

中图分类号: TP242.6
学科分类号: 081104

论文编号: 1028703 17-SX175

硕士学位论文

未知环境下无人机的双目视觉 SLAM 方法研究

研究生姓名	钟家跃
学科、专业	模式识别与智能系统
研究方向	视觉导航
指导教师	王从庆教授

南京航空航天大学

研究生院 自动化学院

二〇一七年六月

Nanjing University of Aeronautics and Astronautics
The Graduate School
College of Automation

Research on Binocular Vision SLAM of UAV in Unknown Environment

A Thesis in

Pattern Recognition and Intelligent System

by

Zhong Jiayue

Advised by

Prof. Wang Congqing

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Engineering

May, 2017

承诺书

本人声明所呈交的硕士学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南京航空航天大学或其他教育机构的学位或证书而使用过的材料。

本人授权南京航空航天大学可以将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、汇编学位论文。

（保密的学位论文在解密后适用本承诺书）

作者签名：钟家跃
日期：2017.6.9
王从庆

摘 要

随着无人机在军事、民用两方面的广泛应用,实现无人机的自主导航成为人们关注的焦点。在室内、缺失 GPS 信号的室外且没有先验信息的环境下,即未知环境下,解决无人机的自主导航的关键是同时定位与地图构建 (Simultaneous Localization and Mapping, SLAM)。由于相机的成本低,获取关于环境的信息量大等优点,本文研究无人机的双目视觉 SLAM,对双目视觉 SLAM 进行了理论分析和综合性实验。

首先,本文对 SIFT 和 ORB 两种特征提取算法进行了理论研究与实验分析,搭建双目视觉 SLAM 的无人机硬件平台,在室内外环境下,通过实验对比了 SIFT 和 ORB 两种算法。实验结果表明,ORB 算法相比 SIFT 算法具有快速性,但是 SIFT 算法的鲁棒性更强。

然后,设计了 SLAM 的前端。实现了室内外基于 SIFT 和 ORB 算法的双目的视觉里程计 (Visual Odometry, VO),获得了初始位置轨迹,再此基础上获得了局部三维点云地图。实验结果表明,在室外的情况下,基于 SIFT 算法的视觉里程计在位置估计的精度上优于基于 ORB 算法的视觉里程计,而在室内,基于 ORB 算法的视觉里程计在位置估计的精度上,稍微占优。

其次,搭建了 SLAM 的后端,完成了单无人机的双目视觉 SLAM 实验。对关键帧、回环检测进行了研究,利用非线性优化的方法,采用基于位姿图优化的方式,对 SLAM 前端获得的初始位置轨迹进行优化。通过实验对比分析,得出双目视觉 SLAM 相比于视觉里程计获得更精确的位置轨迹和更精确的地图。

最后,研究多无人机双目视觉 SLAM,对提出的两种两架无人机双目视觉 SLAM 方案进行理论研究和仿真实验。结果表明,在方案二中两架无人机双目视觉 SLAM 相比于单无人机双目视觉 SLAM 具有更精确的位置轨迹和地图,同时在快速性方面具有优势。

关键词: SLAM, 双目视觉, 视觉里程计, 非线性优化, 位姿图优化, 无人机

ABSTRACT

With the widely used of Unmanned Aerial Vehicle (UAV) in the area of military and civil, achieving the autonomous navigation of UAV is the topic. In the absence of GPS and without the previous information about the environment, namely, under the unknown environment, the key of indoor and outdoor autonomous navigation is Simultaneous Localization and Mapping (SLAM). Cameras are into the field of SLAM, because of its cheap and capturing large information from the environment. The theme of thesis is the SLAM of UAV based on binocular vision. This thesis analyzes the basic theory of SLAM based on binocular vision, and conducts relative experiments.

First of all, SIFT and ORB which are the two feature extraction algorithms of image are analyzed theoretically. The hardware platform of SLAM for UAV based on binocular vision is built. Both indoor and outdoor, the experiments based on the SIFT and ORB are carried out. As a result, ORB takes less time than SIFT, otherwise, the robustness of SIFT is stronger than ORB.

Then, the front end of binocular vision SLAM based on feature is accomplished. In the situation of indoor and outdoor, the experiment of Visual Odometry (VO) is carried out. The features still are SIFT and ORB. As a result, the initial position and pose are got. So the local 3D point cloud map is obtained. The result of experiment indicates that the VO based on SIFT has an advantage of the accuracy in the estimation of position outside, comparing with the VO based on ORB. Both are well-matched indoors. However, the accuracy of VO based on ORB is higher than VO based on SIFT.

In addition, the back end of binocular vision SLAM based on feature is completed. The experiment of SLAM based on binocular vision is carried out. Under the conditions that keyframe and Loop Closure are considered, the Nonlinear Optimization with pose graph is used for the optimization of the initial position and pose. The experimental result shows that the position of binocular vision SLAM and map are more accuracy than VO.

Finally, the SLAM of UAVs based on binocular vision is studied. Two schemes are studied and applied to two UAVS based on binocular vision SLAM. The simulation of two UAVS based on binocular vision SLAM is carried out. The simulation experiments from the second schemes show that the SLAM of two UAVs is not only better than the SLAM of single UAV in the accuracy of position and map, but takes less time.

Key Words: SLAM, Binocular Vision, Vision Odometry, Nonlinear Optimization, Pose Graph Optimization, Unmanned Aerial Vehicle.

目录

第一章 绪论	1
1.1 无人机研究背景和意义.....	1
1.2 SLAM 技术的概述	2
1.2.1 SLAM 的研究意义和目的.....	2
1.2.2 SLAM 的发展	2
1.2.3 无人机视觉 SLAM 研究的意义.....	4
1.3 本文主要研究内容.....	5
第二章 双目立体视觉成像系统及特征提取与匹配算法研究	7
2.1 引言	7
2.2 双目立体视觉成像系统.....	7
2.2.1 ZED 立体相机介绍.....	7
2.2.2 ZED 双目立体视觉成像原理.....	8
2.2.3 ZED 立体相机标定.....	11
2.2.4 立体校正.....	16
2.2.5 立体匹配.....	16
2.2.6 三角测量.....	16
2.3 特征提取与匹配算法.....	17
2.3.1 SIFT 算法	17
2.3.2 ORB 算法	22
2.3.3 特征匹配策略.....	24
2.4 特征提取与匹配实验.....	25
2.4.1 室外 SIFT、ORB 特征提取与匹配	25
2.4.2 室内 SIFT、ORB 的特征提取与匹配	27
2.5 本章小结	28
第三章 无人机的双目视觉里程计研究	29
3.1 引言	29
3.2 双目视觉的运动估计.....	30
3.2.1 3D-2D 运动估计方法.....	30
3.2.2 3D-3D 运动估计方法.....	33
3.3 局部优化	33
3.4 初始位姿与地图融合.....	34
3.5 VO 实验结果及分析.....	35
3.6 本章小结	40
第四章 无人机的双目视觉 SLAM 的后端优化	41
4.1 引言	41
4.2 SLAM 优化模型	41
4.3 回环检测	47
4.4 位姿图优化	48

4.5 SLAM 实验结果及分析	49
4.6 本章小结	57
第五章 多无人机双目视觉 SLAM 方法研究	58
5.1 引言	58
5.2 多无人机双目视觉 SLAM 系统.....	58
5.2.1 两架无人机双目视觉 SLAM 方法.....	58
5.2.2 多无人机视觉 SLAM 方法.....	59
5.3 Octomap 地图研究.....	59
5.4 仿真实验	61
5.5 本章小结	66
第六章 总结和展望	67
6.1 本文的主要工作.....	67
6.2 前景展望	67
参考文献	69
致谢	75
在学期间的研究成果及发表的学术论文	76

