固定路线中的视觉惯性定位

系统研究

(申请清华大学电子信息硕士专业学位论文)

培 养 单 位 ： 深圳国际研究生院

专 业 领 域 ： 大数据技术与工程

申 请 人 ： 薛 林 松

指 导 教 师 ： 张 凯 副教授

二〇二五年三月

固定路线中的视觉惯性定位系统研究 薛 林 松

**Research** **on** **Visual-Inertial** **Localization** **System** **for** **Fixed** **Routes**

Thesis submitted to

**Tsinghua** **University**

in partial fulfillment of the requirement for the professional degree of

**Master** **of** **Electronic** **and** **Information** **Engineering**

by

**Xue** **Linsong**

**(Big** **Data** **Technology** **and** **Engineering)**

Thesis Supervisor : Associate Professor Zhang Kai

**March,** **2025**

学位论文指导小组、公开评阅人和答辩委员会名单

指导小组名单

|  |  |  |
| --- | --- | --- |
| 李 XX | 教授 | 清华大学 |
| 王 XX | 副教授 | 清华大学 |
| 张 XX | 助理教授 | 清华大学 |

公开评阅人名单

|  |  |  |
| --- | --- | --- |
| 刘 XX | 教授 | 清华大学 |
| 陈 XX | 副教授 | XXXX 大学 |
| 杨 XX | 研究员 | 中国 XXXX 科学院 XXXXXXX 研究所 |

答辩委员会名单

|  |  |  |  |
| --- | --- | --- | --- |
| 主席 | 赵 XX | 教授 | 清华大学 |
| 委员 | 刘 XX 杨 XX  黄 XX 周 XX | 教授 研究员  教授 副教授 | 清华大学  中国 XXXX 科学院  XXXXXXX 研究所  XXXX 大学  XXXX 大学 |
| 秘书 | 吴 XX | 助理研究员 | 清华大学 |

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包 括：（1） 已获学位的研究生必须按学校规定提交学位论文， 学校可以 采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为 教学和科研目的， 学校可以将公开的学位论文作为资料在图书馆、资 料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容； （3）按照上级教育主管部门督导、抽查等要求，报送相应的学位论文。

本人保证遵守上述规定。

作者签名： 导师签名：

日 期： 日 期：

摘 要

近年来，固定路线中运行的车辆、机器人等智能体越来越受到研究者关注， 而 定位是这类智能体的基础性功能之一， 因此固定路线中的定位有着重要的研究意 义和应用价值。以往的定位方法一般需要使用高精度的卫星信号或高成本的传感 器来实现精确的定位功能，这些昂贵且需要精心维护的设备限制了低成本固定路 线智能体的应用。

为了使用低成本传感器完成较高精度的定位任务，本文提出了一种使用视觉 惯性信息的固定路线定位系统，该系统涵盖了从建图到定位的完整流程， 并分别 从离线建图、里程计、地图定位三个方面针对现有方法中的设计问题进行改进：

(1) 现有的建图方法中存在着精度与成本的矛盾，低成本的建图方法，例如同 步定位与建图，往往因为在线建图的局部优化限制而存在精度劣势。针对这一问 题本文设计了一种以运动结构恢复 (Structure from Motion,SfM) 为基础并融合了高 精度全局信息的离线建图模块。这一模块结合了 SfM 的全局优化精度优势和全局 信息所提供的尺度信息，能够恢复出高精度且具有真实尺度的视觉点云地图。

(2) 现有的基于通用场景设计的视觉惯性里程计 (Visual-Inertial Odometry, VIO)，忽略了车辆和轮式机器人运动模式中的先验知识。针对这一问题本文设计 了一种基于车身运动模式感知的伪观测视觉惯性里程计 (Pseudo Obsevation Visual- Inertial Odometry, PO-VIO) 模块。这一模块根据车身运动模式来构建合理的伪观测 约束，基于伪观测约束来估计车身与惯性传感器的标定参数，联合优化标定参数 与车身状态量，能够估计出更为合理且更高精度的车身状态。

(3) 现有的地图定位方法普遍将定位问题看作是基于地图观测的最大似然估计 问题，这一做法忽略了地图本身存在的误差。针对这一问题本文设计了一种基于 最大后验概率估计的地图定位模块。这一模块将地图点的误差建模为以其空间坐 标为中心的三维高斯分布， 并基于这一先验概率分布进行位置和姿态的最大后验 概率估计，有效减小了建图误差带来的定位误差。

本文所提出的系统在 3 个公开数据集上进行了测试，论证了本文所提出系统 的有效性， 并通过消融实验分析了各种设计的效果。实验表明，本文所提出的固 定路线中的视觉惯性定位系统有着较高的定位精度，在理想场景下可以达到厘米 级的定位精度，并且可以适应天气、光照变化等环境因素改变的场景。

关键词： 固定路线；状态估计；同步定位与建图；视觉惯性里程计

Abstract

In recent years, intelligent agents operating on fixed routes—such as vehicles and robots—have attracted growing research interest. Localization, as one of their fundamen- tal functions, is of considerable research significance and practical value. Conventional localization methods typically rely on high-precision satellite signals or expensive sensors to achieve accurate positioning; however, these costly and maintenance-intensive devices hinder the application of low-cost fixed-route intelligent agents.

To reconcile the trade-off between low-cost sensors and high-accuracy localization, this paper introduces a visual–inertial localization method specifically designed for fixed- route applications. The proposed framework offers a comprehensive pipeline spanning from mapping to localization, with key enhancements in three areas: offline mapping, Visual–Inertial Odometry (VIO), and map-based localization.

(1) Conventional Simultaneous Localization And Mapping (SLAM) methods often suffer from a lack of global optimization, leading to suboptimal accuracy. To address this, we introduce an offline mapping method based on Structure from Motion (SfM) that integrates high-precision global information. This method leverages the global optimiza- tion advantages of SfM and the scale information provided by global data to reconstruct a high-precision, metrically accurate visual point cloud map.

(2) Existing VIO systems designed for general scenarios tend to overlook the kine- matic priors inherent in vehicles and wheeled robots. In response, we propose a VIO approach that incorporates pseudo-observation constraints based on vehicle kinematics. The method first classifies and estimates the vehicle state, performs an initial coarse cal- ibration between the vehicle body and the inertial sensor, and then applies zero-velocity constraints along the lateral and gravitational axes to jointly optimize the vehicle state and calibration parameters, thereby improving VIO accuracy.

(3) Most map-based localization methods treat the localization problem as a maxi- mum likelihood estimation based solely on map observations, neglecting potential map- ping errors. To overcome this limitation, we propose a maximum a posteriori estimation framework that integrates prior map information. By modeling map points as Gaussian distributions centered on the coordinates obtained during mapping, our method effectively mitigates localization errors caused by mapping inaccuracies.

Extensive experiments on three publicly available datasets—covering offline map- ping,odometry, and map localization—demonstrate the effectiveness of the proposed ap- proach. Ablation studies further confirm the contributions of each module. Under ideal conditions, our method achieves centimeter-level localization accuracy and exhibits ro- bustness to variations in weather and lighting conditions.

**Keywords:** fixed route; state estimation; SLAM; VIO

目 录

[摘 要](#bookmark1) [I](#bookmark1)

[Abstract](#bookmark2) [II](#bookmark2)

[目 录](#bookmark3) [IV](#bookmark3)

[插图清单](#bookmark8) [VII](#bookmark8)

[附表清单](#bookmark10) [IX](#bookmark10)

[符号和缩略语说明](#bookmark12) [X](#bookmark12)

第 1 章 [绪论](#bookmark14) [1](#bookmark14)

1.1 [课题背景](#bookmark16) [1](#bookmark16)

1.2 [国内外研究现状](#bookmark18) [3](#bookmark18)

1.2.1 [视觉惯性定位](#bookmark20) [3](#bookmark20)

1.2.2 [地图辅助的定位方法](#bookmark22) [6](#bookmark22)

1.3 [研究内容](#bookmark24) [9](#bookmark24)

1.3.1 [解决高精度先验地图构建问题](#bookmark26) [9](#bookmark26)

1.3.2 [解决符合车辆运动学的视觉惯性融合问题](#bookmark28) [10](#bookmark28)

1.3.3 [解决地图观测与定位优化问题](#bookmark30) [10](#bookmark30)

1.4 [研究创新点](#bookmark32) [11](#bookmark32)

1.5 [固定路线中的视觉惯性定位系统整体设计](#bookmark34) [11](#bookmark34)

1.6 [论文组织结构](#bookmark36) [13](#bookmark36)

第 2 章 [融合 SfM 与 GNSS 的离线建图模块设计](#bookmark38) [14](#bookmark38)

2.1 [坐标系定义](#bookmark40) [14](#bookmark40)

2.2 [模块整体设计](#bookmark42) [15](#bookmark42)

2.3 [建图预处理方法](#bookmark44) [15](#bookmark44)

2.3.1 [地图关键帧筛选算法](#bookmark46) [16](#bookmark46)

2.3.2 [基于语义分割算法的地图动态元素剔除](#bookmark48) [16](#bookmark48)

2.4 [基于 SfM 的无尺度建图方法](#bookmark50) [17](#bookmark50)

2.5 [基于三维相似变换的建图对齐方法](#bookmark52) [19](#bookmark52)

2.6 [基于 GNSS 增强的建图融合方法](#bookmark54) [19](#bookmark54)

2.7 [本章总结](#bookmark56) [20](#bookmark56)

第 3 章 [基于车身运动模式感知的 PO-VIO 设计](#bookmark58) [21](#bookmark58)

3.1 [PO-VIO 整体设计](#bookmark60) [21](#bookmark60)

3.2 [基于神经网络的车身运动模式判别方法](#bookmark62) [23](#bookmark62)

3.2.1 [网络输出定义](#bookmark64) [24](#bookmark64)

3.2.2 [网络输入定义](#bookmark66) [24](#bookmark66)

3.2.3 [网络结构设计](#bookmark68) [24](#bookmark68)

3.2.4 [网络训练方法](#bookmark70) [25](#bookmark70)

3.3 [基于车身运动学假设的伪观测预处理方法](#bookmark72) [26](#bookmark72)

3.4 [融合伪观测约束的非线性优化方法](#bookmark74) [29](#bookmark74)

3.4.1 [非线性优化的状态量与目标函数](#bookmark76) [29](#bookmark76)

3.4.2 [优化残差的雅可比矩阵计算](#bookmark78) [30](#bookmark78)

3.5 [本章总结](#bookmark80) [31](#bookmark80)

第 4 章 [基于最大后验概率估计的地图定位模块设计](#bookmark82) [32](#bookmark82)

4.1 [模块整体设计](#bookmark84) [32](#bookmark84)

4.2 [粗到细视觉定位方法](#bookmark86) [33](#bookmark86)

4.2.1 [基于神经网络的粗到细特征匹配算法](#bookmark88) [33](#bookmark88)

4.2.2 [基于概率投票的鲁棒定位算法](#bookmark90) [35](#bookmark90)

4.3 [紧耦合优化预处理方法](#bookmark92) [36](#bookmark92)

4.3.1 [坐标系转换矩阵初始化算法](#bookmark94) [36](#bookmark94)

4.3.2 [初始位姿的自适应选择算法](#bookmark96) [37](#bookmark96)

4.4 [基于最大后验概率估计的紧耦合优化](#bookmark98) [37](#bookmark98)

4.4.1 [最大后验概率估计问题建模](#bookmark100) [37](#bookmark100)

4.4.2 [使用位姿图求解紧耦合优化问题](#bookmark102) [39](#bookmark102)

4.5 [转换矩阵更新方法](#bookmark104) [41](#bookmark104)

4.6 [本章总结](#bookmark106) [42](#bookmark106)

第 5 章 [视觉惯性定位系统搭建与测试](#bookmark108) [43](#bookmark108)

5.1 [系统与测试细节](#bookmark110) [43](#bookmark110)

5.1.1 [硬件与软件环境](#bookmark112) [43](#bookmark112)

5.1.2 [测试数据集与评价指标](#bookmark114) [44](#bookmark114)

5.2 [离线建图模块测试](#bookmark116) [49](#bookmark116)

5.3 [PO-VIO 模块测试](#bookmark118) [51](#bookmark118)

5.3.1 [车身状态判别方法的性能测试](#bookmark120) [51](#bookmark120)

5.3.2 [PO-VIO 的性能测试](#bookmark122) [52](#bookmark122)

5.3.3 [PO-VIO 性能提升的来源分析](#bookmark124) [55](#bookmark124)

5.3.4 [消融实验](#bookmark126) [56](#bookmark126)

5.4 [紧耦合地图定位模块测试](#bookmark128) [58](#bookmark128)

5.4.1 [精度测试](#bookmark130) [58](#bookmark130)

5.4.2 [消融实验](#bookmark132) [68](#bookmark132)

5.5 [本章总结](#bookmark134) [69](#bookmark134)

第 6 章 [总结与展望](#bookmark136) [70](#bookmark136)

6.1 [论文总结](#bookmark138) [70](#bookmark138)

6.2 [未来展望](#bookmark140) [71](#bookmark140)

[参考文献](#bookmark142) [72](#bookmark142)

附录 A [补充内容](#bookmark144) [77](#bookmark144)

[致 谢](#bookmark146) [81](#bookmark146)

[声 明](#bookmark148) [82](#bookmark148)

[个人简历、在学期间完成的相关学术成果](#bookmark150) [83](#bookmark150)

[指导教师评语](#bookmark152) [84](#bookmark152)

[答辩委员会决议书](#bookmark154) [85](#bookmark154)

插图清单

图 1.1 [固定路线智能体的应用场景](#bookmark156) [1](#bookmark156)

图 1.2 [不同类型的地图](#bookmark158) [2](#bookmark158)

图 1.3 [视觉定位中的关键帧和滑动窗口](#bookmark160) [4](#bookmark160)

图 1.4 [视觉惯性定位中的关键帧和滑动窗口](#bookmark162) [5](#bookmark162)

图 1.5 [端到端定位效果](#bookmark164) [9](#bookmark164)

图 1.6 [系统整体设计示意图](#bookmark166) [12](#bookmark166)

图 2.1 [坐标系定义与转换](#bookmark168) [14](#bookmark168)

图 2.2 [离线建图模块整体流程](#bookmark170) [15](#bookmark170)

图 2.3 [关键帧视差对地图点精度的影响](#bookmark172) [16](#bookmark172)

图 2.4 [语义分割效果](#bookmark174) [17](#bookmark174)

图 3.1 [车身坐标系和 IMU 体坐标系分解示意](#bookmark176) [21](#bookmark176)

图 3.2 [视觉惯性里程计框架](#bookmark178) [22](#bookmark178)

图 3.3 [车身状态分布情况](#bookmark180) [27](#bookmark180)

图 4.1 [粗到细匹配过程](#bookmark182) [32](#bookmark182)

图 4.2 [粗到细匹配过程](#bookmark184) [34](#bookmark184)

图 4.3 [独立粗到细定位结果中的跳变](#bookmark186) [36](#bookmark186)

图 4.4 [紧耦合优化位姿图](#bookmark188) [39](#bookmark188)

图 5.1 [软件部署环境](#bookmark190) [43](#bookmark190)

图 5.2 [KITTI 数据集场景](#bookmark192) [44](#bookmark192)

图 5.3 [Office Loop 场景](#bookmark194) [45](#bookmark194)

图 5.4 [Neighborhood 场景](#bookmark196) [46](#bookmark196)

图 5.5 [Business Campus 场景](#bookmark198) [46](#bookmark198)

图 5.6 [Old Town 场景](#bookmark200) [46](#bookmark200)

图 5.7 [KAIST 数据集场景](#bookmark202) [47](#bookmark202)

图 5.8 [KAIST 数据集上 RP E r↓ 分布示意图](#bookmark204) [55](#bookmark204)

图 5.9 [车身速度展示](#bookmark206) [56](#bookmark206)

图 5.10 [KITTI 数据集测试轨迹示意图](#bookmark208) [60](#bookmark208)

图 5.11 [4Seasons 数据集定位结果及误差分析](#bookmark210) [60](#bookmark210)

图 5.12 [4Seasons 数据集 OF3-5 场景差异展示](#bookmark212) [62](#bookmark212)

图 5.13 [4Seasons 数据集 NH3 与 NH7 场景差异展示](#bookmark214) [64](#bookmark214)

图 5.14 [不同定位方法在 OF5 序列上的定位误差分布](#bookmark216) [65](#bookmark216)

图 5.15 [不同定位方法在 OF6 序列上的定位误差分布](#bookmark218) [66](#bookmark218)

图 5.16 [4Seasons 数据集 NH3 与 NH7 场景差异展示](#bookmark220) [66](#bookmark220)

附表清单

表 2.1 [语义分割类别分布与屏蔽情况](#bookmark222) [16](#bookmark222)

表 3.1 [车身判别网络设计参数](#bookmark224) [25](#bookmark224)

表 5.1 [本文中所涉及的运行超参数](#bookmark226) [44](#bookmark226)

表 5.2 [数据集使用划分](#bookmark228) [47](#bookmark228)

表 5.3 [KITTI 数据集建图 AT E t↓ 对比](#bookmark230) [49](#bookmark230)

表 5.4 [KITTI 数据集建图 RPE↓ 对比](#bookmark232) [50](#bookmark232)

表 5.5 [车身状态预测效果](#bookmark234) [51](#bookmark234)

表 5.6 [视觉惯性里程计的 AT E t↓ 对比](#bookmark236) [53](#bookmark236)

表 5.7 [视觉惯性里程计的 RPE↓ 对比](#bookmark238) [53](#bookmark238)

表 5.8 [KAIST 数据集视觉惯性里程计效果对比](#bookmark240) [54](#bookmark240)

表 5.9 [视觉惯性里程计消融实验](#bookmark242) [57](#bookmark242)

表 5.10 [KITTI 数据集多指标对比](#bookmark244) [59](#bookmark244)

表 5.11 [4Seasons 数据集 Office Loop 场景第 2 序列指标对比](#bookmark246) [61](#bookmark246)

表 5.12 [4Seasons 数据集 Office Loop 场景部分序列指标对比](#bookmark248) [63](#bookmark248)

表 5.13 [4Seasons 数据集 Neighborhood 场景第 2 与第 3 序列指标对比](#bookmark250) [64](#bookmark250)

表 5.14 [4Seasons 数据集 Neighborhood 场景第 4 与第 7 序列指标对比](#bookmark252) [65](#bookmark252)

表 5.15 [4Seasons 数据集 Business Campus 场景部分序列指标对比](#bookmark254) [67](#bookmark254)

表 5.16 [4Seasons 数据集 Old Town 场景第 2 序列指标对比](#bookmark256) [67](#bookmark256)

表 5.17 [消融实验结果对比](#bookmark258) [68](#bookmark258)

符号和缩略语说明

GNSS 卫星导航系统 (Global Navigation Satellite System)

PPP 精密单点定位 (Precise Point Positioning)

RTK 实时动态定位 (Real-Time Kinematic)

SLAM 同步定位与地图创建 (Simultaneous Localization And Mapping)

EKF 扩展卡尔曼滤波器 (Extended Kalman Filter)

BA 光束法平差 (Bundle Adjustment)

IMU 惯性测量单元 (Inertial Measurement Unit)

VO 视觉里程计 (Visual Odometry)

VIO 视觉惯性里程计 (Visual-Inertial Odometry)

SfM 运动恢复结构 (Structure from Motion)

Sim3 三维相似变换 (Similarity Transformation in 3D)

SO(3) 三维旋转群 (Special Orthogonal Group)

SE(3) 三维特殊欧氏群 (Special Euclidean Group)

CNN 卷积神经网络 (Convolutional Neural Network)

TCN 时间卷积神经网络 (Temporal Convolutional Neural Network)

MAP 最大后验概率 (Maximum A Posteriori)

ML 最大似然 (Maximum Likelihood)

第 1 章 绪论

本章节分析了现有的固定路线定位技术的研究背景和研究意义，介绍了国内 外关于视觉惯性定位和地图辅助定位的研究现状，并指出了现有研究的不足之处。 本章通过论述核心研究内容，提炼出了文章的创新之处； 简要介绍了整个系统的 总体设计与其对应的研究内容；最后，给出了本文的组织结构。

1.1 课题背景

近年来，汽车和机器人的智能化是炙手可热的研究领域，而固定路线下运行 的智能汽车和机器人是其中的一个重要研究和应用方向。相较于通用智能，在固 定路线下应用的智能化技术更容易实现， 因此固定路线场景下的应用涌现，例如 图 [1.1](#bookmark156) 中所示的无人驾驶公交车[[1](#bookmark259)]、智驾汽车通勤模式[[2](#bookmark260)[]、变电站巡检车[3](#bookmark261)]、矿山 无人运输机器人[[4](#bookmark262)]等，都是目前固定路线智能体的典型应用场景。



(a) 无人驾驶公交车[[1](#bookmark259)] (b) 智驾汽车通勤模式[[2](#bookmark260)]



(c) 变电站巡检车[[3](#bookmark261)] (d) 矿山无人运输机器人[[4](#bookmark262)]

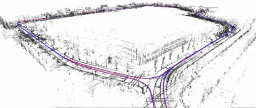
图 1.1 固定路线智能体的应用场景

在固定路线场景中，汽车和机器人运行在提前设置好的运行线路上，通过自 身所具备感知能力对环境做出反馈，完成规划、导航等更复杂的任务。而在感知

能力中， 定位是十分基础但却十分关键的感知能力之一，定位系统是智能汽车和 机器人运行的最基础系统之一。因此， 固定路线中的精确定位是一项具有较高实 用价值的基础性研究， 只有实现了高精度的定位，才能确保复杂任务的可靠实施， 才能保证整个智能体的安全。

在定位方法中，视觉惯性定位 (Visual-Inertial Localization) 是一种应用广泛的 定位形式，它通过融合视觉和惯性传感器的数据，实现对自身姿态与位置的估计。 视觉传感器可以获取环境的语义信息，而惯性传感器，例如惯性测量单元 (Inertial Measurement Unit, IMU)，可以获取自身的运动信息，两者结合可以实现对自身位 置的估计。视觉惯性定位方法具有定位精度上限高、成本低、易于实现等优点， 因 此在固定路线定位中得到了广泛的应用。然而，视觉惯性定位方法也存在一些问 题，例如视觉传感器易受光照、天气等环境因素影响， 惯性传感器易受积分漂移等 因素影响，这些因素都会影响视觉惯性定位的精度。因此，如何提高视觉惯性定 位的精度，是当前固定路线定位研究的一个重要问题。

一种主要的改进方式是通过引入卫星导航系统 (Global Navigation Satellite Sys- tem, GNSS) 信息来提高视觉惯性定位的精度。 一般的商用低成本单频 GNSS 的精 度在 10 米左右，这对于需要精确位置的智能化车辆和机器人来说是远远不够的。 为了改进单频 GNSS 的精度，近年来出现了一些高精度的增强 GNSS 方法，例如 精密单点定位 (Precise Point Positioning, PPP)[[5](#bookmark263)]技术和实时动态载波相位差分 (Real Time Kinematic, RTK)[[6](#bookmark264)]技术。PPP 和 RTK 可以将 GNSS 的定位误差缩减至厘米 级，是非常理想的高精度定位方法。但是 PPP 需要较长的初始化时间[[7](#bookmark265)]， 所以不 适合应用在实时定位场景上。RTK 虽然有不错的实时性，但是却经常容易受到天 [气、温度或者遮挡等问题的影响而产生单次误差较大的定位结果[8](#bookmark266)]。因此，直接 引入 GNSS 观测信息会受到多种限制，并且高精度的 GNSS 服务也需要额外的费 用，这对于一些低成本、大批量的固定路线应用来说是难以接受的。



(a) 人类可理解语义地图 (b) 视觉点云地图

图 1.2 不同类型的地图

另一种改进方式是通过引入地图信息来提高视觉惯性定位的精度。此处所指

的地图并非是人类理解的语义地图 [1.2(a)](#bookmark158)，而是稀疏点云地图，如图 [1.2(b)](#bookmark158)所示。点

云地图由图像、拍摄图像的相机位置与姿态， 图像特征点及其空间坐标所构成，是 一种适合计算机存储和使用的地图。地图信息可以提供给视觉惯性定位系统一个 先验的位置信息，从而可以减小定位误差。在固定路线条件下， 由于车辆或机器 人运行的路线是固定的， 因此可以提前获取到路线的地图信息，这方便于视觉惯 性定位系统使用地图先验信息。除此之外， 因为地图信息的收集过程没有实时性 要求，可以在精心选择的条件下进行，所以还可以排除天气、温度等环境因素的 干扰而使用高精度 GNSS 信息来辅助提高建图精度。从经济性方面考虑，先验地 图可以一次建立、多次使用，而使用高精度 GNSS 往往需要购买连续运行参考站 (Continuously Operating Reference Stations, CORS) 服务商的持续服务[[9](#bookmark267)]， 会产生持 续的运营成本。综合来看， 引入地图信息有着较高的可实施性和经济性，是一种 理想的提高固定路线视觉惯性定位精度的方法。

总的来说， 固定路线中的视觉惯性定位是一项有着广泛应用的基础技术，但 目前受限于视觉惯性定位技术本身的局限性，其精度还有待提高。在目前可选的 改进方法中， 引入地图信息是一种比较理想的提高视觉惯性定位精度的选择。因 此，本文将围绕固定路线中的视觉惯性定位系统展开研究，通过解决地图构造、地 图识别和地图使用等问题来提高定位的精度，设计一种精度满足需求，且具有实 际操作性的固定路线中的视觉惯性定位系统。

1.2 国内外研究现状

1.2.1 视觉惯性定位

视觉惯性定位是一种融合了视觉、惯性信息的定位方法，其中视觉信息一般 来源于图像、视频等， 而惯性信息则一般依靠 IMU 采集。视觉惯性定位方法以视 觉定位和惯性定位技术为基础，但是目前的视觉惯性定位技术一般以视觉定位技 术为核心，而惯性定位技术作为补充或约束信息加入到视觉定位中， 因此介绍视 觉惯性定位的发展有必要从视觉定位技术的发展说起。

