度图的宽可知，扫描范围的最小角度为 α ，最大角度为 β ，将这个范围内的深度像素细分为 N 份，那么可以通过一个数组 `laser[N]` 来保存深度值，那么空间中一点 P 投影到 `laser` 数组中的索引值 n 计算方法如下：

$$n = \frac{(\theta - \alpha)}{(\beta - \alpha) / N} \quad (4-18)$$

以空间一点 P 在 $X_c O_c Z_c$ 平面上投影的点 B 到光心 O_c 的距离 r 即为在 `laser` 数组中的值计算方式如下：

$$r = OB = \sqrt{x^2 + z^2} \quad (4-19)$$

则在深度图中一定高度范围内，如果有比当前 r 还要小的点，则将替换原来的距离 r ，这就保证了可以检测到在一定高度范围内最近的障碍物。

4.4.3 基于 ORB-SLAM2 的二维占据网格地图构建算法流程

本文采用双目相机采集图像通过 ORB-SLAM2 提供视觉定位,该相机通过一个 ROS 节点发布左目、右目图像与深度图像等主题, ORB-SLAM2 系统订阅这些主题作为系统的输入,并输出相机的位姿 x 。将双目相机生成的深度图像通过 DepthToScan 功能包转化为 ROS 平台下 Scan 格式的数据。ISM 逆传感器模型通过相机位姿 x 与 Scan 数据计算得到占据概率,并通过 ROS 平台下 Occupancy Grid Map 构建二维网格地图,算法流程图如下所示:

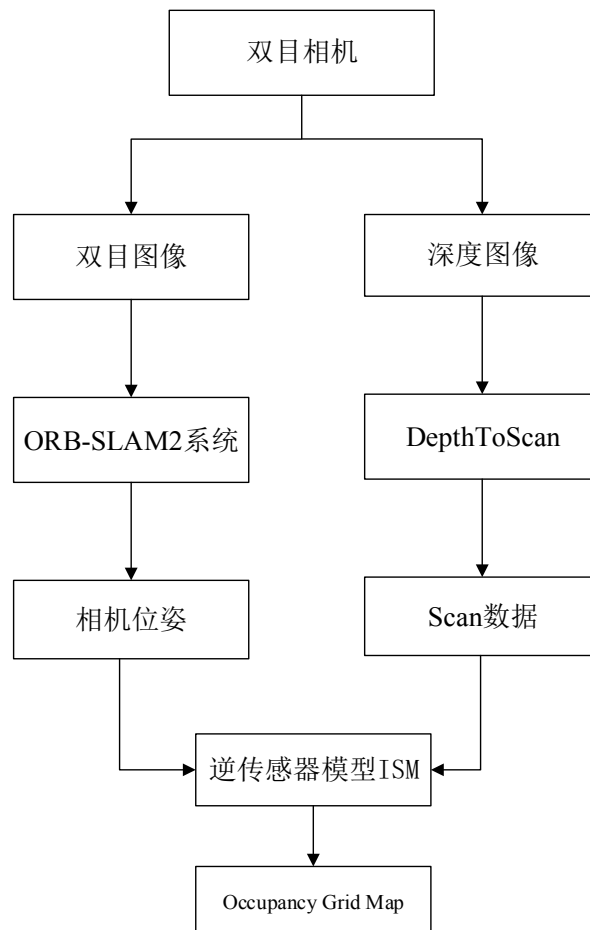
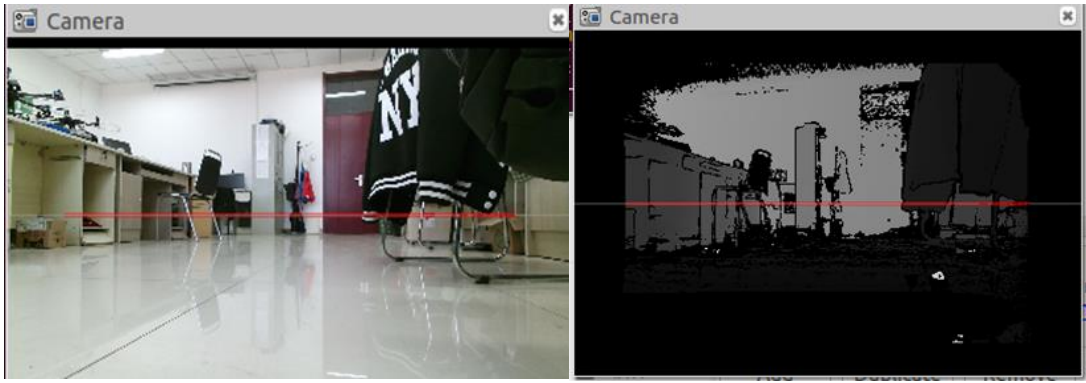


