



浙江工业大学

# 硕士学位论文

论文题目： 基于双目视觉与惯导信息融合的 SLAM 系统实现

作者姓名 万薇

指导教师 邢科新副教授

学科专业 控制工程

培养类别 全日制专业学位硕士

所在学院 信息工程学院

提交日期 2019 年 1 月

浙江工业大学硕士学位论文

基于双目视觉与惯导信息融合的 SLAM 系统实现

作者姓名：万薇

指导教师：邢科新副教授

浙江工业大学信息工程学院

2019 年 1 月

**Dissertation Submitted to Zhejiang University of Technology  
for the Degree of Master**

**Implementation on SLAM System Based on Binocular  
Vision and inertial Information Fusion**

**Candidate: Wan Wei**

**Advisor: Associate Prof. Xing Ke-Xin**

**College of Information Engineering  
Zhejiang University of Technology  
January 2019**

## 浙江工业大学

### 学位论文原创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下，独立进行研究工作所取得的研究成果。除文中已经加以标注引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不含为获得浙江工业大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

作者签名：万薇

日期：2019年 1 月

### 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权浙江工业大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

- 1、保密□，在一年解密后适用本授权书。
- 2、保密□，在二年解密后适用本授权书。
- 3、保密□，在三年解密后适用本授权书。
- 4、不保密☒。

(请在以上相应方框内打“√”)

作者签名：万薇

日期：2019年 1 月

导师签名：

日期：2019年 1 月

# 基于双目视觉与惯导信息融合的 SLAM 系统实现

## 摘 要

SLAM(Simultaneous localization and mapping, 即时定位与地图构建)是指利用自身所携带的传感器获取移动机器人所在环境的三维空间模型和机器人的运动轨迹。视觉 SLAM 技术随着图像处理与机器视觉的发展, 逐渐成为实现移动机器人自主导航的关键技术之一。

本文针对移动机器人同时定位与地图构建中存在累积误差的问题, 提出利用双目视觉与惯导(Inertial measurement unit, 即 IMU)信息融合的位姿估计与优化算法, 提高了移动机器人位姿估计和地图构建的准确性与鲁棒性。本文的主要工作和成果如下:

1. 针对双目视觉与惯导信息融合的 SLAM 系统中特征点的尺度问题, 本文利用双目立体相机的视差原理获取空间特征点的深度信息。同时, 利用 ORB 特征提取与匹配算法对图像的特征点进行描述, 最后采用随机抽样一致性算法剔除误匹配, 该方法不仅提高了特征描述与提取的速度, 而且提高了特征匹配的准确率。

2. 针对双目视觉与惯导信息融合的 SLAM 系统中惯导的偏差及噪声问题, 本文提出利用双目视觉计算的初始位姿信息与 IMU 的预积分模型, 初始化 IMU 的偏差与速度。由于使用双目相机可计算特征点的深度信息, 因此 IMU 初始化过程中的计算复杂度将会减小。

3. 针对双目视觉与惯导信息融合的 SLAM 系统中移动机器人位姿的准确性问题, 本文提出利用双目视觉位姿估计算法 EPnP(An Accurate  $O(n)$  Solution to the PnP Problem)算法, 估计移动机器人的位姿, 并在位姿优化中增加 IMU 预积分的约束, 使移动机器人估计的位姿更加准确。

4. 针对检测回环及回环优化的问题, 本文采用基于视觉词袋模型(bag-of-words, 即 BoW)的闭环检测算法检测关键帧间的相似度, 该方法提高了回环检测的效率与准确度。同时, 根据回环检测的结果, 利用图优化算法进行全局优化减小移动机器人累积的漂移误差。

最后, 对全文进行总结, 并对进一步的研究提出一些展望。

**关键词:** SLAM, 双目, 惯导, 融合, 位姿估计, 回环检测

# **IMPLEMENTATION ON SLAM SYSTEM BASED ON BINOCULAR VISION AND INERTIAL INFORMATION FUSION**

## **ABSTRACT**

For robots in complex environment, using their own sensors to obtain the three-dimensional space model of the robot's environment and trajectory, which is the Simultaneous localization and mapping (SLAM). With the development of image processing and machine vision, visual SLAM technology has gradually become one of the important technologies for autonomous navigation of mobile robots.

In order to solve the problem of accumulated errors in the simultaneous localization and mapping construction of mobile robots, using the tightly-coupled stereo-inertial sensors to improve the accuracy of pose estimation and robustness of mobile robots. The main work and achievements of this paper are as follows:

1. In order to meet the depth of feature points of the tightly-coupled stereo-inertial SLAM, this paper uses the parallax principle of binocular camera obtain the depth of feature points. At the same time, use the ORB algorithm to detect and describe the feature points of the image, and we adopt feature matching algorithm based on Random Sampling Consistency (RANSAC) algorithm to perform feature matching to eliminate false matches. This not only improves the speed of image feature extraction and description, but also improves the accuracy of feature matching.

2. In order to solve the bias and noise of IMU in the tightly-coupled stereo-inertial SLAM, this paper proposes to use the initial binocular pose information and the pre-integration model of IMU to initialize the bias and speed of IMU. Because the binocular camera provides feature point depth information, reducing the computational complexity during initialization.

3. Aiming at the accuracy of the mobile robot's position in the front tightly-coupled

stereo-inertial SLAM, this paper proposes a pose estimation algorithm EPnP to estimates the pose of robot, and the IMU pre-integration constraints are added to pose estimation. This methods improve the accuracy of estimated pose of the mobile robot.

4. Aiming at the problem of how to detect closed loop and how to optimize the detected loopback, a closed loop detection algorithm based on bag of words is used to detect the similarity between the current key frame and all the previous key frames. The algorithm can improves the accuracy of loopback detection Meanwhile, according the result of the loop, this paper use graph optimization to optimize the detected loopback. And, this approach reduces the accumulated drift error.

Finally, summarize the full text and put forward the prospect of further research.

**Key Words:** slam, stereo, imu, tightly-coupled, pose estimation, loop detection

# 目 录

摘 要.....	i
第 1 章 绪 论.....	1
1.1 课题研究背景与意义.....	1
1.2 视觉 SLAM 国内外研究现状.....	2
1.2.1 基于视觉的位姿估计的研究现状.....	2
1.2.2 视觉 SLAM 中优化研究现状.....	4
1.2.3 视觉 SLAM 整体方案研究现状.....	5
1.2.4 视觉与惯导融合的 SLAM 整体方案研究现状.....	6
1.3 本文主要工作.....	7
第 2 章 双目视觉模型.....	9
2.1 引言.....	9
2.2 摄像机标定.....	9
2.3 图像特征提取与匹配.....	12
2.3.1 图像特征提取.....	13
2.3.2 图像特征匹配.....	15
2.4 双目立体匹配测距.....	16
2.5 本章小结.....	19
第 3 章 基于双目视觉与 IMU 融合的算法.....	20
3.1 引言.....	20
3.2 双目视觉与惯导融合的 SLAM 系统框架.....	20
3.3 IMU 预积分模型.....	21
3.3.1 IMU 的预积分模型.....	21
3.3.2 IMU 测量的数据噪声.....	24
3.3.3 IMU 测量的偏差.....	24
3.4 IMU 偏差估计算法.....	25
3.4.1 陀螺仪偏差估计.....	25
3.4.2 加速度偏差及重力加速度估计.....	25
3.4.3 速度估计.....	27
3.4.4 实验结果.....	27
3.5 基于双目视觉与 IMU 融合的位姿估计.....	29
3.5.1 基于双目视觉的位姿估计算法.....	29
3.5.2 基于视觉与 IMU 融合的位姿优化算法.....	31
3.6 本章小结.....	33
第 4 章 基于图优化的视觉 SLAM 算法.....	34
4.1 引言.....	34
4.2 SLAM 中的图模型.....	34



4.3 回环检测.....35

4.4 位姿图优化.....38

4.4.1 基于滤波优化方法的 SLAM 后端优化..... 39

4.4.2 基于位姿图优化方法的 SLAM 后端优化.....40

4.5 本章小结.....41

**第 5 章 实验设计与分析..... 42**

5.1 引言.....42

5.2 实验平台介绍.....42

5.3 实验算法设计.....43

5.4 实验结果与分析.....44

5.5 本章小结.....47

**第 6 章 总结与展望..... 48**

6.1 研究工作总结.....48

6.2 展望.....49

**参考文献..... 50**

**致谢..... 54**

**攻读学位期间参加的科研项目和成果..... 55**

# 第 1 章 绪 论

## 1.1 课题研究背景与意义

近年来，随着计算机技术和传感器技术的飞速发展，能应用在复杂的环境下的高智能化机器人，不仅在航空航天及军事领域中得到很好的应用，而且还广泛地应用于工业、农业、服务业等各个行业中。除此之外，智能移动机器人是一个集合了环境感知、动态决策与规划、行为控制与执行等多功能于一体的综合系统，随着时代的发展，越来越多的智能移动机器人已逐渐融入到我们的生活中。如图 1-1 所示的谷歌的无人驾驶汽车和图 1-2 所示的中控研发的智能巡检机器人。谷歌的无人驾驶汽车配备有雷达激光测距系统、摄像机与惯性导航系统等，实现无人驾驶情况下汽车的启动、行驶及停止。同样的，智能巡检机器人使用机器视觉、电磁场和惯性传感器等多种传感器融合，实现了移动机器人的自主移动与自动巡检的导航。



图 1-1 谷歌无人驾驶汽车



图 1-2 中控巡检机器人

在各行各业的应用中，对未知环境的导航与探索一直是移动机器人实现智能化的热点与难点。而移动机器人的即时定位与地图构建（即 SLAM）是其中一个重要的基本问题<sup>[1]</sup>。SLAM 为实现移动机器人自主导航与定位、路径规划及目标识别等任务的提供了前提条件。随着社会的不断进步与发展，移动机器人的应用环境更趋向复杂化与非结构化发展。在 SLAM 系统中，实现移动机器人的精确定位与地图构建是移动机器人实现智能化的关键技术。SLAM 技术从 1986 年提出以来<sup>[2]</sup>，经过三十多年的发展，取得了一系列突破性的

研究成果。与此同时，随着人工智能与深度学习的发展，SLAM 技术已经成为移动机器人领域中极富挑战性的研究课题。

SLAM 是由同步的机器人状态估计与环境地图构成，其中 SLAM 所采用的传感器有激光雷达、视觉传感器和多种传感器的融合等。激光传感器可以通过散发的激光束快速并精确地获取移动机器人与周围环境的距离和角度信息，因此基于激光的 SLAM 系统一直是实现移动机器人定位与导航问题最热门的方案之一，同时也是目前应用最广泛的二维 SLAM 算法。比较著名的基于激光 SLAM 的地图构建方法是 Grisetti<sup>yz</sup> 和 Stachniss 于 2007 年提出的 Gmapping<sup>[3-4]</sup>的地图构建算法，该地图构建方法至今一直被人引用。激光传感器虽然拥有精度高，速度快等优点，但也存在价格昂贵、受环境限制等缺点。与激光传感器相比，视觉传感器拥有价格便宜、能获取更丰富的环境信息等优点，使其更适用于构建三维环境地图。通常视觉 SLAM 使用的视觉传感器主要有单目相机、RGB-D 相机与双目相机，由于双目相机不受红外影响且可以获得环境的深度信息，相比其他相机，有更好的适应性，可以广泛应用在室外<sup>[5]</sup>。但视觉传感器容易受到摄像机帧率、光线及旋转快慢等的影响，使其在快速移动的情况下，成像画面模糊。最近几年，惯导与视觉融合的 SLAM 系统成为了热门的研究话题。惯性测量单元具有体积小、频率高及质量轻的特点，在航天、虚拟与现实及地图导航中广泛应用。但由于其会随时间漂移，使其不适合长时间定位，而视觉测量具有精度高，能检测累计误差。因此，利用视觉测量作为 IMU 测量的参照，用视觉测量结果校正当前时刻 IMU 的测量结果。这样不仅利用了 IMU 的瞬时测量精度，也抑制了 IMU 的累计漂移误差，与相机位姿融合后可以得到精度更高的测量值。视觉 SLAM 的功能主要是在未知环境中，移动机器人使用自身携带的传感器，获取移动机器人所在环境的三维模型及机器人的运动轨迹。由于，双目相机不仅可以获得环境的二维图像信息，还可以得到对应环境特征点的深度信息，同时利用惯导的高频特性可对快速运动的情况进行跟踪。因此本文所采用的传感器是双目相机与惯导两者的融合。

本课题主要研究基于双目与惯导融合的 SLAM 系统。即利用双目与 IMU 获取的信息，来计算得到移动机器人的运动轨迹和周围环境的三维空间模型。

## 1.2 视觉 SLAM 国内外研究现状

### 1.2.1 基于视觉的位姿估计的研究现状

视觉 SLAM 主要可以分为前端位姿估计和后端全局优化两部分。前端也称作视觉里程计，主要用于根据相邻两幅图像的投影关系计算出位姿的变化，即位姿估计。视觉里程

计的实现方法根据是否需要提取特征点分为：基于特征点法的位姿估计和基于直接法的位姿估计。

基于特征点法的位姿估计是对图像上的具有一定特征的像素点进行提取与匹配，同时该像素点对光照变化及运动变化不敏感，最后根据得到的二维或三维匹配点计算相机的位姿。特征点方法是目前视觉 SLAM 位姿估计中比较成熟的主流实现方法。通常比较常用的特征提取与描述算法有 SURF<sup>[6]</sup> (Speeded-Up Robust Features)、SIFT<sup>[7]</sup> (Scale-invariant feature transform) 和 ORB (Oriented FAST and Rotated BRIEF)<sup>[8]</sup> 等等。利用特征提取方法提取图像的特征点并建立描述子之后，需要采用基于 RANSAC<sup>[9]</sup> 的特征匹配算法进行特征匹配并剔除其中的误匹配。最后，根据匹配到的特征点对来估计相机的位姿运动。估计位姿时，使用不同的相机类型，将得到不一样的特征点信息，其中单目相机由于无法获取深度信息，得到的是 2D-2D 的匹配点，这样就需要使用对极几何来进行求解，比较经典的解决方法有 Nister 提出的五点法与六点法<sup>[10-11]</sup> 与 R.Hartley 提出的七点法与八点法<sup>[12-13]</sup>。

当使用的传感器为双目相机或者深度相机时，可以计算得到特征点的三维空间坐标。这时可以采用 3D-2D 点的匹配方法或者 3D-3D 点的匹配方法。如果利用 3D-2D 的匹配点计算移动机器人的位姿，可以使用 PnP 的位姿估计算法求解相机的位姿。求解 PnP 的方法有许多种，高小山等人曾提出 P3P 算法<sup>[14]</sup>，该算法主要利用三对匹配点估计相机位姿。算法实现步骤是根据三角形相似的原理求解投影点在摄像机坐标系下的三维坐标。该方法简单高效，但存在一定的缺陷，如果匹配的点数超过三对以上，就无法利用其他匹配点的信息，并且如果在选取得三对匹配点中存在误匹配的错误的信息，将会导致结果出现不可估量的错误。针对这些问题，Lepetit 和 Moreno-Noguer 等人提出了 EPnP<sup>[15]</sup> 算法，该算法可以利用更多的匹配信息，并且采用了迭代的方法对相机位姿进行优化，来减少噪声所产生的影响，类似的算法还有 RPnP<sup>[16]</sup> 算法等。如果采用的是三维到三维的匹配点进行计算，可以采用松弛迭代法与迭代最近点算法 (Iterative Closest Point, ICP)<sup>[17]</sup> 进行求解。目前，大部分 3D 特征匹配位姿求解采用的方法是迭代最近点算法，该算法通过多次迭代运算对运动变化进行优化，每次迭代过程中寻找源点云数据集与目标点云数据集最短的欧式距离，最后求解机器人的运动变化并将该变化对应到目标点云数据集上进行迭代优化。

基于特征点法的位姿轨迹算法是现如今视觉里程计的主流实现方法，虽然该方法有效实用，但仍存在着许多问题如：特征提取与描述子的计算耗时较长，且忽略了图像中除去特征点外可能有用的信息，在特征点稀疏的地方存在没有足够多的匹配点来计算相机运动

等问题。针对这些问题, Silveira 和 Mails 等人提出采用直接法<sup>[18]</sup>来实现视觉 SLAM。直接法是参考光流法<sup>[19]</sup>演变而来的。光流法描述了像素点随时间在图像中运动的方法, 可以被用来跟踪特征点的运动。直接法则是根据图像的像素信息来估计相机的运动。直接法避免了特征提取与匹配的过程, 节约了特征点描述与提取的时间, 在图像特征缺失的场合也可以较好的估计相机的位姿。同时, 该方法在图像模糊、低纹理环境和高频纹理的情况下有着更好的鲁棒性。但是, 直接法在相机相对运动过大时, 容易出现跟踪失败的情况, 因此只有在运动较小时才能成功。此外, 容易受相机曝光时间和外界光线变化的影响等。

### 1.2.2 视觉 SLAM 中优化研究现状

视觉 SLAM 中的后端优化包括回环检测与优化, 主要用于解决位姿估计随时间产生的漂移误差的问题。回环检测算法用于检测经过一段时间的运动, 移动机器人是否回到之前运动过的地方。由于视觉里程计中存在累积误差会导致计算得到的位姿并没有回到原先的位姿, 所以通过回环检测算法可以检测到移动机器人又回到了原来的位姿, 从而将计算得到的位姿与之前的位姿相比较获取现在的位姿, 达到修正错误的计算结果, 减少累积误差的目的。

正确的回环检测可以校正视觉 SLAM 前端位姿估计中产生的累积误差, 然而错误的回环检测会对现有结果造成严重影响, 甚至导致计算结果严重偏离真实值。因此在回环检测中必须保证高准确率。Hahnel 和 Burgard 等人提出了一种基于几何关系的回环检测方法<sup>[20]</sup>。该方法通过检测机器人之前到过相似场景的周围时, 检测当前是否存在回环。该方法对 SLAM 中估算的位姿精度要求较高, 当累积误差较大时回环检测准确度不高。Endres 和 Hess 等人<sup>[21]</sup>提出从之前的关键帧中随机抽取一定数量的关键帧与当前帧进行特征匹配来进行回环检测的方法。该方法虽然取得了一定的效果, 但是偶然性太大, 不能保证检测到所有的回环。尽管有着多种多样的回环检测方法, 但是基于外观的回环检测方法一直是 SLAM 领域的主流方法<sup>[22-23]</sup>。但如何快速地判断两幅图像间的相似度一直是困扰研究者的难题。针对这一难题, Gálvez-López 和 Tardós 提出了基于 bag-of-words (词袋模型) 的回环检测方法<sup>[24-25]</sup>, 该方法将提取到的特征描述子抽象成视觉词汇并构建成视觉词典, 并且采用 TF-IDF(Term Frequency-Inverse Document Frequency)的方法进行图像检索, 取得了十分不错的效果, 大大加快了回环检测的速度, 同时也提高了回环检测的准确率, 该方法是如今 SLAM 领域最好的回环检测方法之一。基于词袋模型的关键技术包括图像特征提取与匹配技术和视觉词典的构造方法。特征提取与匹配应尽量使所提取的特征具有代表性