视觉定位中常用的技术之一是同步定位与地图创建 (Simultaneous Localization And Mapping, SLAM)。Davison et al. [[10](#bookmark268)]于 2007 年首次提出了使用单目图像进行定 位的方法 MonoSLAM， 以扩展卡尔曼滤波器 (Extended Kalman Filter ，EKF) 为核 心，实时对图像观测到的特征点进行位置估计，构建 3D 概率稀疏地图，然后使用 地图点对相机位置和姿态进行估计，这是第一种能够达到实时的视觉定位方法，它 使得利用视觉信息估计姿态成为可能。同年 Klein et al.[[11](#bookmark269)]也提出了一种基于单目 图像的定位方法 PTAM(Parallel Tracking And Mapping) ，与 MonoSLAM 不同的是， PTAM 的核心估计方法从扩展卡尔曼滤波器改变为光束法平差 (Bundle Adjustment,

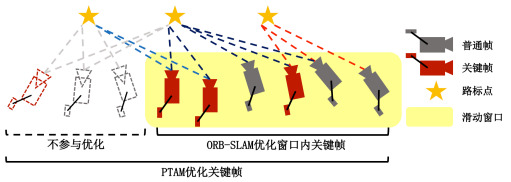


图 1.3 视觉定位中的关键帧和滑动窗口

BA)[[12](#bookmark270)] 。BA 是一种非线性优化方法，其根据同一个地图点在多个图像上的观测位 置， 同时优化图像拍摄时相机的位姿和地图点的空间坐标。因为一次优化使用的 数据更多，所以这种方法相比于滤波器更准确，但是计算量更大，所以 PTAM 引 入了关键帧的概念： 只对关键帧进行光束法平差，而对于普通帧则使用关键帧对 其进行约束和定位估计。

PTAM 的另一大进步是将定位与建图两个功能解耦，使两个功能分别同时进 行，而这一优点被 Mur-Artal et al. [[13](#bookmark271) ,[14](#bookmark272)]吸取，创造出 ORB-SLAM。ORB-SLAM 是 一个包含 3 个主要线程的完整 SLAM 系统： 实时特征点跟踪 (Tracking) 线程、局 部建图优化 (Local Mapping) 线程、回环检测 (Loop Closing) 线程。ORB-SLAM 中 不仅继承了 PTAM [的关键帧思想，还使用了如图1.3](#bookmark160)所示的滑动窗口概念，进一步 提高了系统实时性。滑动窗口设置包含固定数量关键帧的窗口， 随着系统的工作， 窗口随时添加新的关键帧并舍弃距离现在最远的关键帧，BA 只优化当前窗口内关 键帧及其地图点的信息。ORB-SLAM 基本奠定了后来视觉定位方法的范式，其提 出的 3 个功能模块也组成后续视觉定位方式的基本骨架。

虽然视觉定位的发展取得了一些进步，但是视觉定位的系统性缺陷却阻碍其 应用。在视觉定位研究最广泛的单目视觉定位方面，其依靠的单目相机传感器天 然缺少对深度的观测，这造成了单目视觉定位的尺度不确定性，即单目视觉定位不 能获得真实世界尺度下的定位结果。为了解决这个问题，研究者们开始在视觉定位 的基础上加入了可以获取真实世界尺度信息的传感器，例如惯性测量单元 (Inertial measurement unit, IMU)，提出了视觉惯性定位。

视觉惯性定位是视觉定位和惯性定位的结合，涉及到多传感器信息融合，根 据融合方式可以分为松耦合[[15](#bookmark273)[]与紧耦合[16](#bookmark274)]。松耦合方式是将视觉定位和惯性定 位分开处理，让两种定位方式独立运行，然后对结果再进行融合，融合的方式可以 选择扩展卡尔曼滤波或者非线性优化。松耦合的优点是系统简单，估计量少， 因 此不论是运行效率还是灵活度都非常高； 缺点是忽略了两种定位方式之间的约束

关系，所以精度较差。目前主流的视觉惯性定位基本是紧耦合方式， 紧耦合需要 将两种定位方式的估计量作为一个整体同时优化。

紧耦合视觉惯性定位方法最早以滤波器的形式发展，此时视觉惯性定位依旧 是以 IMU 运动模型为核心， 以线性化和观测性为主要研究方向， 最突出的工作 是 Mourikis et al. [[17](#bookmark275)]提出的多状态约束卡尔曼滤波 (Multi-State Constraint Kalman Filter ，MSCKF)，该方法以 IMU 位置、姿态、速度等状态量和固定数量的相机位 置、姿态为主要的估计参数， 并没有将地图点加入到优化列表中。MSCKF 工作的 最主要贡献在于推导出一种测量模型，该模型能够表达从多个相机位姿观察到静 态特征时出现的几何约束。

此后, 许多基于 MSCKF 的工作相继提出， 框架整体的精度和鲁棒性得到了 不断的提升， 例如 Li et al. [[18](#bookmark276)]在 MSCKF 的基础上提出的 MSCKF 2.0 以及 Sun et al. [[19](#bookmark277)]提出的双目视觉版本 MSCKF 。Bloesch et al. [[20](#bookmark278)]提出了一种基于迭代扩展 卡尔曼滤波器 (Iterated extended Kalman filter ，IEKF) 的直接法单目视觉惯性里程 计 ROVIO(RObust Visual Inertial Odometry)。在视觉方面，该方法将地图点在图像 中对应点周围的图像块作为路标点的描述子，从而得到光度误差，然后将光度误 差进行变换得到 IEKF 中的启发项，进而进行滤波状态的更新。整体的滤波方程的 构造是以 IMU 为中心进行构造的，保证能观状态不受不断增长的全局协方差的影 响，这样可以减小因非线性而造成的误差。

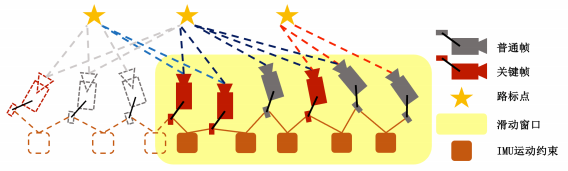


图 1.4 视觉惯性定位中的关键帧和滑动窗口

近年来随着非线性优化在视觉定位中的广泛应用，视觉惯性定位方法也尝试 使用这种优化方式。参考视觉定位中将地图点和相机位姿以图 (Graph) 的形式联 系起来，然后使用光束法平差进行统一优化，视觉惯性定位在相机位姿之间加入 由 [IMU 数据组成的运动约束项，如图1.4](#bookmark162)所示。IMU 运动约束指的是 IMU 预积分， Forster et al. [[21](#bookmark279)]提出了应用于视觉惯性里程计中的基于四元数的预积分公式，将传 统的 IMU 运动积分公式转化为帧间增量的形式， 并且将运动积分中的 IMU 偏置 (Bias) 线性化，避免在优化过程中IMU 偏置发生变化时重新积分，提高运行效率。 Leutenegger et al. [[22](#bookmark280)]提出的 OKVIS(Open Keyframe-based Visual-Inertial SLAM) 利

用基于关键帧的滑动窗口进行批量非线性优化，先于滑动窗口的关键帧被边缘化， 不用来进行估计。系统前端使用多尺度 Harris[[23](#bookmark281)]特征检测器来提取特征点，然后 在其基础上计算 BRISK(Binary Robust Invariant Scalable Keypoint)[[24](#bookmark282)]描述子， 以便 在帧与帧之间进行数据关联。

Qin et al. [[25](#bookmark283)]提出的 VINS-Mono(Monocular Visual-INertial System) 类似于 OKVIS，但引入了几个全新的功能，其完整系统包括观测值预处理、初始化、局部 视觉惯性联合优化、全局图优化和回环检测 5 个部分，前端提取 Harri 特征点，并 采用 LK 光流 (Lucas-Kanade opticalflow)[[26](#bookmark284)]法跟踪相邻帧。VINS-Mono 方法只计 算特征点，不计算描述子，同时使用光流法跟踪特征点的运动，这样就减少了计算 和匹配描述子的时间和资源。系统采用与 OKVIS 相似的基于滑动窗口的紧耦合位 姿估计方法，并且加入了基于词袋模型 (Bag of Binary Words，DBoW)[[27](#bookmark285)]的回环检 测线程，使系统具有重定位功能。Liu et al. [[28](#bookmark286)]提出 ICE-BA(Incremental, Consistent and Efficient Bundle Adjustment)，沿用 VINS-Mono 中基于 LK 光流的特征点跟踪 技术，后端则是提出了增量式 BA，主要分为 3 个部分： 局部 BA、全局 BA 以及 相对边缘化 (Relative-Marginalization)，前两者采用增量式方法提升了后端速度，后 者保证了局部 BA 和全局 BA 的一致性。

一些视觉定位方法也发展出了融合惯性信息的版本，Campos et al. [[29](#bookmark287)]在 ORB- SLAM 的基础上提出了 ORB-SLAM 3，引入 IMU 尝试解决在快速运动时丢失特征 点的问题：ORB-SLAM 3 分别对 ORB-SLAM 的 3 个线程作出修改，用以融合 IMU 信息：在跟踪线程，基于重投影误差和 IMU 预积分，建立帧与帧之间的约束关系 来构造代价函数，从而得到当前帧位姿的最优估计； 在局部建图线程，有了新的 关键帧之后，将会对前 N 个关键帧进行优化，当前的关键帧 (第 N+1 帧) 将固定不 变； 在全局回环检测线程， 由于 IMU 提供了尺度信息， 因此全局优化将从 7 个自 由度下降到 6 个自由度，全局位姿优化将忽略 IMU 信息， 因此不再优化速度和偏 差，当完成全局位姿优化后，再根据矫正后的位姿对速度进行矫正。

**总结而言，** **视觉惯性定位系统的组成部分至今已基本定型，** **其主要由视觉** **SLAM** **系统的功能组件和预积分的惯性测量处理组成。因此，** **目前的视觉惯性** **定位方法仍遵循定位与建图同步的工作逻辑，** **并没有考虑过如何将先验地图运用** **到定位工作中。本文为了补充这一研究，希望提出一种地图辅助的视觉惯性定位** **方法，改变以往定位与建图高度同步的范式，这与此前的工作有较明显的差异。**

1.2.2 地图辅助的定位方法

地图辅助定位是一种车辆定位中常用的技术， 常用的地图类型有高精 (High- Definition，HD) 地图、激光雷达点云地图、视觉重建地图等。 Jeong et al. [[30](#bookmark288)]开发了

一种高效的 HD 地图表示方法，并利用粒子滤波器估计车辆的六自由度 (6-Degree Of Freedom，DOF) 位姿，该方法实现了分米级精度。Xiao et al.[[31](#bookmark289)]和 Guo et al.[[32](#bookmark290)]提 出了一种基于 HD 地图的低成本视觉定位方法。这两种方法综合利用了低级地图 特征 (如点和线特征) 以及结构特征 (如地图元素的语义和类别)，从而能够精确估 计车辆的位姿。尽管上述方法在定位方面取得了令人印象深刻的成果，但它们都 依赖于成本高昂的高清地图，这些地图的构建需要细致的数据采集和人工标注。

为了降低对高清地图的依赖，研究者们开始探索基于图像和激光雷达扫描的 地图构建方法。Stewart et al. [[33](#bookmark291)]开发了一种利用单目相机在由搭载激光雷达传感器 的测绘车辆预先生成的三维地图中进行定位的方法。该方法通过将激光雷达反射 图像与图像强度进行匹配， 并选择 NMI(Normalized Mutual Information) 最高的合 成图像来确定相机位姿。Zuo et al. [[34](#bookmark292)]采用基于 MSCKF 的视觉惯性里程计 (Visual- Inertial Odometry, VIO) 技术，在由激光雷达生成的点云地图中进行定位。具体而 言，他们将点云地图作为观测值以更新 MSCKF，从而提升其性能。Lin et al. [[35](#bookmark293)]则 将图像引入到先验地图信息中，其方法通过激光雷达估计位姿， 同时利用车载图 像构建稀疏点云地图。

近年来， 随着视觉三维重建技术的进步， 涌现了以 SLAM 和运动恢复结构 (Structure from Motion ，SfM)[[36](#bookmark294)]为代表的地图重建方法。通过 SLAM 和 SfM 获得 的重建地图一般包含地图关键帧的位姿、稀疏点云结构、稀疏点云特征等要素， 这 些要素可以在定位时快速提供粗略的位置参考，然后选择该位置附近的局部点云 地图与当前定位图像进行细匹配和优化定位。具体来说，重建地图的地图关键帧 的位姿可以通过特征点匹配、局部地图优化和全局地图优化等手段获得； 重建地 图的稀疏点云结构一般是通过提取地图关键帧的特征点， 并根据地图关键帧的相 对位姿三角化获得； 重建地图的稀疏点云特征则是由地图关键帧的特征点描述子 组成。在定位过程中， 首先将待定位图像的特征点与地图的稀疏点云进行匹配，匹 配到足够的地图点后可以使用 PnP(Perspective-n-Point) 算法直接计算待定位图像 的位置和姿态。

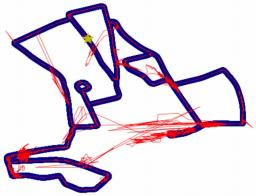
在具体的重建地图使用方面，VINS-Mono 、VINS-Fusion 和 ORB-SLAM 3 等 一些 SLAM 系统都已具有保存和加载先验地图的功能，这些 SLAM 系统可以提前 在指定场景下初始运行、建图，而在此后的运行中依靠初始运行的结果进行定位。 Surber et al. [[37](#bookmark295)]提出了一种针对无人机的定位系统， 旨在利用相机和 IMU，在初始 飞行中构建参考地图。在后续飞行中，系统通过 VIO 和图像匹配技术，将当前观 测与参考地图进行对比，实现精确的自我定位。Hao et al. [[38](#bookmark296)]提出了一种两阶段的 视觉惯性定位系统： 在初始运行的过程中加入了额外的 GNSS 信息建图，而在后

续运行中保持纯净的视觉惯性定位。在此系统中，初始运行阶段的 GNSS 信息提 供了较强的位置约束， 因此其所建地图精度得到了提升。这些地图使用方法思考 了如何建立先验地图和实时场景中的特征点匹配关系，从而使用先验地图的精确 地图点为当前定位提供参考。

然而，上述方法在匹配过程中依旧使用着传统的手工设计特征，这些特征在匹 配精度和鲁棒性上仍有提升空间。除此之外，上述方法的重建地图基本都是基于 SLAM 系统的，这意味着地图的构建和定位是高度耦合的，系统本身的建图误差会 [传导、积累到定位过程中。虽然有工作[38](#bookmark296)]尝试在建图阶段加入额外位置信息， 提 升建图效果，但其仍使用注重实时性的 SLAM 系统进行建图，依然会受到 SLAM 系统误差累积的影响。为了提升特征匹配的精度， 同时降低建图误差对定位的影 响，一些研究者开始尝试使用基于深度学习的特征搭配精度较高建图技术，例如 SfM 或者带 GNSS 增强的 [SfM[39](#bookmark297)] ，来提升整体的定位精度。

Sarlin et al. [[40](#bookmark298)]提出了一种“由粗到细”的地图使用方式：其建图过程使用 SfM 和具有精确位姿的图片进行； 其定位方法首先使用基于深度学习的图片级描述子 [NetVLAD(Vector of Locally Aggregated Descriptors)[41](#bookmark299)]对定位图片进行粗位置查询， 此后使用基于深度学习的像素级描述子 SuperPoint [[42](#bookmark300)]对定位图片的特征点进行细 致匹配。这种方法在匹配过程中使用了基于深度学习的特征，从而提升了匹配的 精度和鲁棒性， 同时具有较高的实时性，为后续许多工作提供了一种地图定位范 式。Yang et al. [[43](#bookmark301)]提出了一种基于视觉重建地图的定位方法， 其方法在“由粗到 细”定位的基础上引入了 VIO，利用实际定位中的帧间相对位姿约束提升了定位 精度。Lin et al. [[44](#bookmark302)]则更是引入了多地图概念，其方法在匹配过程中使用了多个地 图，从而提升了匹配的鲁棒性。近年来， 由于深度学习端到端技术的启发，也出现 了一些使用隐式地图使用的定位方法，Xue et al. [[45](#bookmark303)]提出的 GNNMapNet 与 Wang et al. [[46](#bookmark304)]所提出的 RobustLoc 都是这类方法的代表。这类方法使用带有坐标和姿态 的图像作为训练集进行训练，在推理时可以只输入图像就端到端地输出定位结果。

虽然这些方法在最终的定位精度上有所提升，但是存在一些问题。首先，这 些研究在建图时直接假设地图图片的准确位姿是可获得的，这在实际操作中并不 一定总是成立的。其次， 这些研究聚焦于先验地图和实际定位之间的匹配关系， 并尝试通过各种手段来增加匹配的约束，例如相对位姿、多地图验证等，但是却 并没有考虑当地图和定位场景之间产生误匹配时如何降低误匹配的影响。这无疑 导致了这些定位方法的脆弱性，使得他们在实际应用中表现出不稳定的性能。此 外，Yang et al. [[43](#bookmark301)]在定位中引入了服务器组件， 因此对通信环境的要求较高；Lin et al. [[44](#bookmark302)]的方法使用多地图以及丰富的约束， 导致运行过程内存占用极大 (测试中



GNNMapNet [[45](#bookmark303)] 定位 (红色) 效果 RobustLoc [[46](#bookmark304)] 定位 (红色) 效果

图 1.5 端到端定位效果

使用 4T RAM)； GNNMapNet 和 RobustLoc 这类端到端的定位方法表现极不稳定， 产生的轨迹平滑性不足，如图 [1.5](#bookmark164)所示。这些问题都限制了上述方法在实际应用中 的推广。

**总结而言，地图辅助的定位方法研究较为分散，根据地图类型的不同，定位方** **法差异较大。使用** **HD** **地图和激光雷达点云地图的方法有较高精度，但是高昂的** **地图成本却也阻碍着这类技术在实际生产中的应用。基于重建地图的定位方式有** **较低的成本，但是当前的研究方法中还有一些限制，例如建图假设过强、定位容错** **不足和实际应用限制较多等问题。为了提供一种实施难度低、容错能力强、完全** **部署在汽车或机器人上的高精度定位方法，本文希望从建图方式开始，构建完整** **的固定路线中的视觉惯性定位系统，** **实现完全部署于车载或机器人端，** **并具有较** **高定位精度。**

1.3 研究内容

本文的研究内容是设计一种完整的固定路线视觉惯性定位系统，具体而言，本 文将通过设计 3 个子模块分别解决“高精度先验地图构建问题”、“符合车辆运动 学的视觉惯性融合问题”以及“地图观测与定位优化问题”，此后将 3 个模块整合 到一起，形成一个完整的固定路线视觉惯性定位系统。

1.3.1 解决高精度先验地图构建问题

先验地图的构建是本文首先要解决的问题，是剩余其他研究内容的基础。因 为 SLAM 类建图算法有实时运算要求， 因此只能进行局部优化，所以存在误差累 积，精度较差的问题。为解决这一问题，本文采用 SfM 作为基础建图方法，其全 局优化策略可以有效降低累计误差。

虽然 SfM 建图具有理论精度高的优势，但原始的 SfM 建图不具备现实尺度。 为了恢复出具有现实世界尺度的地图，必须增加有现实尺度的观测信息，本文所 采用的观测信息为高精度 GNSS(例如 RTK) 观测。虽然 RTK 在使用时可能会存在 瞬时测量受环境因素影响而产生波动等问题，但是对于建图环节来说，环境可以 静心挑选， 因此 RTK 完全可以运用在实际建图中，并且其较高精度的定位观测可 以为建图优化提供较好的约束。

本文将着力于融合 GNSS 观测信息与 SfM，提出一种精度较高且稳定性突出 的视觉先验地图构建方式，所构建的地图将具有现实世界尺度，且地图关键帧具 有较高的位姿估计精度，以此来解决先验地图构建问题。

1.3.2 解决符合车辆运动学的视觉惯性融合问题

视觉惯性信息的融合是定位的核心步骤，融合的好坏决定着最终定位的精度。 目前视觉惯性信息的融合方式主要分为卡尔曼滤波器和非线性优化两种，其中卡 尔曼滤波器的优势是在估计矩阵维度较低时计算简单，速度较快， 劣势是精度欠 缺；非线性优化的方法能够提供较高的精度，且随着硬件技术进步，实时运行非线 性优化方法也是可能的。

当下基于非线性优化的视觉惯性融合方法，大多数优先考虑通用场景，对于 固定路线中运行的车辆或轮式机器人缺乏特殊设计。车辆和轮式机器人的运动有 其特有的规律和约束，通用的设计往往难以满足这些约束， 导致估计的车身状态 有时会与车辆运动学产生明显冲突，例如会在车身横向和重力方向产生持续速度 估计误差，这种冲突也降低了状态估计的精度。

本文选择以非线性优化作为融合方案，着力于解决如何在非线性优化中构建 视觉信息和惯性信息的约束关系， 并加入符合固定路线运行的车辆和机器人车辆 运动学的合理约束，提高融合后的状态估计精度。

1.3.3 解决地图观测与定位优化问题

地图观测与定位优化是将先验地图与实际定位过程结合的关键。地图观测首 先要解决的问题是将地图点与待定位图像特征点进行关联，构建图像特征点观测 与图像位置、姿态参数之间的转换关系。基于特征点和描述子的关联方式具有一 定的可行性，但是考虑到建图过程可能与实际定位过程有较大的场景表现差异，需 要选择对场景差异不敏感的特征点及描述子。

定位优化是获得高质量定位结果的必要步骤，由于地图观测中的各种误差，定 位结果通常表现为不稳定的状态，为了维护一个相对稳定、平滑的轨迹， 需要定 位优化过程。在本文的系统中，视觉惯性融合的结果能够提供较为精确的帧间相

对位姿，再结合对地图的观测，可以进一步提高定位精度。但是如何处理两种来 源 (地图观测与视觉惯性融合) 的信息，如何处理两种来源信息的内在误差，是需 要解决的问题。

本文将着力于使用一种基于数据驱动的， 能够对场景差异较为鲁棒的先验地 图观测技术，构建定位与先验地图的关联。在此基础上使用一种考虑多种误差的 定位优化方法，将地图观测与视觉惯性融合的结果进行优化， 以获得高精度的定 位结果。

1.4 研究创新点

本文的创新点主要体现在以下几个方面：

(1) 提出了一种新颖且完整的固定路线视觉惯性定位方法，该方法将“高精度 先验地图构建”、“符合车辆运动学的视觉惯性融合”、“地图观测与定位优化”三 个内容有机结合，形成一个完整的视觉惯性定位系统。

(2) 在高精度先验地图构建方面，本文提出了一种融合 SfM 与 GNSS 的离线建 图模块，采用 SfM 和高精度 GNSS 观测数据相结合的方式，构建具有现实世界尺 度的多层次先验地图，提高地图的精度和稳定性。

(3) 在符合车辆运动学的视觉惯性融合方面，本文提出了一种基于车身运动模 式感知的伪观测视觉惯性里程计 (Pseudo Obsevation Visual-Inertial Odometry, PO- VIO)，该 VIO 使用深度神经网络对车身运动模式进行检测，针对直行状态下的车 辆或轮式机器人进行速度观测约束，提高了里程计的估计精度。

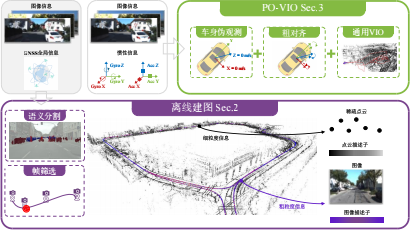
(4) 在地图观测与定位优化方面，本文在模块设计中使用了基于深度学习的地 图特征来完成地图观测，数据驱动的特征相较于传统手工设计特征能更好适应变 化场景。基于最大后验概率估计的定位优化方法，将地图观测与视觉惯性融合的 结果进行优化，同时引入了地图点的误差建模，获得了高精度的定位结果。

1.5 固定路线中的视觉惯性定位系统整体设计

本文所提出方法的整体系统设计如图 [1.6](#bookmark166)所示， 其主要包含 3 个模块：离线建 图模块、PO-VIO 模块和地图定位模块。在 3 个模块中，离线建图模块负责构建先 验地图；PO-VIO 模块和地图定位模块负责实际定位。PO-VIO 使用图片和惯性信 息可以获得连续图片的相对位姿变换； 地图定位则使用视觉惯性里程计结果和地 图观测完成最后的定位，获得相机位姿。整个系统的设计与研究内容的对应关系 是：离线建图模块对应着解决高精度先验地图构建问题的研究内容，PO-VIO 模块

对应着解决符合车辆运动学的视觉惯性融合问题的研究内容，地图定位模块对应 着解决地图观测与定位优化问题的研究内容。

|  |
| --- |
|  |
| |  | | --- | |  | | 粗到细定位 | |  | | 多个坐标系对齐 | |  | |



最大后验概率

地图定位 Sec.