图 4-6 二维网格地图构建流程图

Fig. 4-6 2D grid map construction flow chart

4.5 深度图转 Scan 数据

图 4-7 (a) 是实验室一角彩色图片, 图 4-7 (b) 则是对应的深度图, 图中红线部分即为每一列像素投影的水平面。



(a) 原始图

(b) 深度图

图 4-7 深度图

Fig. 4-7 Depth image

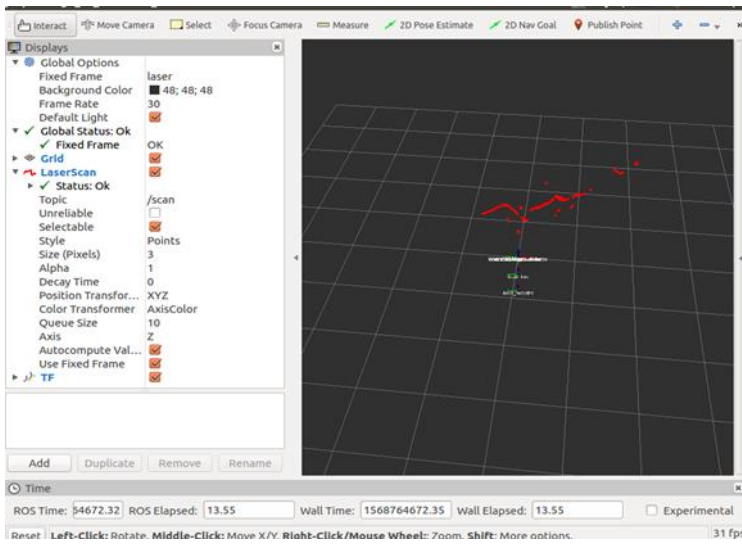


图 4-8 Scan 格式数据

Fig. 4-8 Scan data

图 4-8 是在 Rviz 可视化平台下 scan 数据的表现形式，其中红色部分代表检测到障碍物。

4.6 本章小结

本章主要是针对目前大多数视觉 SLAM 系统只能输出相机轨迹和构建稀疏点云地图，并不能满足机器人导航与障碍物检测，提出了基于双目相机网格地图构建的 VO-Scan 算法。首先，基于双目相机建立一个适用于双目相机的逆传感器模型，并对其进行了详细的推导，同时也对二维占据网格地图模型进行了详细的推导，利用双目相机构建

的深度图像，建立适用于传感器模型的深度数据，最后建立了基于双目立体视觉的二维占据网格地图的方案。

第 5 章 实验与分析

5.1 引言

本章主要是对二维网格地图算法进行详细的实验测试并得到实验结果，针对实验结果进行分析得出结论，本章主要在真实环境中进行测试。

5.2 实验平台

本文实验平台为 Intel(R) Core(TM) i5-9300H CPU@2.40GHZ 四核八线程，8GB 内存 1660Ti 显卡，采用 Ubuntu16.04 操作系统，ROS 操作系统版本为 Kinectic 版本，相机采用 MYNTEYE s1030-IR。

5.2.1 ROS 操作系统

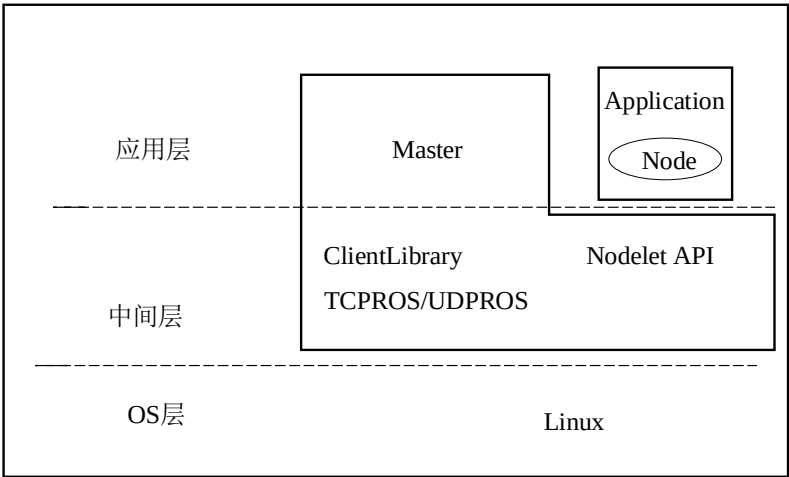


图 5-1 ROS 结构图

Fig. 5-1 ROS structure

ROS 操作系统是用于编写机器人软件程序的一种具有高度灵活性的软件架构，它 2007 年起源于斯坦福大学，ROS 总体设计的目标是提高机器人研发中的软件复用率，它相当于包含通信机制、开发工具，应用功能、机器人开发生态系统与一身。具有以下五个特点：