图表清单

图 1.1 国内外军用先进无人机.....	1
图 1.2 双目视觉 SLAM 方法组成框图.....	4
图 2.1 ZED 立体相机.....	7
图 2.2 透视投影倒立成像几何示意图.....	8
图 2.3 透视投影几何示意图.....	9
图 2.4 世界坐标系和摄像机坐标系之间的变换图.....	9
图 2.5 图像坐标系与像素坐标系变换关系图.....	10
图 2.6 对极几何约束	15
图 2.7 立体校正后立体成像图.....	16
图 2.8 无畸变对准的立体成像投影图.....	17
图 2.9 图像的多尺度表示图.....	18
图 2.10 金字塔高斯差分图.....	19
图 2.11 像素点的空间 26 邻近像素点图.....	19
图 2.12 关键点方向直方图.....	21
图 2.13 关键点生成描述子图.....	21
图 2.14 FAST 算法原理图	22
图 2.15 无人机双目视觉 SLAM 实验硬件平台.....	25
图 2.16 无人机用 ZED 在两个时刻左相机采集到的室外 RGB 图.....	25
图 2.17 室外两时刻的三维点云图.....	26
图 2.18 利用 SIFT 算法获得的室外特征匹配图	26
图 2.19 利用 ORB 算法获得的室外特征匹配图	26
图 2.20 无人机用 ZED 在两个时刻左相机采集到的室内 RGB 图.....	27
图 2.21 室内两时刻对应的三维点云图.....	27
图 2.22 利用 SIFT 算法获得的室内特征匹配图.....	28
图 2.23 利用 ORB 算法获得的室内特征匹配图	28
图 3.1 视觉里程计流程图.....	29
图 3.2 重投影误差图	31
图 3.3 双目视觉里程计流程图.....	35
图 3.4 室外相邻两帧融合后的点云图.....	36
图 3.5 室内相邻两帧融合后的点云图.....	36

图 3.6 VO 估计的室外位置误差图	37
图 3.7 VO 获得的室外三维点云图	38
图 3.8 VO 估计室内位置误差图	39
图 3.9 VO 获得的室内三维点云图	40
图 4.1 运动观测模型图	44
图 4.2 g^2o 优化图	45
图 4.3 基于无人机的双目视觉 SLAM 流程图	48
图 4.4 室外基于 SIFT 的 VO 与 SLAM 的位置误差对比图	50
图 4.5 室外基于 ORB 的 VO 与 SLAM 的位置误差对比图	51
图 4.6 室外 SIFT 和 ORB 的双目视觉 SLAM 系统估计的位置误差对比图	52
图 4.7 室外双目视觉 SLAM 获得的三维点云地图	53
图 4.8 室内基于 SIFT 的 VO 与 SLAM 的位置误差对比图	54
图 4.9 室内基于 ORB 的 VO 与 SLAM 的位置误差对比图	55
图 4.10 室内 SIFT 和 ORB 的双目视觉 SLAM 系统估计的位置误差对比图	56
图 4.11 室内双目视觉 SLAM 获得的三维点云地图	57
图 5.1 无人机双目视觉 SLAM 区域图	58
图 5.2 八叉树表示空间示意图	60
图 5.3 室外单无人机、两无人机双目视觉 SLAM 在三轴的位置误差对比图	62
图 5.4 UAV2 获得的室外地图模型	63
图 5.5 UAV3 获得的室外地图模型	63
图 5.6 室外地图融合后的地图模型	63
图 5.7 室内单无人机、两无人机双目视觉 SLAM 在三轴的位置误差对比图	64
图 5.8 UAV2 获得的室内地图模型	65
图 5.9 UAV3 获得的室内地图模型	65
图 5.10 室内融合后的地图模型	65
表 2.1 ZED 标定后参数	15
表 2.2 室外 ORB 和 SIFT 特征提取与匹配比较	27

注释表

f	相机焦距	O-XYZ	相机坐标系
$O_m - xy$	图像坐标系	$O_w - X_w Y_w Z_w$	世界坐标系
R	旋转矩阵	t	平移矩阵
$O_p - uv$	像素坐标系	u_0, v_0	图像平面主点
f_x, f_y	相机焦距	K	相机内参数矩阵
M_2	相机外参数矩阵	M_3	世界坐标系下 空间点的齐次坐标
r_1, r_2, r_3	相机外参数矩阵的列向量	H	单应性矩阵
h	H 的列向量表示	k_1, k_2	透镜畸变参数
D	视差	b	双目相机基线
Z	深度	t	尺度参数
$L(x, y, t)$	尺度参数 t 的尺度	d	欧式距离
I	图像	T	变换矩阵
ξ	变换矩阵的李代数	J	重投影误差项的雅克比矩阵
P, Q	三维坐标点集合	p	像素点
C_k	表示相机在 k 时刻的位姿	x	无人机状态变量矩阵
u	无人机输入变量矩阵	w_k	无人机运动噪声
v_k	观测噪声	y	路标点
$Q_{k,j}$	观测噪声高斯分布协方差	R_k	运动噪声高斯分布协方差
$\hat{H}$	Hessian 矩阵	z_k	观测矩阵

$\mathbf{z}_{1:t}$	1 到 t 时刻节点 m 的观测数据	$P(m)$	节点 m 的初始占有概率
$\mathbf{z}_t$	t 时刻节点 m 的观测数据	$P(m \mathbf{z}_t)$	在观测 $\mathbf{z}_{1:t}$ 下的概率
$\hat{L}(m \mathbf{z}_t)$	观测 $\mathbf{z}_{1:t}$ 下的概率对数值	$\hat{L}(m)$	节点 m 的初始占有概率的对数值

缩略词

英文简称	英文全称	中文全称
GPS	Global Navigation System	全球定位系统
SLAM	Simultaneous Localization and Mapping	同时定位与地图构建
SIFT	Scale Invariant Feature Transform	尺度不变特征变换
ORB	Oriented FAST and Rotated BRIEF	定向加速段测试特征和旋转二元稳健独立基本特征
VO	Visual Odometry	视觉里程计
UAV	Unmanned Aerial Vehicle	无人机
RGB-D	Red Green Blue-Depth	红色、绿色、蓝色、深度
EKF	Extended Kalman Filter	扩展卡尔曼滤波
PTAM	Parallel Tracking and Mapping	并行追踪和制图
AR	Augment Reality	增强现实
GPU	Graphics Processing Unit	图像处理单元
BM	Bidirectional Matching	双向匹配
SGBM	Semi-global Block Matching	半全局匹配
GC	Graphics Computer	计算机图解
FAST	Features from Accelerated Segment Test	加速段测试特征
BRIEF	Binary Robust Independent Elementary Features	二元稳健独立基本特征
oFAST	FAST Keypoint Orientation	加速段测试特征的关键点定位
HDMI	High Definition Multimedia Interface	高清晰度多媒体接口
VGA	Video Graphics Array	视频图形阵列
FLANN	Fast Library for Approximate Nearest Neighbors	近似最近邻快速库
ICP	Iterative Closest Point	迭代最近点
RANSAC	Random Sample Consensus	随机采样一致性
PnP	Perspective from n Points	n 点透视
PF	Particle Filter	粒子滤波
MAP	Maximize a Posterior	最大后验
g^2o	general graph optimization	基本图优化

第一章 绪论

1.1 无人机研究背景和意义

无人机 (Unmanned Aerial Vehicle, UAV), 是一种依靠自身飞控计算机实现飞行和无线电遥感操纵的不载人飞行器^[1]。相比于有人飞机, 无人机普遍具有体积小、廉价、轻便、易于操作、可在严苛的环境下飞行等优点。基于此, 无人机在军事行业备受青睐。1995 年美国研制的 RQ-4A 全球鹰无人机, 满足了空中防御侦察办公室向联合力量指挥部提供远程侦察能力的需要。1994 年首飞的由美军研制的捕食者无人机, 为军事作战人员提供了情报。2000 年中国研发了彩虹无人机, 用于获取战场情报和对目标进行准确攻击。2013 年, 中国的利剑隐身无人机成功首飞, 标志着我国在军用无人机方面跻身世界顶尖行列。以上所述的先进的用于军事打击和防御的无人机, 如图 1.1 所示。



图 1.1 国内外军用先进无人机

在民用领域, 无人机已逐渐应用于航拍、电影、救灾、植保、快递等。具体体现在: 用于为农作物喷洒农药、化肥的植保无人机; 中国深圳大疆公司研制的用于影像系统的无人机占据着巨大的市场份额; 2013 年, 亚马逊测试将用无人机用于小物件的快递; 我国的顺丰速运公司测试利用无人机对偏远地区进行送货。

1.2 SLAM 技术的概述

1.2.1 SLAM 的研究意义和目的

SLAM (Simultaneous localization and mapping) 被译为即时定位与构图^[2], 由机器人的运动状态和环境的地图模型两部分组成。机器人的运动状态由其三自由度的位置和三自由度的姿态组成, 在其他描述中, 运动状态还包括机器人的速度, 传感器的基线, 标定参数。地图由环境中的地标或障碍物表示, 可以是稀疏的特征点构成的点云地图, 也可以是半稠密或稠密的地图。在下缺失 GPS 的室内环境, 丢失 GPS 的室外复杂环境下, 且在室内、室外都没有先验的环境信息的环境下, 机器人需要 SLAM 进行自身定位。上述的先验环境信息指可以为机器人提供定位的信息, 例如具有位置信息的二维码。地图的构建由以下两方面的意义。第一, 地图可以用于路径规划和提供直观的可视化的环境模型给操作人员。第二, 地图可用于减少机器人的状态误差。在早期, 定位由地面机器人的轮的圈数和周长获得, 这称为里程计, 但是轮式里程计获得的位置估计随着时间很快漂移。SLAM 的目的在于建立一个全局连续的地图, 为此需要进行自身运动的估计和减少位姿漂移的回环检测 (Loop Closure)。如果没有回环检测部分, SLAM 退化为里程计。在 SLAM 的发展中, 将观测到的更多的地标用于位姿估计可以有效地减少, 甚至纠正位姿漂移, 这是一重大的进展^[3]。越来越多的里程计算法基于视觉和惯性测量元件, 这使得获得的轨迹小于百分之零点五的漂移^[4]。但是, SLAM 依旧有存在的必要性, 主要体现在以下三个方面。