**轨迹点**



**地图点**

模块初始化

图 1.6 系统整体设计示意图

离线建图模块利用 GNSS 观测数据和图像构建一个包含粗粒度层级和细粒度 层级组成的分层先验地图。粗粒度层级包括地图关键帧的位姿和关键帧描述子，而 细粒度层级则包含稀疏点云结构及点云点描述子。

GNSS 增强的 SfM 是离线建图模块的核心，使用这一技术有以下的必要性：首 先，SfM 作为一种具有全局优化功能的建图方式，其相较于 SLAM 更适合建立高 精度的地图；其次，SfM 作为一种单目建图技术，其所创建的地图并不具有真实物 理世界的尺度， 因此必须添加具有尺度信息的 GNSS 信息作为补充， 以获得具有 真实世界尺度的地图； 最后，相对精确的 GNSS 信息对于 SfM 的建图过程也有增 益，有助于获得更加精确的地图。

基于分层先验地图，PO-VIO 模块和地图定位模块计算实际定位时所拍摄图像 在全局坐标系中的位姿。

在 PO-VIO 模块中， 图像和惯性信息经过融合后可以相对精确地估算相邻图 片之间的位姿变换， 虽然相对位姿不可以直接用来作为定位结果，但是可以作为 约束条件用于实际的地图定位。作为一种约束条件，视觉惯性里程计的精度对于 整个定位系统的性能也发挥着重要作用， 因此，本文根据车身运动学的合理假设 提出了基于车身运动模式感知的 PO-VIO 模块设计： 车身的运动模式可以在不增 加传感器的前提下提供伪观测约束，以此来提升视觉惯性里程计的估计精度。

地图定位模块是整个系统最后， 也是最重要的模块，其主要功能包括地图观 测与定位、地图定位与 PO-VIO 坐标系对齐、定位优化等。地图的观测与定位将实 际拍摄的图片与地图进行粗到细匹配，逐渐获得一个较为粗糙的位姿观测； 地图 定位与 PO-VIO 坐标系对齐则将视觉惯性里程计的相对位姿转化到地图坐标系中，

方便其后续使用；定位优化是整个模块的核心，其同时参考视觉惯性里程计、地图 观测以及地图先验信息，以位姿图的形式通过非线性优化算法获得精确定位。

1.6 论文组织结构

本文共六章，各章内容安排如下：

第一章为绪论，首先介绍了本文的研究背景、研究意义， 此后根据对国内外研 究的调研介绍相关研究进展，通过与现有研究的对比引出本文研究的侧重点， 并 概括本文的研究内容和系统整体设计，最后对全文结构进行介绍。

第二章为离线建图模块介绍， 首先介绍了系统中所涉及的所有坐标系，为后 文定义各个模块的数学原理提供统一的参考； 此后介绍了离线建图模块的基本组 成，包括建图预处理、无尺度建图、建图对齐和建图融合等， 对每个过程的内部原 理做出详细解释。

第三章为 PO-VIO 模块介绍，本章首先从车身运动模式入手，将其分为 3 种类 别： 静止、直行和转弯。此后介绍了一种以深度学习为基础的车身运动模式检测 方法，该方法使用 Transformer[[47](#bookmark305)]对惯性信息序列进行处理， 以检测车辆处于何种 运动模式； 此后本章介绍了一种车身坐标系和 IMU 体坐标系之间的转换方法； 最 后本章介绍了如何将车身运动模式所附带的伪观测约束融合到视觉惯性里程计后 端优化中，给出了该约束的定义并求解相关的雅可比矩阵，使得姿态估计问题可 以被非线性优化求解。

第四章为地图定位模块介绍，首先介绍了基于深度学习的粗到细定位方法；此 后介绍了坐标系转换矩阵初始化算法和初始位姿的自适应选择算法，这两部分共 同工作，能够完成紧耦合优化预处理，将 PO-VIO 结果和粗到细定位结果的坐标系 统一； 在这之后是基于最大后验概率估计的紧耦合优化介绍，这一部分详细推导 了如何将先验地图建模为高斯概率地图， 以及如何以非线性优化的形式完成最大 后验概率估计；最后介绍了转换矩阵的更新并论述了其必要性。

第五章为实验与结果分析， 首先对实验实施的环境、参数和数据集等信息进 行了介绍； 此后分别针对本文提出的 3 个模块进行实验，再对整个系统的性能进 行了综合评估，并对实验结果进行了详细的分析，分别从定位精度、定位鲁棒性、 定位实时性等方面进行了评估； 最后进行了消融实验，更加细致得验证了本文提 出的方法的有效性。

第六章为总结与展望，首先总结了本文的研究工作，然后对本文的研究工作 进行了展望，提出了未来的改进方向。

第 2 章 融合 SfM 与 GNSS 的离线建图模块设计

本章主要对离线建图模块进行介绍。在离线建图模块介绍部分展示了该模块 的基本运行过程，详细介绍了各个过程的设计目和实际功能。

2.1 坐标系定义

本课题的设计中考虑了 7 个主要参考坐标系：全局坐标系 (.)g、视觉建图世界 坐标系 (.)w、地图关键帧坐标系 (.)k、PO-VIO 世界坐标系 (.)o、IMU 体坐标系 (.)b、 车身坐标系 (.)u 和定位相机坐标系 (.)c，以上均为笛卡尔系，如图 [2.1](#bookmark168)所示。



cj<



ki 、 l

uj 



R, t



Ri , ti





bj 

Ru

b



Rj , tj





z

w



y

o

s, R, t

g 

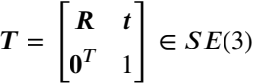
R, t

x

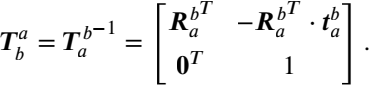
图 2.1 坐标系定义与转换

全局坐标系 (.)g 基于局部切平面的东-北-上 (East-North-Up，ENU) 坐标系。视 觉建图世界坐标系 (.)w 在离线建图阶段初期建立，固定在第一个地图关键帧上。地 图关键帧坐标系 (.)k 和定位相机坐标系 (.)c 附着于相机光学中心，这两个坐标系用 于表示投影误差，而投影误差是优化的主要目标。IMU 体坐标系 (.)b 是 IMU 的读 数坐标系，用来指示加速度、角速度方向。车身坐标系 (.)u 表示车或者机器人 (轮 式) 的理想物理模型质点， 一般设定为车底盘某位置，在本文中由于估计的需求， 将其设定与 IMU 体坐标系 (.)b 原点重合，但是两坐标系之间有所旋转。由于关键 帧坐标系、相机坐标系、 IMU 体坐标系和车身坐标系随时间变化，我们使用下标 i 表示第 i 时刻的地图关键帧坐标系 (.)ki、定位相机坐标系 (.)ci 、IMU 体坐标系 (.)bi 和车身坐标系 (.)ui 。PO-VIO 世界坐标系 (.)o 类似于视觉建图世界坐标系，设定为 里程计的初始状态，但不同的是 PO-VIO 世界坐标系的 z 轴与真实物理世界的重 力方向重合，方向从地心指向受重力物体质心。

旋转和平移分别用矩阵 R ∈ SO(3) 和向量 t ∈ ℝ3 表示。齐次线性变换用矩阵

表示，其中 T 表示从 a 坐标系到 b 坐标系的变换矩阵，其

逆矩阵可以通过以下方式轻松获得

 （2.1）

在第 3 章的 VIO 中还涉及到四元数表示的旋转，四元数 q = [qw , qx , qy, qz]T ，其中 qw 是实部，qx , qy, qz 是虚部， 四元数的模长为 1， 即 qTq = 1。四元数的基本运算 以及四元数与旋转矩阵之间的转换关系可以参考附录 [A.1](#bookmark306)。

2.2 模块整体设计

离线建图模块的整体流程如图 [2.2](#bookmark170)所示：首先通过图片和 GNSS 信息进行预处 理，主要是筛选出地图关键帧以及进行语义分割 (Semantic Segmentation) 等操作； 紧接着纯图像首先进行基于 SfM 的无尺度建图， 以获得在视觉建图世界坐标系下 的相机位姿和地图点云； 然后使用 GNSS 信息和 SfM 结果进行建图对齐， 即实现 视觉建图世界坐标系和全局坐标系的对齐； 此后进行建图融合， 以非线性优化的 方式同时优化视觉重投影误差和 GNSS 提供的位置误差；最后输出分层次的地图。

|  |  |  |
| --- | --- | --- |
|  | 无尺度建图 |  |
|  | | |

|  |
| --- |
| 建图对齐 |
|  |

 建图融合 

|  |
| --- |
|  |
|  |

**图像信息**



**GNSS全局信息**



|  |
| --- |
| 建图预处理 |
| **语义分割**    **关键帧筛选** |

图 2.2 离线建图模块整体流程

值得注意的是，在建图阶段本文并没有使用到惯性信息，这主要由于惯性信 息的使用必须同时满足“时间连续”和“时间间隔合理”两个条件才可以获得较 好的效果，否则，不连续的时间会导致积分中断，过长的时间间隔会导致积分误差 过大，而本文所假设的建图情况并不一定同时满足上述两个条件，故而只使用图 片和 GNSS 信息。

2.3 建图预处理方法

预处理部分的主要目的是提高建图的效率并准备好建图所需数据， 具体工作 是根据图像和 GNSS 信息筛选出地图关键帧并根据建图需求对图片进行一系列处

理，包括语义分割、特征提取和特征匹配。

2.3.1 地图关键帧筛选算法

在建图过程中，关键帧的选择对于建图效率和精度有着重要影响，关键帧的选 择一般遵循以下的原则： (1) 关键帧之间必须有足够的共视区域，这是为了保证有 足够的特征点可以被匹配； (2) 关键帧之间的距离不宜太小，这是为了降低重投影 误差对地图点云精度的影响。如图 [2.3](#bookmark172)所示， 当地图点在相机 2 上偏移了同样的角 度 δθ 时，对地图点的估计误差会因为帧间相对距离的不同而有差异，而当帧间位 移较小的时候，误差明显加大。为了达成原则 (1)，本文使用连续的图像帧作为地 图关键帧，保证了帧之间有足够的共视区域；为了达成原则 (2)，本文使用了 GNSS 提供的位置信息来过滤掉距离过近的地图关键帧，设定过滤阈值为 1.5m， 即如果 两个地图关键帧之间的距离小于 1.5m，则保留其中一个，否则保留两者。

δd

地图点 P

地图点 P



δd ●

δθ

δθ

t

●



t

相机 1 相机 2 相机 1 相机 2

图 2.3 关键帧视差对地图点精度的影响

2.3.2 基于语义分割算法的地图动态元素剔除

在地图关键帧筛选之后，接下来是对图片进行语义分割来剔除地图中的动态 元素。语义分割可以利用图片中的高层语义分析， 协助达成 SfM 的静态环境假设， 即 SfM 所重建出来的地图点云应归属于静态物体。在实际的建图中，静态环境假 设的达成并不容易：例如在城市环境下，建筑物是理想的建图元素，因为其位置相 对固定，短时间内外观变化较小；但是汽车、行人等则不适合作为建图元素， 因为 其位置可能变化，不满足静态假设。因此，本文使用语义分割算法对图片进行处 理，将图片中动态物体剔除，保留静态部分以便建立更加稳定和精确的地图。

表 2.1 语义分割类别分布与屏蔽情况

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别 | road | … | sky | person | rider | car | truck | bus | train | motorcycle | bicycle |
| 标签序号 | 0 | … | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 屏蔽 |  |  |  | √ | √ | √ | √ | √ | √ | √ | √ |

因为语义分割技术只是离线建图工作中的一小部分，所以本文选择了开箱即

用的语义分割模型，并根据建图模块的实际需求进行了微调。本文所选择的语义分 割模型是 SegFormer[[48](#bookmark307)] ，SegFormer 是一个基于 Transformer 的语义分割模型，凭 借其紧凑但高效的网络结构，其在语义分割精度和速度上均取得了较好的效果。本 文使用了 SegFormer 的 B1 [型，即 SegFormer-B1 在 cityscapes[49](#bookmark308)]数据集上的预训练 权重作为初始，在该数据集的灰度图像上进行了微调。cityscapes 数据集是一个城 市道路环境下的语义分割数据集，包含了大量的城市道路图片和对应的标签，标 签中包含了道路、建筑、树木等静态物体的信息， 以及汽车、行人等动态物体。微 调以 1024 × 1024 的分辨率进行了 160,000 轮迭代，使得模型在灰度图上的表现接 近了其在彩色图片上的表现，以便在建图过程中适应建图所使用的灰度图像输入。



图 2.4 语义分割效果

本文的语义分割模型在 cityscapes 数据集上训练，其输出的类别以及每一类别 的标签序号如表 [2.1](#bookmark222)所示。在建图过程中， 本文会将潜在的移动物体类别所包含的 像素位置标记为屏蔽，后续的特征提取和匹配过程将不考虑这些像素位置， 具体 的屏蔽情况也汇报在表 [2.1](#bookmark222) 中。语义分割的实际效果以及屏蔽内容以图 [2.4](#bookmark174) 所示， 左侧为原始图像，右侧为语义分割的结果示意图，其中被白色覆盖的区域为保留 部分，被其他颜色覆盖的区域为屏蔽部分，可以看到被屏蔽的大部分区域均为行 人、自行车等动态物体。

2.4 基于 SfM 的无尺度建图方法

经过建图预处理的图像会使用 SfM 构建无尺度的点云地图。 SfM 是一种摄 影测量技术， 用于从二维图像序列中估计三维结构[[36](#bookmark294)] 。 SfM 以地图关键帧序列 r = {ki ∣ i = 1 ⋯ Nk } 和相机内参矩阵 K 作为输入，输出地图关键帧序列的位姿

集合 丁 = {Ti ∣ i = 1 ⋯ Nk } 以及空间点集 P = {p ∈ ℝ3 ∣ m = 1 ⋯ Np }，称为稀

疏点云结构。需要注意的是，相机内参矩阵数量本应与地图关键帧数量对应，但 是对于固定路线定位来说，建图所用图像均由同一相机拍摄，因此共用一个内参。

SfM 通常以特征检测和匹配为第一步：对于每个地图关键帧 ki， 首先检测出 NFi 个局部特征组成集合 ℱi = {(xj, fj) ∣ j = 1, ⋯ , NFi }，其中 xj ∈ ℝ2 表示特征点

的像素坐标，fj ∈ ℝdl 表示其 dl 维特征描述子。随后，根据关键帧的相邻关系进 行图像匹配，以获得有共视关系的图像对。

在相邻帧之外，为了通过闭环优化减轻漂移的问题，本文在建图时还引入了基 于深度特征的闭环检测方法。通过为每个关键帧提取一个 dg 维的图像描述子 (模 长为 1)Di ∈ ℝdg ，并计算所有地图关键帧之间图像描述子的相似度得分来实现，对 每一帧来说，取得分超过预设阈值 tl 的前 m 个关键帧组成闭环图像对。对于所有 匹配上的地图关键帧对，SfM 通过图像特征点描述子，在已匹配的图像间完成特 [征点匹配。根据图像间的特征点匹配关系， 使用对极几何 (Epipolar geometry)[50](#bookmark309)]可 以粗略估算出两帧图像的相对旋转以及无尺度的相对位移，组成无尺度的相对位 姿变换。

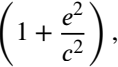
在计算出所有匹配图像间的位姿变换后，SfM 会选择第一个地图关键帧作为 视觉建图世界坐标系的原点， 并通过帧间相对位姿变得叠加换来获得每一帧相对 于第一帧的位姿， 即获得位姿集合 丁。随后，SfM 用三角化方法在视觉建图世界 坐标系中估计稀疏点云结构 P 的粗略空间坐标。位姿估计与三角化过程密切相关： 位姿的不确定性会导致三角化点的可靠性下降，反之亦然。因此，仅通过相对位 姿变换叠加而获得的位姿集合 丁 和稀疏点云结构 P 是不满足精度要求的，为了获 得精确的 丁 和 P，更精确的优化过程必不可少。

BA 是一种高效的位姿和点云结构优化方法，可通过最小化重投影误差来同时 相机位姿和空间点坐标，重投影误差可以表示为：

 （2.2）

其中，p 是图像特征点观测 j 对应的空间点三维坐标，p 是一种损失函数，用于

降低异常值的权重，[一般使用柯西损失函数 (Cauchy/Lorentzian Loss)[51](#bookmark311)]， 其定义 为

 = c2 ln  （2.3）

其中 e 表示误差值，c = 1.0；函数 π(⋅) 将空间点投影到图像空间，定义如下：

π(Ti , p) = [k ⋅ (Ri ⋅ p + ti )/dj]0∶2 , （2.4）

其中，dj 是 p 在相机坐标系中的深度， [⋅]0∶2 表示选择向量的前两个元素。需要

注意的是，Ei 表示单张图像的重投影误差， 因此整个 BA 是对输入图像序列中的 所有重投影误差进行累积优化的结果。BA 的进行需要使用非线性优化方法，例如

高斯-牛顿法或者 Levenberg-Marquardt 法[[52](#bookmark312)]，这需要误差函数（[2.2](#bookmark310)）关于参数 Ti

和 p 的雅可比矩阵，优化过程及雅可比矩阵的推导参考附录 [A.2](#bookmark313)。

2.5 基于三维相似变换的建图对齐方法

SfM 可以为离线建图阶段的图像提供精化的位姿和稀疏点云地图，但这些位 姿和结构并不是在全局坐标系下的， 并且缺乏真实世界的物理尺度， 因此不适用 于实际的定位任务。为了解决这一问题， 本文利用图像的对应 GNSS 观测数据， 将 SfM 的输出与真实世界的物理尺度对齐。标准 GNSS 报文仅提供地理坐标系 中的纬度、经度和海拔信息，地理坐标系是一种墨卡托投影 (Mercator Projection) 坐标系，而地图关键帧位姿和点云结构都是在笛卡尔坐标系中。因此， 需要首先 将原始 GNSS 报文信息从地理坐标系统转换为笛卡尔 ENU 坐标系中的位置序列 L = {i ∈ R3 l i = 1 … Nk}。这一过程遵循 Subirana et al. [[53](#bookmark314)] 提出的转换方法，并

设置  以确保数值稳定性。

由于建图对齐过程不仅需要解决视觉建图世界坐标系与全局坐标系之间的变 换问题， 还需要处理两者之间的尺度差异， 本课题采用三维相似变换 (Similarity

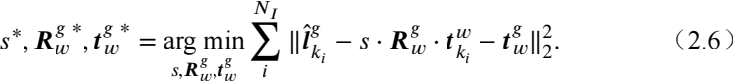
Transformation in 3D, Sim3)[[54](#bookmark315)] 。 Sim3 包括一个比例因子 s、 一个旋转矩阵 R 和

一个平移向量 t。如果第 i 个地图关键帧在视觉建图世界坐标系中的位置为 ti =

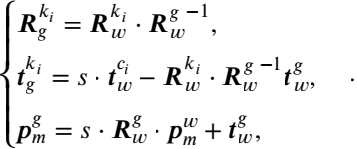
-RiT . ti ，对应的 GNSS 观测值为 i ，则 Sim3 变换可以将它们对齐，具体如下：

i = s . R （2.5）

Sim3 参数的最小二乘解可以通过求解以下优化问题获得：



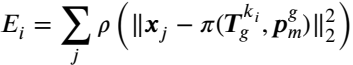
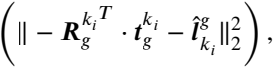
利用 Sim3 参数，可以方便地将地图关键帧位姿 丁 和稀疏点云结构 P 转换到 具有真实物理尺度的全局坐标系中：

 （2.7）

2.6 基于 GNSS 增强的建图融合方法

尽管已经获得了较为精确的地团关键帧位姿和稀疏点云结构， 并成功将它们 转换到全局坐标系中，但在求解 Sim3 参数的最小二乘解时仍可能引入误差。为了 获得更精确的位姿和点云结构，还需要将当前的全局坐标系参数与 GNSS 观测进 行融合优化。该融合过程需要确保地图关键帧位姿和点云结构同时满足 GNSS 和 视觉的观测， 因此需要在视觉 BA 的基础上融合进 GNSS 的观测约束，将全局坐

标系中的重投影误差重新表示为：

 + α . p  （2.8）

其中，α 是一个权重系数，用于平衡视觉重投影误差与位置偏差。然后通过 BA 过 程对式（[2.8](#bookmark316)）中的误差进行最小化。

至此， 已经介绍了地图关键帧位姿和稀疏点云结构的优化，这些空间信息有 效地描述了环境结构。然而，为了在定位过程中充分利用这些信息，还需要视觉 地图关键帧描述子和稀疏点云的描述子，用于地图与实时捕获图像之间的匹配。

在离线建图过程中，直接使用 {Fi l i = 1, … , Nk } 作为地图点的描述子，{Di l i = 1, … , Nk } 作为关键帧描述符。离线建图模块的最终输出可以分为两个层级： 粗粒度层级和细粒度层级。粗粒度层级输出包括关键帧位姿和描述符，用于建立 图像匹配关系；而细粒度层级输出包括点描述符和点结构，用于实现点的匹配。

2.7 本章总结

本章主要介绍了固定路线定位系统中的离线建图模块设计。离线建图模块以 融合 GNSS 的 SfM 为基础， 同时根据建图需求引入了基于深度学习的语义分割、 闭环检测以及特征点提取技术，将更多的人类经验加入到离线建图中。离线建图 模块的另一个特点是在传统 SfM 的基础上引入了GNSS 观测作为约束和融合信息， 具体来说此模块首先使用 SfM 技术构建无尺度的点云地图，然后通过三维相似变 换将地图与全局坐标系对齐，最后通过 GNSS 观测信息融合优化地图关键帧位姿 和点云结构，显著提升建图的精度。

第 3 章 基于车身运动模式感知的 PO-VIO 设计

本章主要对 PO-VIO 模块进行介绍。本文中的 PO-VIO 不同于通用的里程计， 其专门针对车辆或者轮式机器人进行了优化，针对这类平台的运动模式进行了假 设与分类， 并根据分类情况分别提出了相应的速度伪观测约束条件。在运动模式 的检测方面，本章介绍了一种基于深度学习的方法。在伪观测约束的使用方面， 本 章设计了粗对齐过程并在原有的后端优化中加入了惯性-车体因子联合优化。

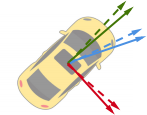
3.1 PO-VIO 整体设计

PO-VIO 是视觉惯性定位的关键组成部分，其工作是为定位提供较为精确的帧 间运动关系约束，这对于最终的定位精度有着不可忽视的作用。通用的 VIO 常用于 无人机、自动驾驶车辆和机器人等领域，其一般假设机体做任意运动，因此以构建惯 性和视觉观测之间的约束关系为主， 并不会根据机体的运动模式做额外设计。而 本章所设计的伪观测视觉惯性里程计 (Pseudo Obsevation Visual-Inertial Odometry, PO- VIO) 则是在通用 VIO 的基础上融合了伪观测约束的改进版本。

在固定路线的车辆或轮式机器人定位中，机体的运动模式往往有规律可循，并 由此引申出部分伪观测约束： 本文考虑的汽车和轮式机器人，其运动往往在平面 上进行， 因此其速度一般可以分解为沿着车辆前进方向的速度和沿着车身横向的 速度； 以车身坐标系为参考，车辆在某些运动模式下不具有沿着车身横向以及重 力垂直方向的速度。换言之， 车辆或者轮式机器人，在某些运动模式下，其横向和 重力方向的速度可以取得 0 值，这种逻辑推导而非实际测量获得的观测即为本文 所述的伪观测。

y

z

uj

j x

b



直线行驶速度方向 转弯速度方向 (黄色)

图 3.1 车身坐标系和 IMU 体坐标系分解示意

如图 [3.1](#bookmark176)所示， 实线代表 IMU 体坐标系 (.)bj ，虚线代表车身坐标系 (.)uj ，车身 坐标系的 y 轴沿着车辆前进方向，x 轴沿着车辆横向方向，z 轴垂直于地面向上。 IMU 体坐标系即 IMU 的测量坐标系，其一般情况下与车身坐标系会存在一定的旋

转关系， 并不一定如图所示的共向。在当前的坐标系下，假设车辆或机器人行驶 在平面 (不考虑上下坡导致的 z 轴速度)，其运动模式可能存在 3 种情况：

(1) 车辆静止，此运动模式下车身坐标系的 y 与 x 轴均无速度；

(2) 直线行驶，此运动模式下车身坐标系的 y 轴有速度，x 轴无速度；

(3) 车辆转弯，此运动模式下车身坐标系的 y 与 x 轴均有速度。

据此分析，可以在车辆或机器人静止和直线行驶时，基于情况 (1) 和情况 (2) 为车身增加两项伪观测，而这两项伪观测的约束条件可以提升通用 VIO 的精度。

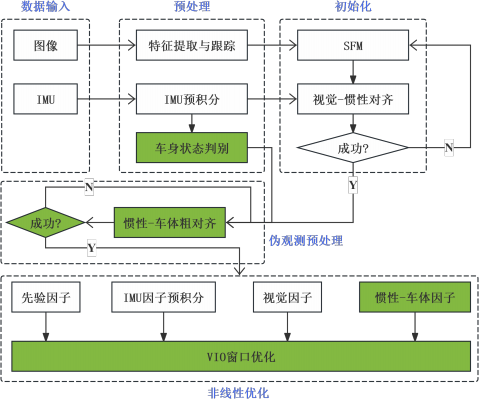


图 3.2 视觉惯性里程计框架

然而，在仅依靠 IMU 信息和视觉信息的情况下，如何区分车辆的运动模式仍 是一个有待解决的问题。此外，这两项约束如何添加到通用 VIO 的优化中，也需 要进一步的研究。为解决上述问题，本章着重介绍如图 [3.2](#bookmark178)所示的 PO-VIO： 其在 通用 VIO 的基础上加入了车身运动模式判别、惯性-车体粗对齐等过程，并在窗口 优化中加入了惯性-车体因子，这些改进 (以绿色标注) 保证了本章对车身运动模式 的分析以及由此引申的伪观测约束能够的合理工作。

3.2 基于神经网络的车身运动模式判别方法

根据对车身运动模式的分析， 只有在车辆和机器人停止或直线行驶时，才能 添加伪观测约束。因此， 需要在 PO-VIO 中设计车身运动模式判别模块，以区分车 辆的运动状态。

以往的工作中，Hu et al. [[55](#bookmark317)]使用传统的统计方法来判别车辆的直行、变道和转 弯等状态；Brossard et al.[[56](#bookmark318)]使用卷积神经网络 (Convolutional Neural Network, CNN) 来生成车身 3 个轴向的测量噪声， 以此来间接判别车辆状态；Huang et al. [[57](#bookmark319)]则使 用时间卷积神经网络 (Temporal Convolutional Neural Network, TCN) 根据 IMU 观测 值来接判别车辆的运动状态。

然而，基于统计的判别方法需要事先采集 IMU 静止状态下的数据进行统计分 析，不够灵活；基于 CNN 和 TCN 的方法针对每一条 IMU 数据进行判别，这在应 用于高频率 IMU(例如 2000Hz 及以上) 时会造成极大的时延。针对以上问题， 本章 希望使用一种更加灵活，并且有望实时进行的方法来进行车身状态判别。

近年来， 因为深度学习技术在泛化性能上展现出的优势，本章希望使用基于 深度学习的方法来进行车身状态判别。在输入信息的选择上，为了使得深度神经 网络能够与 IMU 的处理频率相匹配，就不得不改变以往逐条处理 IMU 数据的模 式，转而寻找一种批量处理 IMU 数据的模式。

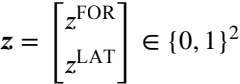
在传统的 VIO 中，IMU 数据的处理采用预积分的形式，这是一种将 IMU 数 据的采集频率与图像数据的采集频率进行同步的处理方式。具体而言，在预积分 的 IMU 处理形式下，两张相邻时间图像之间的 IMU 数据会被整合为一个预积分 因子，这样就可以将 IMU 数据的处理频率降低到与图像数据相同的频率，这个频 率一般在 10Hz 到 30Hz。在这个频率下，基于深度神经网络的方式是有望实时进 行的，因此预积分因子内包含的 IMU 数据序列是一个理想的网络输入。

然而，如果选择预积分因子内的 IMU 序列作为网络输入，那么就会面临一个 问题：因为 IMU 频率和相机频率不一定是整除关系，所以两帧图像之间的 IMU 序 列长度并不固定。而如果使用 CNN 或者 TCN 等必须输入规整数据的网络结构，就 难以处理 IMU 序列。近年来， 基于 Transformer 结构的神经网络在自然语言处理领 域展现出了优异的性能，其注意力机制使得其能够处理任意长度的序列数据，所 以也十分适合处理所有序列化的数据。

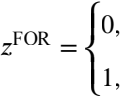
基于以上优势考虑，本章选择以 Transformer 结构为基础，设计一个用于车身 运动模式判别的深度神经网络。

3.2.1 网络输出定义

对于网络的输出定义，本文选择对车身的 y 轴 (车身前向) 速度和 x 轴 (车身横 向) 速度状态进行判别。用向量

 （3.1）

来表示车身是否处于静止的运动模式和车身是否处于直线行驶的运动模式。其中 ZFOR 指示车身是否处于静止的运动模式，其取值为



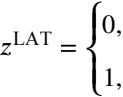
车辆不处于静止的运动模式

.