- (1) 它是一种点对点的设计，包含许多节点单元，支持分布式网络、通信方式采用 RPC 与 TCP/UDP 通信系统，适合多机器协同；
- (2) 多语言支持，接口定义与语言无关；
- (3) 架构精简，集成度高；

- (4) 有丰富的组件化工具包，像 Gazebo 物理仿真环境组件，Rviz 数据可视化平台，rqt 工具箱等极大的方便了开发人员；
- (5) 免费且开源。

5.2.2 MYNTEYE 相机



图 5-2 MYNTEYE s1030-IR

Fig. 5-2 MYNTEYE s1030-IR

小觅双目摄像头标准系列，是视觉 SLAM 研究的优质深度相机方案，帧率为 10FPS 到 60FPS，分辨率为 752*480 深度图分辨率基于 CPU/GPU 可达到 752*480，支持 IR，像素尺寸为 6.0*6.0 微米，基线为 120.0mm，焦距为 2.1mm，深度图工作距离为 0.8m 到 5m，功耗为 2.7w，采用 USB3.0 数据传输方式，支持 ubuntu16.04，ros 操作系统 kinetic 版本，并经过了大多数开源项目的测试，如 ORB-SLAM2、OKVIS、Vins-Mono、Vins-fusion、VIORB。

5.3 场景测试

5.3.1 室内实验

室内实验场景选择在实验室，四周围绕着一圈桌子以及中间有一排桌子，如图 5-3 所示。图 5-4 展示的是通过 ORB-SLAM2 系统构建的实验室的稀疏点云地图与相机的运动轨迹，红色的点代表当前帧下的点云地图，黑色的点代表历史帧下的点云地图，绿色的线代表相机的轨迹信息，蓝色则表示 ORB-SLAM2 的关键帧。本实验是围绕实验室运动一圈，通过稀疏点云地图并不能完全看出实验室的障碍物，而只能看出关键帧下图像匹配后的一些特征点。



图 5-3 室内场景

Fig. 5-3 Indoor scene

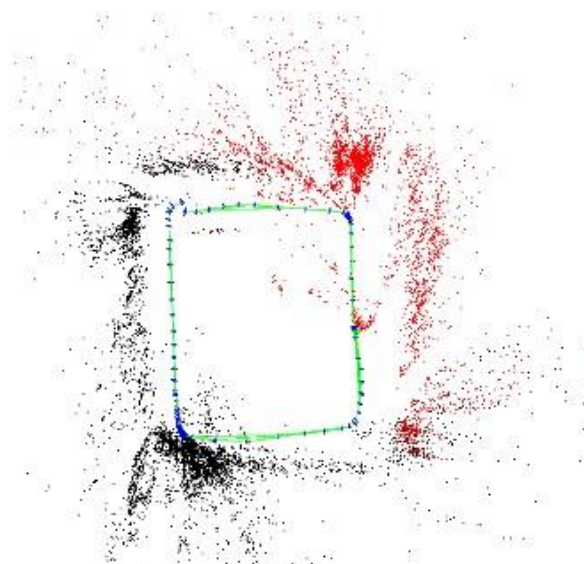


图 5-4 室内场景下的稀疏点云地图

Fig. 5-4 Sparse point cloud map in indoor scene

图 5-5 是通过 RVIZ 平台展示了本文算法构建的室内二维网格地图，黑色部分代表的就是障碍物消息，其他地方则为自由区域，整体上看可以大致扫描出实验室障碍物信息。但是由于深度图的测量范围为 0.8m 至 5m，当障碍物距离障碍物比较近的时候，小于 0.8m 会导致地图生成错误，在图 5-5 中上下两端中可以看出由于过道比较窄，错误地判断出周围存在障碍物。但是在左右两端清晰地标记出障碍物。