第一, 视觉惯性导航系统可以认为是一个退化的 SLAM 系统, 这个退化的 SLAM 系统便是缺少回环检测部分。第二, 里程计将环境理解为一个无尽的模型, 将当前所处的位置当作一个新的位置; 回环检测使得机器人理解真实的环境, 即可能这是一个封闭的场景。因此, SLAM 相对于里程计, 使得机器人理解真实的环境, 获得正确的环境模型; 同时也可以在某两点间找到捷径。第三, SLAM 获得的全局连续一致性地图被现实中很多应用场景需要; 例如, 在军事或民用领域, 机器人的任务是探索未知环境并将该环境地图传输给操作人员; 机器人在对建筑和桥梁进行结构性检查时也需要三维地图。

在 SLAM 技术中, 基于机器人携带的传感器的不同, SLAM 分为激光 SLAM 和视觉 SLAM。根据视觉 SLAM 中摄像头的差异, 视觉 SLAM 可以分为单目视觉 SLAM、立体视觉 SLAM、RGB-D SLAM。视觉 SLAM 方案中又分为主流的基于特征点的视觉 SLAM 和基于直接法的视觉 SLAM。

1.2.2 SLAM 的发展

SLAM 的历史可以粗略地分为三个阶段^[2]。1986-2004 这一阶段称为传统时代, 解决 SLAM 的主流技术是概率表达的形成和滤波技术, 包括 EKF(extend Kalman filters)、Rao-Blackwellized

粒子滤波和最大似然估计,近 20 年的研究发展可参照^{[5][6]}。

下一个时期称为算法分析时代,时间段为 2004-2015,这一时期经历了 SLAM 的基本特性的研究,包括可观察性、收敛性、连续性。基于优化的 SLAM 方法得到快速发展,此方法被认为优于传统的滤波 SLAM 方法^[7]。同时,SLAM 的稀疏性有利于求解被认识到,开源的 SLAM 库得到快速发展,并将 SLAM 利用于实际中。2007 年 A.J.Davison 提出了 MonoSLAM, MonoSLAM 是第一个实时单目视觉 SLAM 系统^[8]。MonoSLAM 运用的滤波技术为 EKF,相机状态量由相机的当前六自由度状态和所有路标点组成,不断地更新相机状态量的均值和协方差矩阵。同年,由 Klein 提出了又一单目视觉 SLAM 开源库—PTAM(Parallel Tracking and Mapping)^[9]。PTAM 首次将视觉 SLAM 问题分为前端(front-end)和后端(back-end)。

前端进行特征提取,数据关联表现为相邻两帧图像的特征匹配和回环检测。前端主要内容是构建视觉里程计,为后端提供机器人状态初值;其次为回环检测,为获得更精确的状态。后端为一个最大后验(Maximize a Posterior, MAP)估计问题,将前端提供的状态初值利用回环检测进行优化,获得更好的状态变量。在获得更好的状态之后,进而获得优化后的全局一致性地图。根据此提出的前后端的概念,传统的 SLAM 方法的后端优化方法是滤波技术。PTAM 的贡献还在于,首次用非线性优化方法处理后端;同时,PTAM 引入关键帧的概念,即选择关键图像进行处理,加快了 SLAM 的处理速度,为在线实时 SLAM 提供了可能。PTAM 是 SLAM 在 AR (Augment Reality) 中的应用。PTAM 的应用场景小,在跟踪过程中容易丢失。2015 年,被提出的 ORB-SLAM^[10]代表着基于特征点的 SLAM 的一个高峰,相比其他 SLAM 方案,ORB-SLAM 具有明显的优势^[11]。同时 ORB-SLAM 将 SLAM 问题分为三部分:跟踪、局部优化位姿、全局优化。ORB-SLAM 由于在回环检测中运用存储的 ORB 字典,导致 SLAM 过程很耗时,计算资源的需求大。ORB-SLAM 依旧是主流的基于特征点的视觉 SLAM 方案,由于特征点的稀疏性,构建的地图为稀疏的特征点地图,不是稠密地图,不能很好的用于避障、导航。2014 年,被提出的 LSD-SLAM (Large Scale Direct monocular SLAM)^[12-13]将直接法应用于单目视觉 SLAM 中。LSD-SLAM 中利用梯度明显的像素位置构建地图,这称为半稠密的地图。LSD-SLAM 的前提是同一位置像素保持不变,在这一强假设下可知,直接法的视觉 SLAM 对外界光照条件严苛。

最后一时期为鲁棒感知时代,这一时期的特点^[2]为:

- 1) 鲁棒的性能: SLAM 系统能长时间在复杂环境下运行,而保持低的失败率。系统能根据环境自动调整参数。
- 2) 高水平的环境感知能力。SLAM 系统构建的环境模型不仅仅是几何结构,还将包含对环境的理解,例如,语义学、物理学。
- 3) 具有感知资源的能力。SLAM 系统感知系统的计算资源,并且根据可利用的计算资源

调整计算任务。

4) 任务驱动感知。为了完成相应的任务, SLAM 系统选择相应的感知信息和滤除无关的传感器数据。SLAM 系统根据任务的需求产生相应的地图表示形式。

1.2.3 无人机视觉 SLAM 研究的意义

鉴于如上所述无人机在多方面的运用, 无人机在执行任务时将可能面临着未知环境的考验, 相机成本低廉, 获取信息丰富, 视觉系统可运用于跟踪、监控等任务中, 因此开展在未知环境下基于视觉的无人机的 SLAM 研究具有重要的意义。在距今为止, 关于地面移动机器人 2D 的激光 SLAM 已经认为被解决。基于无人机的 SLAM 解决了 3D 下六自由度的位姿估计问题, 相比于只需要解决 2D 下三自由度的地面移动机器人 SLAM, 计算量更大。在未知环境下, 无人机只是一个运动平台, 无人机通过相机对自身位姿进行估计, 获取位姿的前提下, 便可以进行导航。当需要利用 SLAM 的地图进行导航时, 无人机便需要利用相应的飞控算法进行飞行控制。实时的 3D SLAM, 尤其是具有鲁棒性的高质量的无人机的视觉 SLAM 依旧是个难题^[14]。无人机的 SLAM 是个难题, 主要是因为无人机对实时处理的严格要求和其有效载荷的限制。HectorSLAM^[15]应用于无人机中具有较强的鲁棒性, 而 Albert 利用 RGBD 相机实现了实时 SLAM^[16]。上述提到的 ORB-SLAM 方案和 LSD-SLAM 方案具有很大可能性应用于无人机中。本文采用的基于特征点的双目视觉 SLAM 流程图如图 1.2 所示, 用双目相机采集图像, 利用获得的 RGB 图像进行特征提取与匹配, 利用深度图获得空间三维坐标。不断利用 2D 的图像点和 3D 的空间点获得两邻两帧之间的变换矩阵。假设无人机初始时刻的位姿为零, 作为参考位姿, 则通过上述获得的一系列相邻两帧之间的变换矩阵, 可以得到该帧对应于参考位姿的变换矩阵, 这些相对变换矩阵便是无人机的初始位姿, 这便是前端, 也称为双目视觉里程计。考虑到运算的实时性而引入关键帧的选择。由于运动过程中无人机会回到曾经到达的区域, 为了获得全局连续一致的地图, 考虑回环检测。由于视觉里程计中位姿存在累积误差, 引入回环检测而描述这个误差, 用概率学中的 MAP 表示, 利用非线性优化的方法使得位姿之间的误差, 这便是后端。后端得到了优化后的位姿, 根据这个位姿将地图融合, 获得全局一致性地图。