车辆处于静止的运动模式

（3.2）

ZFOR 表示车身横向速度为 0 假设是否成立，其分布为



车辆不处于直线行驶的运动模式

.

车辆处于直线行驶的运动模式

（3.3）

3.2.2 网络输入定义

对于网络输入，本章使用一个预积分因子内的 IMU 数据序列作为网络输入。 单条 IMU 采样数据包括加速度计读数 b ∈ ℝ3 和陀螺仪读数 b ∈ ℝ3， 两个读数 的可以分解为：

b = ab + R ⋅ g + ba + na

b = wb + bw + nw

（3.4）

其中 ab , wb 分别表示 IMU 此时真正的加速度和角速度，bw , ba 分别表示陀螺仪和

加速度计的零偏，nw , na 分别表示陀螺仪和加速度计的噪声，R 表示此时全局坐

标系到 IMU 体坐标系的旋转矩阵，g 表示地球重力加速度。

单条读数可以反映某一时刻的 IMU 运动情况，因此如果结合采样时间对 IMU 序列进行积分， 就可以获得一段时间内的 IMU 运动情况， 即旋转和位移。因此， 使用 IMU 序列数据，理论上足以预测车身状态。实践中， 本文将单条加速度计和 单条陀螺仪读数拼接为一个 6 维向量，作为一条完整的 IMU 采样数据。在此之后， 将预积分因子内的多条 IMU 采样数据拼接为一个矩阵 X ∈ ℝn×6 用以表示 IMU 数 据序列，其中 ℝn×6 表示 X 由 n 条 IMU 读数组成。

3.2.3 网络结构设计

对于网络结构，本章希望其能够输入一个可变形状矩阵 X， 稳定输出向量 z。 [因此，本章使用了经典的 Encoder-Decoder 架构 Transformer[47](#bookmark305)]： 其网络的基本参

表 3.1 车身判别网络设计参数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 层 | 层类型 | 输入维度 | 特征维度 | 前馈层维度 | 头数 | 输出维度 |
| 1 | 全连接层 | 6 | 128 | N/A | N/A | 128 |
| 2 | Transformer Encoder | 128 | 128 | 1024 | 8 | 128 |
| 3 | Transformer Encoder | 128 | 128 | 1024 | 8 | 128 |
| 4 | Transformer Decoder | 128 | 128 | 1024 | 8 | 128 |
| 5 | Transformer Decoder | 128 | 128 | 1024 | 8 | 128 |
| 6 | 全连接层 | 128 | 2 | N/A | N/A | 2 |

数如表 [3.1](#bookmark224)所示。除了表中所示的神经网络层以外， 本章网络中还有一个可训练参 数 q ∈ ℝ128 ，其用于和两层 Transformer Decoder 进行交互，并转化为带有车身运 动模式特征的向量 q ′ ∈ ℝ128，此后 q ′ 经过全连接层转化为预测向量 y ∈ ℝ2，y 经 过 Sigmoid 函数激活后得到预测的车身运动模式概率的向量 y ∈ (0, 1)2 ，此后以 τ 为阈值，将 y > τ 作为积极预测，反之为负面预测，因为本文方法对车身状态判断 的查准率要求较高，因此这个阈值将取一个较大的值。

3.2.4 网络训练方法

对于网络训练，需要精心设计的过程主要包括数据清洗、预测真值生成、损失 函数设计和训练策略。

在数据清洗方面， 由于 IMU 采样数据包括加速度和角速度，这两类数据的数 值分布有较大差异： 角速度以 rad/s 为单位， 一般是一个较小的数字，加速度以 m/s2 为单位，其取值范围一般较大。因此如果使用原始数据作为网络输入，往往 会造成网络训练的不稳定， 因此本文使用归一化策略，将所有的 IMU 采样数据元 素调整为值域在 (−1, 1) 之间：首先根据训练数据的分布特征计算 IMU 采样数据的 均值 μ ∈ ℝ6 和逐元素标准差 σ ∈ ℝ6，然后对每一条 IMU 数据根据

 （3.5） 进行转换。x 是一条包涵加速度和角速度的 IMU 数据， 除法是逐元素除法，可以 在相同形状的的数据上进行操作。

在预测真值生成方面，本文需要使用到车辆的真实运动状态。具体来说， 首先 根据车辆的真实运动状态的更新频率切分 IMU 序列，然后根据车身运动状态自动 生成真值 z。根据对预设状态的分类，首先根据相对位置信息判断车辆是否静止， 然后根据相对旋转判断车辆是否直线行驶，具体如算法 [3.1](#bookmark320) 所示，其中第 16 行即 罗德里格斯旋转公式 (Rodrigues’ rotation formula)[[58](#bookmark321)]将旋转矩阵转换为旋转向量，

第 3 章 基于车身运动模式感知的 PO-VIO 设计

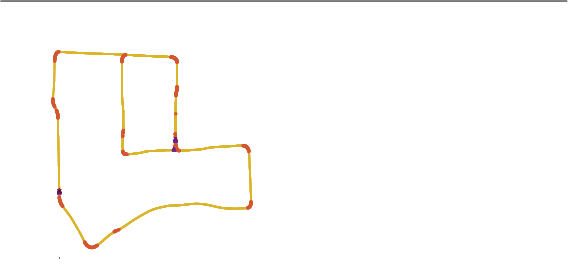
|  |
| --- |
|  |
| **算法** **3.1** Generate training data and ground truth |
| **输入：**IMU 时间戳序列 ts, IMU 采样数据序列 xs, 车辆运动状态时间戳序列 Us, 车辆位置状态 序列 ps, 车辆旋转状态序列 Rs  **输出：**IMU 数据序列列表 Xs, 真值列表 zs 1: Us.insert(0, 0.0); Xs ← []; Zs ← []; i ← 1  2: **while** i < len(Us) **do**  3: j ← i − 1; X ← []; z ← [] 4: **while** ts[0] <= Us[j] **do**  5: ts.pop(0), xs.pop(0) 6: **end** **while**  7: **while** ts[0] <= Us[i] **do** 8: X.append(xs.pop(0))  9: **end** **while**  10: Xs.append(X)  11: **if** ps[i] − ps[j] < ep **then** 12: z.append(1)  13: **else**  14: z.append(0) 15: **end** **if**  16: U ← Rs[j]T ⋅ Rs[i] 17: **if** norm(U)< er **then** 18: z.append(1)  19: **else**  20: z.append(0) 21: **end** **if**  22: zs.append(z) 23: **end** **while** |

其具体过程详见附录 [A.3](#bookmark322) ，norm(⋅) 表示对向量取模操作。

在损失函数和训练策略方面，由于模型的最终输出会使用 Sigmoid 函数进行概 率预测，相对应的损失函数选择二值交叉熵损失函数 (Binary Cross Entropy, BCE)。 [在训练策略方面，本文使用 Adam 优化器[59](#bookmark323)]， 学习率为 1e − 4，训练批次大小为 8，训练轮数为 100。此外， 由于一般情况下车辆直行的情况比静止和转弯的情况 要多， 如图 [3.3](#bookmark180) 所示， 因此如果直接使用全部数据训练， 则会导致样本分布的不 均衡。为了解决这个问题，本文在数据采样方面使用了均衡策略： 在训练的每个 epoch 中，以转弯和静止区间的并集数据量为基准，从直行数据中随机采样等量数 据，然后与转弯和静止区间的并集数据合并。

3.3 基于车身运动学假设的伪观测预处理方法

伪观测可以为 VIO 增加一定的约束条件，但是其并不能直接使用，而是需要 一定的预处理过程，其中最关键的就是惯性-车体粗对齐。如图[3.1](#bookmark176) 中的坐标系示 意图所示， 车身坐标系和 IMU 体坐标系往往并不重合， 有时甚至有较大的差距， 而 VIO 中一般以 IMU 体坐标系作为姿态估计的基准坐标系。因此， 如果使用车

第 3 章 基于车身运动模式感知的 PO-VIO 设计

Y (m)

|  |
| --- |
| · 直行  转弯  静止  起点 |

250 

200 

150 

100 

50 

0 

50 

100 

0 100 200 300 400

X方向 (m)



|  |  |  |
| --- | --- | --- |
| 直行  直行 | 转弯  转弯 | 静止  静止 |

|  |
| --- |
| 82.1%  81.0%  80.2%  .  .  76.1%  .  71.9%  .  67.2%  .      32.6%  ★  28.1%  21.6%  19.4%  15.7%    10.1% 7.7%  3.3%  2.3%  0.1% |

场景1 场景· 2 场景3 场景· 4 场景5 场景· 6

场景



- 40 

2000

4000

5000

3000

1000



0.0%

0.5%

 80

 60

 20



0

- 0

.



图 3.3 车身状态分布情况

身坐标系的速度作为约束，就需要将车身坐标系的速度转换为 IMU 体坐标系的速 度。这就需要进行惯性-车体粗对齐，即将车身坐标系的速度转换为 IMU 体坐标系 的速度。一般来说，这个转换关系需要包含旋转和平移两种变换，但是考虑到车身 整体的刚性结构，IMU 与车身之间也是刚性连接， 因此可以将车身坐标系与 IMU 体坐标系的原点等价为重合状态，在这种状态下可以只考虑车身坐标系与 IMU 坐 标系之间的旋转关系。

在拥有状态判别， 并且只考虑两坐标系之间旋转关系的情况下，求解车身坐 标系和 IMU 体坐标系之间的关系还可以使用一些更强的先验知识：

1. 在车辆直行的情况下，速度方向基本指向车身的前向，即车身坐标系的 y 轴；

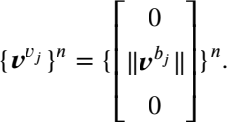
2. 在车辆直行的情况下，速度大小基本为合速度的模长。

基于以上假设，可以使用以下方法求解车身坐标系和 IMU 体坐标之间的转换关系。

首先根据车身状态判别的结果，筛选直行时刻的 IMU 速度估计 {vj }n ，其中 vj 表示第 j 时刻的 IMU 速度估计，n 表示 IMU 速度估计的总数，然后使用 VIO 估计的车身姿态 {Rj }n 可以获得 IMU 体坐标系下的速度估计

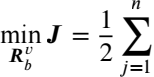
{vbj }n = {Rj −1 ⋅ vj }n . （3.6）

又根据直线行驶时的车身假设可以获得此时车身坐标系下的速度为

 （3.7）

此时，关于车体和 IMU 体坐标系之间的旋转关系 R 可以求解以下带有约束的优

化问题获得：

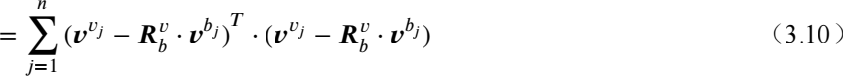
 Ⅱvuj - R . vbj Ⅱ （3.8）

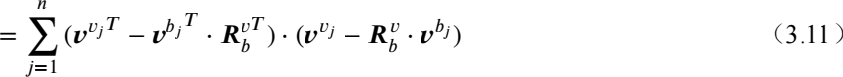
s.t. (R)TR = I

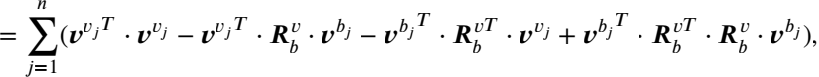
其中的约束条件是为了保证所求旋转矩阵是一个合法的旋转矩阵。针对上述优化 问题，其求解过程如下：

n



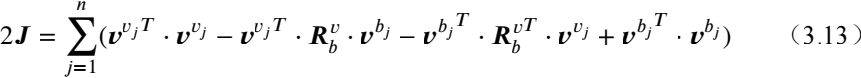


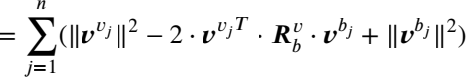


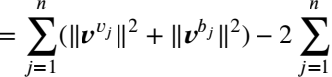


（3.12）

根据约束条件 (R)TR = I 可得



 （3.14）

 vujT . R . vbj , （3.15）

可以观察到此时 J 的最小化与前一常数项无关，仅与后项带 R 的项有关，因此可

以将 J 的最小化问题转化为最大化后一项，即

n

 vujT . R . vbj . （3.16）

注意到

 vujT . R . vbj = trace （3.17）

= trace （3.18） 其中 Vb, VuT 分别是 {vbj }n , {vuj }n 序列的协方差矩阵。对 Vb . VuT 进行奇异值分

解 (Singular Value Decomposition, SVD) 可以获得

Vb . VuT = U . Σ . VT , （3.19）

因此可以对式（[3.18](#bookmark325)）进行变换得到

trace = trace （3.20）

= trace （3.21）

其中 VT . R . U 为三个正交矩阵的乘积，因此仍是一个正交矩阵，而 Σ 为对角矩

阵。根据正交矩阵的性质可知其内部所有元素均小于 1(因为正交矩阵的行列式为

1)，所以对于 trace(Σ . VT . R . U) 来说，其取得最大值时即 VT . R . U 为单位矩

阵。结合式（[3.16](#bookmark324)），式（ [3.18](#bookmark325)），式（ [3.19](#bookmark326)）可知，最大化 J / 时即

R = V . UT . （3.22）

此时获得的 R 可能会因为存在噪声等原因而并不精确，但此过程仅为粗对齐，所

以精度误差有一定的容忍，后续在非线性优化部分仍会对此参数进行持续优化。

3.4 融合伪观测约束的非线性优化方法

经过车身状态判别和惯性-车体粗对齐之后，还需要将伪观测相关的两个因子 加入到 VIO 的窗口优化中，才能发挥伪观测约束对于整个系统的提升作用。由于 伪观测约束是在原有 VIO 的优化项上添加， 因此描述伪观测约束的相关因子，必 须了解 VIO 的状态估计量及相关约束。

3.4.1 非线性优化的状态量与目标函数

本文的 VIO 的估计量可以表示为



x k = [pk , vk , qk , ba , bw ] , （3.23）

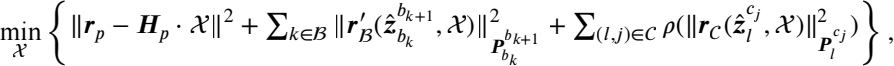
x = [q, q]

其中 x k 表示第 k 帧时刻的 IMU 姿态，x 表示相机到 IMU 的外参，q 表示 IMU

到车身的外参，λk 表示第 k 帧相机的特征点深度。q 是惯性-车体因子，其表示

IMU 体坐标系到车身坐标系的旋转关系。

根据以上的状态估计量，可以得到 VIO 的优化目标函数



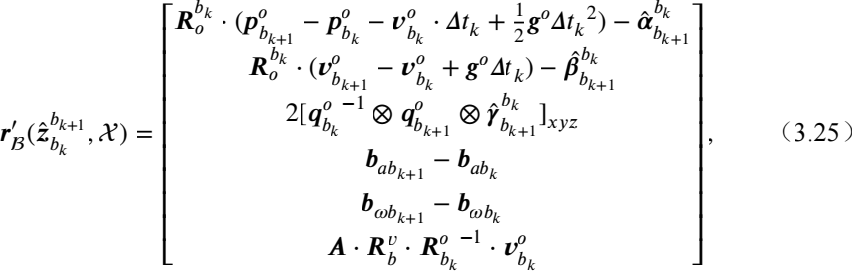
（3.24）

其中 rp , Hp 代表与边缘化相关的观测与矩阵，rc 代表视觉观测的残差函数，r 是

本文提出的、增加了伪观测约束的 IMU 残差函数，plcj 代表视觉观测的协方差矩

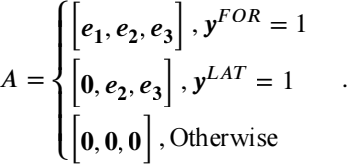
阵，p 代表伪观测约束的协方差矩阵，ℬ 表示存在 IMU 残差的时刻集合，c 则

表示地图点和视觉观测的匹配集合。r 具体可以表示为



其中 Δtk 表示时间间隔，go 表示 VIO 世界坐标系下的重力向量，+1 , β̂+1 , +1

分别表示相对位移、相对速度和相对旋转的预积分观测值， 其具体表示参考 Qin et al. [[25](#bookmark283)]的工作。IMU 残差函数的最后一项是相较于经典 VIO 而增加的速度伪观 测约束。A 是一个指示函数，其取值根据车身状态判别的结果取值

 （3.26）

其中 ei 表示单位矩阵的第 i 列向量，YFOR , YLAT 分别表示预测的车辆前进和横向 的状态。

3.4.2 优化残差的雅可比矩阵计算

若使用伪观测约束， 需要使用其相对于各估计量的雅可比矩阵。为了简化表

达，本处使用 r 代表 IMU 残差函数中的伪观测约束， 即式（[3.25](#bookmark327)） 的最后一项。

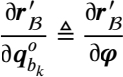
首先给出速度伪观测约束相对于状态量 vk 的雅可比矩阵

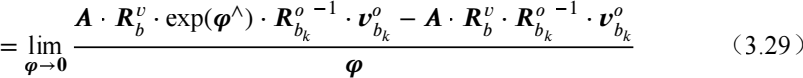
 = A ⋅ Rk −1 . （3.27）

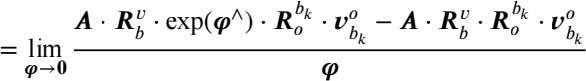
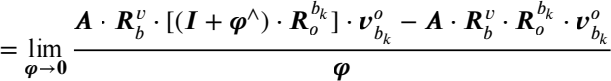
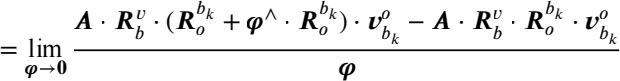
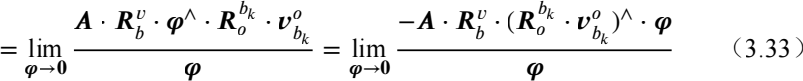
对于 qk 的雅可比矩阵，约束中并没有直接使用四元数 qk ，而是使用等价的旋转矩

阵 Rk ，避免了四元数运算 ⊗ 的繁琐使用。求导方式本文选择使用扰动模型[[60](#bookmark329)]进

行，假设旋转的扰动值为 φ ∈ ℝ3(对于四元数和旋转矩阵有同样的效果)，则：

 （3.28）



 （3.30）  （3.31）  （3.32） 

= -A . R （3.34）

其中 Λ 表示将一个向量转换为反对称矩阵的函数，在推导中直接使用了近似

exp(φΛ ) ≈ I + φΛ . （3.35）

和反对称矩阵的乘法交换规律

aΛ . b = -bΛ . a. （3.36）

同理可得伪观测约束关于惯性-车体因子的雅可比矩阵：

 = -A .  （3.37）

至此已经给出了速度伪观测对于估计量的雅可比矩阵。根据雅可比矩阵，使用高 斯牛顿法或者列文伯格-马夸尔特法进行非线性优化，以获得最终的状态估计量。

3.5 本章总结

本章主要介绍了固定路线定位系统中的 VIO 设计。整体设计考虑了车辆或轮 式机器人的特殊运动状态，首先借助深度学习技术对车辆的运动状态进行判别，这 一部分包括了网络设计和详细的训练过程； 然后根据判别结果进行伪观测预处理， 这一部分主要是使用 VIO 本身估计的速度和假设的车身状态进行粗对齐； 最后将 伪观测约束和惯性-车体因子加入到 VIO 的非线性优化中，以提高定位的精度。

第 4 章 基于最大后验概率估计的地图定位模块设计

本章主要对紧耦合地图定位模块进行介绍。紧耦合定位的第一个关键过程是 粗到细定位，本章中对选择这一过程的原因及如何实施进行了详细的介绍。此外， 由于视觉惯性里程计和粗到细定位的坐标系不一致，无法直接使用， 因此本章引 入转换矩阵初始化和有效性检验两个步骤进行对齐。在此之后， 紧耦合优化是定 位的另一个关键过程，本章对这一过程进行了理论说明和公式推导， 并以非线性 优化技术完成实际优化。最后为了解决视觉惯性里程计的漂移问题，本章提出了 转换矩阵更新过程，并对其进行了详细的介绍。

4.1 模块整体设计

紧耦合地图定位是整个定位系统的核心部分，其主要使用离线建图模块的多 层次先验地图和视觉惯性里程计的相对位姿估计结果来计算相机在全局坐标系中 的精确位姿，整体设计如图 [4.1](#bookmark182)所示。

PO-VIO

|  |  |
| --- | --- |
| **图像信息** | |
| **惯性信息** | |
| Gyro Z  Gyro Y  Gyro X | Acc Z    Acc Y  Acc X |

紧耦合优化

|  |
| --- |
|  |
| **最** **大** **后** **验** **概** **率** **位** **姿** **图** **求** **解** |

转换矩阵更新

外点剔除

旋转矩阵平均

位移向量平均

|  |
| --- |
|  |
|  |
|  |
|  |

优化预处理

|  |
| --- |
|  |
| **转** **换** **初** **始** **化** **自** **适** **应** **选** **择** |

|  |
| --- |
| 粗到细定位 |
| **粗** **到** **细** **定** **位** **概** **率** **投** **票** |

先验地图

图 4.1 粗到细匹配过程

整个模块的第一步是粗到细定位，其主要根据当前的图像观测和地图信息进 行对比以及关联，从而获得相机在全局坐标系中的定位观测。粗到细匹配是粗到 细定位的关键，其含义是首先根据粗粒度的图像级特征匹配结果和先验位置信息 来获得图像的粗略位置范围，然后根据细粒度的像素级特征匹配结果来计算更精 确的位姿结果。

除了粗到细定位，视觉惯性里程计同样可以获得位置观测结果，但是这两种 定位方式的结果却因为分属全局坐标系 (.)g 和 VIO 世界坐标系 (.)o 而无法直接对 齐。因此需要一个转换矩阵初始化过程来获得两个坐标系的对齐参数， 即转换矩 阵 Tog 。此外，由于转换矩阵对于整个定位过程有着重要作用，所以系统中对其成 功有着严格要求，即只有当粗到细定位成功时，才会进入后续阶段。

在获得了两个坐标系的转换矩阵之后，理论上可以直接使用粗到细定位结果 和转换后的视觉惯性里程计结果进行融合，但是考虑到粗到细定位结果可能会因 为误匹配等原因而存在较大误差， 因此需要使用有效性验证过程来判断粗到细定 位结果的可靠性。如果粗到细定位的结果通过了检验， 则优先选择其作为后续优 化的初始值； 否则， 则需要使用经过转换的视觉惯性里程计结果作为初始位姿来 进行后续优化。

为了获得较为精细的相机位姿，还需要进行紧耦合优化。这一步主要使用非 线性优化的方法，将前面步骤中生成的粗到细定位结果、视觉惯性里程计输出以 及先验地图的点云结构， 以非线性优化的形式进行融合。在融合过程中考虑到地 图本身所携带的误差，将先验地图的地图点视作服从高斯分布的概率点云，通过 最大后验概率来求解相机位姿。

经过上述步骤后已经输出相机位姿，但是由于视觉惯性里程计自身在运行期 间会产生漂移的累积， 因此在经过长时间的运行后，视觉惯性里程计的世界坐标 系已经产生了变化。如果此时依旧沿用之前的转换矩阵， 则会导致视觉惯性里程 计的误差传导到整个定位系统，因此转换矩阵的更新是有必要的。

4.2 粗到细视觉定位方法

4.2.1 基于神经网络的粗到细特征匹配算法

粗到细特征匹配是粗到细定位的核心步骤，粗到细匹配基于粗粒度和细粒度 [特征描述子， 即图像描述符和点描述符。以往手工设计的特征描述子[61](#bookmark332)-[62](#bookmark333)]和词袋 模型是图像描述子和特征点描述子的常见选择，在这种形式下，所有关键帧的特 征点描述子被聚类， 聚类中心被汇总成一个词袋，新图像的特征点首先被分配到 其最近的聚类中心描述符中， 并统计词袋中每个中心的描述子的频率，形成一个 词袋向量，作为图像描述子。但是这种方法存在以下三个主要缺点：

(1) 手工设计的描述子主要关注图像的局部特征，当由于季节变化或昼夜循环 导致地图场景外观显著变化时，可能会失效.

(2) 过度依赖特征点描述子匹配可能导致错误匹配，例如单张图像的特征点被 关联到多个场景，或多个场景中的点被匹配到同一张图像，从而降低匹配准确性.