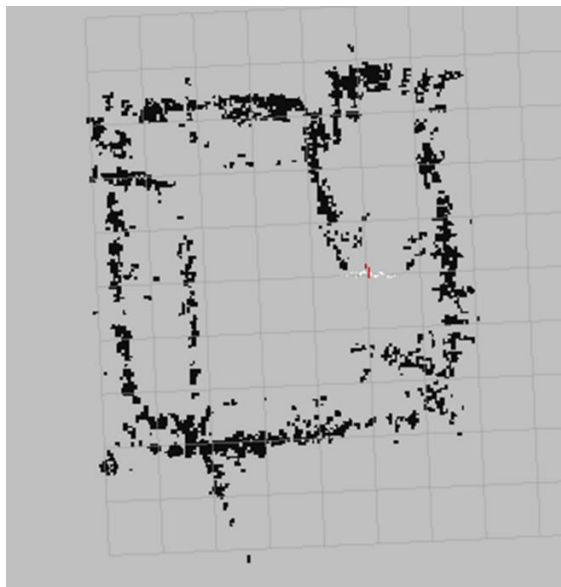


图 5-5 室内场景下的二维网格地图

Fig. 5-5 2D grid map of indoor scene

5.3.2 室外实验

室外实验选择校园一角，路的两旁种有灌木和树木，如图 5-6 所示。图 5-7 是该场景下的稀疏点云地图以及运动轨迹地图。红色的点代表当前帧下的点云地图，黑色的点代表历史帧下的点云地图，绿色的线代表相机的轨迹信息，蓝色则表示 ORB-SLAM2 的关键帧。本次实验是从小路一端直走一段距离，虽然点云地图看上去有很多标记点，但是周围障碍物与小路混在一起，无法判断出小路



图 5-6 室外场景

Fig. 5-6 Outdoor scene

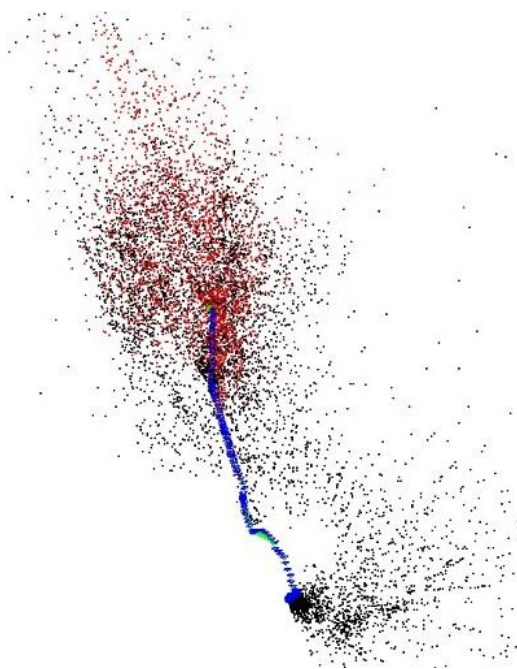


图 5-7 室外场景下的稀疏点云地图

Fig. 5-7 Sparse point cloud map in outdoor scene

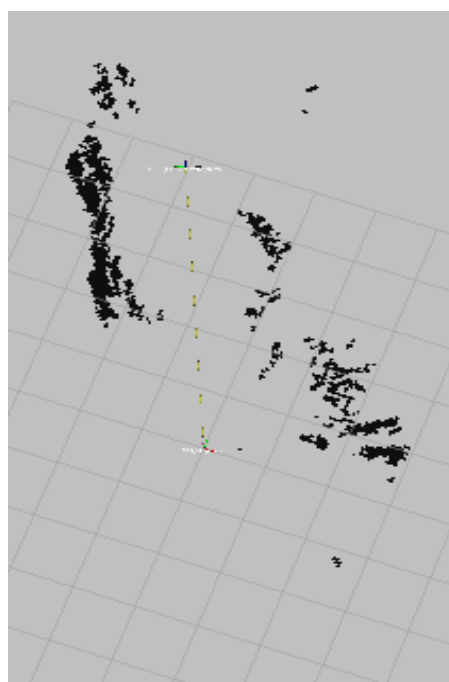


图 5-8 室外场景下的二维网格地图

Fig. 5-8 2D grid map of outdoor scene

图 5-8 展示了基于本文算法构建的二维室外环境网格地图，并通过 ROS 系统下的 RVIZ 可视化平台下显示，图中可以看出小路两边的障碍物被标记出来，而且也可以通过地图清晰地找到小路地位置。适用于机器人判断障碍物。

5.4 对比分析实验

本文将建图算法与 RTAB-map 建图算法选择室外二维网格地图与真实环境进行对比。RTAB-map 是基于实时外观的地图构建方法，自 2013 年以来，它作为一个开源库进行发布，具有内存管理功能来处理大规模和长期在线运行，支持视觉和激光雷达传感器，提供二维和三维地图的构建方案，使用具有 RGB 的深度图像来构造地图。通过 ROS 平台下的 RVIZ 可视化界面中的测量工具对二维地图中选择三处位置测量，如图 5-9 所示。并与真实环境下同样的位置进行对比。