与可区分性,降低提取过程的复杂度。在特征提取完成后,需要对特征提取生成视觉单词。比较典型的方法是 K-means 方法,该方法将训练图像库的大量特征按照相似程度进行聚类。

根据回环检测的结果,为减少累计误差,得到更精确的结果,需要对位姿估计的数据进行优化。在早期,研究学者认为基于非线性优化的 SLAM 优化算法计算十分复杂,因此,传统的滤波器方法在早期的视觉 SLAM 优化中占据重要的地位。基于 EKF (Extended Kalman Filter, 扩展卡尔曼滤波器) 的后端优化方法是其中比较经典也是最常用的,如 Grasa 等<sup>[26]</sup>提出的 EKF-SLAM 方法实现了相机位姿估计的优化与三维空间特征点的计算。此外, Lategahn 等<sup>[27]</sup>提出一种基于双目视觉的 SLAM 算法,实现了较高精度的三维地图的构建,其中就采用了基于 EKF-SLAM 的位姿估计方法,同时利用闭环检测实现了全局的地图构建。但是闭环检测是人工实现的,非算法自动实现。与此同时,研究者们还对各种各样的滤波器在后端优化中的应用进行了研究,例如无迹卡尔曼滤波器<sup>[28]</sup>(Unscented Kalman Filter, 简称 UKF)方法具有较高收敛性与计算简单优点,但在非线性系统中误差较大;信息滤波器<sup>[29]</sup>(Information Filter, 简称 IF)方法具有简单、稳定与准确的优点,但是不能处理数据关联问题;粒子滤波器<sup>[30-31]</sup>(Particle Filter, 简称 PF)方法能够处理非线性与非高斯分布系统,但计算复杂度会随着系统的增加而增加。随着 SLAM 技术的发展,该方法逐渐应用到大规模、非结构化中场景,基于滤波法人方法无法满足要求,非线性优化的方法逐渐被用到视觉 SLAM 的后端优化中,得到了较好的效果。实际上,在计算量相同的情况下,相比于滤波器的方法,非线性优化的方法获得了更好的效果<sup>[32]</sup>。近年来,研究者认识到运用 Bundle Adjustment (译为光束平差法) 算法能很好的解决了 SLAM 问题中的稀疏性<sup>[33-34]</sup>,基于该方法的优化可以精确的优化相机的位姿及特征点位置。基于非线性的优化方法不仅考虑机器人状态之间的关系,同时对机器人运动轨迹的整体观测进行全局优化。因此,Latif 和 Cadena 等人提出利用位姿图<sup>[35-37]</sup>优化相机位姿,舍弃对特征点的优化,极大的减少了计算量,提高了效率。Kaess 和 Ranganathan 等人提出使用因子图<sup>[38-40]</sup>方法优化全局地图,也取得了不错的效果。

### 1.2.3 视觉 SLAM 整体方案研究现状

Davison 于 2007 年提出的 MonoSLAM<sup>[41]</sup>是第一个实时的单目 SLAM 系统,同时也是许多实时 SLAM 方案的发源地。该方法初步实现了视觉 SLAM 的实时运作,但该方法存在路标数量有限,稀疏特征点容易丢失及应用场景有限等不足之处。同年, Klein 提出的

PTAM<sup>[42]</sup> (Parallel Tracking and Mapping) 也是 SLAM 研究领域重大突破。PTAM 首次提出并实现位姿跟踪与地图构建的并行双线程, 将 SLAM 主要分为前端和后端。该方法的前端部分提出了关键帧筛选策略, 减小了计算量, 同时后端优化算法采用非线性优化算法, 提高了视觉 SLAM 的实时性。然而 PTAM 也存在着位姿跟踪容易丢失, 应用场景小等缺陷。Mur-Artal 等在 PTAM 的基础上与 2015 年提出了 ORB-SLAM<sup>[43]</sup>, 这是现代视觉 SLAM 系统的突破性进展。与之前的 SLAM 系统相比, 该系统可支持单目、双目和 RGB-D 三种视觉传感器, 有广泛的适应性。而且, ORB-SLAM 中的多线程结构使整个系统获得了良好的跟踪效果与地图构建效果, 保证了全局地图的一致性。但是 ORB-SLAM 系统也存在不足, 对 CPU 的要求较高且对每幅图像都需要提取特征点较耗时。

与基于特征点生成的稀疏地图不同, Jakob Engle 等人于 2014 年提出了基于直接法的 LSD-SLAM(Large Scale Direct monocular SLAM)系统<sup>[44-45]</sup>。该方法首次将直接法成功应用在 SLAM 系统。直接法与特征点法不同, 它是直接对像素点进行操作, 无需对特征点进行提取与描述。该方法可以通过单目相机实现半稠密地图的构建, 但是 LSD-SLAM 对摄像机的曝光与内参的影响很大, 在相机快速运动时容易出现丢失的情况, 且无法实现回环检测。基于特征点法与直接法的优缺点, Forster 等人提出了基于半直接法的视觉里程计, 即 SVO (Semi-direct Visual Odometry)<sup>[46-47]</sup>。半直接法将特征点法与直接法相结合, 然后根据这些特征点周围像素点的信息对相机的位姿进行估计。该方法的计算速度快, 可以移植到手机等轻便设备, 但该方法无法实现完整的地图构建功能, 只实现了位姿估计。

#### 1.2.4 视觉与惯导融合的 SLAM 整体方案研究现状

近些年, 随着惯导技术的发展, 越来越多的位姿估计算法采用视觉与惯导的融合<sup>[48-49]</sup>, 得到的移动机器人的位姿在精度上有所提高。同时, 利用视觉和 IMU 融合来构建地图已经成为非常热门的研究话题, 融合方法主要有松耦合方法与紧耦合方法。松耦合方法是仅通过视觉算法来独立地估计姿态, 并且只在单独的估计步骤中融合 IMU 的测量。Konolige 等人<sup>[50]</sup>提出了基于双目视觉与 IMU 信息松耦合的优化方法, 虽然该方法提高了里程计的长期精度, 但是, 由于多传感器内部状态之间的相关性被忽略, 所以在位姿估计精度方面松耦合方法不是最优的。紧耦合系统是把 IMU 的状态与相机的状态合并在一起, 共同构建运动方程与观测方程, 然后进行状态估计。其中紧耦合的优化方法主要有基于关键帧非线性优化的方法和基于滤波优化的方法。在早期 SLAM 的研究中, 基于滤波的方法占据了重要的位置, Mouriki 等<sup>[51]</sup>提出一种基于 EKF 的单目视觉实时融合的方法, 称

为多状态约束卡尔曼滤波器(MSCKF)算法。该算法将路标模型与 EKF-SLAM 采用的模型进行完整的估计之间的切换,使性能进一步的提高。然而采用滤波的优化算法只能计算增量式的运动,不能实现回环检测与地图重用的功能。这意味着估计的轨迹会产生漂移,即使传感器一直运行在相同环境中,也无法准确定位。在此基础上,Leutenegger 等<sup>[52]</sup>提出了基于视觉与 IMU 融合的非线性优化方法,适用于单目及双目相机,在视觉与惯导融合的 SLAM 方法中做出了重要贡献。该方法将视觉测量与 IMU 的读数紧密集成在 SLAM 中,IMU 的误差项与路标的重投影误差集成优化,该优化方法优于使用滤波的优化方法,同时提高了精确度与鲁棒性,但计算复杂度较高。2013 年,Forster 等<sup>[53]</sup>提出 IMU 的预积分方法进一步推动了视觉与 IMU 融合的视觉里程计的发展,该方法可以将 IMU 的预积分模型集成到视觉惯导紧耦合的因子图中,避免了特征点云的优化,相比于其他方法更高效。在此之后,Mur-Artal 根据 IMU 的预积分方法,提出了基于单目与惯导融合的 SLAM 系统<sup>[54]</sup>,该方法实现了完整了实时 SLAM 系统,提高了系统的精度与鲁棒性,但是该系统使用单目相机无法直接获取尺度信息,依赖于 IMU 计算的尺度信息。Tong Qin<sup>[55]</sup>也提出了基于单目视觉与惯导的 SLAM 系统(VINS-Mono),该模块不仅提供了精确的局部姿态、速度与方位估计,而且实时实现了摄像机的外部校准与 IMU 偏置校正。

### 1.3 本文主要工作

视觉与惯导融合的 SLAM 是实现移动机器人任务规划、导航和控制等自主运动的前提之一,有很大的研究价值和应用前景。本文针对视觉 SLAM 系统中存在的累计误差及与运动过快导致跟踪丢失的问题,提出一种基于双目与惯导信息融合的 SLAM 算法。首先,利用双目视觉的位姿估计算法求解位姿,然后利用 IMU 的预积分模型与估计的位姿融合优化,最后利用回环检测算法与图优化进行全局地图构建。

全文包含 6 章,各章节的具体内容安排如下。

第一章首先介绍了双目视觉与惯导融合的 SLAM 的研究背景和意义,并结合国内外研究现状分析目前主流的 SLAM 方法所存在的缺点与不足,最后介绍了本文的主要研究内容与结构安排。

第二章介绍了双目相机的涉及的摄像机模型及双目标定的原理,并展示了如何获取相机标定参数。然后,介绍了相关图像特征提取与匹配相关算法,并分析各自优缺点;最后,利用双目视觉的视差原理及立体匹配方法,得到特征点的三维空间坐标。

第三章主要对 IMU 的初始化过程进行研究,首先,介绍了 IMU 的运动模型以及预积



分模型。然后利用预积分模型求取 IMU 中加速度与陀螺仪的偏差及其速度。最后研究双目视觉与 IMU 融合的 SLAM 前端相关算法，一方面，阐述了双目视觉的位姿估计算法，另一方面，利用 IMU 预积分约束优化位姿提高了位姿估计的精度。

第四章阐述双面视觉与 IMU 融合的 SLAM 后端算法，首先介绍了回环检测的相关算法，然后介绍 SLAM 位姿的数学建模，最后采用图优化方法对相机位姿进行优化，减少位姿估计中产生的累计误差。

第五章为本文的实验部分，利用数据集分析对比了本方法与其他方法，同时介绍实验平台，最后通过实验验证本文算法的鲁棒性及精确性。

第六章为总结和展望，总结了本文所提算法的优点以及不足之处，并对未来视觉 SLAM 的发展进行了展望。

## 第 2 章 双目视觉模型

### 2.1 引言

双目视觉与惯导融合的 SLAM 系统需要通过双目相机获取环境的特征点的，与单目相机相比，双目相机不仅能够获取周围图像信息，而且还能够通过视差计算获取特征点的深度信息，得到特征点的空间三维坐标。本章，首先介绍了摄像机相关坐标系及相机标定原理。然后对图像特征点的提取与匹配原理进行论述，最后采用的是基于 ORB 的提取特征点的方法，该方法具有光照不变形，运行稳定等优点。最后，利用双目视差原理计算特征点的深度信息，并进行了相关实验。

### 2.2 摄像机标定

摄像机标定就是根据摄像机所拍摄到的图像获取摄像机相关的参数。摄像机标定的参数精度会对后续利用视觉求取位姿估计的精度有影响，因此摄像机标定是求解移动机器人位姿估计的前提。目前计算机标定的方法主要有：传统标定方法<sup>[56]</sup>、自标定方法<sup>[57]</sup>和主动标定方法<sup>[58]</sup>。传统的标定方法利用已知景物特征点的结构信息，得到摄像机的相关参数。传统标定法的优点是标定精度高，但是需要高精度参照物作为标准。相对传统标定方法，自标定的方法利用多幅图像上对应点的信息进行标定，因此灵活性强，但该方法使用非线性标定鲁棒性不高。而主动视觉标定方法是根据主动摄像机的运动获取图像数据，线性的求解摄像机的参数模型，但该方法对摄像机运动平台的精度要求较高，且无法应用于摄像机运动未知与无法控制的场合。因此，本文采用传统的标定方法，求取摄像机的内部参数。

摄像机标定过程中主要利用的是摄像机的小孔成像模型，其中摄像机的小孔成像模型涉及的坐标关系去如图 2-1 所示。标定过程中涉及的坐标系分别为：图像物理坐标系、图像像素坐标系、摄像机坐标系及世界坐标系，其中摄像机坐标系由  $X_c$  轴、 $Y_c$  轴与  $Z_c$  轴表示，世界坐标系由  $X_w$  轴、 $Y_w$  轴与  $Z_w$  表示，图像的物理坐标系由  $(x, y, z)$  表示，图像像素坐标系由  $(u, v)$  表示， $OO_1$  为摄像机的焦距  $f$ 。

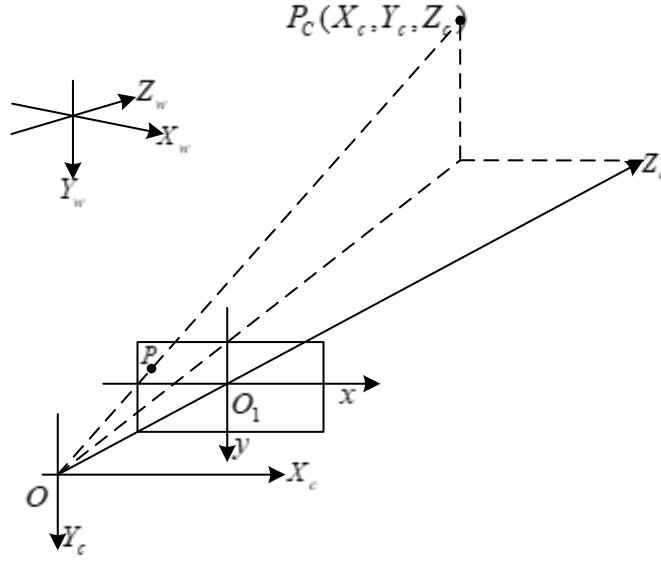


图 2-1 摄像机成像模型坐标系的关系

利用相机的模型可建立出相机坐标系与图像成像坐标系之间的关系，假设空间一点  $P_c$  在图像的成像点为  $p$ ，其中设  $O_1$  在图像像素坐标系中的坐标为  $(u_0, v_0)$ ，每一个像素点在  $x$  轴和  $y$  轴上的物理尺寸大小为  $d_x, d_y$ ，则可得图像的物理坐标与图像像素坐标关系为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-1)$$

以图中的世界坐标系为基准，空间点  $P_c$  在摄像机下的坐标为  $(X_c, Y_c, Z_c, 1)^T$ ，则点  $p$  的图像物理坐标系与摄像机坐标系的关系可以表示为：

$$\begin{cases} x = \frac{fX_c}{Z_c} \\ y = \frac{fY_c}{Z_c} \end{cases} \quad (2-2)$$

将式 (2-2) 表示成矩阵形式为：

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2-3)$$

则图像像素点与摄像机坐标系的关系表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z} A \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}, A = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2-4)$$

其中，矩阵  $A$  为摄像机内部参数矩阵。

设  $P_c$  点在世界坐标系下的坐标为  $(X_w, Y_w, Z_w, 1)^T$ ，则摄像机坐标系与世界坐标系的关系为：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = R_t \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, R_t = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \quad (2-5)$$

式中， $R$  表示摄像机坐标系与世界坐标系的旋转关系， $t$  表示摄像机坐标系与世界坐标系的平移关系， $R_t$  为摄像机的外参矩阵。

根据式 (2-6) 和式 (2-7) 可得， $p$  点的像素坐标与世界坐标的关系为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} R_t \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2-6)$$

上述为单目相机成像模型的关系以及相机的内外参数模型，由于本文使用的是双目相机，根据图 2-1 可知，左右相机之间有一个平移及旋转的关系，在双目标定过程中，以左边相机为基准进行标定。

根据上述的标定原理及方法，本文采用传统的标定方法获取双目相机的内参数与外参数，标定的主要流程如下 2-2 所示：

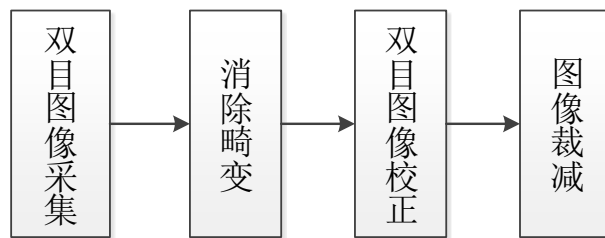


图 2-2 双目标定流程图

标定程序使用基于 MatLab 的立体视觉标定方法，完成双目视觉标定的过程。利用该工具箱可以准确便捷的获取两个摄像机之间的旋转与平移关系。首先单独标定左右相机，利用标定的棋盘图。从不同角度获取的 19 张图片加入标定工具箱，如下图 2-3 所示，利用标定工具箱即可得到标定结果。最后，将得到的左右相机的标定结果加入双目标定工具箱中，可得左右相机之间的旋转平移关系。

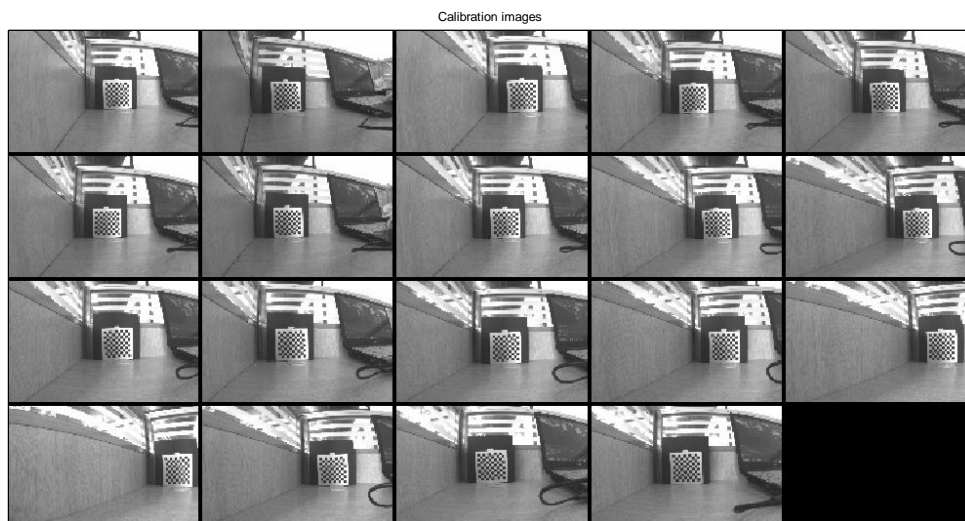


图 2-3 相机获取标定图片

通过标定实验，可得本文使用的双目相机的内参为：

$$A = \begin{bmatrix} 470.48 & 0 & 335.68 \\ 0 & 460.855 & 241.05 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-7)$$