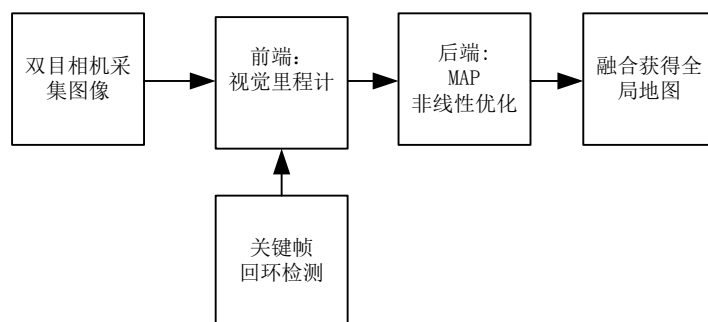


图 1.2 双目视觉 SLAM 方法组成框图

与单无人机视觉 SLAM 相比,多无人机视觉 SLAM^[17-18]有以下优势:

- 1) 在大规模环境下,时间紧迫,单无人机的续航能力有限,多无人机 SLAM 可以缩短任务时间。
- 2) 随着执行 SLAM 的时间增加,单无人机获得的图像信息越来越多,视觉 SLAM 需要处理的图像越多视觉里程计中计算无人机位姿变换矩阵所需的计算量大,加重无人机处理器的计算负担,而多无人机 SLAM 将任务量减少,有效地缓解了这一问题。
- 3) 单无人机绘制大规模环境地图将占用单无人机的大量内存。
- 4) 大规模的环境下,单无人机 SLAM 视觉里程计的累积误差随着时间增加。多无人机 SLAM 中,将目标区域划分为多个子区域,完成各区域的 SLAM,同时将获得的状态和地图进行融合,获得更精确的定位与地图。

1.3 本文主要研究内容

本文研究基于无人机的双目视觉 SLAM 技术,采用主流的基于特征点的方法实现双目视觉 SLAM。主要包括:构建双目视觉 SLAM 前端和后端,实现了单无人机的双目视觉 SLAM;研究多无人机的双目视觉 SLAM,重点研究两无人机的双目视觉 SLAM 的方法。

第一章:简述了本论文的研究背景和意义。介绍了无人机的研究现状、SLAM 的历史、现状和未来发展趋势;对视觉 SLAM 中重要的、具有代表性的开源方案进行了介绍。

第二章:首先,分析了双目相机的成像原理,研究了相机的针孔成像模型、双目相机的对极几何约束、相机的标定,完成了双目相机的标定;其次,基于特征提取中需要提取的特征,研究了两大主流特征提取算法: SIFT 和 ORB;最后,用无人机搭建双目相机 ZED、嵌入式设备构建实验平台,对室内、室外采集实验数据,对采集的图像分别进行基于 SIFT 和 ORB 的特征提取与匹配实验,对结果进行了分析,比较了 SIFT 和 ORB 的性能。

第三章:构建了双目视觉里程计。首先,对运动估计问题进行了研究,对 3D-2D、3D-3D 两种运动估计方法进行了研究与比较;其次,研究了视觉里程计中基于 Bundle Adjustment 的局部优化方法,用于获得更精确的位姿与地图;最后利用 SIFT 和 ORB 构建了基于室内外的双目视觉里程计,获得了无人机的初始位姿,同时生成了融合后的局部三维点云地图,对室内外获得的位置和地图进行了分析与比较。

第四章:构建基于非线性优化的双目视觉 SLAM 的后端。首先,研究并构建了双目视觉 SLAM 的后端优化模型,利用 g^2o 描述了 SLAM 后端优化模型;其次,研究了回环检测;再者,在回环检测的基础上,利用位姿图对前端获得位姿进行优化,获得更精确的位姿和融合后的全局三维点云地图;最后利用 SIFT 和 ORB 构建双目视觉 SLAM 系统,在室内、室外两种环境下进行了实验。

第五章：研究多无人机视觉 SLAM 方法；着重对两无人机视觉 SLAM 进行了理论研究，提出两种针对两无人机双目视觉 SLAM 的方法，并对两无人机双目视觉 SLAM 在室内外环境下进行了仿真实验。

第六章：对全文的研究内容进行了总结，并指出了本文目前存在的不足，说明了值得进一步研究的方向和问题，为后续的研究工作提供参考。

第二章 双目立体视觉成像系统及特征提取与匹配算法研究

2.1 引言

双目立体视觉成像系统主要包括双目立体相机的标定、立体矫正、立体匹配和双目测距原理。针对基于特征的视觉 SLAM 系统，需要对图像进行特征提取与匹配。在双目视觉 SLAM 系统中，能否准确构建双目立体成像系统，即精确实现双目立体相机的标定，立体矫正，立体匹配，从而获得精确的视差图，进一步获得深度图，是搭建双目视觉 SLAM 系统的基础。图像的预处理的结果决定了能否提取良好的特征点。特征提取与匹配的准确与否与速度快慢分别决定了后期能否获得精确的位姿信息、三维地图和实时在线 SLAM。

本章首先介绍了本文所用的双目立体相机 ZED，双目立体视觉成像系统的各个组成部分的原理。最后，对图像的特征提取与匹配算法进行了研究。

2.2 双目立体视觉成像系统

2.2.1 ZED 立体相机介绍

本文构建的双目立体视觉中选择了 ZED 立体相机，如图 2.1 所示。



图 2.1 ZED 立体相机

ZED 立体相机是一种重量轻，基于被动立体视觉的深度传感器。ZED 立体相机输出包含两个同步化的左右图像的视频。ZED 的驱动芯片的图像处理单元（GPU）从两个左右图像中计算深度图。因此 ZED 可以获得环境的 3D 地图。ZED 立体相机主要包含以下三方面优点。

- (1) 远距离的深度感知能力。以三维的方式感知自然，利用双目视觉和高分辨率的传感器，可以以 100FPS 获得室内或室外 70cm 至 20m 之间物体的深度。
- (2) 六轴位置追踪。可用于追踪相机在三维空间的运动；在不需要额外传感器的情况下，获得位置和姿态毫米的精确率。
- (3) 大规模的 3D 地图绘制。提供大规模和实时的 3D 重建，以秒的速度获得环境的 3D 模型。

2.2.2 ZED 双目立体视觉成像原理

本文所采用的 ZED 立体相机将三维场景变换为二维彩色图像。这种变换关系由三维到二维的映射描述^[19]:

$$(X, Y, Z) \rightarrow (x, y) \quad (2.1)$$

本文只考虑三维空间到二维空间的透视投影变化, 针孔成像模型近似的透视投影模型是最常用的成像模型。如图 2.2 所示, 其中 $P(X, Y, Z)$ 为 3D 空间点, O 为透镜中心, 相机坐标系 $O\text{-}XYZ$ 满足右手定则, 图像平面坐标系为 $O_m\text{-}xy$ 。图像坐标系的 x 轴, y 轴分别与相机坐标系的 X 轴, Y 轴平行, 但反向相反。经过投影中心与图像平面垂直的 Z 轴为光轴。 O 到图像平面的距离为焦距 f 。 P 通过透镜中心 O , 投影于图像平面上一点 $p(x, y)$, 投影为倒立的。为了不发生这种情况, 假设图像平面在投影中心前距离为 f 处, 如图 2.3 所示。过点 $P(X, Y, Z)$ 作平面 XOZ 的垂线于一点 M , 过点 M 做 OZ 轴的垂线与 OZ 轴交于一点 N , 连接点 P , N 。过 $p(x, y)$ 作 $O_m x$ 轴的垂线交 $O_m x$ 于一点 e , 连接点 p , O_m 。由三角形 ΔpOO_m 相似于三角形 ΔPON , 三角形 $\Delta pO_m e$ 相似于三角形 ΔPNM 可得透视投影方程:

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z} \quad (2.2)$$

式(2.2)使用含有齐次坐标的矩阵形式表示如下:

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.3)$$

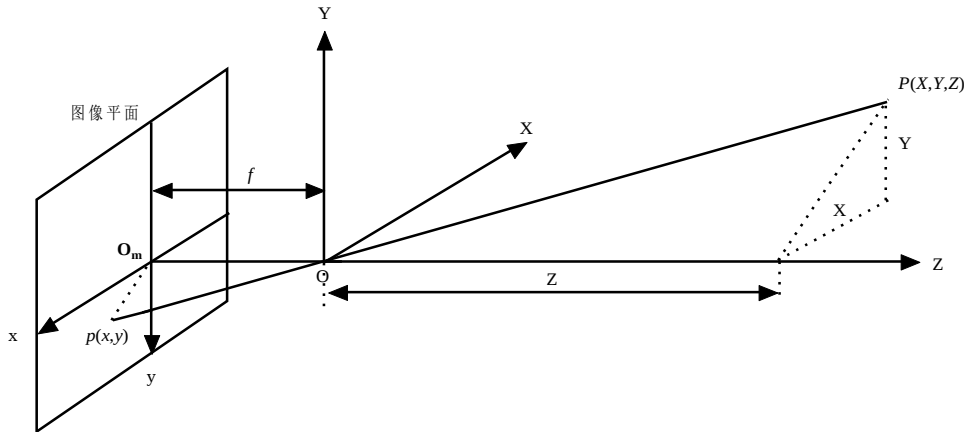


图 2.2 透视投影倒立成像几何示意图

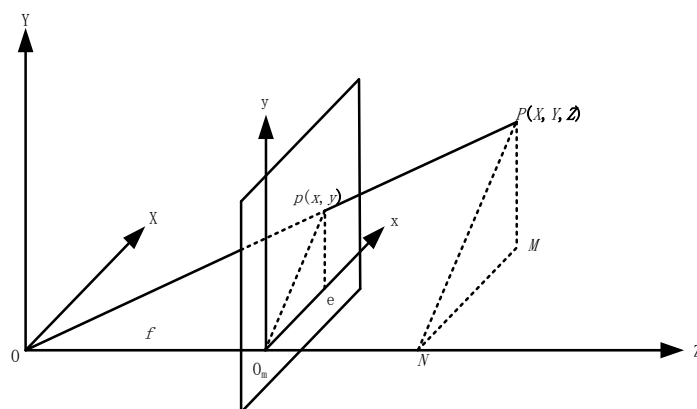


图 2.3 透视投影几何示意图

视觉成像系统中有四种坐标系：世界坐标系、摄像机坐标系、图像坐标系，像素坐标系。以上四者间可相互转化。世界坐标系即绝对坐标系；摄像机坐标系为以相机为中心的坐标；图像坐标系表示场景点在图像平面的投影；像素坐标系表示像素位置。以下内容讲述了四者之间的变换关系。

相机的平移、旋转运动时，相机不发生形变，为刚性变换。如图 2.4 所示，表示了世界坐标系和摄像机坐标系的变换关系。世界坐标系 $O_w - X_w Y_w Z_w$ 经过 3 自由度旋转变换、3 自由度平移变换为相机坐标系 $O - XYZ$ ，相应的变换矩阵分别为 R ， t ，即两者之间的关系可用方程表示：

$$P_c = RP_w + t \quad (2.4)$$

其中 P_w 代表世界坐标系某点坐标， P_c 代表相机坐标系下与 P_w 对应的点坐标。 R ， t 称为相机的外参数。

相机坐标系与图像坐标系的关系如上述图 2.2 所示，两个坐标系中对应点之间的变换关系由式 (2.2) 表达。

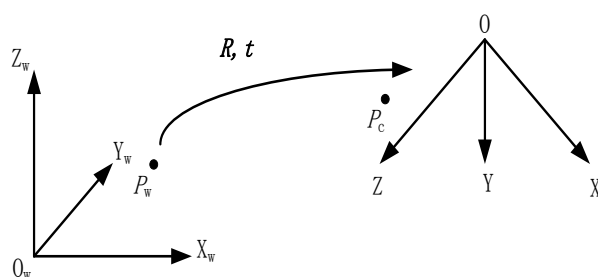


图 2.4 世界坐标系和摄像机坐标系之间的变换图

图像坐标系和像素坐标系之间的变换，如图 2.5 所示，其中 u_0 ， v_0 是图像平面中心，

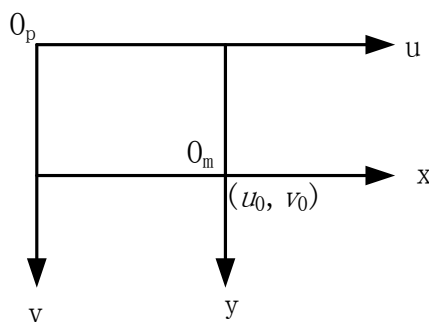


图 2.5 图像坐标系与像素坐标系变换关系图

$O_p - uv$ 表示像素坐标系， $O_m - xy$ 表示图像坐标系。图像坐标系和像素坐标系之间的关系可用方程表示如下：

$$\begin{aligned} u &= \frac{x}{dx} + u_0 \\ v &= \frac{y}{dy} + v_0 \end{aligned} \quad (2.5)$$

其中， dx 、 dy 表示相机感光芯片上的实际的像素大小。将式(2.5)表示成矩阵形式，如下所示：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.6)$$

同理，通过式(2.5)得到像素坐标转化为图像坐标的矩阵形式，如下所示：

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} dx & 0 & -u_0 dx \\ 0 & dy & -v_0 dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2.7)$$

综合式(2.3)、式(2.4)、式(2.5)、式(2.6)可得世界坐标与像素坐标之间的转换，如下所示：

$$Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \quad (2.8)$$

$$\begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{M}_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{M}_2 \mathbf{M}_3$$

式(2.8)中, u_0 、 v_0 、 f_x 、 f_y 为相机内参, $\mathbf{K}_1$ 称为内参数矩阵, $\mathbf{K} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$,

$\mathbf{M}_2 = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$ 称为相机的外参数矩阵, $\mathbf{M}_3 = \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$ 。

双目系统通过模拟人眼的立体成像功能,实现立体成像^[20]。双目立体成像系统采用透视投影成像原理,经过双目立体标定、立体矫正、立体匹配过程获得视差图,利用三角测量原理获得深度信息,从而获得深度感知能力。

2.2.3 ZED 立体相机标定

相机标定的目的是为了获得相机的内外参数,本文使用张正友的棋盘格标定方法^[21]。张正友标定法的具体算法如下:

(1) 由式(2.8)可知像素点坐标与世界坐标系下坐标之间的关系如下:

$$Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K} \mathbf{M}_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2.9)$$

式(2.9)将外参数矩阵表示成4个列向量的形式 $[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]$ 。

(2) 为方便计算,令 $Z_w = 0$,式(2.9)变为:

$$Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = H \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (2.10)$$

平面的单应性的含义为两平面的投影映射。式(2.10)中矩阵 $H = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}$ 称为单应性矩阵，表示 $Z_w = 0$ 时空间点和像素点之间的映射关系。

设

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad (2.11)$$

由式(2.9)、式(2.11)可得：

$$\begin{cases} Zu = h_{11}X_w + h_{12}Y_w + h_{13} \\ Zv = h_{21}X_w + h_{22}Y_w + h_{23} \\ Z = h_{31}X_w + h_{32}Y_w + 1 \end{cases} \quad (2.12)$$

由式(2.12)得到：

$$\begin{cases} u = \frac{h_{11}X_w + h_{12}Y_w + h_{13}}{h_{31}X_w + h_{32}Y_w + 1} \\ v = \frac{h_{21}X_w + h_{22}Y_w + h_{23}}{h_{31}X_w + h_{32}Y_w + 1} \end{cases} \quad (2.13)$$

由式(2.13)可得：

$$\begin{cases} h_{11}X_w + h_{12}Y_w + h_{13} = uh_{31}X_w + uh_{32}Y_w + u \\ h_{21}X_w + h_{22}Y_w + h_{23} = vh_{31}X_w + vh_{32}Y_w + v \end{cases} \quad (2.14)$$

令

$$\mathbf{h} = [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ h_{31} \ h_{32} \ 1]^T \quad (2.15)$$

则

$$\begin{bmatrix} X_w & Y_w & 1 & 0 & 0 & 0 & -uX_w & -uY_w & -u \\ 0 & 0 & 0 & X_w & Y_w & 1 & -vX_w & -vY_w & -v \end{bmatrix} \mathbf{h} = 0 \quad (2.16)$$

为求解含 8 个未知数的向量 $\mathbf{h}$ ，由式(2.16)可知，至少需要 4 组对应点。基于误差的考虑，实际标定过程中会多选择几组对应点，并利用最小二乘方法求出向量 $\mathbf{h}$ ，因此得到单应性矩阵 H 。

(3) 令 $\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$, 由式(2.10)可知, 存在比例因子 $s = \frac{1}{Z}$, 因此可得包含缩放因子的单应性矩阵, 即有:

$$\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = s\mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \quad (2.17)$$

因为 $\mathbf{R}$ 为旋转矩阵, 所以 $\mathbf{r}_1$ 和 $\mathbf{r}_2$ 是单位正交向量, 则:

$$\begin{cases} \mathbf{r}_1^T \mathbf{r}_2 = 0 \\ \mathbf{r}_1^T \mathbf{r}_1 = \mathbf{r}_2^T \mathbf{r}_2 \end{cases} \quad (2.18)$$

综合式(2.17)和式(2.18)可得:

$$\begin{cases} \mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 = 0 \\ \mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 \end{cases} \quad (2.19)$$