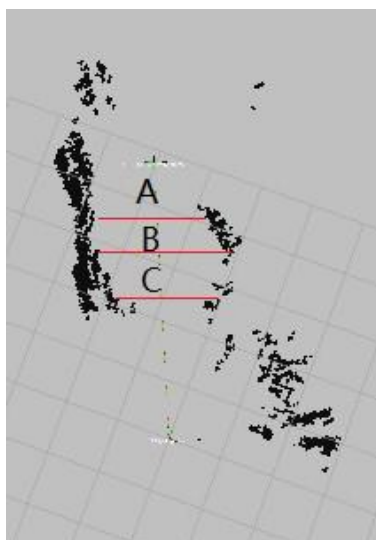


图 5-9 室外场景下的二维网格地图标记图

Fig.5-9 2D grid map marker for an outdoor scene

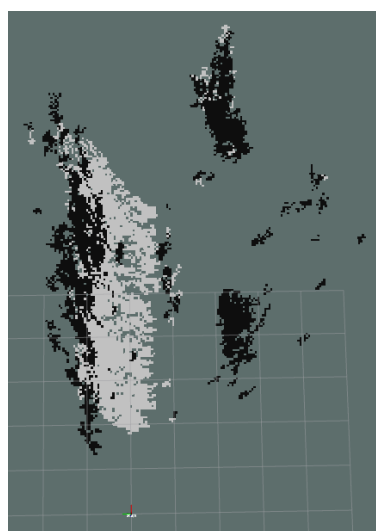


图 5-10 RTAB-map 算法构建二维网格地图

Fig. 5-10 RTAB- MAP algorithm to construct two-dimensional grid map

由于本文算法是基于双目相机作为视觉传感器，所以 RTAB-map 同样采用双目传感

器在同样地场景下构建二维网格地图。结果如表 5-1 所示，表中实际值是通过卷尺对真实环境做的测量，而测量值是在 RVIZ 可视化平台下提供的测量工具所得，所以所有数据均有一定范围的误差，但从表中结果可知本文算法整体上在障碍物判断性能上优于 RTAB-map。由于本文采用 ORB-SLAM2 算法提供姿态信息，实时性方面却没有 RTAB-map 好^[53]。

表 5-1 2D 网格地图与真实环境对比

Table 5-1 2D grid map vs. real environment

测量对象	实际值/m	本文算法		RTAB-map	
		测量值/m	误差/m	测量值/m	误差/m
线段 A	3.57	3.42	0.15	3.31	0.26
线段 B	4.32	4.19	0.13	4.03	0.29
线段 C	3.28	3.12	0.16	3.04	0.24

5.5 本章小结

本章首先对实验平台进行了描述，在 ORB-SLAM2 的基础上进行了二维网格地图的构建，分别选择了室内与室外环境进行实验，在室内环境的某些地方障碍物与机器人的距离较近，超出了深度测量范围，所以导致局部区域对障碍物判断错误，而室外环境，障碍物基本在深度测量范围之内，所以比较清晰的判断出障碍物的位置。并与 RTAB-map 对比分析，本文可以构建满足机器人判断障碍物的二维网格地图。

第6章 结论与展望

6.1 结论

智能机器人技术发展迅速,单纯的视觉定位与建图并不能够满足复杂场景的应用。不同的场景可以采用不同的传感器。纯视觉定位在快速运动与纹理特征差的环境下容易导致定位失败,而与 IMU 和 GPS 的融合就可以应对复杂的场景,如 IMU 能够在快速运动下获得加速度与角速度, GPS 能够在空旷的地方获得绝对的位置信息。有了位姿信息,利用视觉传感器采集到的图像信息可以对周围环境进行描述。

本文通过权衡目前视觉传感器的优缺点,选定双目立体相机为视觉传感器,它既能满足室外使用,也能通过两镜头之间的基线得到真实尺度,而且现在双目相机内部已经通过 GPU 加速处理计算深度图像,对整个系统减少了计算负担。本文研究了 VO/IMU、IMU/GPS 数据融合方法、以及利用 ORB-SLAM2 的姿态信息构建二维网格地图的方法。并做了实验验证,具体研究内容如下:

(1) 首先研究了双目视觉传感器的基本模型,以及视觉里程计的基本原理,图像匹配与姿态计算方法。IMU 通过加速度与角速度积分计算刚体姿态信息的基本原理,并对其误差模型,运动模型进行了研究。以及对 GPS 的基本定位原理,定位误差进行分析。