双目相机之间的旋转平移矩阵为：

$$R = \begin{bmatrix} 0.992 & 0.005 & -0.123 \\ -0.004 & 0.999 & 0.005 \\ 0.123 & -0.004 & 0.999 \end{bmatrix} \quad (2-8)$$

$$T = [-122.408 \quad 0.252 \quad -9.181]^T \quad (2-9)$$

## 2.3 图像特征提取与匹配

图像特征提取与匹配是根据获取图像中的特征信息，确定两幅图像之间的相同像素点的信息。图像中的特征提取是利用计算机分析和处理图像信息，然后确定图像中局部的不

变特征，最后对局部特征进行描述。根据对图形特征点的描述，可以通过比较特征描述符的距离进行匹配，得到两幅图像之间特征点的对应关系。图像特征提取与匹配是研究立体匹配与位姿估计的重要组成部分，选择精度高与鲁棒性好的特征提取方法，一方面提高特征匹配的效率，另一方面可以提高匹配的精确度。

### 2.3.1 图像特征提取

图像的特征是图像中一些具有高辨识度的像素点，如图像中的角点、线边缘与区块等区域。图像角点的特征相比于其他方法有更高的辨识度，所以图像角点的提取应用广泛。然而优秀的特征点必须具有高重复性、局部性、准确性及独特性等特点。常用的特征描述子有：SIFT、SURF、ORB 等。其中 SIFT(Scale-invariant feature transform，即尺度不变特征变换)，该方法于 1999 年由 David Lowe 首次提出。SIFT 特征具备尺度和旋转不变性，且对视角变化与仿射变换保持一定的稳定性。然而，SIFT 的缺点有计算量大，对特征点的要求高，在低纹理区域的特征点数量较少。于是在此基础上，Herbert Bay 等对其改进提出了 SURF 算法，该算法改进了 SIFT 计算复杂度高的缺点，提高了算法的执行效率。但该算法对局部像素的梯度方向十分依赖，使得特征点的匹配成功率下降，另外在尺度上也有较大误差。2011 年 ICCV 会议上 Ethan Rublee 等首次提出 ORB 特征点检测与描述算法。ORB 算法的特征点由关键点和描述子组成，特征点提取部分采用的 Oriented-FAST 方法<sup>[59]</sup>，特征点描述部分采用的是 BRIEF 特征描述子<sup>[60]</sup>。

FAST 算法用于检测局部像素灰度变化明显的地方，其优点是提取角点的速度很快。如图 2-4 所示 FAST 角点的提取是通过判断该像素点与周围相邻足够多的像素点是否处于不同区域，若处于不同区域，则该像素点为角点。FAST 角点的检测步骤为：

1. 在图像中任选一点  $p$ ，设其像素的亮度值为  $I_p$ 。
2. 以像素点  $p$  为圆心画圆，取半径为 3 的周围的 16 个像素点。
3. 设定阈值  $T$ ，若这周围的 16 个像素点中有连续的  $N$  个像素点的像素值都小于  $(I_p - T)$  或大于  $(I_p + T)$ ，那么认为像素点  $p$  为特征点。
4. 对每个像素点重复执行上述步骤，提取出一幅图像中的所有特征点。

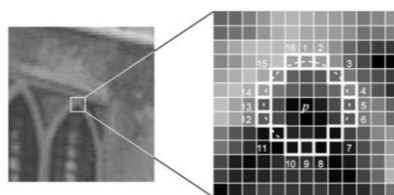


图 2-4 FAST 特征点提取示意图

ORB 特征点的尺度不变性是通过图像金字塔构建的，其中特征的旋转不变性是由灰度质心法实现，该方法的步骤为：

在一个半径为  $r$  的图像块中，定义图像块的矩为：

$$m_{pq} = \sum_{x,y \in r} x^p y^q I(x,y) \quad (2-11)$$

其中  $p,q=\{0,1\}$ ， $I(x,y)$  为图像灰度，可得该图像块的质心为：

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (2-12)$$

设角点坐标（即此图像块的几何中心）为  $O$ ，得方向向量  $\overrightarrow{OC}$ ，于是可得特征点的方向：

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}} / \frac{m_{00}}{m_{00}}\right) = \arctan(m_{01} / m_{10}) \quad (2-13)$$

BRIEF 描述子是一个二进制描述子，相比于其他描述子具有简单快速的优点。该描述子的原理主要分为三步：

1. 以特征点  $P$  为中心，取一个  $S \times S$  大小的 Patch 邻域；
2. 在这个邻域内随机选取  $N$  对点，然后对这  $2N$  个点做高斯平滑，比较  $N$  对像素点的灰度值的大小：

$$T(p; x, y) := \begin{cases} 1: p(x) < p(y) \\ 0: p(x) \geq p(y) \end{cases} \quad (2-14)$$

其中， $p(x)$  是图像块中位于点  $x$  的灰度值，即对于特征点邻域内的像素点  $x$  和  $y$  比较大小，若果  $p(x)$  小于  $p(y)$  则取 1，反之则取 0。

3. 重复上述步骤，对  $N$  组二进制点进行比较后，形成一个二进制编码，这个编码就是对特征点的描述。

由于 BRIEF 描述子不具备旋转不变性，因此需要给 BRIEF 提供旋转不变性。利用之前的 ORB 特征提取阶段已经计算特征点的方向角信息，对其提取的特征点的领域进行旋转，然后在提取 BRIEF 描述子，这样可以保证特征描述子的旋转不变性。但是该方法计算量大，因此 ORB 算法的旋转不变性是对每个特征点，随机产生的  $n$  对点进行旋转，这样将会更加高效。设这  $n$  对随机点为  $(x_i, y_i)$ ，一个  $2 \times n$  的矩阵定义如下：

$$S = \begin{pmatrix} x_1 & \dots & x_n \\ y_1 & \dots & y_n \end{pmatrix} \quad (2-15)$$

将这  $n$  对随机点按照 FAST 得到的特征点的方向角  $\theta$  旋转可得：

$$S_{\theta} = R_{\theta} S \quad (2-16)$$

其中， $S_{\theta}$ 表示旋转后随机点的位置， $R_{\theta}$ 表示对应的旋转矩阵，然后在新得到的点集  $S_{\theta}$ 中提取描述符即可。这样就使得 ORB 特征描述子具备了旋转不变性。ORB 算法在鲁棒性上略低于 SIFT，但是运行速度远快于 SIFT，考虑 SLAM 中需要对大量的特征点进行提取与匹配，提高系统的实时性需求，本文将选取 ORB 算法作为特征提取算法。

### 2.3.2 图像特征匹配

图像的特征匹配是双目视觉位姿估算中关键的一步。特征匹配解决的是前一帧图像特征点与当前帧图像特征点之间的相对应的关系。通过特征提取算法对图像进行特征提取之后，可以根据特征点匹配的信息对摄像机的位姿进行估计，并剔除其中的误匹配。

图像特征匹配中最简单的方法是暴力匹配法，即计算当前图像每一个特征点与匹配图像特征点之间描述子的距离。描述子间的距离表示两个特征点之间的相似程度。本文得到的是二进制的描述子，使用汉明距离表示描述子的距离。得到一系列描述子的汉明距离之后，通过选取距离最近并小于一定阈值的点作为当前的匹配点。这种方法虽然简单可行，但是随着匹配点数目的增多，计算量也随之增大，影响该算法的实时性。因此当匹配点数量较多时，一般采用快速近似最近邻搜索算法进行特征匹配。

本文采用的是基于 OpenCV 集成的最近邻搜索匹配方法，同时使用 RANSAC(Random Sample Consensus) 算法剔除样本中噪声数据，该算法能高效得到样本数据，同时计算得到正确的数学模型。其基本思想如下：

假设给定一组由  $N$  个数据点组成的集合  $P$ ，集合中包含正确数据（内点：可以被模型描述的数据），也包含异常数据，即噪声数据（外点：无法适应数学模型的数据）。且初始化模型参数所需要的最小样本数为  $n$  ( $n < N$ )，执行以下步骤来拟合模型：

- 1) 假设从集合  $P$  中随机抽取  $n$  个数据点都是内点；
- 2) 利用这  $n$  个数据点可以拟合得到数学模型  $M$ ；
- 3) 用得到的数学模型  $M$  去测试所有其它数据，与  $M$  的距离超过阈值的点认定为外点，不超过阈值的认定为内点；
- 4) 如果有足够多的点被归纳为内点，则认为该模型足够合理；
- 5) 用得到的所有内点重新估计出数学模型。

重复执行上述步骤固定次数，每次产生的数学模型要么因为对应的内点数太少而被舍



弃，要么因为比现有的模型更好（包含更多的内点数）而被选用。如下图 2-5 所示，仍存在有误匹配的情况。最后采用随机抽样一致算法剔除误匹配，减少了误匹配的数量，该算法有效地剔除了误匹配，得到正确的特征匹配结果如下图 2-6 所示：

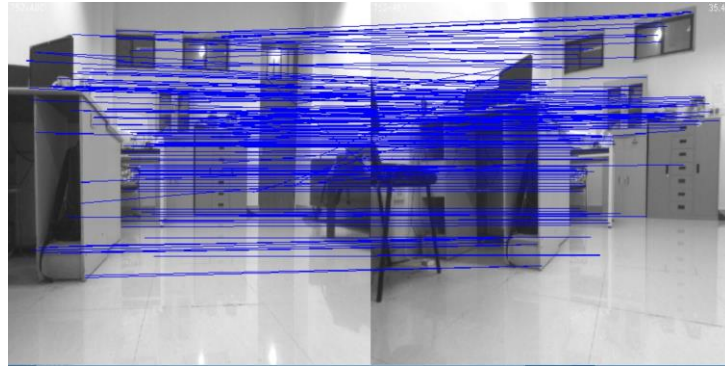


图 2-5 未剔除误匹配的匹配情况

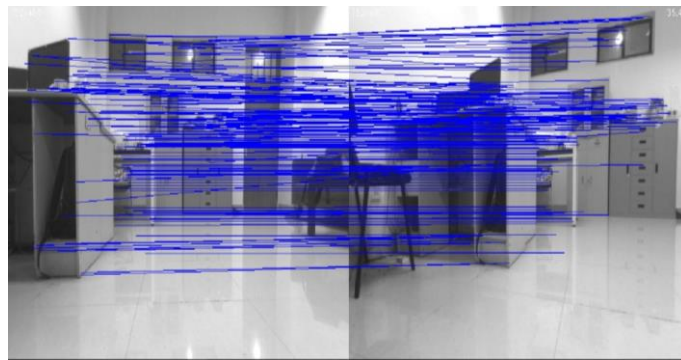


图 2-6 RANSAC 算法剔除误匹配

## 2.4 双目立体匹配测距

双目立体相机可以通过左右相机分别获取左右两幅图像，再通过极线校正利用双目视差原理，可以获取图像像素点对应的空间点  $P_c$  的深度信息，如图 2-7 所示，展示了双目立体视觉测距原理：

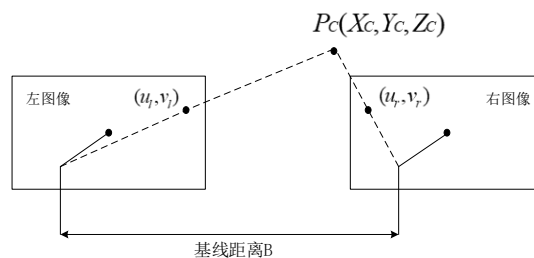


图 2-7 双目视觉测距原理图

图 2-7 中，左右相机光心的距离为基线长度  $B$ ，环境中的特征点  $P_c$  在左右两个相机投影的像素坐标分别为  $(u_l, v_l)$  与  $(u_r, v_r)$ ，物理坐标为  $p_l=(x_l, y_l)$  与  $p_r=(x_r, y_r)$ ，当左右两幅图像经极线校正后， $P_c$  在两平面内的投影点的  $y$  坐标相等，即  $y_l=y_r=y$ ，则由三角相似关系可得：

$$\begin{cases} x_l = f \frac{X_c}{Z_c} \\ x_r = f \frac{X_c - B}{Z_c} \\ y = f \frac{Y_c}{Z_c} \end{cases} \quad (2-17)$$

其中，同一极线上左右像素点之间的间距，称为视差  $\text{Disparity}=x_l-x_r$ ，因此可计算出点  $P_c$  在摄像机坐标系下的坐标为：

$$\begin{cases} X_c = f \frac{B \cdot x_l}{\text{Disparity}} \\ Y_c = f \frac{B \cdot x_r}{\text{Disparity}} \\ Z_c = f \frac{B}{\text{Disparity}} \end{cases} \quad (2-18)$$

立体匹配的目的在于估计空间特征点在双目图像对中两投影点之间的匹配关系。为得到准确的匹配点，首先要确定在相似性最大化区域进行搜索，目前大部分研究文献的匹配点搜索区域都限制在由双目图像对极几何约束的直线上。如果左右两个摄像机的坐标系平行且共面，则此双目视觉成像系统是最理想的成像系统，所得图片为理想的双目图像对，因此无需对双目图像进行校正。但真实情况下双目相机之间经常会存在旋转和平移的关系，使得所成的图像极线并不平行，需经过一定的校正才能进行立体匹配。其中，双目图像匹配对是两个摄像机从不同角度对同一场景的映射，特征点在两个成像平面上的像素点为一组对应匹配点。

一般极线校正方法主要分为两种基于相机标定的算法与不需要标定相机参数的算法。基于相机参数的极线校正算法，首先左右摄像机的内外参数矩阵及畸变参数已经通过标定得到，然后利用这些参数可以计算出图像的投影变换矩阵，最后可以校正图像得到校正后的图像。这类校正方法简单易行，但是需要对摄像机进行标定。然而，无需相机参数的标定方法根据双目图像对的几何约束关系，求出匹配点之间的本质矩阵和投影矩阵，经过重映射得到校正的图像。这类方法的应用范围广，但是匹配点的精度对校正的结果有很大的影响。因此，目前大多数的立体提匹配方法都是基于相机标定的方法进行匹配。

对图像进行极线校正后，利用特征提取与匹配算法对双目立体图像对进行匹配，如下

图 2-8 所示。立体匹配是计算机视觉中非常重要又复杂的问题。主要是利用二维图像信息获取三维特征点的深度信息，恢复真实的三维场景。如上图 2-7 所示的原理，可以计算得到特征点的三维信息。由于图像经极线校正，因此可在同一极线上搜索图像的特征匹配点，节约了图像的特征匹配时间。同一极线上得到的匹配点匹配结果如下图 2-8 所示：

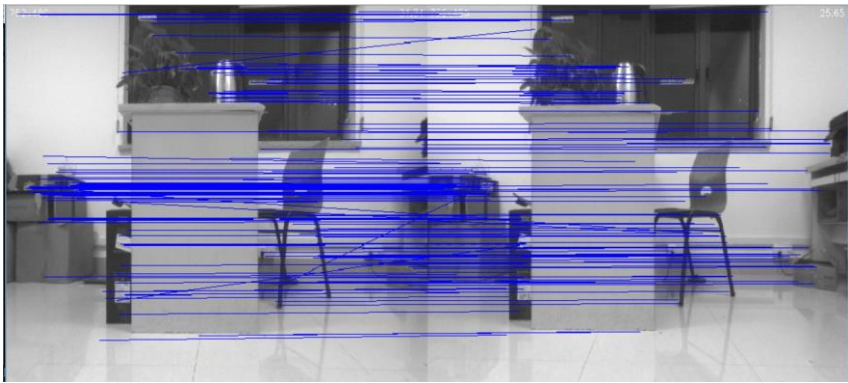


图 2-8 双目极线校正的立体匹配结果

根据立体匹配的效果，可以得到双目的匹配点的像素点坐标，然后根据上式（2-17），可以计算出图像特征点的深度信息，实验过程中可测出物体与实际双目相机中心的位置，然后利用计算得到的深度值与实际测量的深度值进行比较，选取图像中的某一特征点即可以得到视差值如下图 2-9 所示：

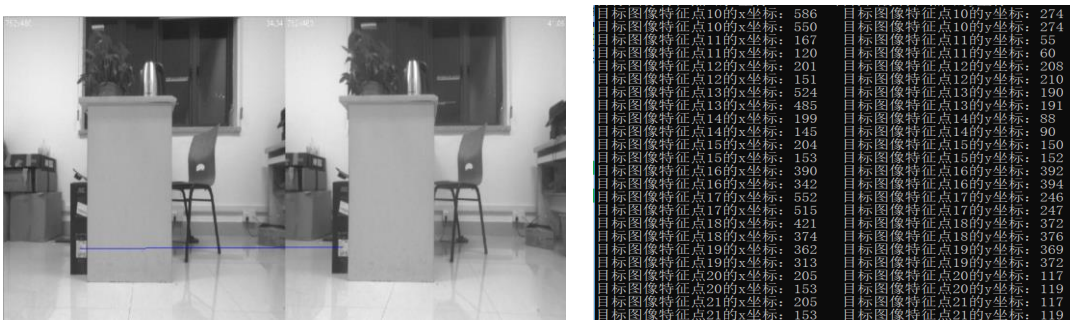


图 2-9 双目极线校正特征点立体匹配视差图

根据上图所示可求出图像特征点 20 的像素差值为 52px，根据式（2-17）与相机标定的结果可计算出深度为 108.57cm，实际测量的距离为 108.3cm，误差为 2.7mm。通过大量的特点的计算，实验结果表明，该方法可以较准确的计算特征点的深度。

## 2.5 本章小结

本章首先介绍了双目视觉中的标定原理,然后对摄像机标定的方法即标定过程进行的详细的论述,通过实验得到了双目相机的相关参数。然后,阐述了图像的特征提取与匹配,根据实验的实时性,选择了基于 ORB 的特征提取与匹配方法。最后详细介绍双目视觉的极线校正以及双目视觉的测距模型,得到特征点的三维空间坐标。

## 第3章 基于双目视觉与IMU融合的算法

### 3.1 引言

双目视觉与惯导融合的 SLAM 系统主要分为前端的视觉里程计部分和后端回环检测与图优化两部分。前端视觉里程计的功能是根据相邻两帧图像之间的匹配信息估计相机的运动，并将其估计的轨迹作为后端全局优化的初始值。首先，阐述了 IMU 的运动模型及其预积分模型，由于 IMU 测量得到的陀螺仪与加速度值存在一定的偏差，为了能更好的融合视觉获得较精准的估计值，将 IMU 的运动估计与其偏差估计分离。然后根据双目视觉初始化 IMU 的偏差，最后将 IMU 的约束加入双目视觉的位姿估计中，提高了系统位姿估计的精度，使移动机器人的位姿更加可靠。