(3) 匹配过程需要比较新图像的所有描述符，效率较低。

近年来，深度神经网络在提取视觉特征方面展现出了巨大的潜力， 一些基于 [深度学习的特征提取和匹配方法[41](#bookmark299)[-42](#bookmark300),[63](#bookmark334)] 被引入到视觉定位，这些方法即使在外观 发生显著变化的环境中也表现出优越的性能。因此，针对第一个问题本文采用基 [于深度学习的特征提取器 SuperPoint [42](#bookmark300)]检测局部特征， 即像素级特征及其描述子；

针对第二个和第三个问题，本文使用由 [NetVLAD[41](#bookmark299)]生成的图像级深度特征描述 子，此方法有效减轻了局部点匹配错误对图像描述子的负面影响， 并消除了遍历 新图像局部描述子的需求。

**粗粒度匹配**

**细粒度匹配**

**图像** **描述子**

**特征点** **描述子**

**地图关键帧** **待定位图像**





|  |
| --- |
|  |

**特征** **提取器**



图 4.2 粗到细匹配过程

如图 [4.2](#bookmark184) 所示，粗到细匹配过程首先从图像中提取图像描述子和特征点描述 子。本文中的图像描述子使用 NetVLAD[[41](#bookmark299)]获取， 这是一种基于 CNN 的图片特征 提取网络，常被用来进行大规模的图像检索，证明其强大的特征判别能力，适合进 行本文中的粗粒度匹配。本文中的特征点描述子使用 Superpoint [[42](#bookmark300)]获取，这是一 种能够同时进行特征点提取和特征点描述子编码的神经网络， 由于其经过了大量 真实图片的训练，因此在真实场景下的表现依旧突出，适合本文的固定路线环境。

在提取了粗粒度和细粒度的特征描述子之后，还需要利用这两种粒度的描述 子进行匹配。首先，粗匹配除了需要使用图像级别的粗粒度特征描述子以外，还 需要使用一些位置先验信息。因为一段固定路线上可能会大量的关键帧，如果使 用图像描述子对每一个关键帧都进行比对，会导致匹配效率较低，所以本文使用 位置先验信息，将一部分距离过远的地图关键帧排除在匹配之外。这个位置先验 信息可以通过以往的定位结果获得，如果处于定位开始阶段，没有可用的定位结 果作为先验，那么可以使用全局坐标系下的原点为先验位置信息，这仍符合本文 所探讨的国定路线定位场景。为了使粗粒度匹配能够为后续步骤提供充足的信息， 匹配结果往往并不会只包含一个地图关键帧，本文选择一次匹配 Nc 个候选地图关 键帧给后续步骤判断、使用。

细粒度匹配相较于粗粒度匹配过程更加直接， 只需要将粗匹配的地图关键帧 中的特征点及其描述子与当前图像的特征点及其描述子直接匹配即可。由于提取 的特征点及其描述子都是基于深度神经网络的， 因此匹配过程也选择基于深度神 经网络的 SuperGlue[[63](#bookmark334)]方法，其输入为特征点坐标和描述子，输出为它们之间的 匹配关系。一旦新图像的特征点与地图关键帧的特征点匹配完成，细匹配即结束。

4.2.2 基于概率投票的鲁棒定位算法

粗到细匹配过程在图像与地图关键帧之间建立了联系，这方便了将相机位姿 的计算与全局坐标系对齐。在离线建图阶段，地图关键帧中的点已经与全局坐标 系对齐，而新图像匹配的地图关键帧特征点也可以被赋予其对应的空间坐标，令

这些点的空间坐标表示为 {p l m ∈ M}，其在新图像中对应的匹配点像素坐标观

测为 {m l m ∈ M}，其中 M 表示匹配关系。则新的位姿 T 可通过求解以下投影

优化问题获得：

 \* = arg  （4.1）

若 干 现 有 方 法[[64](#bookmark335)-[65](#bookmark336)] 可 用 于 求 解 最 优 位 姿 T， 本 课 题 采 用 EPnP(Efficient

[Perspective-n-Point) 方法[64](#bookmark335)]来高效地计算位姿，其最少仅需要 4 对匹配点即可完 成计算。

此外，由于粗到细的匹配结果会给出 Nc 个候选地图关键帧，所以可以获得 Nc

个位姿估计结果 {T,j}c1(其中可能存在因为匹配点过少而匹配失败的情况)，这些

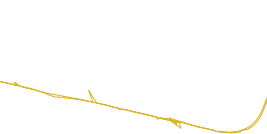
结果可以通过投票机制来获得较为鲁棒的位姿估计结果， 具体过程如算法 [4.1](#bookmark337) 所 示，其中的 relang(., .),relpos(., .) 分别表示计算两个转换矩阵的相对旋转角度和相 对位移，相对角度和相对位移阈值分别设置为 tr 和 tp，其具体取值参考表 [5.1](#bookmark226)。

|  |  |
| --- | --- |
| **算法** **4.1** Vote for visual localization result | |
| **输入：**EPnP 估计结果序列 TS,EPnP 估计成功标志序列 BS, 相对角度阈值 tr, 相似位移阈值 **输出：**最优位姿估计结果 Tr  1: Tr ← I4  2: matchmax ← 0 3: i ← 0  4: **while** i < len(TS) **do** 5: match ← 0  6: **if** BS[i] indicates success **then** 7: j ← 0  8: **while** j < len(TS) Λ j ≠ i **do** 9: r ← relang(TS[i], TS[j]) 10: p ← relpos(TS[i], TS[j]) 11: **if** r < tr Λ p < tp **then**  12: match ← match + 1 13: **end** **if**  14: **end** **while**  15: **if** match > matchmax **then** 16: Tr ← TS[i]  17: **end** **if** 18: **end** **if**  19: **end** **while** | tp |

4.3 紧耦合优化预处理方法

4.3.1 坐标系转换矩阵初始化算法

大多数情况下，粗到细定位是一种较为可信的定位信息来源，但是由于真实 场景和建图场景可能存在较大差异， 以及匹配等过程中的误差，所以仅使用粗到 细定位的结果对于高精度要求的场景仍然是不足的。此外，每次粗到细定位的结 果是相对独立的，因此其连续的定位结果也可能会存在跳变，如图 [4.3](#bookmark186)所示， 其中 图的左侧是整个轨迹示意图，红色框选中局部区域，右侧是该区域详细展示。

100 

视觉定位的轨迹局部展示

视觉定位的轨迹 二 局部区域

|  |  |
| --- | --- |
| |  | | --- | |  | |

0 

100 

Y (m)

 - 200



300 

400 

500 

800  600  400  200 0

X方向 (m)

图 4.3 独立粗到细定位结果中的跳变

为了解决这种跳变，增加最终定位结果的平滑性， 需要使用视觉惯性里程计 的相对位姿估计。除此之外，在粗到细定位偶尔失败的时候，还可以使用视觉惯 性里程计的定位结果作为预备， 能有效防止系统的整体崩溃。但是视觉惯性里程 计和粗到细定位的坐标系并不一致，想要使用视觉惯性里程计的结果， 需要将其 转换到全局坐标系中。因此， 需要一个转换矩阵初始化过程来获得两个坐标系的 对齐参数，即转换矩阵 Tog 。对于每个相机 ci，其存在一个粗到细定位结果 cgi，还

有一个视觉惯性里程计的定位结果 i ，那么定位 VIO 世界坐标系与全局坐标系

之间的变换为：

og,i = cgi i （4.2）

其中，(̂) 表示带噪测量值。为了确定一个鲁棒的初始化变换 Tog ，本文使用了类 [似4.1](#bookmark337) 的投票类算法，从前 20 个关键帧中收集的所有og,i 进行评估，选择内点样本 数量最多的变换, 该变换随后被用作对齐参数，将 VIO 获得的位姿映射到全局坐标 系中。内点的选择同样使用相对角度和相对位移阈值 tr , tp 来进行判断，取值等同 [于4.2](#bookmark86)节中的同名变量。

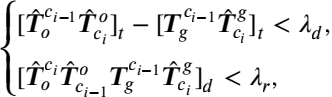
由于转换矩阵相当于 VIO 世界坐标系和全局坐标系之间的对齐参数，所以其 成功对齐对于整个定位系统的有效性至关重要。因此， 只有当粗到细定位成功时，

才会进入后续阶段，否则，将会使用滑窗形式不断丢弃旧的帧，更新最新的候选对 齐参数，重新进行投票算法来估计 Tog 。

4.3.2 初始位姿的自适应选择算法

虽然粗到细定位使用了基于深度神经网络的特征提取和匹配方法来获得相对 准确的地图点和当前观测之间联系， 并使用基于投票的方式来鲁棒得获得新图像 在全局坐标系中的粗略位姿 Tcg，但粗匹配和细匹配仍可能存在错误，因此定位观 测的误差仍存在不确定性。虽然较小的误差可以通过后端优化来消除，但是如果 出现了较大的突变， 则可能对系统的稳定性产生严重的负面影响。因此， 需要一 个有效性验证过程来判断粗到细定位结果的可信度。

对于第 i 时刻的粗到细定位位姿测量值 cgi，首先将其与第 i − 1 时刻已经优化 的位姿 T 进行比较，如果满足以下条件，则认为新粗到细定位测量值是有效的：

 （4.3）

其中， [⋅]t 表示提取变换矩阵的平移分量， [⋅]d 表示提取变换矩阵的旋转角度： 首 先将变换矩阵的旋转部分提出，然后通过罗德里格斯旋转公式获取相对旋转角度。 阈值 λt 和 λd 分别是判断新粗到细定位观测的相对平移和相对旋转角度是否符合 有效性的条件阈值，具体取值参考[5.1](#bookmark226)。

如果粗到细定位的位姿测量值未能满足条件（[4.3](#bookmark338)），则被视为不可接受，在这 种情况下，使用从 VIO 转换得到的测量值

i = Tog i （4.4）

作为优化的初始位姿。

4.4 基于最大后验概率估计的紧耦合优化

4.4.1 最大后验概率估计问题建模

粗到细定位和视觉惯性里程计的定位测量结果都无法单独达到较高的精度： 粗到细定位的测量值由于与先验地图的匹配误差而存在噪声，而 VIO 的测量值则 受到累积漂移的影响。因此，融合这两种测量值的优化方法对最终精度至关重要。 本文提出使用与先验地图紧耦合的位姿图来优化最终位姿。紧耦合指的是优化过 程中，相机位姿的优化与和地图点云结构的优化同时进行。在这种优化模式下， 地 图的先验地图结构不再是固定不变的，而转化为一种概率地图，基于这种概率地

图，后段的位姿优化也从最大似然估计转化为一种最大后验概率估计。这样做的 优势是将点云结构的不确定性考虑到优化过程中，从而提高了优化的鲁棒性。

如果当前已有的相机帧观测到的点结构表示为 P = {p|m = 1, ⋯ , M}，它们

的观测表示为 七 = {im |i = 1, ⋯ , N}，其中 xim 表示 i-th 相机帧对点 p 的观测，

则位姿估计过程可以表示为最大化概率：

 （4.5）

其中，丁 = {Tcgi |i = 1, ⋯ , N} 表示位姿的集合。又因为：

p(丁, P|七) ∝ p(七|丁, P) ⋅ p(丁, P), （4.6）

即概率（ [4.5](#bookmark339)） 所表示的最大后验概率可以分解为似然和先验的乘积。其中， p(七|丁, P) 表示似然，p(丁, P) 表示先验。假设位姿和点结构的先验分布相互独立，

可以将先验分解为：

p(丁, P) = p(丁) ⋅ p(P).

由于我们主要关注点结构的先验，可以将后验分布重写为： p(丁, P|七) ∝ p(七|丁, P) ⋅ p(丁) ⋅ p(P)

∝ p(七|丁, P) ⋅ p(P). 因此，位姿和点结构的估计可以表示为：

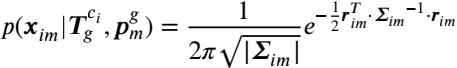


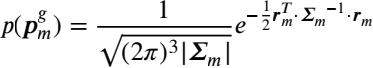
其中各项展开为：





假设似然和先验均为高斯误差，则两个概率可以表示为：





（4.7）

（4.8）

（4.9）

（4.10）

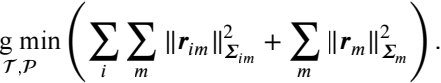
（4.11）

（4.12）

（4.13）

其中 Σim ∈ ℝ2×2 表示视觉观测的误差的协方差矩阵，Σm ∈ ℝ3×3 表示地图点的先 验误差的协方差矩阵，rim 和 rm 分别为视觉观测误差和地图点云先验误差，其定

义在后文详细介绍。因此式（[4.9](#bookmark340)）的概率取负对数并省略常量值可重写为：

丁\* , P\* = ar （4.14）

至此， 已经将求解概率最大化问题转化为优化非线性残差问题。其中，rm 表示的 先验约束是本文区别于一般定位建图方法的关键点： 经典的位姿优化将地图点云 结构视为固定不变的参数，而本文将地图点云结构建模为以点云三维坐标为中心 的高斯概率分布。在这种建模下， 点云的空间坐标也作为待优化的参数。这种基 于地图点的先验值进行优化的方式，也可以看作是一种最大后验概率 (Maximum A Posteriori Probability, MAP) 优化。

4.4.2 使用位姿图求解紧耦合优化问题

式（[4.14](#bookmark341)）是一个非线性优化问题，可以通过优化位姿图的形式进行优化。位姿 图是一种表示优化变量及其相互关系的图数据结构，其将优化问题转化为一个图， 图中的节点表示待优化的参数，边表示约束，通过最小化边上的残差来求解最优参

数。因此，如果要求解位姿 {Tcgi li = 1, … , N} 和地图点云结构 {plm = 1, … , M}，

可以将其视作图优化的顶点，而残差 rim, rm 可以作为图的边，此处定义其为地图 观测边和地图点先验边。

|  |
| --- |
|  |

g T

cj

T+2

|  |
| --- |
|  |

|  |
| --- |
|  |

……

T+1

|  |
| --- |
|  |

p | m = 1, … , Np

T+n

● **相机位姿顶点**

 **地图点云结构顶点**  **相邻帧边**

 **地图关键帧边**  ~~回~~  **地图观测边**

~~口~~  **地图点先验边**

图 4.4 紧耦合优化位姿图

除了上述两种残差，根据视觉惯性里程计和粗到细定位给出的位置观测，还 可以构造相邻帧边和地图关键帧边，整个紧耦合优化位姿图如图 [4.4](#bookmark188)所示。首先已 经定义的地图观测边 rim 可以表示为：

rim = im - π(Tcgi , p), （4.15） 其中，im 是地图点 p 在相机位姿 Tcgi 下的视觉观测值。地图点先验边 rm 可以表

示为：

rim = im - π(Tcgi , p), （4.16）

其中， 是点 p 的先验位置，在建图阶段即确立。 相邻边用于约束位姿的平滑性，其表达式为：

ri = [(i )-1 (TiT)]lie , （4.17）

其中，i 是视觉惯性里程计的测量值，[.]lie [表示求解矩阵的李代数 (Lie algebra)[66](#bookmark342)] 。

虽然视觉惯性里程计在运行一段时间后会产生世界坐标系的漂移，但是其获得的 帧间相对运动还是相对准确的，因此可以用于约束相邻帧的位姿。

如果位姿图中存在由粗到细定位初始化的相机位姿， 则必定存与之相对应的

参考地图关键帧。假设相机位姿 Tcgi 对应于地图关键帧 T， 则地图关键帧边可以

构造为：

rij = [(ckij )-1 (Tgkj Tcgi )]lie ,

其中，ckij 是使用地图关键帧和当前图像计算得到的地图定位测量值。

最终，所有需要优化的残差结合起来可以作为整体优化目标：

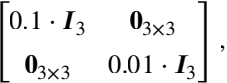




（4.18）

（4.19）

其中，S 表示相邻帧的集合，c 表示有参考地图关键帧的位姿集合。协方差矩阵 Σi , Σij, Σim , Σm 一般作为超参数来平衡各种误差在数量级上的不平衡。因为在矩 阵的李代数中有部分表示旋转误差，部分表示平移误差，旋转误差的模等于转动 角，这个值一般较小，而平移误差的模一般表示位移距离，这个值一般较大，所以 设置

Σi = Σij =  （4.20）

其中 I3 表示 3 维的单位矩阵，03 表示 3 维的零矩阵。 Σim 一般表示相机的观测协 方差，这个值一般根据相机的焦距 f 而确定，本文设置为

Σim =  . I2 . （4.21）

最后，对于地图点的先验协方差，此处根据其建图的像素误差确定

Σm = diag(Rm-1 . K-1 . zm . um . I3 - Rm-1 . tm), （4.22） 其中 diag(.) 表示将向量作为对角线元素构造对角矩阵，Rm 和 tm 分别表示观测

到地图点 p 最大误差的地图关键帧在全局坐标系下的旋转矩阵和平移向量，K 表 示相机的内参矩阵，zm 和 um 分别表示地图点的深度和像素误差，深度根据旋转和 平移计算，像素误差值在建图阶段记录。

除此之外，为了约束优化的规模，提升系统的实时性，本文使用滑动窗口的方 法来保证只有定量的位姿参与优化， 即确定 丁 = {Tcgi |i = 1, ⋯ , N} 中的 N 是一个 合理值。另外针对每一帧的地图点优化数量，本文选择只将单帧所能观测到的地 图点按照重投影误差排序， 只选择其中的前 K 个点进入位姿图优化，保证了点云

结构 P = {p|m = 1, ⋯ , M} 的优化规模

M ⩽ K ⋅ N. （4.23）

4.5 转换矩阵更新方法

转换矩阵控制着从 VIO 世界坐标系到全局坐标系的转换，而在某些情况，例 如粗到细定位偶然失效时，其关系着视觉惯性里程计的结果能否被紧耦合优化正 确初始化，所以有必要对其进行实时维护。

|  |
| --- |
| **算法** **4.2** Update the transformation matrix |
| **输入：**旧转换矩阵 Told , 当前相机的转换矩阵序列 TS **输出：**新转换矩阵序列 Tnew  1: accinlier ← 0  2: Tnew ← Told  3: **for** T in TS **do**  4: r ← relang(Told , T) 5: **if** r < 60∘ **then**  6: accinlier ← accinlier + 1 7: **end** **if**  8: **end** **for**  9: **if** accinlier ⩾ 0.5 ⋅ len(TS) **then**  10: Racc ← I3 11: tacc ← 03  12: **for** T in TS **do**  13: R, t ← decompose(T)  14: Racc ← Racc + R  15: tacc ← tacc + t  16: **end** **for**  17: U, Σ, V = SVD(Racc )  18: Rnew ← U ⋅ transpose(V) 19: tnew ←  ⋅ tacc  20: Tnew ← compose(Rnew , tnew ) 21: **end** **if** |

由于视觉惯性里程计本身存在的漂移累积，所以在运行一段时间之后，转换 矩阵必然会出现无法将 VIO 世界坐标系和全局坐标系对齐的情况。因此， 需要一

个实时的转换矩阵更新过程来保证定位的准确性。

在本文中，更新矩阵的实时维护就由转换矩阵更新步骤来完成。其主要工作是 在每次紧耦合优化完成之后，根据优化结果更新转换矩阵。假设优化后的相机位 姿为 {Tcgi |i = 1, ⋯ , N }，则可以计算出当前每个相机的转换矩阵 {og,i |i = 1, ⋯ , N }， 假设待更新的转换矩阵为 og,i，则更新算法如算法 [4.2](#bookmark343)所示。

算法中 decompose(⋅) 表示将转换矩阵分解为旋卷矩阵和位移向量，compose(⋅, ⋅) 则表示将旋卷矩阵和位移向量组合为转换矩阵， SVD(⋅) 表示奇异值分解函数， tranpose(⋅) 表示矩阵转置操作。

4.6 本章总结

本章主要介绍了紧耦合地图定位模块的设计，其输入包括当前的图片、先验 地图信息以及视觉惯性里程计，其主要工作就是使用前面工作的信息综合估计当 前相机的位置和姿态。首先，本章介绍了粗到细定位的设计，其主要工作是使用 深度神经网络提取特征，然后使用基于投票的方法估计相机的位置和姿态。接着， 本章介绍了转换矩阵初始化和有效性验证的设计，这两部配合工作可以确保有一 个正确的初始化位姿可以被紧耦合优化使用。然后，本章介绍了紧耦合优化的设 计，其主要工作是将粗到细定位和视觉惯性里程计的结果融合，然后使用图优化 的方法估计相机的位置和姿态。最后，本章介绍了转换矩阵更新的设计，其主要 工作是根据优化后的结果更新转换矩阵。

第 5 章 视觉惯性定位系统搭建与测试

本章主要对定位系统的搭建即实施进行介绍。首先介绍了定位系统运行所依 靠的硬件平台以及软件环境，不同于一般的定位系统，本文方法的实施依赖于深 度学习及图形处理单元 (Graphics processing unit, GPU)。然后介绍了本文所使用的 数据集以及评价指标，本文实验主要建立在 3 个数据集上，分别是 KITTI 数据集、 4Seasons 数据集和 KAIST 数据集，其中KITTI 数据集场景相对单一，适合验证离 线建图模块的性能，以及初步验证定位模块的性能；4Seasons 数据集提供的场景丰 富，适合验证 VIO 和紧耦合地图定位模块的性能；KAIST 数据集提供了超长轨迹， 适合验证 VIO 模块的性能。最后本章介绍了定位系统三个模块的测试结果，并通 过相应的消融实验来验证各部分的真实作用。

5.1 系统与测试细节

5.1.1 硬件与软件环境

本文所设计的定位系统的运行环境依赖于同时具有 CPU(Central processing unit) 和 GPU(Graphics processing unit) 的计算机， 具体来说本文所实验的平台使 用 12th Gen Intel(R) Core(TM) i7-12700 CPU 和 NVIDIA GeForce RTX 3080 GPU。



Ubuntu 20.04 LTS ROS Noetic VINS-Mono TensorRT

图 5.1 软件部署环境

系统的软件部署环境如图 [5.1](#bookmark190) 所示， 其中 Ubuntu 20.04 LTS 是本文所使用的 操作系统；ROS Noetic 是本文所使用的机器人操作系统，其主要作用是将多个子 系统的结果方便快捷地统一处理；VINS-Mono 是本文所使用的视觉惯性定位系统 的基线，其具有较为简洁的逻辑划分，方便快速改进、验证想法； TensorRT 是本 文所使用的深度学习推理框架，其主要负责系统中的神经网络推理任务，其针对 NVIDIA 显卡的速度优化使得整个系统可以实时运行。此外， 本文所设计系统中还 有大量超参数，其具体取值如表 [5.1](#bookmark226)所示。

表 5.1 本文中所涉及的运行超参数

|  |  |  |  |
| --- | --- | --- | --- |
| 参数名 | 首次出现章节 | 作用 | 取值 |
| tl | [2.4](#bookmark50) | SfM 中闭环检测相似度阈值 | 0.8 |
| m | [2.4](#bookmark50) | SfM 中闭环帧数量 | 5 |
| α | [2.6](#bookmark54) | 建图融合权重系数 | 1000 |
| τ | [3.2](#bookmark62) | 车辆状态正例预测阈值 | 0.9 |
| Nc | [4.2.2](#bookmark90) | 粗到细匹配候选关键帧数量 | 7 |
| tr | [4.2.2](#bookmark90) | 投票过程相对旋转角度阈值 | 15∘ |
| tp | [4.2.2](#bookmark90) | 投票过程相对位移阈值 | 5m |
| λr | [4.3.2](#bookmark96) | 有效性验证角度阈值 | 60∘ |
| λp | [4.3.2](#bookmark96) | 有效性验证位移阈值 | 15m |
| N | [4.4](#bookmark98) | 滑动窗口容量 | 20 |
| K | [4.4.2](#bookmark102) | 单帧图像所考虑的点云比例 | 10 |

5.1.2 测试数据集与评价指标

<5.1.2.1> KITTI 数据集

KITTI 数据集[[67](#bookmark344)]是 SLAM 和自动驾驶研究中最受欢迎的数据集之一，它包含 多个小时的交通场景数据，这些数据通过多种传感器采集，包括高分辨率 RGB 立 体相机、灰度立体相机以及 3D 激光扫描仪。该数据集提供了 11 条经过精心选择 的序列，构成视觉里程计基准。每条序列包含来自四个相机的数据： 两个灰度相 机和两个彩色相机。此外， KITTI 数据集还为这 11 条视觉里程计基准序列中的 10 条提供了原始 GNSS 观测数据和IMU 测量数据。



序列 00-城镇环境 序列 01-高架环境



序列 09-郊外环境 序列 10-农村环境

图 5.2 KITTI 数据集场景

本文使用 KITTI 视觉里程计基准中包含 GNSS 观测数据的 10 条序列，评估本

文所提出系统的建图性能以及整体性能。这 10 条序列包括了城镇、高架、郊外以 及农村等常见的车辆行驶环境，如图 [5.2](#bookmark192)所示。

[5.1.2.2 4](5.1.2.24)Seasons 数据集

4Seasons 数据集[[68](#bookmark345)]是一个专为自动驾驶研究设计的数据集，提供了涵盖季节 变化和复杂感知条件的图像、IMU 等数据。同一路径在不同季节和多种感知条件 下被多次采集，使其特别适合于评估针对固定路径设计的方法。该数据集使用的 传感器包括立体相机、IMU 和 RTK-GNSS， 同时还提供了精度达到厘米级的高精 度真值位姿。

与 KITTI 数据集相比，4Seasons 数据集能够提供更具挑战性和说服力的实验 结果，突出了本文方法的优势。本文选择了四个具有代表性的场景，分别是

**(1)Office** **Loop** **(OF):** 位于城郊工业区的环形路线，长度约为 3776 米，包含了 多种感知条件， 如建筑物、树木和道路标志等， 如图 [5.3](#bookmark194) 所示。此路线在不同季 节采集了 5 条序列 (以 OF1-5 代称)，每条序列包含了来自灰度立体相机、IMU 和 RTK-GNSS 的数据。



场景感知条件 路径卫星图

图 5.3 Office Loop 场景

**(2)** **Neighborhood** **(NH):** 位于城镇环境的环形路线，长度约为 2106 米, 包含了 多种感知条件，如树木和房屋等，包含了更多的植被，接近于农村环境，如图 [5.4](#bookmark196) 所示。此路线在不同季节采集了 7 条序列 (以 NH1-7 代称)，每条序列包含了来自 灰度立体相机、IMU 和 RTK-GNSS 的数据。

**(3)** **Business** **Campus** **(BC):** 位于校园和商业区的多条环形路线，长度约为 3011 米，包含了更多的建筑物，如图 [5.5](#bookmark198)所示。此路线在不同季节采集了 3 条序列 (以 BC1-3 代称)，每条序列包含了来自灰度立体相机、IMU 和 RTK-GNSS 的数据。

**(4)** **Old** **Town** **(OT):** 位于城市中心的商业区环线，长度约为 5034 米，包含了 多种感知条件，此外还有大量的移动物体，如汽车、行人和自行车等， 如图 [5.6](#bookmark200)所 示。此路线在不同季节采集了 2 条序列 (以 OT1-3 代称)，每条序列包含了来自灰