(2) 针对纯视觉定位在快速运动状态下会导致定位失败的情况,提出了基于 VO 相对状态更新的 EKF 的 VI-rel 方法。为了给视觉刚体运动提供一个很好的刚体先验约束,基于此研究了视觉重投影误差, IMU 与积分以及残差模型,最后通过 VO 与 IMU 通过紧耦合的方法建立联合优化目标函数,求解姿态并保存。从保存的姿态里面找一个先前的姿态添加到 IMU 预测的状态中,建立增广状态矩阵,并对增广状态雅各比矩阵进行推导。并利用此刻的 VO 测量姿态与先前添加进状态矩阵的姿态构建相对测量模型作为 EKF 的观测模型进行更新,并推导了观测模型的雅各比矩阵,并通过数据集进行实验验证,结果表明整体轨迹误差比 VINS-fusion 要小。

(3) 针对视觉传感器在空旷环境下无法提供视觉定位的情况,提出了基于 GPS 绝对状态更新 EKF 的 IG-abs 方法。GPS 与 IMU 采用松耦合的 EKF 数据融合方法, IMU 对刚体运动状态预测,提供预测方程, GPS 提供观测方程更新预测状态,但是由于 GPS 测量会由于受到多径效应等干扰出现异常值,导致观测值误差很大,一般的方法是通过当前卫星数量来判断 GPS 测量值是否可靠,但该方法效果并不明显。故可以通过预测残差平方和与观测残差平方和来判断更新后的状态是否可靠。

(4) 针对目前大多数视觉 SLAM 算法只输出姿态信息与稀疏点云地图,无法充分的描述周围环境,提出了基于 ORB-SLAM2 的二维网格地图构建的 VO-Scan 方法。目

前大多数视觉 SLAM 都只输出姿态信息与周围环境的稀疏点云地图，对周围环境的描述效果有限，但是可以利用输出的姿态信息与环境深度信息构建二维网格地图来对周围环境进行描述。首先，将双目立体相机的深度图转换为 Scan 格式的数据，通过双目视觉传感器与 ORB-SLAM2 算法提供姿态信息，二者相结合来构建二维网格地图，并通过逆传感其模型来更新网格的占据概率，最后经 ROS 平台实验验证，可以构建供机器人导航、避障的二维占据网格地图。

6.2 展望

限于个人能力与研究时间限制，本文对 VO/IMU、IMU/GPS 数据融合以及二维网格地图构建做了初步的研究，基于本文还可以做进一步工作。

（1）由于 GPS 对于高度的测量误差较大，本文并未对高度引入计算，可以引入气压计等其他传感器辅助测量高度信息。

（2）本文只在二维网格地图构建方法的过程中，只采用了视觉的定位信息，并未将其他传感器的融合定位信息引入，可以将数据融合后的传感器姿态信息引入到网格地图构建算法中来提高建图精度。

参考文献

- [1] 王霞, 左一凡. 视觉 SLAM 研究进展[J]. 智能系统学报, 2020, 15(5): 825-834.
- [2] Cobb H S. GPS pseudolites: Theory, design, and applications[D]. Citeseer, 1997.
- [3] Yousif K, Bab-Hadiashar A, Hoseinnezhad R. An overview to visual odometry and visual SLAM: Applications to mobile robotics[J]. Intelligent Industrial Systems, 2015, 1(4): 289-311.
- [4] Ryu K, Dantanarayana L, Furukawa T, et al. Grid-based scan-to-map matching for accurate 2D map building[J]. Advanced Robotics, 2016, 30(7): 431-448.
- [5] Sun S, Xu B, Sun Y, et al. Sparse Pointcloud Map Fusion of Multi-Robot System[C]. 2018 International Conference on Control, Automation and Information Sciences (ICCAIS), 2018: 270-274.
- [6] Vatansever S, Butun I. A broad overview of GPS fundamentals: Now and future[C]. 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), 2017: 1-6.
- [7] Mirzaei F M, Roumeliotis S I. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation[J]. IEEE transactions on robotics, 2008, 24(5): 1143-1156.
- [8] Caron F, Duflos E, Pomorski D, et al. GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects[J]. Information fusion, 2006, 7(2): 221-230.
- [9] Melbouci K, Collette S N, Gay-Bellile V, et al. Model based RGBD SLAM[C]. 2016 IEEE International Conference on Image Processing (ICIP), 2016: 2618-2622.
- [10] Younes G, Asmar D, Shamma E, et al. Keyframe-based monocular SLAM: design, survey, and future directions[J]. Robotics and Autonomous Systems, 2017, 98: 67-88.
- [11] Sanchez-Rodriguez J-P, Aceves-Lopez A. A survey on stereo vision-based autonomous navigation for multi-rotor MUAVs[J]. Robotica, 2018, 36(8): 1225-1243.
- [12] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]. IEEE & Acm International Symposium on Mixed & Augmented Reality, 2008.
- [13] Engel J, Schps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]. European Conference on Computer Vision, 2014.
- [14] Pire T, Fischer T, Civera J, et al. Stereo parallel tracking and mapping for robot localization[C]. IEEE/RSJ International Conference on Intelligent Robots & Systems, 2015.
- [15] Engel J, Koltun V, Cremers D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016: 1-1.
- [16] 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, 2017, 33(5): 1255-1262.
- [17] Yun S, Ji S, Shi Z, et al. GPS-Supported Visual SLAM with a Rigorous Sensor Model for a Panoramic Camera in Outdoor Environments[J]. Sensors, 2013, 13(1): 119-136.
- [18] Konolige K, Agrawal M. FrameSLAM: From bundle adjustment to real-time visual mapping[J]. IEEE Transactions on Robotics, 2008, 24(5): 1066-1077.
- [19] Afia A B, Escher A C, Macabiau C. A Low-cost GNSS/IMU/Visual monoSLAM/WSS Integration Based on Federated Kalman Filtering for Navigation in Urban Environments[J], 2015.
- [20] Amedeo V, Giancarmine F, Domenico A, et al. Differential GNSS and Vision-Based Tracking to Improve Navigation Performance in Cooperative Multi-UAV Systems[J]. Sensors (Basel, Switzerland), 2016, 16(12).
- [21] Yu S, Nuske S, Scherer S. A Multi-Sensor Fusion MAV State Estimation from Long-Range Stereo, IMU, GPS and Barometric Sensors[J]. Sensors, 2017, 17(12): 11.
- [22] VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on