### 3.2 双目视觉与惯导融合的 SLAM 系统框架

双目视觉与惯导融合的 SLAM 系统主要分为前端的视觉里程计部分和后端回环检测与图优化两部分。前端的视觉里程计部分，首先利用双目相机与惯导获取的采集信息，然后利用 ORB 特征提取与匹配算法获取特征点的深度信息，然后利用 IMU 的预积分模型，计算 IMU 的初始偏差，最后在双目视觉估算的位姿中加入 IMU 预积分的约束，提高位姿估计的精度。后端部分主要使对前端的估计进行回环检测与全局优化，使用基于词袋树的模型对移动机器人曾到过的地方进行识别，然后根据回环检测的结果对全局地图进行优化，最终实现双目与 IMU 融合的 SLAM 系统，系统的主要框架如下图 3-1 所示：

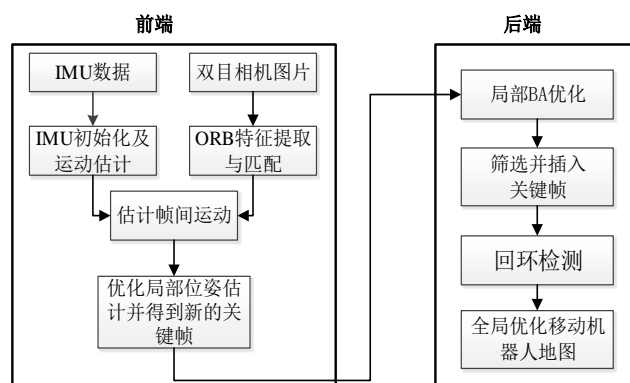


图 3-1 双目与惯导融合 SLAM 系统框架图

### 3.3 IMU 预积分模型

本系统使用的传感器为双目相机与 IMU，IMU 可以在高频运动情况下测量角速度与加速度的信息，使其在物体高速运动的情况下获取较好的估计，但受 IMU 的偏差及噪声的影响，长时间的位姿估计会产生明显的漂移误差。相反的，相机在高速运行的情况下，获取的图片模糊，无法对图像的特征信息进行描述。因此，利用双目视觉与惯导融合，可以解决移动机器人在旋转较快情况下的不足，同时提高系统的鲁棒性。IMU 与相机之间的位姿关系可以由标定获取，然后利用预积分模型将 IMU 的观测值与实际值分开，利于 IMU 中偏差的估计与优化。

#### 3.3.1 IMU 的预积分模型

本系统使用的传感器为双目相机与惯导，这两者对应的空间关系如图 3-2 所示，通常情况下，IMU 的坐标系用  $B$  表示，相机坐标系用  $C$  表示，世界坐标系用  $W$  表示。IMU 坐标系转换到世界坐标系的变换关系是  $T_{WB}=(R_{WB}, p_W)$ ， $R_{WB}$  与  $p_W$  分别是旋转矩阵与平移向量， $T_{BC}$  表示摄像机坐标系到 IMU 坐标系的变换，而这两者之间的变换关系是通过标定校准获得。

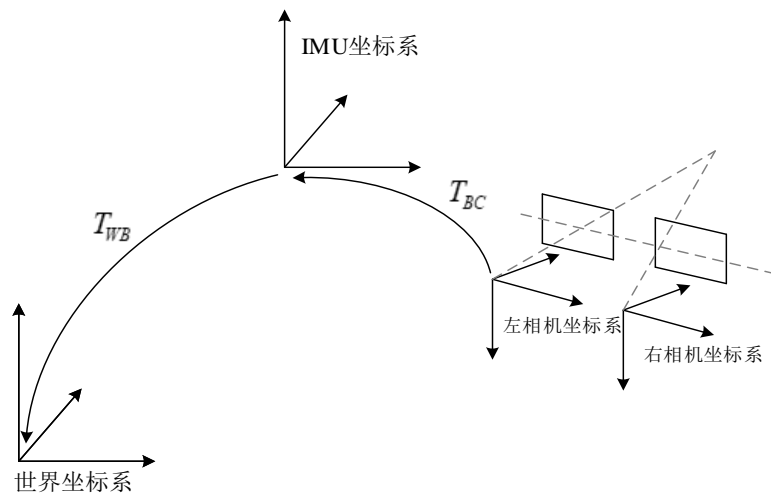


图 3-2 坐标系之间的关系

惯性测量单元最基本的元器件是陀螺仪与加速度计，所以本文中采用的是六轴惯性测量传感器，该传感器以一定的时间间隔  $\Delta t$  来采样，测量可得 IMU 的加速度与角速率分别

表示为  ${}_B\tilde{\alpha}_{WB}(t)$  与  ${}_B\tilde{\omega}_{WB}(t)$ , 其中  ${}_B\omega_{WB}(t)$  表示在 IMU 坐标系下  $B$  相对于  $W$  的瞬时角速度,  ${}_W\alpha(t)$  表示在世界坐标系下的瞬时加速度,  $R_{WB}^T$  表示世界坐标系到 IMU 坐标系的旋转,  $g_w$  表示重力加速度,  $b_g(t)$  与  $b_a(t)$  分别表示陀螺仪与加速度的偏差,  $\eta_g(t)$  与  $\eta_a(t)$  表示随机噪声, 则 IMU 的测量模型为:

$${}_B\tilde{\omega}_{WB}(t) = {}_B\omega_{WB}(t) + b_g(t) + \eta_g(t) \quad (3-1)$$

$${}_B\tilde{\alpha}_{WB}(t) = R_{WB}^T(t)({}_W\alpha(t) - g_w) + b_a(t) + \eta_a(t) \quad (3-2)$$

根据 IMU 的测量模型, 可推导传感器的运动状态, 引入运动学模型如下:

$$\begin{aligned} \dot{R}_{WB} &= R_{WB} \hat{\omega}_{WB} \\ \dot{v}_{WB} &= \alpha_{WB} \\ \dot{p}_{WB} &= v_{WB} \end{aligned} \quad (3-3)$$

由上式 (3-3) 可以推导其积分形式为:

$$R_{WB}(t + \Delta t) = R_{WB}(t) \text{Exp}\left(\int_t^{t+\Delta t} {}_B\omega_{WB}(\tau) d\tau\right) \quad (3-4)$$

$${}_Wv(t + \Delta t) = {}_Wv(t) + \int_t^{t+\Delta t} {}_W\alpha(\tau) d\tau \quad (3-5)$$

$${}_Wp(t + \Delta t) = {}_Wp(t) + \int_t^{t+\Delta t} {}_Wv(\tau) d\tau + \int_t^{t+\Delta t} \int_t^{\tau} {}_W\alpha(\tau) d\tau \quad (3-6)$$

假设  $\alpha_{WB}$  与  $\omega_{WB}$  在  $[t, t+\Delta t]$  时间内保持不变, 则上式 (3-4) - (3-6) 可以采用离散形式表示为:

$$R_{WB}(t + \Delta t) = R_{WB}(t) \text{Exp}({}_B\omega_{WB}(t) \Delta t) \quad (3-7)$$

$${}_Wv(t + \Delta t) = {}_Wv(t) + {}_W\alpha(t) \Delta t \quad (3-8)$$

$${}_Wp(t + \Delta t) = {}_Wp(t) + {}_Wv(t) \Delta t + \frac{1}{2} {}_W\alpha(t) \Delta t^2 \quad (3-9)$$

将式 (3-1) 与式 (3-2) 代入式 (3-7)、(3-8) 及式 (3-9) 可得, 其中  $\eta_{gd}(t)$  与  $\eta_{ad}(t)$  为  $\eta_g(t)$ 、 $\eta_a(t)$  的离散值:

$$R_{WB}(t + \Delta t) = R(t) \text{Exp}({}_B\tilde{\omega}_{WB}(t) - b_g(t) - \eta_{gd}(t)) \Delta t \quad (3-10)$$

$${}_Wv(t + \Delta t) = {}_Wv(t) + g_w \Delta t + R(t)(\tilde{\alpha}(t) - b_a(t) - \eta_{ad}(t)) \Delta t \quad (3-11)$$

$${}_Wp(t + \Delta t) = {}_Wp(t) + {}_Wv(t) \Delta t + \frac{1}{2} g_w \Delta t^2 + \frac{1}{2} R_{WB}(t)({}_B\tilde{\alpha}(t) - b_a(t) - \eta_{ad}(t)) \Delta t^2 \quad (3-12)$$

根据上面所述的 IMU 观测模型, 可知无法直接计算出 IMU 的偏差。由于 IMU 的频

率较快, 获取图像数据的时间内可获取多帧 IMU 数据, 因此利用两个关键帧之间的 IMU 数据求积分, 以两个关键帧的位姿变换作为约束, 计算 IMU 测量数据的偏差, 其关系如图 3-3 所示。根据式 (3-10) - (3-12) 可以计算出  $j$  时刻的状态为:

$$R_{WB}^j = R_{WB}^i \prod_{k=i}^{j-1} \text{Exp}(({}_B\tilde{\omega}_{WB}^k - b_g^k - \eta_{gd}^k)\Delta t) \quad (3-13)$$

$${}_w v_B^j = {}_w v_B^i + g_w \Delta t + \sum_{k=i}^{j-1} R_{WB}^k (\tilde{\alpha}^k - b_{\alpha}^k - \eta_{ad}^k) \Delta t \quad (3-14)$$

$${}_w p_B^j = {}_w p_B^i + \sum_{k=i}^{j-1} \left[ {}_w v_B^k \Delta t + \frac{1}{2} g_w \Delta t^2 + \frac{1}{2} R_{WB}^k (\tilde{\alpha}^k - b_{\alpha}^k - \eta_{ad}^k) \Delta t^2 \right] \quad (3-15)$$

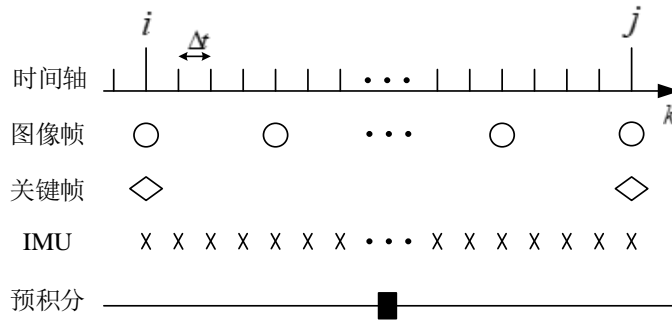


图 3-3 视觉与 IMU 预积分的运动关系

上式计算了从  $i$  时刻到  $j$  时刻由 IMU 测量得到的运动估计, 然而当  $R_{WB}^i$  改变时, 所有旋转矩阵  $R_{WB}^k$  也改变, 因此需要重新计算上式 (3-13) - (3-15), 因此相对运动增量可以定义为:

$$\Delta R_{ij} = R_{Bw}^i R_{WB}^j = \prod_{k=i}^{j-1} \text{Exp}(({}_B\tilde{\omega}_{WB}^k - b_g^k - \eta_{gd}^k)\Delta t) \quad (3-16)$$

$$\Delta v_{ij} = R_{WB}^i ({}_w v_B^j - {}_w v_B^i - g_w \Delta t_{ij}) = \sum_{k=i}^{j-1} \Delta R_{ik} (\tilde{\alpha}^k - b_{\alpha}^k - \eta_{ad}^k) \Delta t \quad (3-17)$$

$$\Delta p_{ij} = R_i^T ({}_w p_B^j - {}_w p_B^i - {}_w v_B^i \Delta t_{ij} - \frac{1}{2} \sum_{k=i}^{j-1} g_w \Delta t^2) = \sum_{k=i}^{j-1} \left[ \Delta v_{ik} \Delta t + \frac{1}{2} \Delta R_{ik} (\tilde{\alpha}^k - b_{\alpha}^k - \eta_{ad}^k) \Delta t^2 \right] \quad (3-18)$$

其中,  $\Delta R_{ik} = R_{WB}^i R_{WB}^k$ ,  $\Delta v_{ik} = R_{WB}^i ({}_w v_B^j - {}_w v_B^i - g_w \Delta t_{ij})$ ,  $\Delta v_{ij}$  与  $\Delta p_{ij}$  并不代表真正的速度与位姿的增量, 这样的表示方式使得式 (3-16) - (3-18) 中等号右边与  $i$  时刻的状态与重力都无关。上式左侧为两个关键帧( $i, j$ )的相对运动增量, 右侧为 IMU 测量数据, 因此可以理解为 IMU



测量模型，所以两个关键帧的关系可由相对运动增量表示：

$$R_{WB}^j = R_{WB}^i \Delta R_{ij} \text{Exp}(J_{\Delta R}^g b_g^i) \quad (3-19)$$

$$v_{WB}^j = v_{WB}^i + g_w \Delta t_{ij} + R_{WB}^i (\Delta v_{ij} + J_{\Delta V}^g b_g^i + J_{\Delta V}^\alpha b_\alpha^i) \quad (3-20)$$

$$p_{WB}^j = p_{WB}^i + v_{WB}^i \Delta t_{ij} + \frac{1}{2} g_w \Delta t_{ij}^2 + R_{WB}^i (\Delta p_{ij} + J_{\Delta P}^g b_g^i + J_{\Delta P}^\alpha b_\alpha^i) \Delta t^2 \quad (3-21)$$

### 3.3.2 IMU 测量的数据噪声

假设两关键帧之间 IMU 的偏差保持不变，即：

$$b_g^i = b_g^{i+1} = \dots = b_g^{j-2} = b_g^{j-1}, b_\alpha^i = b_\alpha^{i+1} = \dots = b_\alpha^{j-2} = b_\alpha^{j-1} \quad (3-22)$$

根据泰勒的一阶近似估计与李代数的相关性质，将式 (3-16) 的相对运动增量的噪声偏差分离到式子后边，则旋转的相对运动增量的近似表示为：

$$\begin{aligned} \Delta R_{ij} &\approx \prod_{k=i}^{j-1} \text{Exp}((\tilde{\omega}_{WB}^k - b_g^i) \Delta t) \text{Exp}(-J_r^k \eta_{gd}^k \Delta t) \\ &= \Delta \tilde{R}_{ij} \prod_{k=i}^{j-1} \text{Exp}(-\Delta \tilde{R}_{k+1j}^T J_r^k \eta_{gd}^k \Delta t) = \Delta \tilde{R}_{ij} \text{Exp}(\delta \varphi_{ij}) \end{aligned} \quad (3-23)$$

其中： $J_r^k = J_r^k((\tilde{\omega}^k - b_g^i) \Delta t)$ ， $\Delta \tilde{R}_{ij} = \prod_{k=i}^{j-1} \text{Exp}((\tilde{\omega}^k - b_g^i) \Delta t)$ ， $\Delta \tilde{R}_{ij}$  定义为旋转测量的预积分表达式，

同理可将  $\Delta v_{ij}$  与  $\Delta p_{ij}$  的近似化简为：

$$\begin{aligned} \Delta v_{ij} &\approx \sum_{k=i}^{j-1} \Delta \tilde{R}_{ik} (I - \delta \varphi_{ik}^\wedge) (\tilde{\alpha}_k - b_i^\alpha) \Delta t - \Delta \tilde{R}_{ik} \eta_k^{\alpha d} \Delta t \\ &\simeq \Delta \tilde{v}_{ij} - \sum_{k=i}^{j-1} [\Delta \tilde{R}_{ik} (\tilde{\alpha}_k - b_i^\alpha)^\wedge \delta \varphi_{ik} \Delta t - \Delta \tilde{R}_{ik} \eta_k^{\alpha d} \Delta t] = \Delta \tilde{v}_{ij} - \delta v_{ij} \end{aligned} \quad (3-24)$$

$$\begin{aligned} \Delta p_{ij} &\approx \sum_{k=i}^{j-1} [(\Delta \tilde{v}_{ik} - \delta v_{ik}) \Delta t + \frac{1}{2} \Delta \tilde{R}_{ik} (I - \delta \varphi_{ik}^\wedge) (\tilde{\alpha}_k - b_i^\alpha) \Delta t^2 - \frac{1}{2} \Delta \tilde{R}_{ik} \eta_k^{\alpha d} \Delta t^2] \\ &\simeq \Delta \tilde{p}_{ij} - \sum_{k=i}^{j-1} [-\delta v_{ik} \Delta t + \frac{1}{2} \Delta \tilde{R}_{ik} (\tilde{\alpha}_k - b_i^\alpha)^\wedge \delta \varphi_{ik} \Delta t^2 - \frac{1}{2} \Delta \tilde{R}_{ik} \eta_k^{\alpha d} \Delta t^2] = \Delta \tilde{p}_{ij} - \delta p_{ij} \end{aligned} \quad (3-25)$$

其中，噪声向量  $\eta_{ij}^A = [\delta \varphi_{ij}^T, \delta v_{ij}^T, \delta p_{ij}^T]^T$  近似为零均值正态分布，需要精确的计算噪声协方差。其中首先将  $\delta \varphi_{ij}$ 、 $\delta v_{ij}$  和  $\delta p_{ij}$  式子展开可迭代可以将  $\eta_{ij}^A$  表达为 IMU 的测量噪声  $\eta_{ad}^k = [\eta_{gd}^k, \eta_{ad}^k]^T$  的线性函数，因此可以根据  $\eta_{ad}^k$  的协方差，通过简单的线性迭代求出  $\eta_{ij}^A$  的协方差。

### 3.3.3 IMU 测量的偏差

假设两个关键帧之间的 IMU 偏差不变，然而在优化偏差的时候是对偏差的微小扰动

量进行优化。偏差在迭代更新的时候，需要扰动偏差不断计算，计算量大，所以采用偏差的一阶泰勒展开式进行迭代更新，因此 IMU 的相对运动增量可表示为：

$$\Delta \tilde{R}_{ij}(b_i^g) \simeq \Delta \tilde{R}_{ij}(\bar{b}_i^g) \text{Exp}\left(\frac{\partial \Delta \tilde{R}_{ij}}{\partial b_i^g} \delta b_i^g\right) \quad (3-26)$$

$$\Delta \tilde{v}_{ij}(b_i^g, b_i^\alpha) \simeq \Delta \tilde{v}_{ij}(\bar{b}_i^g, \bar{b}_i^\alpha) + \frac{\partial \Delta \tilde{v}_{ij}}{\partial b_i^g} \delta b_i^g + \frac{\partial \Delta \tilde{v}_{ij}}{\partial b_i^\alpha} \delta b_i^\alpha \quad (3-27)$$

$$\Delta \tilde{p}_{ij}(b_i^g, b_i^\alpha) \simeq \Delta \tilde{p}_{ij}(\bar{b}_i^g, \bar{b}_i^\alpha) + \frac{\partial \Delta \tilde{p}_{ij}}{\partial b_i^g} \delta b_i^g + \frac{\partial \Delta \tilde{p}_{ij}}{\partial b_i^\alpha} \delta b_i^\alpha \quad (3-28)$$