场景感知条件 路径卫星图

图 5.4 Neighborhood 场景



场景感知条件 路径卫星图

图 5.5 Business Campus 场景

度立体相机、IMU 和 RTK-GNSS 的数据。



场景感知条件 路径卫星图

图 5.6 Old Town 场景

<5.1.2.3> KAIST 数据集

[KAIST 数据集[69](#bookmark346)]同时提供了LiDAR(Light Detection And Ranging)、立体相机、 GPS、IMU、编码器和气压计等多种传感器数据， 使得用户可以针对不同传感器组 合进行多模态信息处理。采集环境涵盖了从多车道宽阔道路到高密度城市街区等 多种复杂场景，这不仅体现了实际城市环境中 GPS 信号不稳定、遮挡严重等问题， 也为算法在动态、遮挡等极端条件下的鲁棒性验证提供了真实数据。此外，数据 集通过高精度传感器与消费者级传感器的双层数据采集， 以及精细的时间同步和 标定处理，确保了数据的高准确性和一致性，并提供了基准轨迹及 ROS 工具，极 大地方便了机器人定位、SLAM 及环境感知等相关领域的研究与测试。

本文中主要选择了 KAIST 数据集中的三个场景， 分别是 20-Highway 、26- Dongtan 和 28-Pangyo，如图 [5.7](#bookmark202) 所示。这三个场景分别代表了高速公路、郊区和 城市环境，适合于测试本文方法在不同场景下的性能。其中 28-Pangyo 是一段超过 10 公里的城市环境，包含了大量的建筑物、行人和车辆，是一个非常具有挑战性 的场景。KAIST 数据集主要被使用来评估本文的 VIO 模块性能。



序列 20-Highway 序列 20-Dongtan 序列 20-Pangyo

图 5.7 KAIST 数据集场景

<5.1.2.4> 数据集使用划分

由于本文需要分别验证 3 个模块的效果， 因此本文将数据集的划分总结如 表 [5.2](#bookmark228)所示。表中的每一列代表了一个测试模块，每一行代表了一个数据集，表中 的符号代表了数据集的使用方法，其中○代表使用该序列进行训练或者建图，△代 表使用该序列进行测试。

表 5.2 数据集使用划分

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 测试模块 | KITTI 4Seasons | | | | | | KAIST |
| 灰色 | 彩色 | OF1 | OF2-5 | NH1 BC1 OT1 | NH2-7 BC2-3 OT2-3 | 20-Highway 26-Dongtan 28-Pangyo |
| 离线建图 VIO  紧耦合地图定位 | ○△  △ | ○ | △  ○ | △ △ | ○ ○ | △ | △ |

a ○ 代表使用该序列进行训练或者建图 b △ 代表使用该序列进行测试

KITTI 数据集参与离线建图和紧耦合地图定位两个模块的测试： 在评估建图 过程时，使用来自单目灰度相机的完整图像集； 在评估整体性能时由于 KITTI 数

据集未直接提供沿同一路径的多次行驶数据，本文中通过使用不同的相机模拟多 次行驶。具体而言， 使用部分彩色相机拍摄的图像进行离线建图过程，然后使用灰 度相机拍摄的图像测试紧耦合地图定位的精度。建图图像的比例设置为 0.25， 即 只使用四分之一的彩色图像来构建先验地图，并在完整的灰度图像集上进行定位。

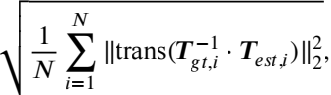
4Seasons 数据集参与 VIO 和紧耦合地图定位两个模块的测试：在评估 VIO 时， 使用 NH1 、BC1 和 OT1 的数据作为训练数据，OF1-5 作为测试数据；在评估紧耦 合地图定位性能时，使用 OF1、NH1、BC1 和 OT1 作为建图数据，OF2-5、NH2-7、 BC2-3 和 OT2-3 作为测试数据。此外由于紧耦合地图定位模块需要使用 VIO 结 果，而 VIO 的训练数据与紧耦合地图定位的测试数据没有重合，所以可以直接使 用 VIO 的结果作为紧耦合地图定位的输入而不需要担心训练数据中混入了测试数 据。

KAIST 数据集参与 VIO 模块的测试： 使用序列 20-Highway 、26-Dongtan 和 28-Pangyo 作为测试数据，而此处的车身状态判别神经网络则直接使用 4Seasons 数 据集上的预训练参数。

<5.1.2.5> 评价指标

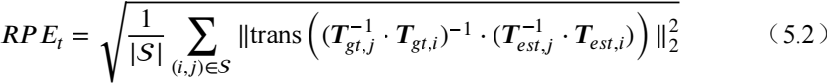
本文实验中使用三种不同的指标来评估定位性能： 绝对轨迹误差 (Absolute Translation Error, ATE)[[70](#bookmark347)]和相对位姿误差 (Relative Pose Error, RPE)[[71](#bookmark348)]。其中绝对 轨迹误差只评估位移分量，记做 AT Et， 单位为米 (m)； 相对位姿误差评估了相对 平移和相对旋转的误差，分别记做 RPEt 和 RPEr，单位分别为米/100 米 (m/100m) 和度/100 米 (。/100m) ，RPEt 的计算是在每 100 米的间隔中计算当前定位结果与真 值 (Ground Truth) 的平移误差，RPEr 的计算是在每 100 米的间隔中计算当前定位 结果与真值之间的旋转角度误差。

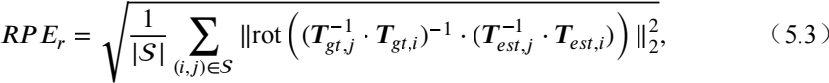
AT Et 的计算方式为可以表示为：

AT Et =  （5.1）

其中 trans(.) 表示提取转换矩阵的平移部分，Tgt,i, Test,i 分别表示位姿的真值和预测 值。由于绝对轨迹误差的计算对于坐标系的选择较为敏感， 因此每次计算前会首 [先使用 Umeyama 算法[72](#bookmark349)]将估计的轨迹与真值轨迹对齐。

RPEt 和 RPEr 的计算方式分别为：





其中 rot(⋅) 表示提取转换矩阵的旋转部分，S 表示相对位姿误差计算的间隔为 100 米的两帧集合。相对位姿误差的计算方式更加追求描述相对变换， 能更好反映出 里程计的性能。此外相对位姿误差的计算方式对于坐标系的选择不如绝对轨迹误 差敏感，但也有一定要求，本文使用 Li et al.[[73](#bookmark350)]提出的方法首先对估计的轨迹与真 值轨迹对齐，然后再计算相应数值。

在实验结果汇报中，由于 KITTI 数据集提供的真值格式与其他数据集不同，所 以在计算方法中，KITTI 由官方提供的计算工具进行计算，数值精度最高保留 2 位 小数；4Seasons 和 KAIST 数据集上的实验结果则使用 Grupp[[74](#bookmark351)]提出的开源工具包 EVO 进行计算，精度相对较高，可以保留6 位小数，方便更精确的对比。

5.2 离线建图模块测试

对于离线建图性能，本文以地图关键帧的位姿评估来定量对比建图精度， 实 验中选择了 KITTI 数据集的所有序列来评估本文的建图方法以及其他几种建图方 法，并使用三种指标进行对比，对比方法包括 GNSS 测量值、COLMAP[[36](#bookmark294)] 、VINS- Stereo[[75](#bookmark352)] 以及本文的离线建图方法。建图结果如表 [5.3](#bookmark230) 和表 [5.4](#bookmark232) 所示。

表 5.3 KITTI 数据集建图 AT Et ↓ 对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序列 | GNSS(上限) | COLMAP\* [[36](#bookmark294)] | VINS-Stereo [[75](#bookmark352)] | 离线建图 |
| 00 | 0.03 | 5.61 | 5.95 | **0.33** |
| 01 | 0.06 | N/A | 6.49 | **0.17** |
| 02 | 0.03 | N/A | 5.99 | **0.61** |
| 04 | 0.02 | 4.99 | 1.08 | **0.54** |
| 05 | 0.02 | 8.28 | 5.98 | **0.24** |
| 06 | 0.04 | 1.78 | 3.48 | **1.55** |
| 07 | 0.03 | 3.41 | 2.41 | **0.20** |
| 08 | 0.02 | N/A | 3.85 | **0.71** |
| 09 | 0.03 | 8.50 | 1.78 | **0.50** |
| 10 | 0.03 | N/A | 3.78 | **0.79** |

a 星号 \* 表示 COLMAP 的结果经过了尺度调整

b VINS-Stereo 表示 VINS-Fusion 的双目版本，使用双目图片和 GNSS 测量

表中使用粗体和红底标出非 GNSS 观测值中的最佳性能结果，GNSS 观测值

表 5.4 KITTI 数据集建图 RPE↓ 对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序列 |  | RPEt ↓ |  |  | RPEr ↓ |  |
| COLMAP | VINS-Stereo | 离线建图 | COLMAP | VINS-Stereo | 离线建图 |
| 00 | 1.91 | 3.45 | **1.20** | 0.82 | 1.45 | **0.69** |
| 01 | N/A | 12.18 | **0.84** | N/A | 3.54 | **0.13** |
| 02 | N/A | 2.65 | **0.63** | N/A | 1.26 | **0.27** |
| 04 | 4.12 | 2.71 | **0.71** | **0.11** | 0.62 | 0.14 |
| 05 | 1.75 | 1.41 | **1.10** | 0.22 | 0.69 | **0.19** |
| 06 | 1.37 | 1.33 | **0.86** | 0.20 | 0.72 | **0.17** |
| 07 | 1.16 | 1.17 | **0.39** | 0.26 | 0.87 | **0.24** |
| 08 | N/A | 3.46 | **1.87** | N/A | 1.77 | **1.17** |
| 09 | 0.86 | 4.45 | **0.62** | 0.31 | 1.87 | **0.25** |
| 10 | N/A | 2.38 | **0.68** | N/A | 0.94 | **0.19** |

a 星号 \* 表示 COLMAP 的结果经过了尺度调整

b VINS-Stereo 表示 VINS-Fusion 的双目版本，使用双目图片和 GNSS 测量

可被视为位置建图精度的上限。此外， 由于 GNSS 测量值仅提供平移信息而不包 含旋转，实验仅在 AT Et 指标下与 GNSS 测量值进行对比。

COLMAP 方法作为一个流行的SfM 框架，代表了仅使用图像的建图性能。基 础的 COLMAP 使用 SIFT[[61](#bookmark332)]描述符进行建图，而本文的建图特征是基于深度神经 网络。VINS-Stereo 是 VINS-Fusion 的双目版本，VINS-Fusion 是一种广泛使用的 松耦合状态估计器，集成了视觉、惯性和 GNSS 测量。实验中使用双目图像输入， 并结合 GNSS 测量，代表典型的 GNSS 辅助视觉建图方法。本文所提出的离线建 图模块也是一种 GNSS 辅助视觉建图方法，其性能将与上述方法进行比较。

VINS-Stereo 和本文的离线建图模块均能在真实世界尺度上建立地图，因此可 以在所有指标上进行对比。但是 COLMAP 无法在没有额外位置先验的情况下重建 真实世界尺度的地图，这使得直接比较绝对平移结果不公平。因此， 实验中首先 对 COLMAP 的轨迹应用 Sim3 变换调整框架和尺度，将其与真值对齐。Sim3 变换 参数通过最小化 COLMAP 位置信息与真值位置之间的最小二乘误差计算得到。对 于相对旋转误差，COLMAP 的结果可以直接与其他方法对比。

[表5.3](#bookmark230)和表[5.4](#bookmark232)比较了 ATE 和RPE 指标，结果表明本文的离线建图方法在所有 候选方法中表现最佳。与 COLMAP 的对比表明，本文的建图方法在平移精度上显 著提升了视觉建图性能。作为一种增量式 SfM 方法，COLMAP 受到累计漂移的影 响，导致平移精度较差。本文的方法通过结合 GNSS 位置先验约束了漂移，实验结

果验证了其在降低绝对平移误差和相对平移误差方面的有效性。此外，通过表中数 据可以发现 COLMAP 方法在许多序列上由于无法初始化前两帧图像而建图失败。 而本文中基于深度学习的特征点可靠地完成初始化，体现了深度学习特征优势。

与 VINS-Stereo 的对比也表明，本文提出的方法优于松耦合的 GNSS 辅助视觉 建图方法。VINS-Stereo 尝试利用GNSS 测量校正视觉 SLAM 的位置估计，但这种 校正不直接影响位姿优化过程，而本文方法将 GNSS 与视觉观测融合，使得平移 先验能够直接影响位姿估计。因此，本文提出的建图方法在整体性能上表现最佳。

在相对旋转误差 RPEr 方面，本文方法在大多数序列中也表现出优越性能。相 对旋转误差反映了位姿估计的准确性，本文方法和 COLMAP 都基于离线的三维重 建方法，能够实现全局优化，所以两种方的位姿估计精度相较于 VINS-Stereo 有明 显提升。由于 GNSS 测量不提供旋转约束，本文的方法与 COLMAP 在旋转精度上 的差异较小。

5.3 PO-VIO 模块测试

5.3.1 车身状态判别方法的性能测试

VIO 的运行过程中使用到了车身状态结果， 并且车身状态判别的结果对里程 计精度有着重要的影响， 因此针对 VIO 的效果展示首先从车身状态的判别结果开 始。

车身状态的判别结果如表 [5.5](#bookmark234)所示，表中展示了对于车身两个状态 ZFOR , ZLAT 判别的查准率、召回率、F1 得分以及准确率。结果表明本文所提出的车身状态判 别方法在所有指标上均表现出色，具有较高的准确性和鲁棒性。

表 5.5 车身状态预测效果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数据集 | 状态量 | 查准率 | 召回率 | F1 得分 | 准确率 |
| 4Seasons | ZFOR | 0.97 | 0.76 | 0.85 | 0.96 |
| ZLAT | 0.98 | 0.97 | 0.98 | 0.96 |
| KAIST | ZFOR | 0.65 | 0.74 | 0.69 | 0.98 |
| ZLAT | 0.99 | 0.88 | 0.93 | 0.87 |

需要注意的是， 由于在 KAIST 数据集上的车身状态判别对于车身状态 ZFOR 的查准率较低， 因此在后续 KAIST 数据集测试中，本文放弃了这一车身状态的选 择，仅使用车身状态 ZLAT 来进行车身运动学约束。

5.3.2 PO-VIO 的性能测试

[5.3.2.1 4](5.3.2.14)Seasons 数据集测试

在验证过车身状态判别的效果后， 本文将车身状态判别加入到 VIO 中，在 [其余序列上与其他几种 VIO 方法进行对比，对比方法包括 DSO[76](#bookmark353)] 、OKVIS2[[77](#bookmark354)] 、 BASALT[[78](#bookmark355)] 、OpenVINS[[79](#bookmark356)] 、VINS-Mono[[25](#bookmark283)] 、ORB-SLAM3[[29](#bookmark287)] 以及本文的 VIO 方 法。需要说明的，除了 DSO 和 ORB-SLAM3 以外，其他方法均为单目相机 +IMU 运行，DSO 和 ORB-SLAM3 的单目初始化过程无法成功，所以本文选择其双目版 本进行初始化，在表中以星号 \* 标出。具体的效果对比如表[5.6](#bookmark236)和表 [5.7](#bookmark238)所示， 其 中使用粗体和红底标出最佳表现。

在 AT Et 的比较中， 除了单纯的里程计对比，本文还加入了里程计配合回环 (Loop Closure) 检测优化的效果， 以 VIO+LC 表示。加入了回环检测优化的方法会 有更高的精度，但是由于回环检测是一种对过去累计误差的修正，具有滞后性，所 以其高精度结果一般难以被实时使用。此处汇报 VIO+LC 的结果是为了展示本文 提出的 VIO 在配合回环检测之后依旧可以取得较好的效果。

另外，有些方法在某些序列上运行失败，所以在表中使用 N/A 表示； 有些方 法可以运行成功，但是中途崩溃，产生的轨迹与真值误差过大，所以在表中我们仅 汇报误差小于 60 米的结果，对于误差大于 60 米的结果我们使用>60 表示。

从表 [5.6](#bookmark236) 中可以看出，在单纯的 VIO 对比中，本文所提出的 VIO 方法在大多 数序列上 AT Et 表现出色，相较于其他方法具有更高的精度。在 VIO+LC 的对比 中，本文的方法在大多数序列上仍然有最好的表现，但是在 OL3 和 OL5 序列上 的表现相对较差，ORB-SLAM3 在这两个序列上表现较好。此外，本文 VIO 基于 VINS-Mono 改进而来，所以在与 VINS-Mono 的对比中全面领先，说明本文 VIO 中 的改进有一定效果。

从表 [5.7](#bookmark238) 中的结果可以看出，本文 VIO 在相对位姿误差 RPEt 和 RPEr 上均 表现有较大差异：在相对平移误差 RPEt 上，本文 VIO 在 OF2 上表现最后，但在 其他序列上的表现均落后于 ORB-SLAM3，甚至 VINS-Mono； 但是在相对旋转误 差 RPEr 上，本文 VIO 在所有序列上均表现出色，具有更高的精度。

产生这一现象的原因可能是：本文所提出的伪观测约束项（[3.25](#bookmark327)）中仅存在有 关旋转和速度的约束， 并不直接对平移进行约束，并且在推导雅可比矩阵（[3.27](#bookmark328)）、 （[3.34](#bookmark330)）和（ [3.37](#bookmark331)）时也只对旋转向量产生梯度，所以在相对平移误差上表现较差， 而在相对旋转误差上表现较好。

表 5.6 视觉惯性里程计的 AT Et ↓ 对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 方法 | 类型 | OL1 | OL2 | OL3 | OL4 | OL5 |
| OKVIS2  BASALT  OpenVINS  VINS-Mono  ORB-SLAM3\* 本文 VIO | VIO | 37.917951 15.696389 30.141148 | 52.712329 12.439054 26.845080 | >60  >60  >60  20.552186 31.387398 | N/A  7.972945 15.335389 | N/A  13.211903 |
| **6.931078** |
| **8.847395** | **4.343367** | **17.193425** | **7.653040** |
| 11.009781 |
| DSO\*  OKVIS2  BASALT  OpenVINS  VINS-Mono  ORB-SLAM3\* 本文 VIO | VIO+LC | 5.575564  37.917951 9.368280 4.902441 | 6.889360  52.712329 7.878626 15.621815 | 13.183826  >60  >60  >60  10.664015 | 8.048476  N/A  7.661955 5.073350 | 6.681058  N/A  8.083566 |
| **6.288431** | **5.520636** |
| **3.834251** | **3.321656** | **4.972996** |
| 11.670275 | 7.652689 |

a 星号 \* 表示 ORB-SLAM3 和 DSO 因单目版本初始化困难而使用双目版本 b LC 表示回环检测 (Loop Closure)

c N/A 表示方法在该序列上无法运行

d >60 表示方法在该序列上的误差大于 60 米

表 5.7 视觉惯性里程计的 RPE↓ 对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 方法 | 指标 | OF1 | OF2 | OF3 | OF4 | OF5 |
| OpenVINS  VINS-Mono  ORB-SLAM3\* 本文 VIO | RPEt | 7.576759 3.921487 | 8.893709 2.330876 2.361792 | 12.631715 5.648737 | N/A  3.447067 | N/A  2.598062 |
| **1.962480** | **4.987597** | **2.963017** | **1.176156** |
| **1.986364** |
| 3.459246 | 5.887755 | 3.731419 | 3.227312 |
| OpenVINS  VINS-Mono  ORB-SLAM3\* 本文 VIO | RPEr | 2.122706 1.062495 0.883776 | 1.752747 0.783557 0.894313 | 3.904525 3.411987 3.002619 | N/A  1.747575 1.764101 | N/A  0.990526 0.981966 |
| **0.818242** | **0.776345** | **2.543815** | **1.732620** | **0.882806** |

a 星号 \* 表示 ORB-SLAM3 因单目版本初始化困难而使用双目版本 b N/A 表示方法在该序列上无法运行

c 由于 OKVIS2 和 BASALT 在 AT Et 指标下误差过大，此处不再汇报 RPE 指标

<5.3.2.2> KAIST 数据集测试

KAIST 数据集上的测试如表[5.8](#bookmark240)所示，表中展示了本文设计的 VIO 在 KAIST 数据集上的效果对比，对比方法包括 Baseline(VINS-Mono) 和本文设计的 VIO。

此外为了方便对比，表中还展示了本文设计相对于 Baseline 的改进变化量 Δ , 用红底表示在该指标下进步，绿底表示在该指标下有所退化。在指标方面，此处 将常规的 AT Et 改变为 AT Et/km， 即每公里的绝对轨迹误差， 以平衡里程计在不 同行驶距离下的精度表现。

表 5.8 KAIST 数据集视觉惯性里程计效果对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序列 | 长度 | 方法 | AT Et/km↓ | RPEt ↓ | RPEr ↓ |
| 20-Highway | 3.2 km | Baseline  本文 VIO  Δ | 27.551458 26.003915 | 19.496478 20.164449 | 1.165007 0.689197 |
| **-1.547543** | **+0.667971** | **-0.475810** |
| 26-Dongtan | 4.0 km | Baseline  本文 VIO  Δ | 6.917041 6.457796 | 8.912866 6.543866 | 1.581715 0.972692 |
| **-0.459245** | **-2.369000** | **-0.609023** |
| 28-Pangyo | 11.47 km | Baseline  本文 VIO  Δ | 5.119043 4.592121 | 11.795006 8.789568 | 0.809538 0.705091 |
| **-0.526922** | **-3.005438** | **-0.104447** |

a Δ 表示本文设计相对于 Baseline 的改进变化量

从表中可以看出， 本文设计的 VIO 在大部分序列上均取得了误差上的改进。 对于 20-Highway 序列，经过改进后，绝对轨迹误差 (AT Et) 从 27.55 降至 26.00，每 公里降低约 1.55；旋转误差 (RPEr) 也由 1.17 降低到 0.69，减少了约 0.48，但相对 平移误差 (RPEt) 则略有增加，从 19.50 上升至 20.16。对于 26-Dongtan 序列，改进 后的 AT Et 由 6.92 降至 6.46 ，RPEt 从 8.91 降低到 6.54 ，RPEr 也由 1.58 降低到 0.97 ，三项指标均有明显改善。28-Pangyo 序列中，AT Et 由 5.12 降至 4.59 ，RPEt 则显著从 11.80 下降到 8.79，而 RPEr 变化较小，从 0.81 略降至 0.78。

总体来看，尽管 20-Highway 序列中 RPEt 有所上升，但本文设计在大部分场 景下均能有效降低 VIO 的绝对及相对误差，整体性能得到了提升。

为了更好对比本文设计的 VIO 与 Baseline 的效果， 本文在图 [5.8](#bookmark204) 中展示了 KAIST 数据集上 RPEr ↓ 的分布情况。从图中可以看出， 本文设计的 VIO 在 KAIST 数据集上的 RPEr ↓ 分布更加集中，且均值更低，相较于 Baseline 有着更好的效果。

20



26

28

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | |  |  | | --- | --- | |  | Baseline  本文VIO | |  | |  | |  | |
|  | | |
|  |  |  |

0 2 4 6 8 10

RPE角度误差(/100米)

图 5.8 KAIST 数据集上 RPEr ↓ 分布示意图

5.3.3 PO-VIO 性能提升的来源分析

为了验证这一猜想，本文对车身坐标系的速度量进行分析。分析建立在 OF2 和 OF3 两条序列上， 这两条序列有着比较突出的特点：OF2 序列上本文 VIO 的 RPEt 和 RPEr 均有所提升，但是整体幅度较小；OF3 序列上本文 VIO 的 RPEt 有 所下降，但是 RPEr 有所提升，而且每 100 米的旋转误差减小了接近 1 度。因此 选择这两条序列进行分析， 能够更好发现本文的速度伪观测约束在何种条件下会 对里程计的效果产生积极影响。两条序列上的速度量在车身坐标系的变化如图 [5.9](#bookmark206) 所示。

在每个序列上分别展示了“VIO”、“VIO+ 伪观测”、“VIO+ 伪观测-因子” 三种情况下的速度变化，其中“VIO”表示使用通用 VIO 方法获得的速度估计结 果；“VIO+ 伪观测”表示在通用 VIO 中加入本文所提出的车身状态判断以及伪观 测约束后的速度估计结果；“VIO+ 伪观测-因子”表示在 VIO 和伪观测约束的基础 上，不使用惯性-车体因子进行后端优化。此外，为了将速度变化与车身状态区间 直观展示，本文中还将对应车身停止、转弯和直行的区间标注在了速度图中。

从两个序列中的速度变化可以看到：相较于 OF2 序列，OF3 序列的“VIO”在 车身横向 (X 轴方向) 和垂直 (Z 轴方向) 速度上有着更大的取值，这说明在 OF3 序 列上基础 VIO 的速度估计有更大的误差，而在该场景下，经过本文所提出的伪观 测约束进行优化后，“VIO+ 伪观测”横向和垂直速度有了更合理的估计，从而使 得 OF3 上的 RPEr 有了明显的提升。相较而言， OF2 序列上的速度变化较为平缓， 所以在该场景下，伪观测约束的加入对于速度估计的影响较小，使得“VIO+ 伪观 测”在 OF2 序列上的 RPEr 仅有微弱提升。

VIO  VIO+  VIO+ -    



4

5

3



 2

 1   0

 1

0 100 200 300 400

 (s)

Y (m/s)

10

5

0

5



0 100 200 300 400

 (s)

1.0

Z (m/s)

0.5

0.0

 0.5

1.0

0 100 200 300 400

 (s)

OF2 序列

VIO  VIO+  VIO+ -    



5

X (m/s)

4

3



 2

 1   0

 1

0 100 200 300 400 500

 (s)

Y (m/s)

10

5

0

5



0 100 200 300 400 500

 (s)

1.0

Z (m/s)

0.5

0.0

 0.5

1.0

0

100

200

 (s)

OF3 序列

300

400

500

图 5.9 车身速度展示

X (m/s)

此外，观察两个序列下的速度变化，可以发现在 OF2 上，“VIO+ 伪观测”的车 身横向和垂直速度全过程都有较为平滑的估计结果，相比之下 OF3 序列上“VIO+ 伪观测”在最后阶段出现了较明显的起伏， 因此其出现了 OF3 序列上本文方法的 RPEt 有所下降的现象，而 OF2 序列上本文方法的 RPEt 达到同类方法的最优。

5.3.4 消融实验

为了探究本文提出的 VIO 方法中各个模块的效果，本文进行了消融实验，消 融实验在 OF3 序列上进行，选择控制的变量包括“VIO”、“惯性-车体因子”、“伪 观测约束”、“车身状态判别”和“车身状态真值”，其中“VIO”是基础配置，在本 文中即为 VINS-Mono，其他变量的加入均是在 VIO 的基础上进行的；“惯性-车体 因子”表示是否在后端窗口优化时将惯性-车体对齐参数作为优化变量；“伪观测约 束”表示是否在后端窗口优化时加入含有伪观测约束的 IMU 残差； “车身状态判 别”表示是否根据车身状态判别神经网络的结果而调整后端优化量；“车身状态真 值”表示是否根据车身状态的真值而调整后端优化量，其可以看做是车身状态判 别的上限。实验结果如表 [5.9](#bookmark242)所示。表中展示了在不同消融实验下的 AT Et、RPEt 和 RPEr 的结果，其中 √ 表示使用了对应的模块。