Robotics, 2018.

[23] Chen X, Hu W, Zhang L, et al. Integration of Low-Cost GNSS and Monocular Cameras for Simultaneous Localization and Mapping[J]. *Sensors*, 2018, 18(7): 2193.

[24] Gakne P V. Improving the Accuracy of GNSS Receivers in Urban Canyons using an Upward-Facing Camera[J], 2018.

[25] Qin T, Cao S, Pan J, et al. A general optimization-based framework for global pose estimation with multiple sensors[J]. *arXiv preprint arXiv:1901.03642*, 2019.

[26] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]. *European conference on computer vision*, 2014: 834-849.

[27] Santana A M, Aires K R, Veras R M, et al. An approach for 2d visual occupancy grid map using monocular vision[J]. *Electronic Notes in Theoretical Computer Science*, 2011, 281: 175-191.

[28] Kamarudin K, Mamduh S M, Shakaff A Y M, et al. Method to convert Kinect's 3D depth data to a 2D map for indoor SLAM[C]. *2013 IEEE 9th international colloquium on signal processing and its applications*, 2013: 247-251.

[29] Goeddel R, Kershaw C, Serafin J, et al. FLAT2D: Fast localization from approximate transformation into 2D[C]. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016: 1932-1939.

[30] Dia R, Mottin J, Rakotovo T, et al. Evaluation of occupancy grid resolution through a novel approach for inverse sensor modeling[J]. *IFAC-PapersOnLine*, 2017, 50(1): 13841-13847.

[31] Hull G. Real-time occupancy grid mapping using LSD-SLAM[D]. Stellenbosch: Stellenbosch University, 2017.

[32] Clemens J, Kluth T, Reineking T. β -SLAM: Simultaneous localization and grid mapping with beta distributions[J]. *Information Fusion*, 2019, 52: 62-75.

[33] Xu L, Feng C, Kamat V R, et al. An occupancy grid mapping enhanced visual SLAM for real-time locating applications in indoor GPS-denied environments[J]. *Automation in Construction*, 2019, 104: 230-245.

[34] Quigley M, Conley K, Gerkey B, et al. ROS: an open-source Robot Operating System[C]. *ICRA workshop on open source software*, 2009: 5.

[35] Da Silva B M, Xavier R S, Gonçalves L M. Mapping and Navigation for Indoor Robots under ROS: An Experimental Analysis[J], 2019.

[36] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters[J]. *IEEE transactions on Robotics*, 2007, 23(1): 34-46.

[37] Labbé M, Michaud F. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation[J]. *Journal of Field Robotics*, 2019, 36(2): 416-446.

[38] Lemaire T, Berger C, Jung I-K, et al. Vision-based slam: Stereo and monocular approaches[J]. *International Journal of Computer Vision*, 2007, 74(3): 343-364.

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

[40] Viswanathan D G. Features from accelerated segment test (fast)[C]. *Proceedings of the 10th workshop on Image Analysis for Multimedia Interactive Services*, London, UK, 2009: 6-8.

[41] Calonder M, Lepetit V, Strecha C, et al. Binary robust independent elementary features[C]. *Proceedings of the European Conference on Computer Vision*, 2010: 778-792.