### 3.4 IMU 偏差估计算法

在双目视觉 SLAM 的初始化过程中，首先执行双目视觉的初始化，获取得到一系列关键帧后，再利用初始化的视觉信息初始化 IMU 的偏差。这样，一方面，可以使传感器充分运动观测到所有变量。另一方面，可以利用视觉信息初始化 IMU 的偏差。IMU 的偏差估计与初始化主要包括以下四个方面。

#### 3.4.1 陀螺仪偏差估计

根据视觉得到的两个连续的关键帧估计陀螺仪的偏差，假设连续关键帧的偏差保持不变，通过最小化 IMU 陀螺仪的预积分与视觉相对旋转的差异，可得陀螺仪的偏差为：

$$\arg \min_{b_g} \sum_{i=1}^{N-1} \left\| \log \left( \left( \Delta R_{ij} \text{Exp}\left(\frac{\partial \Delta R_{ij}}{\partial b_g} b_g\right) \right)^T R_{BW}^i R_{WB}^j \right) \right\|^2 \quad (3-29)$$

式中， $N$  是当前关键帧数量， $R^{(\cdot)}_{WB} = R^{(\cdot)}_{WC} R_{CB}$  是双目视觉 SLAM 算法计算得到的摄像机中心坐标  $R^{(\cdot)}_{WC}$  与摄像机与 IMU 中心之间的标定矩阵  $R_{CB}$  相乘得到的。利用高斯牛顿方法，可求解出上式。

#### 3.4.2 加速度偏差及重力加速度估计

由于双目相机可以获得特征点的深度信息，直接利用连续三个关键帧的关系与 IMU 预积分速度的测量，可得到重力加速度的初略估计。首先在世界坐标系下，左边相机坐标系与 IMU 坐标系之间的关系为：

$$P_{WB} = P_{WC} + R_{WC} P_{CB} \quad (3-30)$$

将 (3-30) 式带入 (3-20) 式得到:

$$P_{WC}^{i+1} = P_{WC}^i + v_{WB}^i \Delta t_{i,i+1} + \frac{1}{2} g_w \Delta t_{i,i+1}^2 + R_{WB}^i \Delta P_{i,i+1} + (R_{WC}^i - R_{WC}^{i+1}) P_{CB} \quad (3-31)$$

利用三个连续的关键帧线性方程, 消除关键帧的速度量, 得到方程为:

$$A(i) g_w = B(i) \quad (3-32)$$

将上式 (3-32) 展开得:

$$A(i) = \frac{1}{2} I_{3 \times 3} (\Delta t_{12}^2 \Delta t_{23} + \Delta t_{23}^2 \Delta t_{12}) \quad (3-33)$$

$$B(i) = (R_{WC}^2 - R_{WC}^1) P_{CB} \Delta t_{23} - (R_{WC}^3 - R_{WC}^2) P_{CB} \Delta t_{12} + (R_{WB}^2 \Delta P_{23} + s P_{WC}^3 - s P_{WC}^2) \Delta t_{12} \\ + R_{WB}^1 \Delta V_{12} \Delta t_{12} \Delta t_{23} - (R_{WB}^1 \Delta P_{12} + s P_{WC}^2 - s P_{WC}^1) \Delta t_{23} \quad (3-34)$$

式中,  $R_{WC}$  是摄像机中心  $C$  相对于世界坐标  $W$  的旋转矩阵,  $R_{WB}$  为 IMU 中心  $B$  相对于世界坐标的旋转矩阵,  $P_{CB}$  代表 IMU 中心  $B$  在摄像机坐标下的位置,  $\Delta t$  为两帧图像的时间差,  $v_{WB}^i$  为两帧间的相机速度。对式 (3-32) 进行奇异值分解, 即可求得重力加速度估计值  $g_w^*$ 。

由式 (3-32) 可得重力加速度的粗略估计, 但没有考虑加速度的偏差。为了提高可观测性, 将重力加速度矢量的幅值  $G$  考虑进去。假设惯性坐标系  $I$  的重力加速度方向  $\hat{g}_I = \{0, 0, 1\}$ , 则可计算出重力方向为  $\hat{g}_w = g_w^* / \|g_w^*\|$ , 利用  $R_{WI}$  来表示惯性坐标到世界坐标系的旋转矩阵, 则重力加速度可以表示为:

$$g_w = R_{WI} \hat{g}_I G \quad (3-35)$$

在惯性坐标系下, 由于  $Z$  轴和  $g_I$  平行, 所以  $R_{WI}$  旋转可以用两个角度来参数化, 同时使用微小扰动来优化旋转矩阵, 可得:

$$g_w = R_{WI} \text{Exp}(\delta\theta) \hat{g}_I G \quad (3-36)$$

$$\delta\theta = \begin{bmatrix} \delta\theta_{xy}^T & 0 \end{bmatrix}^T, \delta\theta_{xy} = \begin{bmatrix} \delta\theta_x^T & \delta\theta_y^T \end{bmatrix}^T \quad (3-37)$$

当  $\|\delta\theta\| \rightarrow 0$  时, 采取一阶近似:

$$g_w \approx R_{WI} \hat{g}_I G - R_{WI} (\hat{g}_I)_\times G \delta\theta \quad (3-38)$$

将式 (3-38) 带入式 (3-31) 得:

$$P_{WC}^{i+1} = s P_{WC}^i + V_{WB}^i \Delta t_{i,i+1} - \frac{1}{2} R_{WI} (\hat{g}_I)_\times G \Delta t_{i,i+1}^2 \delta\theta \\ + R_{WB}^i (\Delta P_{i,i+1} + J_{\Delta P}^a b_\alpha) + (R_{WC}^i - R_{WC}^{i+1}) P_{CB} + \frac{1}{2} R_{WI} \hat{g}_I G \Delta t_{i,i+1}^2 \quad (3-39)$$

同样用三个连续的关键帧消除速度变量得到方程,利用奇异值分解可求得重力加速度方向与加速度偏差,如下所示:

$$\begin{bmatrix} \phi(i) & \zeta(i) \end{bmatrix} \begin{bmatrix} \delta\theta_{xy} \\ b_\alpha \end{bmatrix} = \varphi(i) \quad (3-40)$$

其中:

$$\phi(i) = \left[ \frac{1}{2} R_{WI} \begin{pmatrix} \hat{g}_I \end{pmatrix}_X G \left( \Delta t_{12}^2 \Delta t_{23} + \Delta t_{23}^2 \Delta t_{12} \right) \right]_{(:,1:2)} \quad (3-41)$$

$$\zeta(i) = R_{WB}^2 J_{\Delta p_{23}}^a \Delta t_{12} + R_{WB}^1 J_{\Delta v_{23}}^a \Delta t_{12} \Delta t_{23} - R_{WB}^1 J_{\Delta p_{12}}^a \Delta t_{23} \quad (3-42)$$

$$\begin{aligned} \varphi(i) = & (R_{WC}^2 - R_{WC}^1) p_{CB} \Delta t_{23} - (R_{WC}^3 - R_{WC}^2) p_{CB} \Delta t_{12} + R_{WB}^2 \Delta p_{23} \Delta t_{12} + R_{WB}^1 \Delta v_{12} \Delta t_{12} \Delta t_{23} \\ & - R_{WB}^1 \Delta p_{12} \Delta t_{23} + \frac{1}{2} R_{WI} \begin{pmatrix} \hat{g}_I \end{pmatrix}_X G \Delta t_{ij}^2 - s \left( (p_{WC}^2 - p_{WC}^1) \Delta t_{23} - (p_{WC}^3 - p_{WC}^2) \Delta t_{12} \right) \end{aligned} \quad (3-43)$$

式中  $\square_{(:,1:2)}$  表示使用矩阵的前两列,利用三个连续的关键帧关系可构造线性方程  $A_{3(N-2) \times 5} X_{5 \times 1} = B_{3(N-2) \times 1}$ 。通过奇异值分解求得重力加速度方向的偏差量  $\delta\theta_{xy}$  以及加速度的偏差  $b_\alpha$ 。

### 3.4.3 速度估计

根据前两节估计出的陀螺仪与加速度的偏差,结合式(3-20),可以计算出关键帧的速度,最后一帧的关键帧速度为:

$$v_{WB}^{k+1} = v_{WB}^k + g_w \Delta t_{i,i+1} + R_{WB}^k \Delta v_{k,k+1} \quad (3-44)$$

同样可得之前关键帧的速度为:

$$v_{WB}^k = \frac{1}{\Delta t_{k,k+1}} \left( p_{WB}^{k+1} - p_{WB}^k - R_{WB}^k \Delta p_{k,k+1} - \frac{1}{2} g_w \Delta t_{k,k+1}^2 \right) \quad (3-45)$$

### 3.4.4 实验结果

根据上述的 IMU 初始化算法,利用 EuRoC 提供的数据集<sup>[61]</sup>提供的双目与 IMU 的传感器数据验证上述算法的有效性。该序列根据光照、纹理、运动快慢和运动模糊分为 11 个不同的层次的数据序列,如序列 V1\_03\_difficult (以下简称 V103) 表示快速运动及运动模糊的情况。数据集由双目相机提供的图片与 IMU 提供的角速度与加速度数据组成。本实验运行在配置为 Intel(R) Core(TM) i5 CPU 4GB RAM 的 Ubuntu 16.04 平台上。利用数据集集中的序列 V202 对 IMU 初始化过程进行分析。IMU 的初始化运行在双目视觉生成一

定关键帧的时候。初始化的过程中，计算出了陀螺仪的偏差与加速度偏差，如下图 3-4、3-5 所示，其中实线、虚线、双虚线分别代表  $x$ 、 $y$  与  $z$  轴三个方向的偏差。从图中可知，由于前两秒时间移动机器人还没有运动，导致 IMU 的加速度偏差噪声较大，同时几秒时间之后加速度偏差短时间内收敛到一定值，确保了偏差在局部范围的准确性。

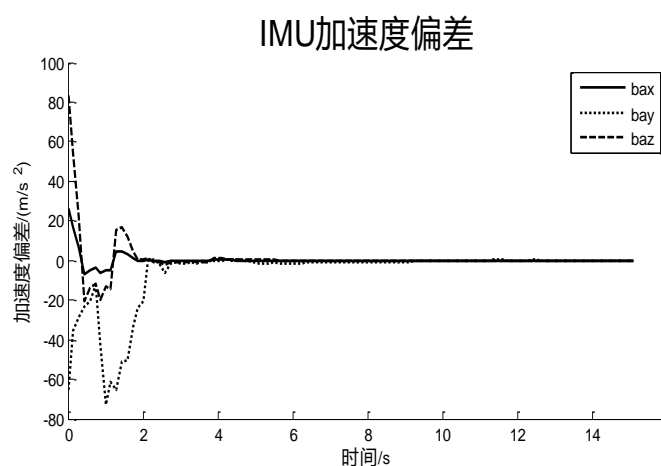


图 3-4 序列 V202 IMU 的加速度初始化偏差

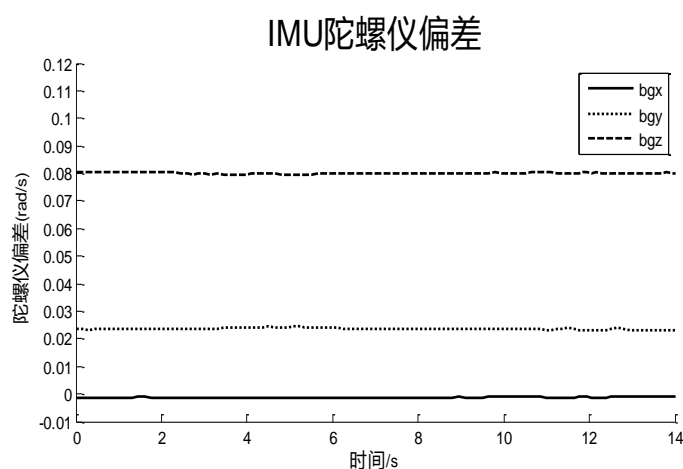


图 3-5 序列 V202 IMU 的陀螺仪初始化偏差

利用本文方法与单目视觉与惯导 SLAM 系统（以下简称 VIOSLAM）方法对 EuRoC 提供的数据集序列对 IMU 初始化求偏差过程进行分析。由于本文方法可以获取特征点深度信息，相比于本文方法在初始化过程中降低了计算复杂度，缩短了初始化时间，具体时间如下表 3-1 所示。通过下面的表格可知，本文方法在快速运动及模糊的情况下，IMU 初始化的时间相比于 VIOSLAM 系统大致缩减了 44%。

表 3-1 各系统耗时比较（单位：秒）

方法	V101	V102	V103	V201	V202	V202
本文方法	15.0	15.0	15.1	15.0	15.0	15.0
VIOSLAM	15.05	15.05	27.3	15.05	15.05	27.5

### 3.5 基于双目视觉与 IMU 融合的位姿估计

通过双目相机前后帧之间的特征点匹配,可以获取几组特征匹配点以及对应的特征点匹配的图像像素坐标信息。由于本文采用的是双目相机,因此可以得到特征点的三维位置信息。根据得到的三维特征点的以及图像的二维投影坐标可以通过 PnP 方法求解相机之间的位姿估计。对于 PnP 问题的求解有多种方法,如直接线性变换(DLT)、P3P、RPnP、EPnP(Efficient PnP)等方法,本文双目视觉的位姿估计利用的就是 EPnP 的求解方法。

#### 3.5.1 基于双目视觉的位姿估计算法

根据特征匹配算法匹配得到的匹配点,选取四对不共面的匹配点可以求解移动机器人之间的位姿变化。根据相机标定模型可以得到摄像机的图像像素坐标与对应的三维点的坐标关系可以表示为:

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i^c \\ y_i^c \\ z_i^c \end{bmatrix} = A p_i^c \quad (3-46)$$

其中 $\lambda_i$ 是标量投影参数, $A$ 为对应的摄像机内参, $u_i$ 与 $v_i$ 为对应的二维图像像素坐标,其中 $i=0,1,\dots,n$ , $p_i^c$ 为摄像机坐标系下的坐标,设匹配点在世界坐标系下的坐标为 $p_i^w$ ,利用在世界坐标系下四个控制点 $c_j^w$ 来表示匹配点的世界坐标:

$$p_i^w = \sum_{j=1}^4 \beta_{ij} c_j^w, \text{ with } \sum_{j=1}^4 \gamma_{ij} = 1 \quad (3-47)$$

其中,向量 $[\beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}]$ 为参考点 $p_i^w$ 以控制点 $\{C_i^c | i=1,2,3,4\}$ 为基的欧氏空间的坐标, $j=1,2,3,4$ 。同理,在摄像机坐标系下,匹配点的摄像机坐标也可以由摄像机的四个参考点坐标 $c_j^c$ 表示:

$$p_i^c = \sum_{j=1}^4 \beta_{ij} c_j^c, \text{ with } \sum_{j=1}^4 \gamma_{ij} = 1 \quad (3-48)$$

将式(3-48)代入式(3-46)可得:

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^4 \gamma_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \quad (3-49)$$

将上式 (3-49) 展开, 可得:

$$\lambda_i = \sum_{j=1}^4 \gamma_{ij} z_j^c \quad (3-50)$$

$$\sum_{j=1}^4 \gamma_{ij} f_u x_j^c + \gamma_{ij} (u_c - u_i) z_j^c = 0 \quad (3-51)$$

$$\sum_{j=1}^4 \gamma_{ij} f_v y_j^c + \gamma_{ij} (v_c - v_i) z_j^c = 0 \quad (3-52)$$

将相机坐标系下的四个控制点展开成列向量, 同时将  $n$  个参考点坐标代入上式可得到方程为:

$$Mx = 0 \quad (3-53)$$

式中,  $M$  是  $2n \times 12$  的矩阵, 则可将  $x$  向量的解, 表示为:

$$x = \sum_{i=1}^N \beta_i v_i \quad (3-54)$$

其中  $v_i$  为  $M$  的右奇异向量, 可以通过求解矩阵  $M^T M$  的零空间特征值取得。因为使用  $M^T M$  求解的计算复杂度只有  $O(n)$  比求解  $M$  小很多, 所以用  $M^T M$  求解, 而不使用  $M$  求解。

$$\text{res} = \sum_i \text{dist}^2 \left( A[R|t] \begin{bmatrix} P_i^w \\ 1 \end{bmatrix}, u_i \right) \quad (3-55)$$

要求解  $x$ , 需要确定  $\beta_i$ , 即需要确定合适的线性组合, 根据得到的参考点不同, 矩阵  $M^T M$  的零空间维度可能为  $N=1,2,3,4$ , 根据维度的不同, 利用两两控制点在世界坐标系或者相机坐标系下的距离相同, 可以求解出  $\beta_i$ 。利用求解出的  $\beta_i$ , 可以得到摄像机坐标系下的控制点坐标, 利用式 (3-26) 可以求解出匹配点在相机坐标系下的坐标  $p_i^c$ 。这样可以通过求解世界坐标系下的匹配点的坐标  $p_i^w$  与相机坐标系下匹配点的坐标  $p_i^c$  的旋转与平移变换表示两帧图像之间的位姿变换, 具体的求解步骤如下:

- 1) 求中心点:  $p_c^c = \frac{\sum p_c^i}{N}, p_w^c = \frac{\sum p_w^i}{N};$
- 2) 去中心:  $q_c^i = p_c^i - p_c^c, q_w^i = p_w^i - p_w^c;$
- 3) 计算  $H$  矩阵:  $H = \sum_{i=1}^N q_c^i q_w^{iT};$

- 4) 对  $H$  矩阵进行 SVD 分解:  $H=U\Sigma V^T$ ,  $R=VU^T$ ,  $t=p_c^c-Rp_w^c$ , 如果  $\det(R)<0$ ,  $R(2,:) = -R(2,.)$ 。

### 3.5.2 基于视觉与 IMU 融合的位姿优化算法

通过 IMU 预积分的模型可以预测当前帧的位姿, 并将该预测值作为位姿优化的初始值。预估过程需要上一帧的位姿以及对应 IMU 的速度、陀螺仪偏差和加速度计偏差, 然后通过当前帧位姿估计的初始值对旋转、位姿及速度进一步的位姿估计和优化, 所以 IMU 中待估计的状态可表示为:

$$\theta = \{R_{WB}^j, p_{WB}^j, v_{WB}^j, b_g^j, b_a^j\} \quad (3-56)$$

根据 IMU 的状态量与相机的位姿, 得到待优化的状态量为:

$$\theta^* = \arg \min_{\theta} (\sum_k E_{proj}(k, j) + E_{IMU}(i, j)) \quad (3-57)$$

式中:  $E_{proj}(k, j)$  表示特征点的重投影误差;  $E_{IMU}(i, j)$  表示 IMU 状态估计的误差。

重投影误差优化是利用双目位姿估计算法计算出的位姿初始值后, 对这些位姿进行优化减小或消除其中的误差。优化过程利用的是 Bundle Adjustment, 该方法主要是对真实三维特征点在图像平面上的投影与双目计算出的三维特征点的重投影虚拟像素之间的差值进行优化。如图 3-6 所示, 双目匹配得到的三维点  $P_C$ , 与三维点在双目左边图片前后帧的投影点位  $p_1$  与  $p_2$ ,  $p'$  是重投影计算出的二维像素点坐标:

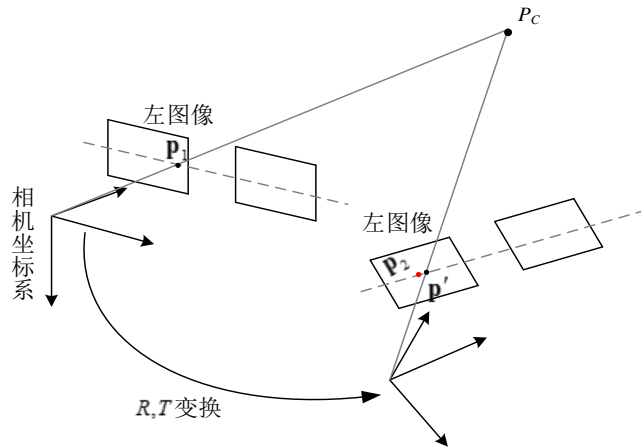


图 3-6 前后两帧图像重投影误差

重投影的具体优化过程如下所述, 根据 IMU 坐标系与相机坐标系及世界坐标系的关系



系，可得第  $k$  个特征点的坐标表示为：

$$X_C^k = R_{CB} R_{BW}^i (X_W^k - p_{WB}^i) + p_{CB} \quad (3-58)$$

根据得到的匹配点的空间坐标，可利用投影函数  $\pi$  函数表示第  $i$  个匹配点的在左相机的投影方程，则第  $i$  个点的投影方程表示为：

$$\pi_{lr}(x_c^i) = \begin{bmatrix} f_x \frac{X}{Z} + u_0 \\ f_x \frac{Y}{Z} + v_0 \\ f_x \frac{(X - B)}{Z} + u_0 \end{bmatrix} \quad (3-59)$$

根据重投影误差的概念，可以得到误差函数为：

$$E_{proj}(k, j) = \sum_n \rho \left( \|p^i - \pi_{lr}(X_C^i)\|^2 \right) \quad (3-60)$$

根据 IMU 带优化的状态量，可得两个关键帧之间的 IMU 数据的误差项可以表示为：

$$E_{IMU}(i, j) = \rho \left( [e_R^T e_v^T e_p^T] \sum_l [e_R^T e_v^T e_p^T]^T \right) + \rho \left( e_b^T \sum_R e_b \right) \quad (3-61)$$

式中， $\sum_l$  表示预积分的信息矩阵； $\sum_R$  表示偏差的随机游走矩阵。其中根据 IMU 预积分模型可以得到 IMU 状态的估计误差，其中旋转的误差表示为：

$$e_R = \log \left( \left( \Delta R_{i,j} \text{Exp}(J_{\Delta R}^g b_g^i) \right)^T R_{BW}^i R_{WB}^j \right) \quad (3-62)$$

速度的误差表示为：

$$e_v = R_{BW}^i \left( {}_w v_B^j - {}_w v_B^i - g_w \Delta t_{ij} \right) - \left( \Delta v_{ij} + J_{\Delta v}^g b_g^j + J_{\Delta v}^\alpha b_\alpha^j \right) \quad (3-63)$$

位姿的误差表示为：

$$e_p = R_{BW}^i \left( {}_w p_B^j - {}_w p_B^i - \frac{1}{2} g_w \Delta t_{ij}^2 \right) - \left( \Delta p_{ij} + J_{\Delta p}^g b_g^j + J_{\Delta p}^\alpha b_\alpha^j \right) \quad (3-64)$$

偏差的误差为：

$$e_b = b^j - b^i \quad (3-65)$$

对上述问题的求解采用图优化过程中的高斯牛顿法进行求解。由于在双目视觉位姿估计的基础上，将 IMU 预积分的约束加入了到了优化的过程中，提高了位姿估计的精度。同时，利用 IMU 的预积分模型估计出的当前帧的位姿，在移动机器人旋转较快或者跟踪失败的情况下，可以利用 IMU 的预测位姿重新对当前帧的特征点进行提取，使得系统的跟踪能鲁棒的进行下去。

### 3.6 本章小结

本章首先介绍了 IMU 的运动模型及预积分模型，同时对利用 IMU 的与积分模型将 IMU 中的偏差与噪声的推导模型进行了详细的说明。然后利用双目视觉获取的关键帧初始了 IMU 中陀螺仪的偏差与加速度的偏差及 IMU 的速度。接着详细的介绍了双目视觉的位姿估计算法，利用双目视觉位姿估计估计算法可以得到视觉估计的初始位姿。最后，利用重投影优化位姿估计，同时加入预积分获得的 IMU 约束，进一步优化了移动机器人的位姿估计。

## 第4章 基于图优化的视觉 SLAM 算法

### 4.1 引言

在视觉 SLAM 系统中，后端的回环检测与全局优化具有重要意义。回环检测通过识别移动机器人之前曾到过的地方，来保证长时间运动过程中移动机器人的准确性。正确的回环检测能为全局优化提供更准确的数据，保证全局位姿估计的一致性。同时，回环检测还能在跟踪丢失后进行重定位。回环检测后通过图优化优化位姿之间的约束关系，减小长时间运动的移动机器人的轨迹的累计误差。最初，SLAM 的优化问题的大多采用基于滤波的方法。然而在认识到 SLAM 问题稀疏性以后，图优化的出现为大规模的 SLAM 优化问题提供了高效的求解方法。

### 4.2 SLAM 中的图模型

在 SLAM 问题中移动机器人的定位与环境地图的构建是相辅相成的，地图构建是根据传感器采集的信息构建移动机器人环境所处的环境的地图。目前移动机器人构建的地图主要分为概率网格地图<sup>[62]</sup>、特征地图以及位姿图<sup>[63]</sup>。概率地图将机器人所在环境分割成许多独立的区域与网格，这类地图适合室内的密集场合，广泛应用在避障与路径规划算法。但该地图与实际环境地图有较大的不一致性，不能完全表现移动机器人所处的环境。特征地图通过移动机器人观测的路标模型来表示。该方法适合路标明显的环境，利用相机获取的视觉特征，可以建立环境的稀疏地图。除了上面所描述的两种地图，还有一种更适合于抽象的地图表示方式——位姿图。位姿图通过节点与边来表示移动机器人的运动轨迹，其中的节点表示移动机器人在关键帧的位姿，边表示节点之间的约束条件，如图 4-1 所示：

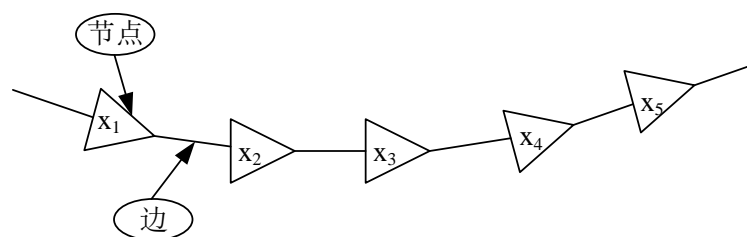


图 4-1 SLAM 中的位姿图模型

根据上面的描述与前端位姿估计计算出的相邻两帧的位姿变换,可以将位姿变换表示成位姿图的形式。这样可以将所有时刻的机器人的位姿加入到图模型中,但是这样将会导致节点的增多,数据量增大,实时性不好。因此,引用关键帧的方法作为机器人的位姿节点,加入位姿图模型中。

关键帧的选取大致需要满足三个原则:保证较好的定位精度、位姿轨迹的精度以及根据不同场景确定帧间的距离。本文由于融合了 IMU 的数据,因此关键帧选取的间隔时间较短,在不超过一定阈值的时间内将该帧判定为关键帧。

### 4.3 回环检测

在相机姿态计算过程中,由于只用到上一帧的数据进行计算,所以计算的结果会存在误差。典型的表现是相机运动了一圈之后又回到原来到过的地方,估计的相机位姿的轨迹会与初始估计的位姿存一定的误差。而回环检测的任务是让系统在回到原来曾到过的地方时认出这个曾经来过的地点,利用图像帧之间的约束关系,消除系统所存在的累计误差,如图 4-2 所示。传统的方法是利用图像检索的方式,搜索对比图像信息,若图像一致,则认为该点被重复访问。SLAM 中的回模型主要包括三种:无回环、局部回环及大回环,如下图 4-3 所示,但是在移动机器人的运动过程中,要尽量避免无回环的情况。因为,无回环的情况下,无法准确的消除移动机器人的累计误差,使估计轨迹的轨迹不能进行全局的优化。

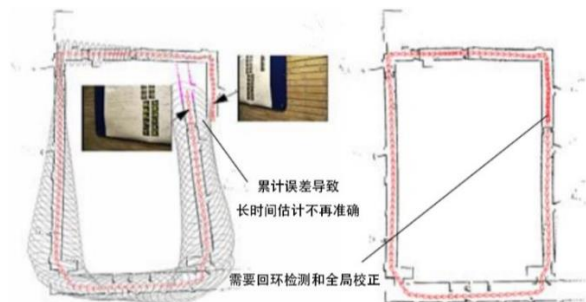


图 4-2 回环检测

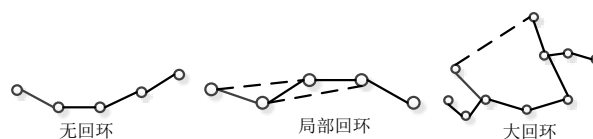


图 4-3 SLAM 中的回环模型

目前主流的闭环检测方法主要有基于图像特征匹配的方法与基于词袋模型的方法。基于图像特征匹配的方法是将当前帧与之前获取的关键帧进行特征匹配,然后根据图像匹配的结果,判断是否发生了闭环。然而该算法随着图像帧数的增多,匹配数量也随之增大,但这样会导致算法运行时间增大,实时性较差。为了解决实时性问题,可以从关键帧中选取一个较小的子集进行匹配,但是这依赖于子集的选取。因此, Nister 等人<sup>[64]</sup>于 2006 年提出了一种基于词袋模型的图像检索方法,该方法使用树形结构加速 BoW 的检索过程,提高了搜索性能。现在主流的 SLAM 系统,大都采用基于词袋模型的回环检测方法。

本文采用的是基于 DBow2 (Bags of binary words for fast place recognition in image sequence) 的回环检测算法,该算法使用 ORB 算法提取图像特征,然后对这些图像特征采用 K-means 的算法聚类,生成词典树模型,建立离线词典。词典树构建后将每一帧中的特征点描述子与词典中的描述子相比较,计算两幅图像帧的相似度,来判断是否存在闭环,具体的实现过程如下:

1. 首先采用 FAST 特征检测算法检测特征向量与 BRIEF 描述子描述这个特征向量 (在 2.3 节有详细介绍);
2. 利用 K-means 聚类方法将获取的特征点样本进行聚类,将聚类得到的节点样本,再各自聚集成  $k$  类,依次类推直到达到设定的层数,形成词典树,如图 4-4 所示:

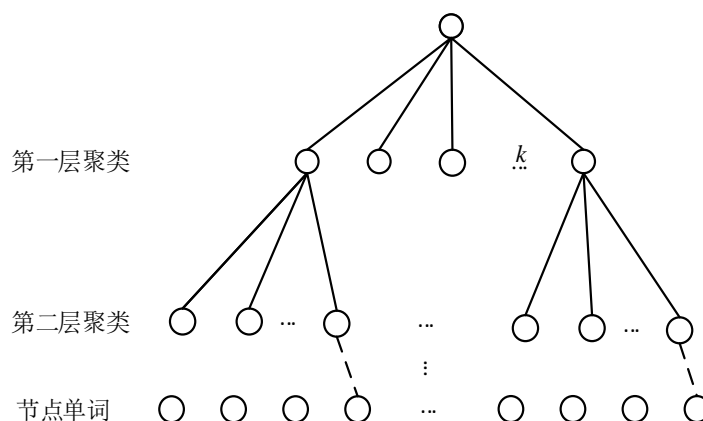


图 4-4 词典树

3. 词典树构建过程中,需要对每个叶子节点赋予一个权重  $\omega$ ,该权重代表了单词出现的频率,权重越低,代表该词出现频率越低区分度越小,其中权重的计算公式为:

$$\omega_i = \log \frac{N}{n_i} \quad (4-1)$$

式中： $N$  代表图像的数量， $n_i$  为图像中单词出现的次数；

4. 将新得到的关键帧与之前一定数量间隔的关键帧进行相似度计算，其中，使用  $\mu_1$ 、 $\mu_2$  为图像的描述向量；两幅图像之间的相似度计算为：

$$s(\mu_1, \mu_2) = 1 - \frac{1}{2} \left| \frac{\mu_1}{|\mu_1|} - \frac{\mu_2}{|\mu_2|} \right| \quad (4-2)$$

5. 如果相似度超过一定阈值，则将该帧与之前的关键帧都进行特征匹配，匹配的内点数超过一定阈值则说明检测到回环，具体流程如下图 4-5 所示：

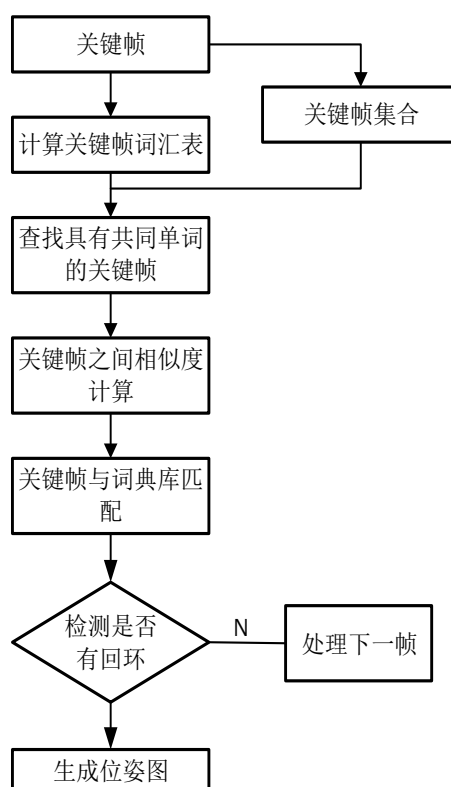


图 4-5 回环检测流程

为了检测移动机器人到了之前曾到过的地方，通过将视觉词典图像词典转换为视觉词袋向量，然后利用图像数据库对图像进行索引，利用图像数据库的方法提高图像的检索速度。图像数据库是由视觉词典树、顺序索引及逆序索引组成，视觉单词是位于上述视觉词典的叶子节点上，而顺序索引是存了图像特征和它们关联的节点，逆序索引存储的则是每幅图像的单词的权重。这样加速了图像帧之间的相似度的测量，从而加快了检测是否有回环的效率。

图像数据库存储了所以给定的图像，通过检索可以得到与给定图像相似的图像。在获取到当前的图像之后，首先将图像转换为词袋向量，然后利用图像数据库查询词袋向量，根据相似度的计算结果，可以得到一系列的候选匹配，然后采用期望得到的最好分数去标准化相似度得分，得到归一化的相似度得分。得到一组最佳匹配分组后，需要通过之前查询的结果检测其时间一致性。即如果在某一时刻存在回环，则在下一段时间内，回环是稳定存在的。时间一致性检测后，需要对得到的回环的候选图像进行几何一致性检测。几何一致性检测是判断两幅图像之间的内点数是否超过 12 个，超过说明候选图像可以作为回环检测图像。通过上述的检验与步骤，可以有效准确地检测出闭环。

#### 4.4 位姿图优化

基于图优化的方法是通过传感器的数据与传感器数据之间的空间约束关系，估计出移动机器人的位姿关系构建出位姿图，然后对位姿图中的节点进行优化。在对位姿图的优化中，最后归结为非线性误差函数最小化的问题。通常求解最小二乘的方法有高斯牛顿法（Gauss-Newton，即 GN）与列文伯格马奎尔特（Levenberg-Marquardt, 即 LM）方法最为常用。然而，仅仅利用这些方法使得 SLAM 的优化性能需要具备专业的领域的知识。然而，通用图优化 g2o 的出现很好的解决了这个问题<sup>[65]</sup>。g2o 提供了基于位姿图的节点与边的基类函数与各类求解器，使得表示成图的优化问题可以通过该框架进行求解，通用图优化的实现框架如下图 4-6 所示。

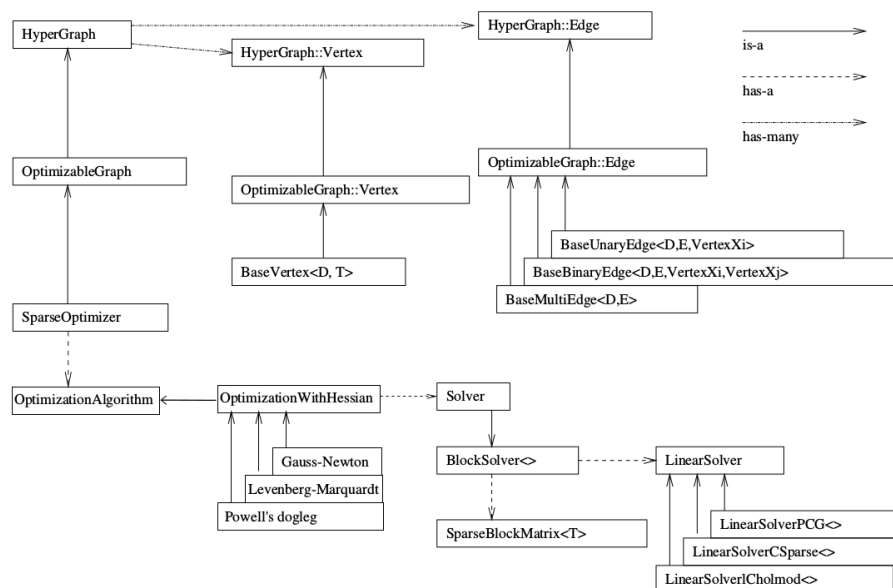


图 4-6 g2o 实现框架

#### 4.4.1 基于滤波优化方法的 SLAM 后端优化

SLAM 中经典的数学模型可以定义为：

$$\begin{cases} x_k = f(x_{k-1}, u_k) + w_k \\ z_{k,j} = h(y_j, x_k) + v_{k,j} \end{cases} \quad (4-3)$$

其中  $x_k$  代表相机位姿，也为图优化的一个节点， $y_j$  代表  $k$  时刻观察到的特征点坐标， $z_{k,j}$  表示位姿节点之间的约束关系， $w_k$  和  $v_{k,j}$  分别表示过程和测量的噪声，且都服从高斯分布。为简化表示，将  $k$  时刻下的未知量都表示为  $x_k$ ，则式 (4-3) 可表示为：

$$\begin{cases} x_k = f(x_{k-1}, u_k) + w_k \\ z_k = h(x_k) + v_k \end{cases} \quad (4-4)$$