表中第二行的含义是不论车身处于何种状态，均使用直行时的车身速度假设，

表 5.9 视觉惯性里程计消融实验

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| VIO | 惯性-车体 因子 | 伪观测 约束 | 车身状态 判别 | 车身状态 真值 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| √ √ √ √ √ | √ √ √ | √ √ √ √ | √ | √ √ | 20.552186 23.269810 17.193425 14.685789 17.361096 | 5.648737 5.594315 5.887755 5.277349 5.099256 | 3.411987 2.588779 2.543815 2.522826 2.577359 |

因此从前两行的对比中可以发现， 当使用了惯性车体因子和伪观测约束后，AT Et 产生了明显的下降，但是 RPEt 有了轻微的改善，RPEr 有了明显的改善。这说明 了惯性车体因子和伪观测约束的加入对于车身旋转的估计有着重要的影响，但是 这会对平移的估计产生一定负面影响。即说明如果全程使用直行时的车身速度假 设，会对里程计的平移估计精度产生负面影响，但对旋转的估计精度有积极影响。

表中第三行在第二行的基础上加入了对车身状态的预测， 虽然预测与真实情 况可能存在偏差，但是可以从数据上得出：当加入了车身状态判别后，AT Et 产生 明显的下降，RPEt 有所增加，RPEr 有轻微改善，这说明了车身状态判别对于绝 对定位精度的提升有着重要的影响。虽然其本身存在的状态预测误差则会对相对 位移的估计产生一定的负面影响，但是综合来看其对于绝对位置和旋转角度估计 的精度提升可以使其忽略对相对位移估计产生的负面影响。

表中第四行在第三行的基础上将车身状态的预测值替换为真值，这一行为对 AT Et 、RPEt 和 RPEr 三个指标都产生了增益效果，这侧面证明了： 当使用预测 值时产生的相对位移估计精度损失是由于预测值的误差导致的，而当使用真值时， 这一误差被消除，所以在 AT Et、RPEt 和 RPEr 三个指标上都取得了较大的进步。

表中最后一行在第四行的基础上去掉了惯性车体因子， 即在优化过程中不优 化惯性-车体因子，而仅使用初始量。这一行相较于第四行的结果是：AT Et和 RPEr 产生了一定程度的损失，但是 RPEt 却达到了最好效果。这一行的结果表明： 惯 性-车体因子的实时优化对于绝对位置精度和旋转估计有着重要的积极影响，但是 对于相对平移估计则有着负面影响。但是综合来看，优化惯性-车体因子对于绝对 位置和相对旋转的估计的提升明显，而对于相对位置估计精度的影响有限， 因此 有保留的必要。

5.4 紧耦合地图定位模块测试

紧耦合地图定位模块的测试在 KITTI 数据集和 4Seasons 数据集上进行，本节 将分别对两个数据集上的测试结果进行展示。其中 KITTI 数据集上的测试使用模 拟的多趟行驶数据，即彩色相机图像建图，灰度图像定位的模式，建图场景和测试 场景的变化较小； 而 4Seasons 数据集上的测试使用真实的多趟行驶数据，定位和 建图的场景变化较大，具有较强的挑战性。

5.4.1 精度测试

<5.4.1.1> KITTI 数据集测试

紧耦合地图定位模块的测试在两个数据集上进行， 分别是 KITTI 数据集和 4Seasons 数据集。在 KITTI 数据集上，本文的紧耦合地图定位模块和 VINS-Fusion、 VINS-Fusion-Map 、ORB-SLAM3 和 ORB—SLAM3-Map 进行对比， 其中 VINS- Fusion-Map 和 ORB—SLAM3-Map 是 VINS-Fusion 和 ORB-SLAM3 的地图版本。 测试结果如表 [5.10](#bookmark244)和图 [5.10](#bookmark208)所示，在表 [5.10](#bookmark244) 中以加粗字体和红底标出最佳表现。

从表中对 KITTI 数据集上多项指标的对比结果来看，不同方法在不同序列上 表现存在较大差异。对于绝对轨迹误差 AT Et 指标，本文方法在序列 00 、01 、02、 05 、06 、07 、08 和 10 上均取得了最低值，说明在绝对位姿精度上具有较明显的优 势； 仅在序列 04 上，ORB-SLAM3 表现更优，而序列 09 的结果略有争议，尽管 VINS 在此处显示出稍低的误差，但总体来看，本文方法在大部分序列中均展现出 更好的平稳性。

对于相对平移误差 RPEt 而言，各方法的表现相对分散。本文方法在序列 01、 02 、07 和 08 上表现突出，但在其他序列中，ORB-SLAM3 、ORB-SLAM3-Map 或 其他方法有时能获得更低的误差，表明在平移精度方面，各算法对场景和运动状 态的敏感性不同。

而在相对旋转误差 RPEr 方面， 本文方法在部分序列， 如 00 、07 、08 、09， 中取得了较低的值，体现出在角度估计上的一定优势，但在序列 01 、02 和 04 上， ORB-SLAM3 或 ORB-SLAM3-Map 的表现更佳，说明在旋转精度上，算法之间仍 存在细微差距。

综合来看，本文方法在大部分场景下均能保持较高的翻译精度， 并在旋转估 计上具有竞争力，这表明所提出的算法具有较好的鲁棒性和适应性。

表 5.10 KITTI 数据集多指标对比

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 指标 | 00 | 01 | 02 | 04 | 05 | 06 | 07 | 08 | 09 | 10 |
| ORB  ORBM  VINS  VINSM  本文方法 | AT Et ↓ | 3.51 7.19 5.95 3.92 | 4.92 10.11 6.49 8.01 | 5.70 10.32 5.99 13.08 | **0.27** | 1.84 1.54 5.98 3.88 | 1.68 2.35 3.48 1.79 | 0.96 1.67 2.41 0.65 | 3.32 5.36 3.85 9.89 | 3.37 7.61 | 0.90 0.95 3.78 3.63 |
| 0.98 1.08 1.23 0.53 |
| **1.78** |
| 6.17 1.87 |
| **0.97** | **4.09** | **0.70** | **1.24** | **1.56** | **0.21** | **1.68** | **0.82** |
| ORB  ORBM  VINS  VINSM  本文方法 | RPEt ↓ | 0.88 0.69 3.45 1.23 | 1.44 1.46 12.18 2.44 | 0.77 0.77 2.65 1.49 | **0.46** | 0.62 | 0.89 | 0.89 0.52 1.17 0.71 | 1.03 1.04 3.46 1.66 | **0.86** | 0.62 |
| **0.40** | **0.46** | **0.61** |
| 0.49 2.71 1.37 1.87 | 0.87 4.45 1.97 1.03 |
| 1.41 1.46 0.75 | 1.33 0.88 0.84 | 2.38 2.45 0.82 |
| **0.74** | **1.44** | **0.68** | **0.39** | **0.96** |
| ORB  ORBM  VINS  VINSM  本文方法 | RPEr ↓ | 0.31 0.26 1.45 0.73 | **0.19** **0.19** | **0.28** **0.28** | 0.19 | 0.26 | 0.27 | 0.50 0.30 0.87 0.57 | 0.31 0.31 1.77 0.65 | 0.25 0.25 1.87 0.59 | 0.29 |
| **0.08** | **0.16** | **0.15** | **0.28** |
| 3.54 0.48 0.34 | 1.26 0.56 0.30 | 0.62 0.69 0.30 | 0.69 0.67 0.19 | 0.72 0.58 0.17 | 0.94 0.99 0.30 |
| **0.27** | **0.26** | **0.29** | **0.19** |

a ORB 和 ORBM 分别表示 ORB-SLAM3 和 ORB-SLAM3-Map b VINS 和 VINSM 分别表示 VINS-Fusion 和 VINS-Fusion-Map

<5.4.1.2> 综合 (Office Loop) 场景测试

4Seasons 数据集实在同一条线路上多次采集获得的，采集的时间跨度从几十 分钟到一年，所以其场景相对于 KITTI 数据集有更多的变化，包括季节、天气、晨 昏等。本文首先在 Office Loop 的一般性场景下进行测试。一般性场景测时的建图 序列为 2020.03.24 17:36 拍摄，而定位序列为 2020.03.24 17:45 拍摄，两个场景差 异较小，因此可以同时比较多个方法的表现。

测试结果如表 [5.11](#bookmark246)所示，其中包括了 VINS-Mono、VINS-Fusion w/ RTK、ORB- SLAM3、VLS[[43](#bookmark301)](2022)、GPLD[[38](#bookmark296)](2023)、[MVIL[80](#bookmark357)](2023) 等方法与本文方法的对 比，其中 VINS-Fusion w/ RTK 使用了高精度 (厘米级)GNSS 信号，而 VLS、GPLD、 MVIL 则是近期涌现的三种基于地图的定位方法。

首先在 OF1 和 OF2 两个序列上进行实验，这两个序列由于拍摄时间接近，所以 场景变化不大，因此在其中进行的实验还包括了 VINS-Monow/ Map1、VINS-Mono

第 5 章 视觉惯性定位系统搭建与测试

真值 本文方法 ORB-SLAM3 ORB-SLAM3-Map VINS-Fusion VINS-Fusion-Map

KITTI序列 01 定位轨迹

KITTI序列 00 定位轨迹

500 -

|  |
| --- |
|  |

|  |
| --- |
|  |

0

400 -

-200

-400

Y(m)

Y(m)

300 -

-600

200 -

-800

100 -

-1,000

0 -

-1,200

-200 -100 0 100 200 300

0 500 1,000 1,500

X(m) X(m)

KITTI序列 08 定位轨迹

KITTI序列 07 定位轨迹

400 -

|  |
| --- |
|  |

100 -

|  |
| --- |
|  |

300 -

50 -

Y(m)

Y(m)

200 -

0 -

100 -

-50 -

0 -

-100 -

-400 -200 0 200 400

-150 -100 -50 0

X(m)

X(m)

KITTI序列 05 定位轨迹

300 -

|  |
| --- |
|  |

200 -

Y(m) Y(m)

100 -

0 -



-200 -100 0 100 200

X(m)

KITTI序列 09 定位轨迹

500 -

|  |
| --- |
|  |

400 -

300 -

200 -

100

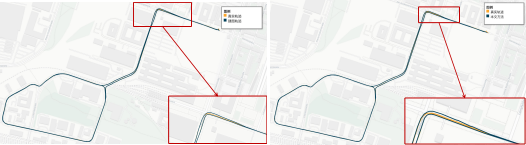
0 -

-100 0 100 200 300

X(m)

图 5.10 KITTI 数据集测试轨迹示意图



(a) OF1 序列建图轨迹 (b) OF2 序列定位轨迹

图 5.11 4Seasons 数据集定位结果及误差分析

w/ Map2 、ORB-SLAM3 w/ Map3 等方法的对比。从表 [5.11](#bookmark246) 中可以看出，本文方法 在 AT Et 和 RPEt 上均取得了最佳表现，而在 RPEr 上略逊于 VLS 方法，此外本 文方法是第一个将定位精度提升到 1 米以内的方法，本文将其在卫星图上的定位 结果展示如图 [5.11(b)](#bookmark210) 所示。

表 5.11 4Seasons 数据集 Office Loop 场景第 2 序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| 离线建图 | 03.24 17:36 (OF1) | 春季 | 晴朗 | 下午 | 0.814881 | 0.929358 | 1.014480 |
| VINS-Mono  VINS-Mono w/ Map1 VINS-Monow/ Map2 ORB-SLAM3  ORB-SLAM3 w/ Map3 VINS-Fusion w/ RTK  VLS  GPLD  MVIL  本文方法 | 03.24 17:45 (OF2) | 春季 | 晴朗 | 下午 | 7.878626 3.092430 2.440850 15.621815 19.249121 1.801669 1.715343 8.720000 6.450000 | 2.330876  N/A  N/A  3.222846  N/A  1.500405  1.630718  N/A  N/A | 0.783557  N/A  N/A  0.890233  N/A  1.085638 |
| **0.668220** |
| N/A  N/A  0.759267 |
| **0.777091** | **1.124076** |

a Map1 表示使用 VINS-Mono 建图结果 b Map2 表示使用本文离线建图结果

c Map3 表示使用 ORB-SLAM3 建图结果

d RTK 表示使用了RTK-GPS 数据，其定位精度达到厘米级

在图 [5.11(b)](#bookmark210) 中可以看到在放大区域有着明显的定位误差，但这并非由于定位 方法引入，而是由建图误差引入： 如图 [5.11(a)](#bookmark210)所示，可以看到在定位误差较大的 区域，建图时本身就存在明显的偏差，这导致了定位过程中引入了这一误差。因 此，本文的定位方法在 4Seasons 数据集上的表现是受到了建图质量的影响，但是 相较于其他方法，本文的定位方法在相同的建图条件下取得了更好的定位精度。

在 Office Loop 的环境下还有更多的具有挑战性的序列，本文在这些序列上也 进行了测试。这些序列包括了 OF3 、OF4 和OF5，具体的场景变化如图 [5.12](#bookmark212)所示，

这些序列的场景相较于 OF2，有了晨昏 [5.12(b)](#bookmark212)，季节 [5.12(c)](#bookmark212)和 [5.12(d)](#bookmark212)，天气 [5.12(d)](#bookmark212) 的变化。定位的结果如表 [5.12](#bookmark248)所示， 需要说明的是， 由于从 OF3 开始场景出现了 较大的变化，因此一些使用传统方法的地图定位方法无法再依靠地图信息运行，因



(a) OF1 序列建图场景



(b) OF3 序列清晨场景 (c) OF4 序列夏季场景 (d) OF5 序列冬季场景

图 5.12 4Seasons 数据集 OF3-5 场景差异展示

此此处仅汇报了可以运行的方法。

具体分析此场景下的不同序列可以看出，OF3 和 OF1 两个序列的分别是在下 午和清晨拍摄， 所以图片中的光照强度首先就有一定的不同， 此外图 [5.12(a)](#bookmark212) 和 图 [5.12(b)](#bookmark212)中道路的车况也有一定区别，在图 [5.12(b)](#bookmark212) 中出现了建图过程未出现的车 辆，为定位的细粒度特征匹配增加了较大困难。OF4 和 OF1 两个序列存在季节的 变化，从图 [5.12(a)](#bookmark212)和图 [5.12(c)](#bookmark212) 中可以看到，OF4 中的场景中出现了更繁茂的植被， 甚至遮盖了部分建筑物，这对于细粒度特征匹配和定位都带来了一定的困难。OF5 和 OF1 两个序列存在天气的变化，从图 [5.12(a)](#bookmark212) 和图 [5.12(d)](#bookmark212) 中可以看到，OF5 中 的场景中出现了降雪， 改变了整个场景的整体感知条件，这对于定位的粗粒度和 细粒度匹配都带来的挑战。

从表中信息可以看出： 本文方法在所有序列的 AT Et 和 RPEt 指标上均取得 了明显的优势，是所有方法中唯一误差稳定在 1 米以内的方法；在 OF5 上的 RPEr 指标上略逊于 VLS 方法，但是差距极小。

在表 [5.11](#bookmark246)和表 [5.12](#bookmark248) 中还可以发现本文方法在地图本身具有误差的情况下，定 位结果却出现了部分超越地图精度的情况。这一现象结合后文的消融实验可以看 到本文将地图建模为概率地图的做法具有一定作用：在地图中引入了不确定性后， 紧耦合优化可以更好地利用地图信息，从而提升定位精度。

<5.4.1.3> 低结构化 (Neighborhood) 场景测试

Neighborhood 场景中有更多的植被，接近于农村环境，因此在这个场景下的测 试能表现出本文方法在较低纹理环境下的表现。在此场景下又增加了几种 Office Loop 场景中没有出现的条件，例如降雨天气 (NH3) 和傍晚环境 (NH7)。两种情况 如图 [5.13](#bookmark214) 所示，左图的降雨条件可以看到明显的水雾出现在图像上，而右图的傍

表 5.12 4Seasons 数据集 Office Loop 场景部分序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et | RPEt | RPEr |
| 离线建图 | 03.24 17:36 (OF1) | 春季 | 晴朗 | 下午 | 0.814881 | 0.929358 | 1.014480 |
| VINS-Mono ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 04.07 10:20 (OF3) | 春季 | 晴朗 | 清晨 | 10.664015 6.288431 2.000104 2.913034 | 5.305404 5.000592 3.678773 3.978592 | 5.305404 5.000592 3.678773 3.519618 |
| **0.938238** | **2.947031** | **2.529328** |
| VINS-Mono ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 06.12 10:10 (OF4) | 夏季 | 晴朗 | 清晨 | 7.661955 5.073350 1.809687 4.515426 | 3.447067 3.207118 2.158715 8.142403 | 1.747575 1.760357 1.577418 2.389229 |
| **0.792201** | **1.402346** | **1.047381** |
| VINS-Mono ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 01.07 12:04 (OF5) | 冬季 | 降雪 | 下午 | 8.083566 5.520636 2.448607 1.862401 | 2.598062 1.878826 1.333865 1.835533 | **0.990526** |
| 1.000615 0.934714 0.999361 0.999785 |
| **0.767040** | **1.251208** |

晚环境，可以看到明显的亮度降低。



NH3 序列降雨场景 NH7 序列傍晚场景

图 5.13 4Seasons 数据集 NH3 与 NH7 场景差异展示

测试结果如表 [5.13](#bookmark250)和表 [5.14](#bookmark252)所示，其中本文方法在所有序列的 AT Et和 RPEt 指标上取得了最好的效果，但是在 RPEr 依旧没有比其他方法有明显优势，但是 所有方法在该指标上的差距都非常小。

在此场景下由于离线建图的精度较高， 因此本文方法在定位过程中能够更好 地利用地图信息，从而提升定位精度，AT Et 指标下都获得了 30 厘米左右的误差， 已经达到了厘米级的定位精度。但是此场景下，本文定位方法并没有出现超越地 图精度的情况，这主要是受限于本文定位方法的精度上限，这一上限也将在后文 的消融实验中进行验证。

表 5.13 4Seasons 数据集 Neighborhood 场景第 2 与第 3 序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| 离线建图 | 03.26 13:32 (NH1) | 春季 | 多云 | 下午 | 0.133943 | 0.400299 | 0.451671 |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 10.07 14:47 (NH2) | 秋季 | 多云 | 下午 | 3.896760 0.584860 0.385877 | 1.177058 2.081065 | **0.957432** |
| 1.764747 1.005846 1.022361 |
| **1.015101** |
| **0.325610** |
| 1.079635 |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 10.07 14:53 (NH3) | 秋季 | 降雨 | 下午 | 1.074461 0.557219 0.338285 | 1.193234 1.922972 1.033000 | **0.980104** |
| 1.199615 1.002556 1.006227 |
| **0.323113** | **1.006227** |

由于 NH5 和 NH6 所代表的场景已经出现在了表[5.13](#bookmark250)和表 [5.14](#bookmark252) 中， 因此不再 重复描述其具体数值，而是以 APEt 的云雨图 [5.14](#bookmark216)和图 [5.15](#bookmark218)形式展示不同方法的 定位结果。云雨图图中展示了不同方法在 NH5 和 NH6 两个序列上的定位误差分

表 5.14 4Seasons 数据集 Neighborhood 场景第 4 与第 7 序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| 离线建图 | 03.26 13:32 (NH1) | 春季 | 多云 | 下午 | 0.133943 | 0.400299 | 0.451671 |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 12.22 11:54 (NH4) | 冬季 | 多云 | 清晨 | 3.326053 0.584881 0.336795 | 1.721227 1.273095 0.828517 | 1.014069 1.337608 |
| **0.722960** |
| **0.303237** | **0.787774** |
| 0.793884 |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 05.10 18:32 (NH7) | 春季 | 多云 | 傍晚 | 1.770878 1.098182 0.304587 | 1.461675 3.455674 1.125579 | 1.239126 2.230523 1.007488 |
| **0.291957** | **1.033805** | **1.007318** |

布，三条横线由高到低分别表示“最大值”，“中位数”和“最小值”，可以看到本 文方法在这两个序列上的定位误差分布更加集中，且中位数更低，这说明本文方 法在这两个序列上的定位精度更高且更稳定。

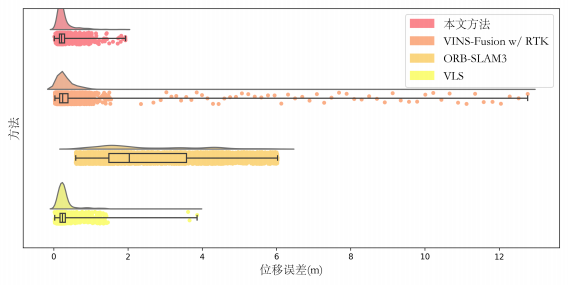


图 5.14 不同定位方法在 OF5 序列上的定位误差分布

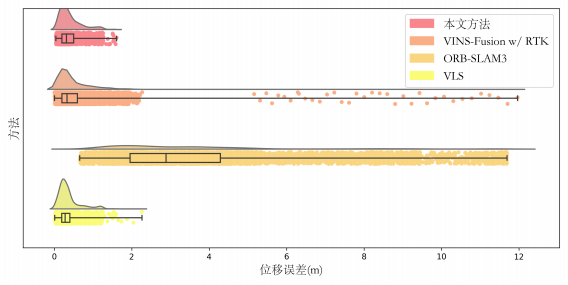


图 5.15 不同定位方法在 OF6 序列上的定位误差分布

<5.4.1.4> 复杂 (Business Campus 与 Old Town) 场景测试

Business Campus 场景在校园环境中，Old Town 场景在城市中心区，这两个场 景的有着更多的建筑物和道路，但也有着更多的移动物体和复杂的动态场景，如 图 [5.16(a)](#bookmark220)和图 [5.16(c)](#bookmark220)所示， 因此在这个场景下的测试能够更好地表现出本文方法 在复杂环境下的表现。此外，在场景感知条件变化方面，这两个场景也覆盖了多 种季节和天气条件，在 Old Twon 的建图序列中还有降雨天气，如图[5.16(b)](#bookmark220) 所示， 也可以反映出本文方法在较差天气建图条件下的定位表现。



(a) BC2 序列复杂场景 (b) OT1 序列降雨场景 (c) OT2 序列复杂场景

图 5.16 4Seasons 数据集 NH3 与 NH7 场景差异展示

在两个场景下的对比如表 [5.15](#bookmark254) 和表 [5.16](#bookmark256)所示， 需要注意的是， 由于 Business Campus 场景中的建图轨迹在进行 RPEt 和 RPEr 指标计算的时候，对齐过程数值 不稳定产生了不合理结果，因此表中只给出了建图轨迹的 AT Et 指标，可以看到该 指标非常小，足以说明建图轨迹较为精确， 因此另外两个指标的缺失也是可以接 受的。本文方法依旧在 AT Et和 RPEt 上取得了最佳表现，但是在 RPEr 上表现不 够稳定，但是与其他方法差距非常小。此外， 在这两个场景下，本文方法在定位精 度上依旧取得了 1 米以内的误差，这不仅说明本文方法在复杂场景下的表现优秀， 也说明了本文方法具有较强的稳定性。

表 5.15 4Seasons 数据集 Business Campus 场景部分序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| 离线建图 | 10.08 09:30 (BC1) | 秋季 | 晴朗 | 清晨 | 0.095337 | N/A | N/A |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 01.07 13:12 (BC2) | 冬季 | 降雪 | 下午 | 7.376041 1.115755 1.092215 | 1.668179 2.313352 1.938506 | **1.313357** |
| 2.588801 1.367412 1.385713 |
| **0.639282** | **1.576658** |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 02.25 14:16 (BC3) | 冬季 | 晴朗 | 下午 | 7.372700 1.076147 0.495447 | 1.760015 2.596570 2.025172 | 1.202648 2.926574 1.786391 |
| **0.446566** | **1.380626** | **1.109240** |

在 Business Campus 场景中，由于离线建图模块的精度较高，因此本文方法在 定位过程中能够更好地利用地图信息，从而提升定位精度，AT Et 指标下都获得了 50 厘米左右的误差，已经有了较高的定位精度。

表 5.16 4Seasons 数据集 Old Town 场景第 2 序列指标对比

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 时间 | 季节 | 天气 | 晨昏 | AT Et ↓ | RPEt ↓ | RPEr ↓ |
| 离线建图 | 10.08 11:53 (OT1) | 冬季 | 降雨 | 清晨 | 1.016104 | 0.456781 | 1.116236 |
| ORB-SLAM3  VINS-Fusion w/ RTK VLS  本文方法 | 02.25 12:24 (OT2) | 冬季 | 晴朗 | 清晨 | 38.930786 0.965686 1.143163 | 4.441040 2.682027 2.318242 | **1.078154** |
| 1.172853 1.443211 1.521039 |
| **0.846396** | **2.318242** |

但是在 Old Town 场景中， 由于建图序列的天气情况较差，所以轨迹的精度较 低，而在这种条件下本文的定位方法依旧达到了 1 米以内的定位精度，这说明本 文方法对于建图质量的要求并不十分苛刻，在较差的建图精度下还可以通过定位 给出更加精确的结果。

5.4.2 消融实验

为了验证紧耦合地图定位中的各个模块对定位精度的影响，本文进行了一系 列的消融实验。在这些实验中， 本文逐步增加紧耦合地图定位中的各个模块，从而 验证了这些模块对定位精度的影响。具体实验结果如表 [5.17](#bookmark258) 所示，表中主要研究 了“VIO”、“粗到细定位”、“转换矩阵更新”、“相邻帧边”、“地图关键帧边”、“地 图观测边”、“地图点先验边”和“真值地图”这 8 个变量对紧耦合地图定位的影 响，实验设置在 Office Loop 的第 3 条序列 (OF3) 上，以 AT Et 为主要报告指标。