[42] Lepetit V, Moreno-Noguer F, Fua P. Epnp: An accurate o(n) solution to the pnp problem[J]. *International journal of computer vision*, 2009, 81(2): 155.

[43] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm[C]. *Proceedings third international conference on 3-D digital imaging and modeling*, 2001: 145-152.

- [44] Kos T, Markezic I, Pokrajcic J. Effects of multipath reception on GPS positioning performance[C]. Proceedings ELMAR-2010, 2010: 399-402.
- [45] Chu T, Guo N, Backén S, et al. Monocular camera/IMU/GNSS integration for ground vehicle navigation in challenging GNSS environments[J]. Sensors, 2012, 12(3): 3162-3185.
- [46] Lourakis M, Argyros A A. Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment?[C]. Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005: 1526-1531.
- [47] Yuan Z, Zhu D, Chi C, et al. Visual-inertial state estimation with pre-integration correction for robust mobile augmented reality[C]. Proceedings of the 27th ACM International Conference on Multimedia, 2019: 1410-1418.
- [48] Fakharian A, Gustafsson T, Mehrfam M. Adaptive Kalman filtering based navigation: An IMU/GPS integration approach[C]. 2011 International Conference on Networking, Sensing and Control, 2011: 181-185.
- [49] Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets[J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [50] Grupp M. evo: Python package for the evaluation of odometry and slam[J]. Note: <https://github.com/MichaelGrupp/evo> Cited by: Table, 2017, 7.
- [51] Yan Z, Sun L, Krajnik T, et al. EU long-term dataset with multiple sensors for autonomous driving[J]. arXiv preprint arXiv:1909.03330, 2019.
- [52] Andert F. Drawing stereo disparity images into occupancy grids: Measurement model and fast implementation[C]. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009: 5191-5197.
- [53] Ragot N, Khemmar R, Pokala A, et al. Benchmark of visual slam algorithms: Orb-slam2 vs rtab-map[C]. 2019 Eighth International Conference on Emerging Security Technologies (EST), 2019: 1-6.

攻读学位期间取得的研究成果

一、学术论文

- [1] 王永祥, 常青, 黄良红, 孙永明. 基于立体视觉的 2D 占据网格地图构建. 电子设计工程 (已录用) (对应论文第 4 章)

致 谢

三年的硕士研究生时光已经接近尾声，在太原理工大学的学习生活也即将结束。回顾过去，收获颇多。老师的谆谆教导，实验室的欢声笑语，学术创作的绞尽脑汁都成为我人生中最宝贵的财富。在今后的生活工作中，我相信这些经历会助我披荆斩棘，勇攀高峰。下面我想对这三年里面帮助过我的人表达我最真诚的谢意。

首先，我要感谢我的指导老师常青老师三年研究生期间对我的关怀与指导。学习中，他倾囊相授，使我的科研能力和理论水平得到了很大的提升，从刚入学对专业懵懂无知到对专业价值理念的熟知，开拓了眼界。在生活中，虽然他是一位师长，但是更像朋友一样给予了我很大的关心和照顾，当然，更要感谢他在我撰写论文的过程中，一丝不苟地为我的论文提出修改意见和指导方案。在此，我对常青老师表达我衷心地感谢和由衷的敬意，祝愿他今后身体健康，工作顺心。

其次，我还要感谢智能感知与安全架构实验室的负责人王耀力老师，王老师以他扎实的专业能力以及过人的学术素养在学习中给予我很大帮助，每当我困惑的时候，王老师的都会为我指点迷津，令我茅塞顿开。他不仅在学习中耐心指导，而且也会教我为人处世的道理。感谢王老师对智能感知与安全架构实验室的付出。

除了两位老师的帮助以外，师兄师姐们给我树立了一个很好的榜样，姜海涛，陆建伟，刘安睿劼和王飞师兄在我的研究方向上给予我指点，以及他们刻苦求知的精神让我备受鼓舞。尤其要感谢我同届小伙伴董潇毅、田歌、刘晓慧、李宇佳，我们互相学习，共同进步。在我遇到困难的时候，总是能第一时间帮助我，祝大家前程似锦。还要感谢我的舍友燕鹏云、孙建其、陈波、王杰和景柱，在紧张的学习之余，带给我无数的欢乐。

特别地，我要感谢我的父母，虽然身处异地，但一直在关心我的学习和生活，无条件地支持我，帮助我，是我最坚强地后盾，让我能无所顾虑地投入到学业之中。

最后感谢在百忙之中抽出宝贵时间评审本论文地专家教授，感谢你们对我的论文提出的宝贵意见。