将  $k-1$  时刻的均值和协方差矩阵记作  $\hat{x}_{k-1}$  和  $\hat{P}_{k-1}$ ，将运动和观测方程进行线性化处理，可得：

$$x_k \approx f(\hat{x}_{k-1}, u_k) + \left. \frac{\partial f}{\partial x_{k-1}} \right|_{\hat{x}_{k-1}} (x_{k-1} - \hat{x}_{k-1}) + w_k \quad (4-5)$$

记偏导数为：

$$F = \left. \frac{\partial f}{\partial x_{k-1}} \right|_{\hat{x}_{k-1}} \quad (4-6)$$

同理可得观测方程：

$$z_k \approx h(\bar{x}_k) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}_k} (x_k - \bar{x}_k) + n_k \quad (4-7)$$

记偏导数为：

$$H = \left. \frac{\partial h}{\partial x} \right|_{\bar{x}_k} \quad (4-8)$$

可得协方差  $P$  为：

$$P(x_k | x_0, u_{1:k}, z_{0:k-1}) = N(f(\hat{x}_{k-1}, u_k), F\hat{P}_{k-1}F^T + R_k) \quad (4-9)$$

记先验的均值为：

$$\bar{x}_k = f(\hat{x}_{k-1}, u_k) \quad (4-10)$$

协方差的均值为：

$$\bar{P}_k = F\hat{P}_{k-1}F^T + R_k \quad (4-11)$$

在观测中可得



$$P(z_k | x_k) = N(h(\bar{x}_k) + H(x_k - \bar{x}_k), Q_k) \quad (4-12)$$

定义卡尔曼增益  $K_k$  为:

$$K_k = \bar{P}_k H^T (H \bar{P}_k H^T + Q_k)^{-1} \quad (4-13)$$

推导可得后验概率的形式为:

$$\begin{cases} \hat{x}_k = \bar{x}_k + K_k(z_k - h(\bar{x}_k)) \\ \hat{P}_k = (I - K_k H) \bar{P}_k \end{cases} \quad (4-14)$$

虽然相比于非线性优化方法, 基于滤波器的方法存在, 但是在机器人移动的过程中, 需要不断的估计自身的位置加入系统的状态向量中, 这样增加了计算的复杂度, 需要维护的状态向量的维数也会增加。在系统的非线性情况严重, 算法的估计精度将会降低。

#### 4.4.2 基于位姿图优化方法的 SLAM 后端优化

由前端双目视觉估计出的位姿, 设移动机器人在两个不同时刻的位姿分别为  $x_k$  与  $x_{k+1}$ , 可以构建出机器人在两个不同时刻的误差的目标函数可表示为:

$$e(x_k, x_{k+1}) = z_{k,k+1} - \hat{z}_{k,k+1}(x_k, x_{k+1}) \quad (4-15)$$

根据构建的误差函数, 该误差函数服从高斯分布, 可以得出 SLAM 问题中的优化目标函数为:

$$F(x) = \sum_{x_k, x_{k+1}} e(x_k, x_{k+1})^T \Omega_{k,k+1} e(x_k, x_{k+1}) \quad (4-16)$$

式中,  $\Omega_{k,k+1}$  表示节点之间的信息矩阵 (即协方差的逆), 最后将 SLAM 的优化问题变为求目标函数最小值的问题, 即为:

$$x^* = \arg \min_x F(x) \quad (4-17)$$

假设机器人的初始位姿  $x$  已知, 则误差函数在初始位姿下的一阶泰勒近似的展开式为:

$$e_{k,k+1}(x + \Delta x) \simeq e_{k,k+1}(x) + J_{k,k+1} \Delta x \quad (4-18)$$

式中,  $J_{k,k+1}$  表示为误差在  $x_k$  出的雅克比矩阵, 其中  $J_{k,k+1} = \partial e_{k,k+1}(x) / \partial x$ , 将上式代入式 (4-17) 可得误差项的表示为:

$$\begin{aligned} F_{k,k+1}(x + \Delta x) &= e(x + \Delta x)^T \Omega_{k,k+1} e(x + \Delta x) \\ &= (e_{k,k+1}(x) + J_{k,k+1} \Delta x)^T \Omega_{k,k+1} (e_{k,k+1}(x) + J_{k,k+1} \Delta x) \\ &= (e_{k,k+1}(x))^T \Omega_{k,k+1} e_{k,k+1}(x) + 2(e_{k,k+1}(x))^T \Omega_{k,k+1} J_{k,k+1} \Delta x + \Delta x^T J_{k,k+1}^T \Omega_{k,k+1} J_{k,k+1} \Delta x \end{aligned} \quad (4-19)$$

由上式可知，该式子为 $\Delta x$ 的二次函数，对 $\Delta x$ 求微分后等于0，因此可得 $\Delta x$ 的最小值 $\Delta x^*$ 。

在利用图优化来求解上述优化问题时，使用李代数与李群表示来表示相机在关键帧处的位姿，则两个关键帧之间误差函数可以表示为：

$$\begin{aligned} e_{ij} &= \ln(T_{ij}^{-1}T_i^{-1}T_j)^{\vee} \\ &= \ln(\exp((- \xi_{ij})^{\wedge})\exp((- \xi_i)^{\wedge})\exp(\xi_j^{\wedge}))^{\vee} \end{aligned} \quad (4-20)$$

式中： $\xi_i$ 和 $\xi_j$ 表示李代数中关键帧的位姿， $T_i$ 和 $T_j$ 表四李群中关键帧的位姿，根据李代数中的求导方法，可以关键帧的位姿分别加入扰动： $\delta \xi_i$ 和 $\delta \xi_j$ ，则可得误差为：

$$\hat{e}_{ij} = \ln(T_{ij}^{-1}T_i^{-1}\exp((- \delta \xi_i)^{\wedge})\exp(\delta \xi_j^{\wedge})T_j)^{\vee} \quad (4-21)$$

根据李代数的求导法则，可以得到误差关于两个位姿的雅克比矩阵，这样可以将所有的位姿顶点与位姿之间的转换关系构成的一条边，组成一个图优化问题，各个顶点的位姿作为优化的变量，位姿间的观测作为约束关系，采用高斯牛顿法或者列文伯格—马夸尔特方法对此优化问题进行求解，主要利用之所述的 g2o 进行求解。

## 4.5 本章小结

本章首先介绍了 SLAM 中的几种表示图的模型，为了简明抽象的表示地图，采用了位姿图的形式。接下来介绍了回环检测的含义及检测回环的方法，主要利用了基于词袋模型的回环检测方法，提高了图像的检索效率，同时能准确高效的识别出移动机器人曾到过的地方。最后介绍了通用图优化求解的框架模型与基于图优化问题的 SLAM 求解方法，使用位姿图的建立了位姿优化的数学模型，并利用所述图优化框架对 SLAM 系统的全局优化问题进行求解。

## 第 5 章 实验设计与分析

### 5.1 引言

前几章本文主要介绍了双目测距模型、IMU 的预积分模型及 IMU 初始化的过程、双目视觉融合 SLAM 的前端位姿估计以及后端的回环检测与优化。本章将通过下面的实验验证双目与惯导信息融合的 SLAM 系统算法的有效性。本文设计了两组对比实验，验证本文提出的双目与 IMU 融合的 SLAM 算法，以及双目 SLAM 求解移动机器人位姿的算法进行测试。通过对比体现出本文算法的优越性及实用性。为展现本文算法广泛适用性使用了 EuRoC 提供的开源数据集。

### 5.2 实验平台介绍

为了检测本文所设计的双目与惯导融合的 SLAM 算法的有效性，设计了三组对比实验，其中第一组采用本文提出的基于双目与 IMU 融合的 SLAM 算法，第二组采用的基于双目视觉的 SLAM 算法，第三组采用基于单目与惯导 SLAM 系统。在实验数据集方面，使用了 EuRoC 提供的开源数据集序列中的 V2\_02\_medium（简称“V202”）。该数据集由微型飞行器在两个不同的厂房采集获取的双目图片集与 IMU 数据集，图片由全局立体快门双目相机以 20 Hz 的频率获取，IMU 数据是由 ADISI 传感器以 200Hz 获取，其中真实轨迹由运动系统捕捉。该数据集，根据不同的光照、纹理及飞行的快慢，将数据序列分为简单、中等及困难三个等级，其中部分数据序列最高速度超过了 2m/s，数据集的图片分辨率为 756×480。因此，该数据集可以用来评估本文 SLAM 系统的有效性。

表 5-1 V2\_02\_medium 序列绝对轨迹误差（单位：米）

双目相机参数名称	双目相机参数量
帧率	25fps
图像大小	752x480
基线距离	120mm
IMU 采集频率	200Hz

本文还通过如图 5-1 所示的平台实时采集了双目图片集与 IMU 的惯导数据，该平台

由双目惯导相机如图 5-2 所示、Ubuntu 小车及笔记本电脑组成，双目相机及 IMU 的参数如下表 5-1 所示：



图 5-1 实验平台



图 5-2 双目惯导相机

本实验在配置为 Intel(R) Core(TM) i5 CPU 4GB RAM 的笔记本电脑上进行，运行的操作系统为 64 位的 Ubuntu 16.04。实验算法采用 C++ 语言进行编写，所用到的开源函数库主要包括：开源的 C++ 矩阵运算库 Eigen，用于支持各种矩阵和矢量的运算；开源的跨平台计算机视觉库 OpenCV，实现了计算机视觉与图像处理的相关算法；支持多语言，跨平台的框架 OpenNI，主要用来实现音频和视觉传感器之间的通信；开源点云库 PCL，实现了大量的关于三维点云的高效数据结构和通用算法。

### 5.3 实验算法设计

介绍完实验平台之后，在这里要详细介绍下实验的算法流程。本次实验使用的是 EuRoc 采集的双目与 IMU 传感器的数据集对双目与 IMU 融合的 SLAM 算法与双目 SLAM 算法进行了比较。通过两者的对比试验体现出本文算法在位姿估计精度的准确性及鲁棒性的优势。

首先要介绍的是基于双目与惯导融合的 SLAM 算法：

- 1) 初始化：读取双目图片与 IMU 陀螺仪与加速计的速度。
- 2) 提取左右图片的 ORB 特征并进行特征匹配，计算出特征点的深度。
- 3) 利用双目相机初始获取的一系列关键帧初始化 IMU 的偏差。
- 4) 利用 EpnP 算法求解移动机器人之间的位姿；
- 5) 将 IMU 预积分模型预测的位姿与双目估计的位姿，进一步优化得到的位姿；

- 6) 根据获取的关键帧，采用特征匹配的方法检测局部回环及优化。
- 7) 利用基于视觉词袋模型计算图像间的相似度，即进行全局回环检测。
- 8) 利用图优化对检测到的全局大回环进行优化。
- 9) 更新移动机器人位姿和三维地图
- 10) 返回第一步继续执行，直到读取完所有的实验图像。
- 11) 得到最终的移动机器人运动轨迹和完整的三维地图。

将基于双目与 IMU 融合的 SLAM 算法的第 3 步与第 5 步去掉，便是本文用来进行对照的双目视觉 SLAM 算法。此外，在实际运算过程中，为了提高算法的实时性，基于双目与 IMU 融合的 SLAM 算法能在摄像机快速旋转或者特征点稀疏的情况下位姿估计能成功的运行，而基于双目相机的 SLAM 算法在摄像机旋转过快的情况下，容易导致跟踪失败。

基于双目与惯导融合的 SLAM 算法的流程图如图 5-3 所示：

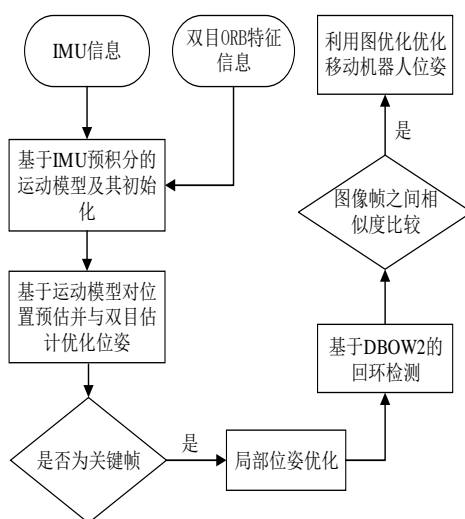


图 5-3 实验算法流程图

## 5.4 实验结果与分析

根据本文的方法，在相同实验条件下，对比了本系统与单目与惯导融合的 SLAM 系统的精度，如图 5-4 与图 5-5 所示。利用了 EuRoC 数据集中的 V202 数据集进行实验，得到的估计轨迹由虚线表示，真实轨迹由实线表示，真实轨迹由数据集捕捉设备提供。

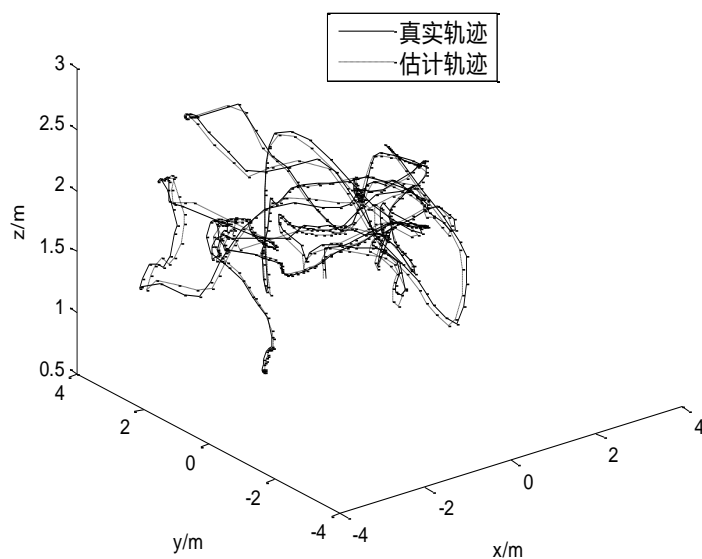


图 5-4 序列 V2\_02\_medium 双目与 IMU 估计轨迹与实际轨迹

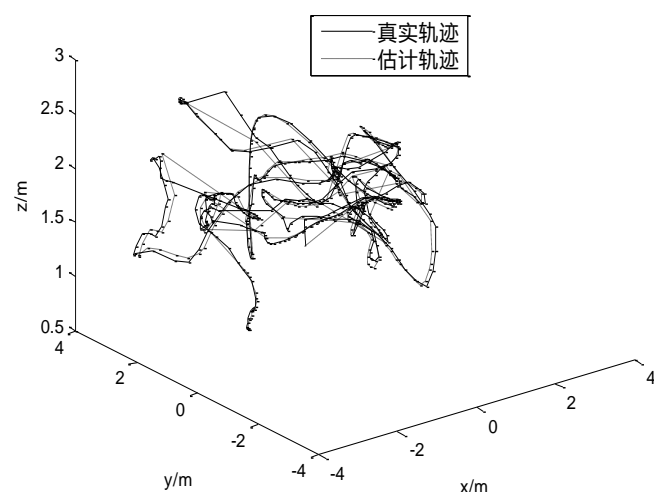


图 5-5 序列 V2\_02\_medium 单目与 IMU 估计轨迹与实际轨迹

为了更好的显示其精度，将图中的绝对轨迹均方根误差(RMSE)、均差(mean)、标准差(std)、中值(median)在表 5-2 中呈现。其中绝对轨迹误差，也称为均方根误差是度量估计值与真实值之间偏差的均值，在 SLAM 系统中表示移动机器人测量位置与实际位置之间的偏差。

表 5-2 V2\_02\_medium 序列绝对轨迹误差（单位：米）

采用方法	RMSE	mean	std	median
本系统	0.180 468	0.163 374	0.069 916	0.156 964
VIOSLAM	0.180 525	0.164 818	0.073 648	0.158 246
双目系统	0.181 228	0.164 373	0.071 518	0.157 424

实验中可得, 由于因为其在前端位姿估计优化中加入了双目的限制, 本方法相比于单目融合 IMU 的 SLAM 系统的精度有所提高。同时, 可以看出图 5-4 中有旋转的时候有较大的误差, 位姿跟踪中有短时间的丢失, 而在本方法中, 旋转的时候误差的精度有明显的提高, 说明本系统的鲁棒性更好。

采用双目与 IMU 融合的方法, 对系同鲁棒性及精度提高的同时增加了计算复杂度。实验中, 得到了平均处理每一帧图像的时间, 如表 5-3 所示。对双目与惯导融合的 SLAM 系统与基于双目的 SLAM 系统进行对比分析, 得到处理每帧平均时间, 表明了本方法时间相比于双目 SLAM 系统的时间有所增加, 平均增加了 2%。因此, 可知本文的方法在 Euroc 数据集上具备实时性, 满足移动机器人的实时性需求。

表 5-3 各系统耗时比较 (单位: 秒)

采用方法	V101	V102	V103	V201	V202
本系统	0.093 67	0.085 48	0.080 87	0.079 39	0.084 76
双目系统	0.093 74	0.086 25	0.080 53	0.079 57	0.084 99

为检测双目与惯导融合方法在实际运行时的效果, 本文用之前的标定好的双目相机与 IMU 的参数, 在图 5-1 所示的平台上进行实验, 其运行过程中的实时效果图如下图 5-6 所示:

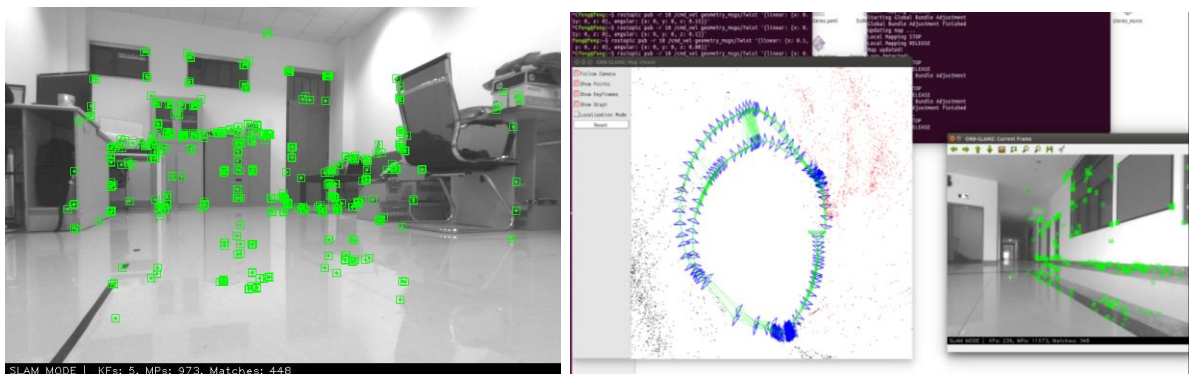


图 5-6 实验环境下双目与惯导融合的 SLAM 的运行状况

由于本文没有定位仪器追踪到移动机器人的实时运动, 因此没有计算得到真实位置与实际位置的误差。从仿真结果及本实验运行结果发现, 双目与惯导融合的 SLAM 系统在初始化阶段运行良好, 并且能实时高效运行。在旋转较快情况下, 移动机器人也能通过 IMU 的预积分模型调整, 重新跟踪到之前的位姿, 具有较强的鲁棒性。此外, 从全局位