表 5.17 消融实验结果对比

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| VIO | 粗到细 定位 | 转换矩阵 更新 | 相邻帧 边 | 地图  关键帧边 | 地图 观测边 | 地图点 先验边 | 真值 地图 | AT Et |
| √  √ √ √ √ √ √ √ | √ √ √ √ √ √ √ √ | √ √ √ √ √ √ | √ √ √ √ √ | √ √ √ √ | √ √ √ | √ √ | √ | 17.193425  >60  6.987066  1.898881  1.445279  1.237693  1.223113  0.938238  0.396039 |

a 实验构建在 Office Loop 的第 3 条序列 (OF3) 上

b 下划线 表示该序列在后期因为转换矩阵未更新而粗到细定位失败

在表 [5.17](#bookmark258) 中只有“粗到细定位”时产生了较大的定位误差，这是因为在 VIO 和转换矩阵为粗到细定位提供先验位置信息的时候，粗到细定位的粗定位仅使用 粗粒度特征产生了误匹配，从而将整个定位误差增长到了一个难以接受的程度。在 第三行， 当使用了 VIO 和转换矩阵后，这一定位误差极大得缩小到了一个可以接 受的程度，这说明本文使用的转换矩阵具有非常重要的作用。但是在仅有转换矩 阵而不维护的情况下，在序列的后段， 由于 VIO 的累积误差越来越大，使得转换 矩阵不再能准确反映当前 VIO 结果和全局坐标系的正确关系，此时粗到细定位的 先验位置信息也失去了意义，所以定位失败， 因此表中的数据以下划线表示其是 不完整序列的定位数据。在加入了转换矩阵更新之后，估计不完整的问题得到了 解决，定位精度得到了进一步的提升。

在第四行到第七行中，逐步加入了相邻帧边、地图关键帧边、地图观测边和地

图点先验边，可以看到这些边的加入都使得定位精度得到了进一步的提升，这说 明这些边对于定位精度的提升都是有益的。这表明本文所提出的紧耦合优化中的 各个残差边都是对定位结果有正向作用的。

在最后一行中，加入了真值地图，可以看到定位精度得到了进一步的提升，这 说明真值地图对于定位精度的提升是有益的。在使用了真值地图后， 定位精度达到 了 30 厘米左右，这可以看作是本文方法的定位精度上限，而这也与 4Seasons 数据集 中实验情况相符：在建图精度不足的情况下，本文方法的精度上限受到地图精度的 限制 (例如 Office Loop 场景和 Old Town 场景)，但是在建图精度较高的情况下，本 文方法的定位精度上限只由紧耦合地图定位模块的精度决定 (例如 Neighborhood 场景和 Business Campus 场景)。

5.5 本章总结

本章主要以实验验证的方式对本文所提出的固定路线定位方法进行了评估。 首先，本章对实验实施的平台进行了介绍，包括实验所使用的数据集、实验所使 用的硬件平台和软件平台以及实验的评估指标包括绝对轨迹误差和相对姿态误差。 然后本章分别针对 3 个子模块的性能进行了测试，并与一些常规方法进行对比，分 析了本文方法的优势所在以及不足之处，在重要模块中还增加了消融实验来详细 验证模块内组成部分对整体精度的影响。最后，本章对实验结果进行了总结，验 证了本文方法的有效性和鲁棒性。

第 6 章 总结与展望

6.1 论文总结

固定路线中的视觉惯性定位方法是一项在自动驾驶、工业巡检等领域有着广 泛需求的技术。但是当前的定位技术依赖于高成本的高精地图或者高精度传感器， 这都严重制约了其在大规模应用中的推广。为了解决这一问题，本文提出了一种 基于视觉惯性的固定路线定位方法，该方法不依赖于高精地图和昂贵的传感器，在 建图阶段只需要图像和高精度的 GNSS 信息，在定位阶段只需要使用低成本的相 机和 IMU 传感器， 即可完成整个从建图到定位的流程。在建图到定位的整个流程 中，本文需要解决的问题包括：(1) 如何利用图像和 GNSS 信息构建具有可接受精 度的地图； (2) 如何提升 IMU 和相机信息进行较高精度的位姿估计； (3) 如何使用 地图和位姿估计进行更精确的定位任务。针对这三个问题，本文提出了一套完整 的固定路线中的视觉惯性定位方法，包括：

(1) 基于 SfM 和 GNSS 融合的离线建图。为了克服以往 SLAM+GNSS 的建图 过程全局优化不足的问题，本文提出了一种基于 SfM 和 GNSS 融合的先验地图构 建方法。首先通过图片和 GNSS 信息进行预处理，主要是筛选出候选关键帧以及 进行语义分割等操作； 紧接着纯图像首先进行 SfM 以获得在视觉建图世界坐标系 下的相机位姿和地图点云； 然后使用 GNSS 信息和 SfM 结果进行建图对齐， 即实 现视觉建图世界坐标系和全局坐标系的对齐； 此后进行建图融合， 以非线性优化 的方式同时优化视觉重投影误差和 GNSS 提供的位置误差； 最后输出分层次的的 地图。

(2) 基于车身运动学假设的视觉惯性里程计。为了解决以往通用视觉惯性里程 计方法在车辆或轮式机器人上的速度估计不合理问题，本文提出了一种基于车身 运行学假设的视觉惯性里程计。在通用 VIO 的基础上，本文引入了车身横向和垂 直速度的 0 状态估计，依据 IMU 的序列数据和基于深度学习的状态预测来检测车 身状态变化，在特定状态下完成惯性-车体对齐并加入伪观测约束，从而获得更精 确的位姿估计。

(3) 基于最大后验概率的紧耦合地图定位方法。为了能够充分使用地图观测信 息和视觉惯性里程计信息，本文提出了一种基于最大后验概率的紧耦合地图定位 方法。整个模块首先通过比较图像观测与地图信息，利用粗粒度的图像特征和位 置先验信息确定初步位置范围，再通过细粒度的像素特征匹配获得更精确的相机 位姿。同时，通过初始化转换对齐视觉惯性里程计和地图观测， 系统也会验证地

图定位结果的有效性。最终，通过紧耦合的非线性优化方法融合粗到细定位、视 觉惯性里程计输出和先验地图信息，获得精细的相机位姿， 并通过不断更新转换 参数来防止误差累积。

本文在公开数据集上对提出的方法进行了大量的实验验证，实验结果表明，本 文提出的固定路线中的视觉惯性定位方法在定位精度和鲁棒性上均优于当前主流 的视觉惯性定位方法，具有较好的实用性和推广性。

6.2 未来展望

在未来，本文考虑在以下几个方向进行进一步的研究：

(1) 当前的研究中使用地图作为重要的观测来源，但是每次使用时并不会对地 图进行更新，这在长时间的使用中可能会导致地图的过时。因此，未来可以考虑 在定位过程中对地图进行增量式更新，以提高地图的实时性和准确性。

(2) 本文在研究过程中发现动态物体对定位精度有较大的影响， 因此未来可以 考虑在定位过程中对动态物体进行检测和处理，以提高定位的鲁棒性。

(3) 基于深度学习的状态检测方法在本文的视觉惯性里程计部分发挥了重要作 用，但是这种使用方法过于简单，未来考虑在更多的环节加入深度学习技术，以提 高定位的精度和鲁棒性。

参考文献

[1] Hall S. Driverless buses are arriving soon in these 3 european cities[EB/OL]. (2023-1-27) [2023-1-27]. <https://www.weforum.org/stories/2023/01/autonomous-buses-geneva-project/>.

[2] 新出行团队. 停车场里的“通勤模式”？阿维塔 11 AVP 代客泊车[EB/OL]. (2023-06-20) [2023-06-20]. <https://www.xchuxing.com/video/5451>.

[3] 宋为伟. 福建实施 1000 千伏变电站无人机多机协同自主巡检[EB/OL]. (2020-09-18) [2020-09-18]. <http://fj.people.com.cn/n2/2020/0918/c181466-34300912.html>.

[4] 李金明陆成宽. 国内首台矿山无人运输机器人投运[EB/OL]. (2023-11-13)[2023-11-13]. <http://fj.people.com.cn/n2/2020/0918/c181466-34300912.html>.

[5] Zumberge J, Heflin M, Jefferson D, et al. Precise point positioning for the efficient and robust analysis of gps data from large networks[J]. Journal of geophysical research: solid earth, 1997, 102(B3): 5005-5017.

[6] Fotopoulos G, Cannon M. An overview of multi-reference station methods for cm-level posi- tioning[J]. GPS solutions, 2001, 4: 1-10.

[7] Bisnath S, Aggrey J, Seepersad G, et al. Innovation: Examining precise point positioning now and in the future[J]. GPS World. March, 2018.

[8] LiX, Huang J,LiX, et al. Review of ppp–rtk: Achievements, challenges, and opportunities[J]. Satellite navigation, 2022, 3(1): 28.

[9] 姜英明, 刘继州, 赵青, 等. 不同服务商网络 RTK 定位性能评估[J]. 城市勘测, 2025(01): 63-66.

[10] 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.

[11] Klein G, Murray D. Parallel tracking and mapping for small ar workspaces[C]//2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE, 2007: 225-234.

[12] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis[C]// Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings. Springer, 2000: 298-372.

[13] Mur-Artal R, Montiel J M M, Tardos J D. Orb-slam: a versatile and accurate monocular slam system[J]. IEEE transactions on robotics, 2015, 31(5): 1147-1163.

[14] Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system formonocular, stereo, and rgb-d cameras[J]. IEEE transactions on robotics, 2017, 33(5): 1255-1262.

[15] Lynen S, Achtelik M W, Weiss S, et al. A robust and modular multi-sensor fusion approach applied to mav navigation[C]//2013 IEEE/RSJ international conference on intelligent robots and systems. IEEE, 2013: 3923-3929.

[16] Falquez J M, Kasper M, Sibley G. Inertial aided dense & semi-dense methods for robust di- rect visual odometry[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016: 3601-3607.

[17] Mourikis A I, Roumeliotis S I. A multi-state constraint kalman filter for vision-aided inertial navigation[C]//Proceedings 2007 IEEE international conference on robotics and automation. IEEE, 2007: 3565-3572.

[18] Li M, Mourikis A I. High-precision, consistent ekf-based visual-inertial odometry[J]. The International Journal of Robotics Research, 2013, 32(6): 690-711.

[19] SunK, MohtaK,Pfrommer B, et al. Robust stereo visual inertial odometry for fast autonomous flight[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 965-972.

[20] BloeschM, BurriM, OmariS, et al. Iterated extended kalman filter based visual-inertial odome- try using direct photometric feedback[J]. The International Journal of Robotics Research, 2017, 36(10): 1053-1072.

[21] Forster C, Carlone L, Dellaert F, et al. On-manifold preintegration for real-time visual–inertial odometry[J]. IEEE Transactions on Robotics, 2016, 33(1): 1-21.

[22] Leutenegger S, Lynen S, Bosse M, et al. Keyframe-based visual–inertial odometry using non- linear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334.

[23] Harris C, Stephens M, et al. A combined corner and edge detector[C]//Alvey vision conference: Vol. 15. Citeseer, 1988: 10-5244.

[24] Leutenegger S, Chli M, Siegwart R Y. Brisk: Binary robust invariant scalable keypoints[C]//

2011 International conference on computer vision. Ieee, 2011: 2548-2555.

[25] Qin T, LiP, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator [J]. IEEE transactions on robotics, 2018, 34(4): 1004-1020.

[26] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//IJCAI’81: 7th international joint conference on Artificial intelligence: Vol. 2. 1981: 674-679.

[27] Gálvez-López D, Tardos JD. Bags of binary words for fast place recognition in image sequences [J]. IEEE Transactions on robotics, 2012, 28(5): 1188-1197.

[28] Liu H, Chen M, Zhang G, et al. Ice-ba: Incremental, consistent and efficient bundle adjust- ment for visual-inertial slam[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1974-1982.

[29] Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[30] Jeong J,ChoY, Kim A. Hdmi-loc: Exploiting high definition map image for precise localization via bitwise particle filter[J]. IEEE Robotics and Automation Letters, 2020, 5(4): 6310-6317.

[31] Xiao Z, Yang D, Wen T, et al. Monocular localization with vector hd map (mlvhm): A low-cost method for commercial ivs[J]. Sensors, 2020, 20(7): 1870.

[32] Guo C, Lin M, Guo H, et al. Coarse-to-fine semantic localization with hd map for autonomous driving in structural scenes[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021: 1146-1153.

[33] Stewart A D, Newman P. Laps-localisation using appearance of prior structure: 6-dof monoc- ular camera localisation using prior pointclouds[C]//2012 IEEE International Conference on Robotics and Automation. IEEE, 2012: 2625-2632.

[34] Zuo X, Geneva P, Yang Y, et al. Visual-inertial localization with prior lidar map constraints[J]. IEEE Robotics and Automation Letters, 2019, 4(4): 3394-3401.

[35] Lin X, Wang F, Yang B, et al. Autonomous vehicle localization with prior visual point cloud map constraintsin gnss-challenged environments[J]. Remote Sensing, 2021, 13(3): 506.

[36] Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.

[37] SurberJ, Teixeira L, Chli M. Robust visual-inertial localization with weak gps priors for repet- itive uav flights[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 6300-6306.

[38] Hao Y, Liu J, Liu Y, et al. Global visual–inertial localization for autonomous vehicles with pre-built map[J]. Sensors, 2023, 23(9): 4510.

[39] Qin V. Colmap with gps position prior[J/OL]. GitHub repository, 2022. [https://github.com/V](https://github.com/Vincentqyw/colmap-gps) [incentqyw/colmap-gps](https://github.com/Vincentqyw/colmap-gps).

[40] Sarlin P E, Cadena C, Siegwart R, et al. From coarse to fine: Robust hierarchical localization at large scale[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 12716-12725.

[41] Arandjelovic R, Gronat P, ToriiA, et al. Netvlad: Cnn architecture for weakly supervised place recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recogni- tion. 2016: 5297-5307.

[42] DeTone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detec- tion and description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 224-236.

[43] Yang Y, Zhang X, Gao S, et al. A real-time fusion framework for long-term visual localization [A]. 2022.

[44] Lin Y, Liu L, Liang X, et al. Visual localization based on multiple maps[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023: 3306-3313.

[45] Xue F, Wu X, CaiS, et al. Learning multi-view camera relocalization with graph neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 11372-11381.

[46] Wang S, Kang Q, She R, et al. Robustloc: Robust camera pose regression in challenging driv- ing environments[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 37. 2023: 6209-6216.

[47] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.

[48] Xie E, Wang W, Yu Z, et al. Segformer: Simple and efficient design for semantic segmentation with transformers[C]//Neural Information Processing Systems (NeurIPS). 2021.

[49] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene un- derstanding[C]//Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[50] Wexler Y, Fitzgibbon A W, Zisserman A. Learning epipolar geometry from image sequences [C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,

2003. Proceedings.: Vol. 2. IEEE, 2003: II-209.

[51] Black M J, Anandan P. The robust estimation of multiple motions: Parametric and piecewise- smooth flow fields[J]. Computer vision and image understanding, 1996, 63(1): 75-104.

[52] Levenberg K. A method for the solution of certain non-linear problems in least squares[J]. Quarterly of applied mathematics, 1944, 2(2): 164-168.

[53] Subirana J S, Zornoza J J, Hernández-Pajares M. Transformations between ecefand enu coor- dinates[J]. Technical University of Catalonia, Spain, 2011.

[54] Horn BK. Closed-form solution of absolute orientation using unit quaternions[J]. Josa a, 1987, 4(4): 629-642.

[55] Hu h, Zhu F, Zhang X. An adaptive method to detect vehicle motion state using mems-imu [J/OL]. Journal of Navigation and Positioning, 2020(005): 008. DOI: [10.3969/j.issn.2095-499](https://doi.org/10.3969/j.issn.2095-4999.2020.05.002) [9.2020.05.002](https://doi.org/10.3969/j.issn.2095-4999.2020.05.002).

[56] Brossard M, Barrau A, Bonnabel S. Ai-imu dead-reckoning[J/OL]. IEEE Transactions on Intelligent Vehicles, 2020, 5(4): 585-595. DOI: [10.1109/TIV.2020.2980758](https://doi.org/10.1109/TIV.2020.2980758).

[57] Huang F, Yi B, Wang X, et al. Vehicle inertial navigation method based on deep learning and motion constraints[J/OL]. Journal of Chinese Inertial Technology, 2022, 30(5): 569-575. DOI: [10.13695/j.cnki.12-1222/o3.2022.05.002](https://doi.org/10.13695/j.cnki.12-1222/o3.2022.05.002).

[58] Dai J S. Euler–rodrigues formula variations, quaternion conjugation and intrinsic connections [J]. Mechanism and Machine Theory, 2015, 92: 144-152.

[59] Kingma D P. Adam: A method for stochastic optimization[A]. 2014.

[60] Forster C, CarloneL, Dellaert F, et al. Imupreintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation[M]. 2015.

[61] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.

[62] Rublee E, Rabaud V, Konolige K, et al. Orb: An efficient alternative to sift or surf[C]//2011 International conference on computer vision. Ieee, 2011: 2564-2571.

[63] Sarlin P E, DeTone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 4938-4947.

[64] Lepetit V, Moreno-Noguer F, Fua P. Ep n p: An accurate o (n) solution to the p n p problem[J]. International journal of computer vision, 2009, 81: 155-166.

[65] Gao X S, Hou X R, Tang J, et al. Complete solution classification for the perspective-three- point problem[J]. IEEE transactions on pattern analysis and machine intelligence, 2003, 25(8): 930-943.

[66] BourbakiN. Lie groups and lie algebras: Chapters 1-3: Vol. 1[M]. Springer Science & Business Media, 1989.

[67] Geiger A, Lenz P, UrtasunR. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//Conference on Computer Vision and Pattern Recognition (CVPR). 2012.

[68] Wenzel P, Wang R, Yang N, et al. 4seasons: A cross-season dataset for multi-weather slam in autonomous driving[C]//Pattern Recognition: 42nd DAGM German Conference, DAGM GCPR 2020, Tübingen, Germany, September 28–October 1, 2020, Proceedings 42. Springer, 2021: 404-417.

[69] Jeong J, Cho Y, Shin Y S, et al. Complex urban dataset with multi-level sensors from highly diverse urban environments[J]. International Journal of Robotics Research, 2019, 38(6): 642- 657.

[70] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of rgb-d slam systems [C]//2012 IEEE/RSJ international conference on intelligent robots and systems. IEEE, 2012: 573-580.

[71] Geiger A, Lenz P, UrtasunR. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 3354- 3361.

[72] Arun K S, Huang T S, Blostein S D. Least-squares fitting of two 3-d point sets[J]. IEEE Transactions on pattern analysis and machine intelligence, 1987(5): 698-700.

[73] Li B, Zou D, Huang Y, et al. Textslam: Visual slam with semantic planar text features[M]// IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2023.

[74] Grupp M. evo: Python package for the evaluation of odometry and slam.[EB/OL]. 2017. <https://github.com/MichaelGrupp/evo>.

[75] Qin T, Pan J, Cao S, et al. A general optimization-based framework for local odometry estima- tion with multiple sensors[A]. 2019. arXiv:1901.03638.

[76] Von Stumberg L, Usenko V, Cremers D. Direct sparse visual-inertial odometry using dynamic marginalization[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 2510-2517.

[77] Leutenegger S. Okvis2: Realtime scalable visual-inertial slam with loop closure[A/OL]. 2022. arXiv: [2202.09199](https://arxiv.org/abs/2202.09199). <https://arxiv.org/abs/2202.09199>.

[78] UsenkoV, DemmelN, Schubert D, et al. Visual-inertial mapping with non-linear factor recovery [J]. IEEE Robotics and Automation Letters, 2019, 5(2): 422-429.

[79] Geneva P, Eckenhoff K, Lee W, et al. OpenVINS: A research platform for visual-inertial esti- mation[C/OL]//Proc. of the IEEE International Conference on Robotics and Automation. Paris, France, 2020. <https://github.com/rpng/open_vins>.

[80] Zhang Z, Jiao Y, Huang S, et al. Map-based visual-inertial localization: Consistency and com- plexity[J]. IEEE Robotics and Automation Letters, 2023, 8(3): 1407-1414.

附录 A 补充内容

A.1 四元数与旋转矩阵转换

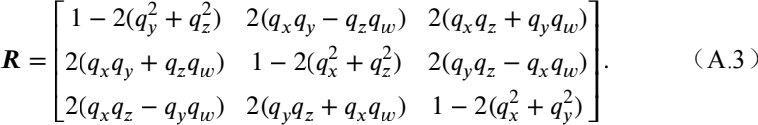
设单位四元数为

q = qw + qx i + qyj + qzk, （A.1）

满足归一化条件

q + q + q + q = 1. （A.2）

对应的旋转矩阵 R 可由下式给出：



给定旋转矩阵 R = [Rij]，可以通过下列方式提取四元数（其中一种常用的方法）：

 （A.4）

 （A.5）

 （A.6）

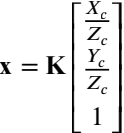
 （A.7）

以上公式仅适用于单位四元数，且在实际应用中可能需要对数值稳定性进行 进一步处理。

A.2 光束法平差雅可比矩阵与优化

假设三维点 **p** = (X, Y, Z)T 在相机坐标系下的坐标为 **p**C ，其投影到图像平面 上的像素坐标为 **p** = (U, U)T 。相机的内参矩阵为 **K**，旋转矩阵为 **R**，平移向量为 **t**。 则有：

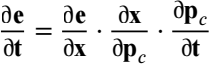
**p**C = **Rp** + **t** （A.8）

 （A.9）

其中，**p**c = (xc , Yc , Zc )T 为三维点在相机坐标系下的坐标。重投影误差 **e** 定义为： **e** = **x** − **x**obs （A.10）

其中，**x**obs为实际观测到的像素坐标。为了进行优化， 需要计算重投影误差对旋转 **R**、平移 **t** 和地图点 **P** 的雅可比矩阵。利用链式法则， 雅可比矩阵可以分解为以下 部分：

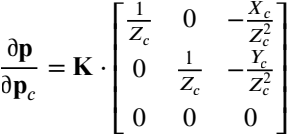
 =  ⋅  ⋅  （A.11）

 （A.12）

 =  ⋅  ⋅  （A.13）

其中，各部分的雅可比矩阵为：

 = **I**2×2 （A.14）

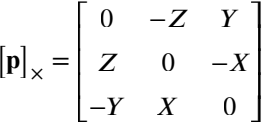
 （A.15）

 = −**R** ⋅  （A.16）

 = **I**3×3 （A.17）

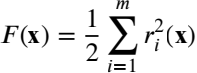
 （A.18）

其中， [**p**]× 表示向量 **P** 的反对称矩阵：

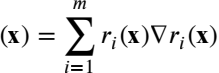
 （A.19）

将上述各部分雅可比矩阵相乘， 即可得到重投影误差对旋转、平移和地图点的雅 可比矩阵。

高斯-牛顿法是一种用于求解非线性最小二乘问题的迭代算法。其目标是找到 参数向量 **x**，使得残差的平方和最小化。具体而言，目标函数可以表示为：

 （A.20）

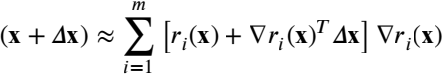
其中，ri (**x**) 表示第 i 个残差函数，m 为残差的数量。为了最小化 F(**x**)，我们需要找 到使其梯度为零的 **x**。首先，计算目标函数的梯度：

▽F  （A.21）

其中，▽ri (**x**) 是残差函数 ri(**x**) 关于 **x** 的梯度。接下来， 对每个残差函数 ri(**x**) 进行 一阶泰勒展开：

ri (**x** + Δ**x**) ≈ ri (**x**) + ▽ri (**x**)T Δ**x** （A.22）

将其代入目标函数的梯度表达式，得到：

▽F  （A.23）

为了使梯度为零，需要解以下线性方程组：

**J**(**x**)T**r**(**x**) + **J**(**x**)T**J**(**x**)Δ**x** = 0 （A.24）

其中，**J**(**x**) 是残差函数的雅可比矩阵，**r**(**x**) 是所有残差的向量表示。解上述方程组， 得到参数更新量：

Δ**x** = - [**J**(**x**)T**J**(**x**)]-1 **J**(**x**)T**r**(**x**) （A.25）

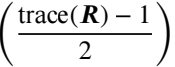
然后，更新参数：

**x** ← **x** + Δ**x** （A.26）

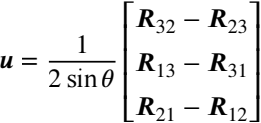
重复上述步骤，直到满足收敛条件。

A.3 罗德里格斯旋转公式

罗德里格斯旋转公式给定一个旋转矩阵 R，可以将其转换为旋转向量 θ = θu， 其中 θ 是旋转角度，u 是旋转轴的单位向量。旋转角度 θ 的计算方式如下：

θ = cos−1  （A.27）

其中，trace(R) 表示矩阵 R 的迹， 即对角线元素之和。旋转轴 u 是一个单位向量， 其计算方式为：

 （A.28）

其中，Rij 表示矩阵 R 的第 i 行第 j 列元素。

致 谢

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作 所取得的成果。尽我所知， 除文中已经注明引用的内容外，本学位论文的研究成 果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其 他个人和集体，均已在文中以明确方式标明。

签 名： 日 期：

个人简历、在学期间完成的相关学术成果

个人简历 1999 年 04 月 21 日出生于山东淄博博山区。

2017 年 9 月考入大连理工大学电信学部计算机科学与技术 (日语强化) 专业， 2022 年 7 月本科毕业并获得工学学士学位。

2022 年 9 月免试进入清华大学深圳国际研究生院攻读大数据技术与工程专业 硕士至今。

在学期间完成的相关学术成果

学术论文：

[1] **Linsong** **Xue**, Luo Qi, Zhang Kai. APM-SLAM: Visual Localization for Fixed Routes with Tightly Coupled A Priori Map. Journal of Intelligent and Connected Vehicles[J]. (审稿中)

[2] **Linsong** **Xue**, Kai Zhang, Guowei Zhu. Visual Localization with Prior Map. ITS World Congress 2025[C]. (审稿中)

[3] Haojie Wu, **Linsong** **Xue**, Kai Zhang. SGAGS: Semantic-Guided Adaptive 3D Gaussian Splatting. In press[C]. (已被 the 5th International Conference on Image, Vision and Intelligent Systems 录用)

专利：

[4] 张凯, 薛林松, 鹿昌义, 等. 一种间断 GNSS 信号下的车辆融合定位方法及相关 设备: 中国, CN118112623A[P]. 2024-05-31.

[5] 张凯, 薛林松, 鹿昌义, 等. 一种基于 NRTK 和视觉信息的巡检车定位系统: 中 国, CN119022939A[P]. 2024-11-26.

指导教师评语

答辩委员会决议书

暂无。