设

$$\mathbf{B} = \mathbf{K}^{-T} \mathbf{K}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \quad (2.20)$$

由式(2.8)得:

$$\mathbf{B} = \begin{bmatrix} \frac{1}{f_x^2} & 0 & \frac{-u_0}{f_x^2} \\ 0 & \frac{1}{f_y^2} & \frac{-v_0}{f_y^2} \\ \frac{-c_x}{f_x^2} & \frac{-c_y}{f_y^2} & \frac{c_x^2}{f_x^2} + \frac{c_y^2}{f_y^2} + 1 \end{bmatrix} \quad (2.21)$$

由上式(2.21)可知: $\mathbf{B}$ 为对称矩阵, 所以得到:

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{12} \\ B_{22} \\ B_{13} \\ B_{23} \\ B_{33} \end{bmatrix} \quad (2.22)$$

由式(2.22)可将式(2.19)表示为:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{12})^T \end{bmatrix} \mathbf{b} = 0 \quad (2.23)$$

设采集 K 个棋盘图像, 则用矩阵形式表示获得方程组, 如下所示:

$$\mathbf{V}\mathbf{b} = \mathbf{0} \quad (2.24)$$

其中, $\mathbf{V}$ 是 $2K \times 6$ 矩阵。由式(2.21)可知矩阵 $\mathbf{B}$ 包含 4 个未知数, 所以当 $K \geq 2$ 时, 方程有解 $\mathbf{b} = [B_{11} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}]^T$ 。根据式(2.21)可得摄像机内参数如下所示:

$$\begin{cases} f_x = \sqrt{\lambda / B_{11}} \\ f_y = \sqrt{\lambda B_{11} / (B_{11}B_{22} - B_{12}^2)} \\ c_x = -B_{13}f_x^2 / \lambda \\ c_y = (B_{12}B_{13} - B_{11}B_{23}) / (B_{11}B_{22} - B_{12}^2) \end{cases} \quad (2.25)$$

其中式(2.25)中:

$$\lambda = B_{33} - [B_{13}^2 + c_y(B_{12}B_{13} - B_{11}B_{23})] / B_{11} \quad (2.26)$$

外参数由式(2.17)可得:

$$\begin{cases} \mathbf{r}_1 = \frac{1}{s} \mathbf{K}^{-1} \mathbf{h}_1 \\ \mathbf{r}_2 = \frac{1}{s} \mathbf{K}^{-1} \mathbf{h}_2 \\ \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{t} = \frac{1}{s} \mathbf{K}^{-1} \mathbf{h}_3 \end{cases} \quad (2.27)$$

其中式(2.27)中 $s = \|\mathbf{K}^{-1} \mathbf{h}_1\|$ 。

透镜畸变是指相机透镜在制作过程中而产生畸变, 主要包含径向畸变和切向畸变。径向畸变来自于透镜的形状; 切向畸变是因为在相机的组装过程中, 透镜本身与图像平面的偏移产生。本文主要考虑径向畸变, 透镜中心的径向畸变为 0, 透镜中心的光线弯曲比偏离中心轻。用 $r = 0$ 位置周围的泰勒级数展开的前两项定量描述, 第一项为 k_1 , 第二项为 k_2 。所以有:

$$\begin{cases} x_c = x(1 + k_1 r^2 + k_2 r^4) \\ y_c = y(1 + k_1 r^2 + k_2 r^4) \end{cases} \quad (2.28)$$

式(2.28)中, (x, y) 是原始位置, (x_c, y_c) 是校正后的新位置。

双目立体相机的标定需要获得各相机的内外参数、两相机间的基线距离和变换矩阵。如图 2.6 所示, O_l 、 O_r 分别为左右相机的投影中心, 空间三维点 P 在左右相机投影平面 I_l 和 I_r 投影点分别为 p_l 和 p_r 。 O_l 在 I_r 上的投影点为 e_r , O_r 在 I_l 上的投影点为 e_l ; e_r , e_l 称为极点。直线 $p_l e_l$ 和 $p_r e_r$ 称为极线。投影平面之间的平移矩阵为 $\mathbf{t}$, 旋转矩阵为 $\mathbf{R}$ 。

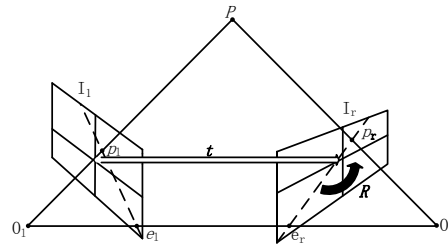


图 2.6 对极几何约束

左右相机坐标系相对于世界坐标系的旋转矩阵分别为 $\mathbf{R}_l$ 、 $\mathbf{R}_r$ ，平移矩阵为 $\mathbf{t}_l$ 、 $\mathbf{t}_r$ 。则有：

$$\begin{cases} \mathbf{X}_L = \mathbf{R}_l \mathbf{X}_w + \mathbf{t}_l \\ \mathbf{X}_R = \mathbf{R}_r \mathbf{X}_w + \mathbf{t}_r \end{cases} \quad (2.28)$$

由式(2.28)可得：

$$\mathbf{X}_R = \mathbf{R}_r \mathbf{R}_l^{-1} \mathbf{X}_L + \mathbf{t}_r - \mathbf{R}_r \mathbf{R}_l^{-1} \mathbf{t}_l \quad (2.29)$$

所以得：

$$\begin{cases} \mathbf{R} = \mathbf{R}_r \mathbf{R}_l^{-1} \\ \mathbf{t} = \mathbf{t}_r - \mathbf{R}_r \mathbf{R}_l^{-1} \mathbf{t}_l \end{cases} \quad (2.30)$$

由式(2.30)可知：获得了左右相机的各自的外参数后，便可以求得两相机间对应的变换矩阵 $\mathbf{R}, \mathbf{t}$ ，即两相机间的外参数。综上所述，本系统利用张正友棋盘标定法完成立体相机的标定。标定结果如表 2.1 所示。

表 2.1 ZED 标定后参数

左相机内参数	$u_0 = 646.21$	$v_0 = 363.868$
	$f_x = 701.129$	$f_y = 701.129$
	$k_1 = -0.168772$	$k_2 = 0.022334$
右相机内参数	$u_0 = 653.858$	$v_0 = 360.087$
	$f_x = 698.875$	$f_y = 698.875$
	$k_1 = -0.169805$	$k_2 = 0.0244331$
立体相机参数	基线等于 120mm	
	$RY = 0.01441$	
	$RX = 0.000326543$	
	$RZ = -0.000329152$	

2.2.4 立体校正

立体校正的目的是将相机在数学上对准到同一个观察平面上，使得摄像机之间的像素行严格互相对准^[22]，即：两相机的主光线平行，图像平面在同一平面且图像的行完全地对准到前向平行的结构上，可利用立体标定所获得的参数进行立体校正。如图 2.6 所示为未校正的立体成像图，左右相机的投影中心分别为 O_l 、 O_r 。图 2.7 所示为校正后的立体成像图，空间点 P 在左右相机图像平面上的投影点分别为 p_l 、 p_r 。直线 $p_l p_r$ 平行于直线 $O_l O_r$ 。

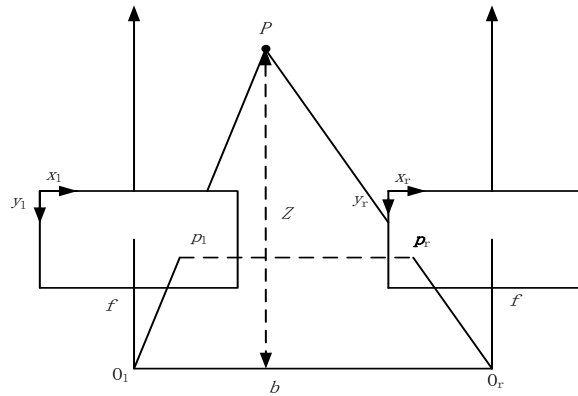


图 2.7 立体校正后立体成像图

2.2.5 立体匹配

在完成立体校正后，找出左右图像中匹配的点对，计算匹配点对在各自图像坐标系中 x 的坐标差值，获得视差图，这便是立体匹配的目的。利用立体成像系统的图像几何约束和场景约束^[23]减少对应点的误匹配和实现匹配。基于各异的约束信息，立体匹配算法包含局部算法和全局算法。局部算法包含区域算法、特征匹配和相位匹配；全局算法包含动态规划算法、图切割算法^[24]。在 opencv 中实现的三种立体匹配算法：双向匹配(Bidirectional Matching, BM)、全局匹配算法(Semi-Global Block Matching, SGBM)^[25]和计算机图解(Graphics Computer, GC)^[26]。