姿优化上，能够有效的检测到回环，并进行全局的调整。因此，纵观全局，实验证明了本文算法的有效性。

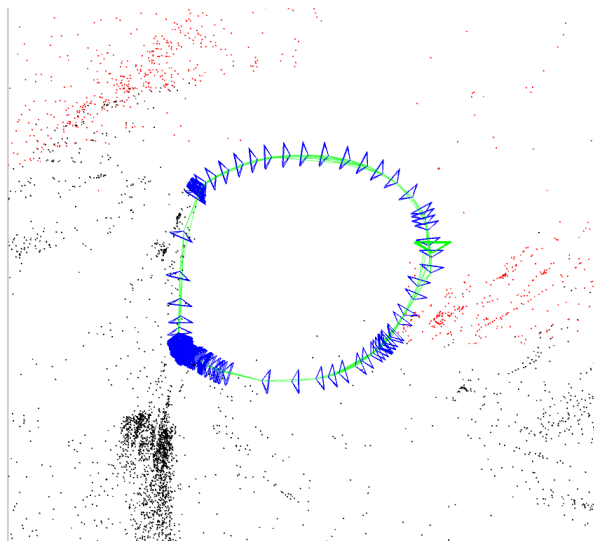


图 5-7 实验生成关键帧及点云数据

## 5.5 本章小结

本章首先介绍了本文所采用的实验平台和数据集，然后详细介绍了实验所采用的两种 SLAM 算法。接着展示了两组对比实验的具体效果，并对实验结果进行了分析。此外，还展示了采用基于双目与 IMU 融合的 SLAM 算法计算得到的三维稀疏地图。实验结果表明，本文所提出的基于双目与 IMU 融合的 SLAM 算法相比于基于双目的 SLAM 算法，拥有更高的准确性与鲁棒性。



## 第 6 章 总结与展望

### 6.1 研究工作总结

双目视觉与惯导信息融合的 SLAM 系统为实现移动机器人的导航与路径规划提供了重要依据，具有广泛的应用价值与广阔的发展前景。双目视觉与 IMU 融合主要包括前端双目视觉与 IMU 融的位姿估计与优化和后端的回环检测与全局优化。这两部分减少了其中存在的非线性误差，提高了系统的精确度与鲁棒性，保证了全局地图的一致性。针对移动机器人中存在的累计误差的问题及旋转过快易跟踪失败的问题，本文将双目视觉与 IMU 融合，以提高系统的鲁棒性及精确度，具体工作和成果如下：

1、查阅国内外相关参考资料，介绍了课题的背景和研究意义，同时结合国内外研究现状，分析了如今各种视觉 SLAM 系统及单目视觉与惯导融合所存在的问题和缺陷。

2、针对 IMU 的观测模型及 IMU 的预积分模型，利用双目视觉初始化获取的一系列关键帧，对 IMU 的偏差及重力方向进行初始化，并估计出 IMU 的速度。同时，利用双目得到的特征点深度信息简化了 IMU 的初始化过程。

3、针对 SLAM 系统的实时性需求，采用 ORB 算法对图像的特征点进行检测与描述，并采用基于随机抽样一致性算法的特征匹配算法进行特征匹配，来剔除其中的误匹配，得到匹配正确率高的匹配结果。这样既提高了图像特征提取与描述的速度，也提高了特征匹配的准确率。

4、针对 SLAM 系统中移动机器人位姿准确求解的问题，本文将 IMU 的预积分约束结合到 EPNP 求解的位姿算法，首先采用 EPNP 算法求解得到移动机器人位姿的初始值，然后利用双目视觉的重投影优化算法及 IMU 预积分模型得到的预测值，同时对位姿进行优化得到更加精确的位姿估计值。

5、针对如何进行准确的回环优化问题，本文首先采用基于视觉词袋模型的闭环检测算法对得到的当前关键帧与所有过往关键帧进行相似度检测，从而检测出是否有回环存在。该算法提高了回环检测的准确性与检测速度。然后采用基于图优化的后端优化方法对检测到的回环进行优化。在图优化模型中只保留位姿之间的约束关系，只优化位姿，这一做法使后端优化变得更加简练，提高了后端优化的效率。

6、最后，通过实验证明本文所提出的双目与惯导融合的 SLAM 算法具有较高的准确性和良好的鲁棒性。

## 6.2 展望

视觉 SLAM 技术拥有着广泛的应用前景，无论是无人机、虚拟与现实技术或者自主移动机器人的发展都离不开位姿估计与实时地图构建等过程。SLAM 技术的发展将推动许多相关领域的飞速发展。本文虽然在技术上取得了一定的成果，但由于时间与条件的限制，本文取得的研究成果尚存在一些缺陷与不足之处，还有一些问题亟待解决与完善。主要包括以下几个方面：

1、为了让 SLAM 系统能够在手机或者嵌入式等小型设备上实时运行，视觉 SLAM 技术越来越趋向于朝轻量化、小型化方向发展。本文采用的是基于特征点法的 SLAM 算法，虽然采用了相对提取速度较快的特征提取与描述算法，但是双目的特征匹配以及前后帧之间的特征匹配的计算量较大。虽然基于直接法的 SLAM 系统计算量相对较小，但是其对光照比较敏感，无法实现回环检测的功能，阻碍了其发展。因此如何进一步提高系统的实时性需要更深入的研究与探讨。

2、双目视觉与惯导的融合只运用在前端位姿估计部分，对于后端全局优化并未用到 IMU 的预积分数据去优化全局位姿，后期可以将 IMU 获取的轨迹同时加入全局优化中。同时，可以利用自标定技术更加完整的实现移动机器人运动过程中相机与 IMU 的标定。

3、到目前为止，SLAM 方案还处于特征点或者像素的层级，随着深度学习的发展，将 SLAM 技术与深度学习技术相结合将是一个重要的发展方向。语义信息可以让 SLAM 得到一个带有标签的地图，同时也可以为回环检测、位姿优化提供更多的条件。反过来 SLAM 也可以帮助语义，在构建大量训练数据时提供便利。因此将 SLAM 与深度学习技术相结合将是接下来一个很有前景的发展方向。

4、本文实现的是静态场景下的 SLAM，没有考虑到动态环境下物体的干扰。在现实生活中虽然也存在许多静态应用场景，但大多数的应用环境都是动态的，这就必须考虑动态环境下物体的遮挡、干扰等问题，需要进一步的研究与完善。

5、本文研究的是单机器人的 SLAM，没有涉及到多机器人，而多机器人的协同工作是未来的发展热点，因此与之对应的多机器人系统的 SLAM 系统，也是一个值得研究的方向，有待研究与拓展。

## 参 考 文 献

- [1] 陈卫东, 张飞. 移动机器人的同步自定位与地图创建研究进展[J]. 控制理论与应用, 2005, 22(3): 455-460.
- [2] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty [J]. The International Journal of Robotics Research, 1987, 5(4): 56-68.
- [3] Grisetti G, Stachniss C, Burgard W. Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling [C]. Proceedings of the 2005 IEEE International Conference on Robotics and Automation. Barcelona: IEEE, 2005, 2432-2437.
- [4] Grisetti G, Stachniss C, Burgard W. Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters [J]. IEEE Transactions on Robotics, 2007, 23(1): 34-46.
- [5] 张国良, 姚二亮, 林志林, 徐慧. 融合直接法与特征法的快速双目 SLAM 算法[J]. 机器人, 2017, (6): 879-888.
- [6] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features [C]. Proceedings of the 9th European conference on Computer Vision. Berlin, Heidelberg: Springer, 2006, 404-417.
- [7] Lowe D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [8] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF [C], International Conference on Computer Vision. IEEE, 2012: 2564-2571.
- [9] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [M]. Association for Computing Machinery, 1981. 24(6): 381-395.
- [10] Nister D. An efficient solution to the five-point relative pose problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 756-770.
- [11] Stewénius H, Engels C, Nistér D. Recent developments on direct relative orientation [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2006, 60(4): 284-294.
- [12] Hartley R I, Zisserman A. Multiple view geometry in computer vision [M]. UK: Cambridge University Press, 2004.
- [13] Hartley R I. In defense of the eight-point algorithm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(6): 580-593.
- [14] Gao X S, Hou X R, Tang J L. Complete solution classification for the perspective-three-point problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 930-943.
- [15] Lepetit V, Moreno-Noguer F, Fua P. EPnP: An accurate O(n) solution to the PnP problem [J]. International Journal of Computer Vision, 2009, 81(2): 155-166.
- [16] Henry P, Krainin M, Herbst E, et al. RGB-D Mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments [J]. International Journal of Robotics Research, 2012, 31(5): 647-663.
- [17] Sharp G C, Sang W L, Wehe D K. ICP Registration Using Invariant Features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): 90-102.
- [18] Silveira G, Malis E, Rives P. An Efficient Direct Approach to Visual SLAM [J]. IEEE Transactions on

- Robotics, 2008, 24(5): 969-979.
- [19] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C]. Proceedings of the 7th international joint conference on Artificial intelligence. San Francisco: Morgan Kaufmann, 1981: 674-679.
  - [20] Hahnel D, Burgard W, Fox D, et al. An efficient fast SLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[C]. Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems[C]. Las Vegas: IEEE, 2003, 206-211.
  - [21] Endres F, Hess J, Sturm J, et al. 3-D Mapping With an RGB-D Camera [J]. IEEE Transactions on Robotics, 2014, 30(1): 177-187.
  - [22] Latif Y, Cadena C, Neira J. Robust loop closing over time for pose graph SLAM [J]. International Journal of Robotics Research, 2013, 32(14): 1611-1626.
  - [23] Ulrich I, Nourbakhsh I. Appearance-based place recognition for topological localization [C]. Proceedings of the 2000 IEEE International Conference on Robotics and Automation. San Francisco: IEEE, 2000, 1023-1029.
  - [24] Gálvez-López D, Tardós J D. Real-time loop detection with bags of binary words [C]. Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco: IEEE, 2011, 51-58.
  - [25] Galvez-López D, Tardos J D. Bags of Binary Words for Fast Place Recognition in Image Sequences [J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
  - [26] Grasa O G, Civera J, Montiel J M M. EKF monocular SLAM with relocalization for laparoscopic sequences [C]. Proceedings of the 2011 IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011, 4816-4821.
  - [27] Lategahn H, Geiger A, Kitt B. Visual SLAM for autonomous ground vehicles [C]. IEEE International Conference on Robotics and Automation, 2011:1732-1737.
  - [28] Sujan V A, Dubowsky S. Efficient Information-based Visual Robotic Mapping in Unstructured Environments [J]. International Journal of Robotics Research, 2005, 24(4): 275-293.
  - [29] Hanani U, Bracha S, Peretz S. Information Filtering: Overview of Issues, Research and Systems [J]. User Modeling and User-Adapted Interaction, 2001, 11(3):203-259.
  - [30] Sim R, Elinas P, Little J J. A Study of the Rao-Blackwellised Particle Filter for Efficient and Accurate Vision-Based SLAM [J]. International Journal of Computer Vision, 2007, 74(3): 303-318.
  - [31] Gil A. Multi-robot visual SLAM using a Rao-Blackwellized particle filter [J]. Robotics and Autonomous Systems, 2010, 58(1): 68-80.
  - [32] Hauke Strasdat, J.M.M. Montiel, Andrew J. Davison. Visual SLAM: Why filter [J]. Image and Vision Computing, 2012, 30(2): 65-77.
  - [33] Lourakis M I A, Argyros A A. SBA: a software package for generic sparse bundle adjustment [J]. ACM Transactions on Mathematical Software, 2009, 36(1): 2.
  - [34] Triggs B, Mclauchlan P F, Hartley R I, et al. Bundle Adjustment-A Modern Synthesis [J]. Vision Algorithms Theory & Practice, 2000.
  - [35] Latif Y, Cadena C, Neira J. Robust loop closing over time for pose graph SLAM [J]. International Journal of Robotics Research, 2013, 32(14): 1611-1626.
  - [36] Lee D, Myung H. Solution to the SLAM Problem in Low Dynamic Environments Using a Pose Graph and an RGB-D Sensor [J]. Sensors, 2014, 14(7): 12467-12469.

- [37] Dubbelman G, Browning B. COP-SLAM: Closed-Form Online Pose-Chain Optimization for Visual SLAM [J]. IEEE Transactions on Robotics, 2015, 31(5): 1194-1213.
- [38] Kaess M, Ranganathan A, Dellaert F. iSAM: Incremental Smoothing and Mapping [J]. Robotics IEEE Transactions on, 2008, 24(6): 1365-1378.
- [39] Kaess M, Johannsson H, Roberts R, et al. iSAM2: Incremental smoothing and mapping using the Bayes tree [J]. International Journal of Robotics Research, 2012, 31(2): 216-235.
- [40] Rosen D M, Kaess M, Leonard J J. RISE: An Incremental Trust-Region Method for Robust Online Sparse Least-Squares Estimation [J]. IEEE Transactions on Robotics, 2014, 30(5): 1091-1108.
- [41] Davison A J. Real-Time Simultaneous Localisation and Mapping with a Single Camera [C]. Proceedings of the Ninth IEEE International Conference on Computer Vision. IEEE Nice, 2003, 1403-1410.
- [42] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]. Proceedings of the 2007 IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara: IEEE, 2007, 1-10.
- [43] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [44] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM [M]. European Conference on Computer Vision. 2014
- [45] Engel J, Sturm J, Cremers D. Semi-dense Visual Odometry for a Monocular Camera [C]. Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE: Sydney, 2013, 1449-1456.
- [46] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Hong Kong: IEEE, 2014, 15-22.
- [47] Forster C, Zhang Z, Gassner M, et al. SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems [J]. IEEE Transactions on Robotics, 2017, 33(2):249-265.
- [48] Rafik Mebarki, Vincenzo Lippiello, Bruno Siciliano. Vision-based and IMU-aided scale factor-free linear velocity estimator [J]. Autonomous Robots, 2016:1-15.
- [49] Usenko V, Engel J, Stückler J, et al. Direct visual-inertial odometry with stereo cameras[C]. IEEE International Conference on Robotics and Automation. IEEE, 2016:1885-1892.
- [50] Konolige K, Agrawal M, Solà J. Large-Scale Visual Odometry for Rough Terrain [J]. Proceeding of International Symposium on Robotics Research, 2010, 66:201-212.
- [51] Mourikis AI, Roumeliotis SI. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation [J].IEEE International Conference on Robotics & Automation, 2007, 22: 3565-3572.
- [52] Stefan Leutenegger, Simon Lynen, Michael Bosse, et al. Keyframe-based visual-inertial odometry using nonlinear optimization [J]. International Journal of Robotics Research, 2015, 34(3):314-334.
- [53] Forster C, Carlone L, Dellaert F, et al. On-Manifold Preintegration for Real-Time Visual--Inertial Odometry [J]. IEEE Transactions on Robotics, 2017, 33(1):1-21.
- [54] Mur-artal R, Tardós J D. Visual-Inertial Monocular SLAM With Map Reuse [J]. IEEE Robotics and Automation Letters, 2016, 2(2):796-803.
- [55] Qin T, Li P, Shen S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator [J]. IEEE Transactions on Robotics, 2017, (99):1-17.
- [56] Z. Zhang. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, 22(11): 1330-1334.
- [57] Mendonca P R S, Cipolla R. A simple technique for self-calibration [C]. Computer Society Conference

- on Computer Vision and Pattern Recognition. 1999:1500.
- [58] 湖战役, 吴福朝. 基于主动视觉摄像机标定方法[J]. 计算机学报, 2002, 25(11): 1149-1156.
- [59] Rosten E, Drummond T. Machine learning for high-speed corner detection [C]. Proceedings of the 9th European Conference on Computer Vision. Berlin Heidelberg: Springer-Verlag, 2006, 430-443.
- [60] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features [C]. Proceedings of the 11th European Conference on Computer Vision. Berlin Heidelberg: Springer-Verlag, 2010: 778-792.
- [61] Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets [J]. International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [62] Moravec, H., Elfes A. High resolution maps from wide angle sonar [C]. In Proceeding of the IEEE International Conference on Robotics and Automation, 1985, 116–121.
- [63] Olson E, Leonard J, Teller S. Fast Iterative Optimization of Pose Graphs with Poor Initial Estimates [J]. 2006:2262-2269.
- [64] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006:2161-2168.
- [65] Kümmerle R, Grisetti G, Strasdat H, et al. G2O: A general framework for graph optimization [C]. Proceedings of the 2011 IEEE International Conference on Robotics and Automation. IEEE Shanghai, 2011, 3607-3613.

## 致 谢

时光荏苒岁月如梭，两年半的研究生生涯即将结束。回顾在浙工大的几年研究生时光感觉十分充实，虽然有苦有泪，但更多的是收获与成长。感到欣慰的是，这两年的学习生涯没有辜负自己辛勤的付出，让自己在学业与生活中都有了质的飞跃，领悟了人生道路都是靠自己的每一分耕耘得来的。同时我也要在这一感谢陪伴我成长的老师与同学，是你们的指导与关心陪我度过了充实的研究生生活。

首先我要感谢的是我的导师邢科新，感谢老师这两年多对我的帮助与支持，老师一直用严谨的认真的态度要求去做好每一件小事。在论文的开题、研究的内容与实验方面给我最大的帮助与指导，本文的每一个环节都是老师倾注了无数的汗水与心血。在这两年多的学习生活中，导师渊博的学识与严谨务实的工作作风以及诲人不倦的奉献精神时刻感染并且激励着我。在这里我要向他表示衷心的感谢。

其次，我要感谢实验室的同学李圣男、江超，师兄竺海光、泰应鹏，师姐陈步华，师弟郭超、冯传涛刘世豪等，在学习上给予的热心帮助。同时我也要感谢我美丽可爱的室友彭柔、侍宇洁、徐偲在生活上的帮助，是你们的陪伴让我的研究生生活变得丰富多彩，并且在科研学习上鼓励支持我不断突破自己。

最后，在这里我也要衷心感谢养育我的父母，感谢他们不仅在生活上给与我的帮助与支持，还给了我精神上的鼓励与关心，更是在我人生的道路上提供了源源的动力。谨以此文献给所有帮助关心过我的人。

## 攻读学位期间参加的科研项目和成果

### 参加的科研项目

[1] 横向项目：AGV 的嵌入式控制器开发.

### 专利

[1] 基于双目视觉特征和 IMU 信息的机器人定位与地图构建系统[发明专利], 申请号 201810218153.1, 公布号 CN 108665540A (已公开)