2.2.6 三角测量

当完成上述立体标定、立体校正、立体匹配后，本文得到无畸变、对准的立体成像图，如图 2.8 所示。空间三维点 P 在左右图像平面的投影点的横坐标分别为 x_l 、 x_r 。

从图 2.8 可由三角形 $\Delta P x_l x_r$ 相似于 $\Delta P O_l O_r$ 得：

$$\frac{b - (x_l - x_r)}{Z - f} = \frac{b}{Z} \quad (2.31)$$

则

$$Z = \frac{fb}{D} \quad (2.32)$$

其中, $D = x_l - x_r$ 称为视差, 由式(2.32)知, 视差 D 与深度 Z 成反比。

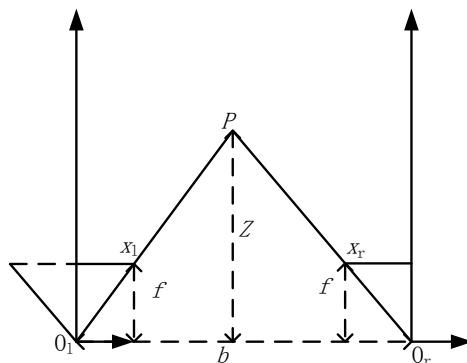


图 2.8 无畸变对准的立体成像投影图

2.3 特征提取与匹配算法

在视觉 SLAM 中, 特征提取与匹配是基本且重要的步骤, 特征提取的数量多少, 特征匹配的正确率的高低决定了视觉 SLAM 中位姿的变换矩阵的正确率, 即决定了位姿的正确率; 同时也决定地图的正确率。因此, 获得良好的特征, 匹配是实现视觉 SLAM 的前提, 是实现在线实时 SLAM 的重要环节, 是基于视觉 SLAM 导航的关键。常见的特征提取算法有尺度不变特征变换算法 (Scale-Invariant Feature Transform, SIFT)^[27]、ORB(Oriented FAST and Rotated BRIEF)^[28], 常见的特征匹配算法有最近邻匹配算法。下文主要介绍 SIFT, ORB 两种特征提取算法。

2.3.1 SIFT 算法

1999 年, 被提出的尺度不变特征可用于图像之间的匹配^[29]。SIFT 算子具有尺度不变、旋转不变、平移不变的特性, 对一定程度的光强的变化具有不变性。SIFT 算法原理是寻找不变的特征点, 进而获得图像的局部特性, 构成局部描述子。通过局部描述子, 对图像进行匹配。SIFT 提取的特征具有很好的独特性, 可用于对准确率需求较高的匹配中, 具有较好的鲁棒性。SIFT 算子可应用于基于图像的诸多应用中, 如基于视觉的识别、跟踪等。

由 SIFT 算法获得的特征点有尺度不变性, 根据尺度空间理论得到尺度不变性。多尺度表示是计算机理解图像的有效方式。多尺度表示的思想是: 用一个信号函数表示图像, 该函数包含一个描述原始图像的尺度参数。多尺度表示中较粗尺度为较细尺度平滑得到, 高斯函数是唯一可用的平滑函数。即: 较细尺度可通过高斯函数获得较粗尺度, 如图 2.9 所示。早期多尺度表示方式有四分树、八分树、图像金字塔。由于上述多尺度表示方法中尺度之间的间隔太大, 因

此后期有尺度空间表示法。尺度空间表示法的尺度参数是连续的，在各个尺度上图像的分辨率是一样。

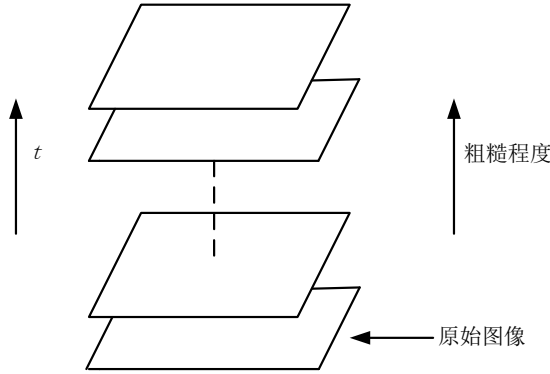


图 2.9 图像的多尺度表示图

在某些前提下，高斯函数是尺度空间唯一的线性的平滑核函数。图像与高斯函数做卷积得到尺度空间，如下所示：

$$L(x, y, t) = \int_{\xi=-\infty}^{\infty} \int_{\eta=-\infty}^{\infty} \frac{1}{2\pi t} e^{-(\xi^2 + \eta^2)/2t} f(x - \xi, y - \eta) d\xi d\eta \quad (2.33)$$

式 2.33 中， $L(x, y, t)$ 为尺度参数 t 时的尺度。

SIFT 算法有以下四个步骤：

1) 检测尺度空间极值点

图像函数 $I(x, y)$ 的尺度函数 $L(x, y, \delta)$ 可由 $I(x, y)$ 与高斯函数 $G(x, y, \sigma)$ 卷积获得，如下式所示：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.34)$$

利用高斯差分函数与图像卷积得到 $D(x, y, \sigma)$ ，在 $D(x, y, \sigma)$ 中寻找的极值点中获取稳定特征点的位置，即关键点的位置。

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2.35)$$

上式 2.35 中， k 为常数。

$D(x, y, \sigma)$ 尺度归一化拉普拉斯高斯函数 $\sigma^2 \nabla^2 G$ 有关系式：

$$\sigma^2 \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (2.36)$$

由式(2.36)得：

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G \quad (2.37)$$

因此， $D(x, y, \sigma)$ 包含因子 $\sigma^2 \nabla^2 G$ 。 $\sigma^2 \nabla^2 G$ 的极值具有稳定的特征，在尺度空间 $D(x, y, \sigma)$ 中寻找极值点转化为在 $\sigma^2 \nabla^2 G$ 中寻找极值点。

为获得尺度无关的特征点，首先利用金字塔高效计算高斯差分图像，如图 2.10 所示。

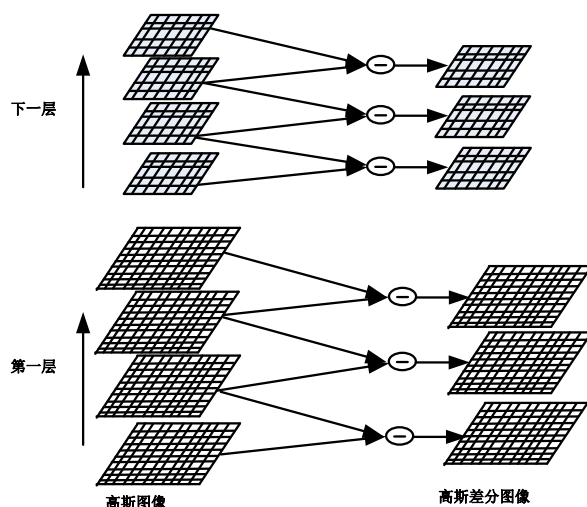


图 2.10 金字塔高斯差分图

由最初图像和高斯函数做卷积得到图像，组成金字塔的第一层。高斯差分图像由金字塔同一层相邻高斯图像差分而来。从第一层中对某图像 a 降采样得到图像 b ， b 为 a 的 0.25。图像 b 成为下一层的初始图像，依旧进行上述图像 a 的操作。图像 a 的尺度空间参数 σ 为初始尺度空间参数的 2 倍。假设需要在 s 个尺度中寻找极值点，则每层 $s+3$ 幅高斯图像， $s+2$ 幅高斯差分图像。

高斯差分图像中的像素点需要与它的 8 个领域像素点、上下两层各 9 个邻近的像素点，共 26 个像素点进行比较。如图 2.11 所示。

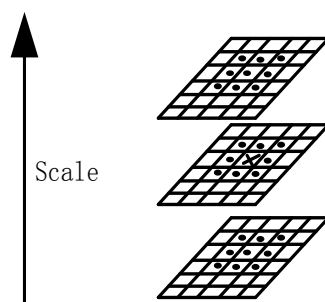


图 2.11 像素点的空间 26 邻近像素点图

2) 提取稳定的关键点

在获得极值点的基础上，去除极值点中不稳定的特征点（低对比度的点和边缘上的点），可提高特征点匹配中抗噪声能力。由于二维函数离散空间得到的极值点与连续空间极值点存在误差，因此，利用获得离散空间的极值点进行曲线拟合，再获得曲线对应的极值点的方法称为子

像素插值(Sub-pixel Interpolation)。

在某极值点 P 对 $D(x, y, \sigma)$ 进行泰勒展开如下式所示：

$$D(x) = D + \frac{\partial D^T}{\partial x} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial x^2} X \quad (2.38)$$

其中， $X = (x, y, \sigma)^T$ 是到点 P 的偏移量。对求式(2.38)导数，令导数为 0，可得：

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (2.39)$$

将式(2.39)代入式(2.38)得：

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad (2.40)$$

若 $|D(\hat{x})| \geq 0.03$ ，保留该极值点；否则，丢弃该极值点。

抗噪性差的位于边缘的极值点，主曲率之比大于非边缘的极值点的主曲率之比。待测极值点的海森矩阵(Hessian matrix)如下：

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.41)$$

$\mathbf{H}$ 的特征值和 D 的主曲率对应成比例，设特征值 α ， β ，且 $\alpha > \beta$ ，由矩阵的性质得：

$$\begin{aligned} \text{Tr}(\mathbf{H}) &= D_{xx} + D_{yy} = \alpha + \beta \\ \text{Det}(\mathbf{H}) &= D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (2.42)$$

设 $r = \alpha / \beta$ ，结合式(2.42)得：

$$\frac{\text{Tr}(\mathbf{H})}{\text{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r + 1)^2}{r} \quad (2.43)$$

式(2.43)可知，当 $|r| \geq 1$ 时，函数 $\frac{\text{Tr}(\mathbf{H})}{\text{Det}(\mathbf{H})}$ 是单调递增函数。选取阈值 $R = 10$ ，则非边缘

点需要满足下式 (2.44)，反之为边缘点。

$$\frac{\text{Tr}(\mathbf{H})}{\text{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r + 1)^2}{r} < \frac{(R + 1)^2}{R} \quad (2.44)$$

3) 提取关键点的局部特性

以提取的每个关键点为中心，确定一圆形区域窗口，窗口大小与关键点的尺度成正比。为窗口内各采样点 $L(x, y)$ 确定方向和模值，如下所示：

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (2.45)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$$

利用窗口中各采样点的梯度方向和模值为关键点确定方向参数，而窗口内的非关键点方向都为相对于关键点方向保持不变，因此，关键点具有旋转不变性。关键点包含信息：位置，尺度，方向。窗口大小与关键点的尺度成正比使得关键点获得的方向信息不因尺度变化而变化，因此关键点具有尺度不变特性。将窗口内所有点的方向和模值确定一个方向直方图，如图 2.12 所示，直方图的横坐标为方向，将圆周划分为 8 份，每份包含 45 度角度范围；纵坐标为对应横坐标的采样点的模值之和。选取纵坐标值最大(峰值)对应的横坐标的方向作为关键点的主方向，当其他纵坐标值大于峰值的 80% 时，表明还有一个该方向的关键点存在。将坐标轴旋转 to 关键点方向，这便确保了具有旋转不变性。

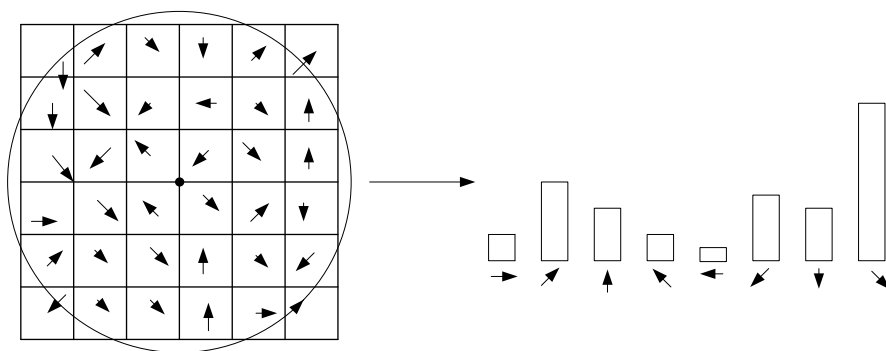


图 2.12 关键点方向直方图

4) 生成特征描述子

将以关键点为中心的邻域内的点的梯度做高斯加权，如图 2.13 所示。将该领域分为四个子区域，其中每个子区域有八个方向。图 2.13 中表示 6×6 的像素矩阵，圆形邻域。邻域分为 2×2 个子区域，每个子区域包含八个方向，将每个子区域的特征点的同方向的梯度模值累加，获得四个特征描述子，每个特征描述子包含八个方向。在实际的实验中，采用 16×16 的像素区域，将邻域划分为 4×4 个子区域，每个子区域获得一个描述子，每个描述子包含八个方向，因此，每个关键点有 $8 \times 4 \times 4 = 128$ 维。获得描述子向量后，为了去除光照变化的影响，对向量进行归一化处理。

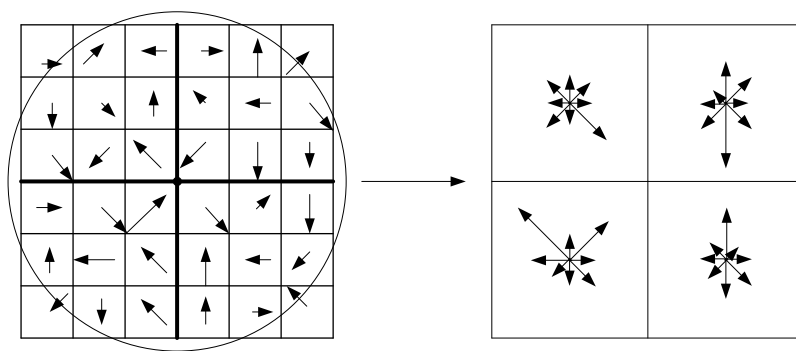


图 2.13 关键点生成描述子图

2.3.2 ORB 算法

1) FAST 算法

图像的角点为图像中感兴趣的像素点，此像素点具有很好的辨识度、能很好地被检测、携带大量有用的信息、在不同图像中有较高的重复率。因此，角点被广泛应用到图像匹配、目标识别、目标跟踪领域中。FAST(Features from Accelerated Segment Test)^[30-31]广泛应用于图像的角点检测。FAST 角点检测的原理：在某待检测的像素点 p 的半径为 r 的圆邻域中，比较 p 的灰度值与其他像素点 p_i 的灰度值的大小， $i=1,2,\dots,n$ ， n 为圆周上的像素点个数。统计 $|p-p_i| > t_1$ 时的次数 m ，如果 m 大于阈值 T ，则像素点 p 可作为角点，实际中 $T=12$ 。如图 2.14^[31]所示，以像素点 p 为圆心的一半径为 3 的圆，圆周上包含 16 个像素点，即 $n=16$ ， $i=1,2,\dots,16$ 。统计 $|p-p_i| > t_1$ 时的次数并与阈值 T 进行比较判断像素点 p 是否为角点。

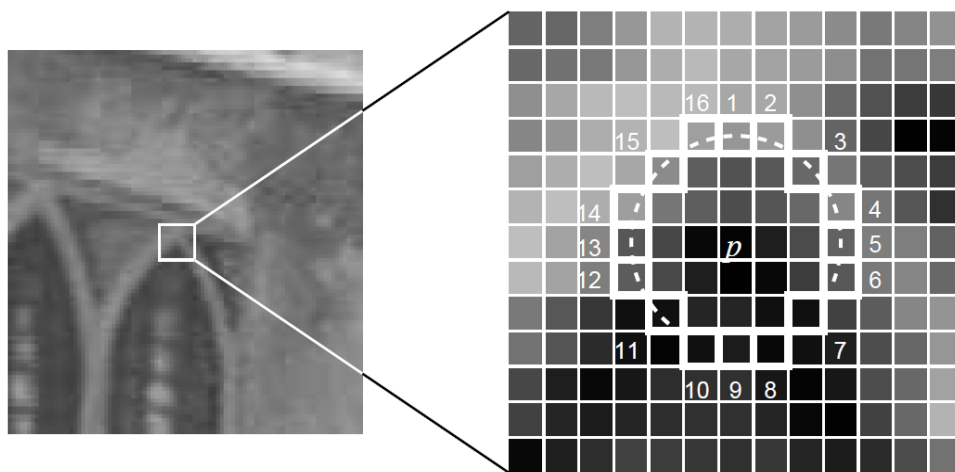


图 2.14 FAST 算法原理图

为了提升 FAST 算法的快速性，简单的方法为先进进行一次粗略的判断，对图 2.13 中下标 $i=1,5,9,13$ 的像素点 p_i 进行 $|p-p_i| > t_1$ 的比较。当此次比较次数至少为三时，再对余下的像素点进行上述操作；否则，判定该像素点 p 不是角点，不再对其余像素点进行于像素点 p 比较。

因此,可以在一定程度上提升 FAST 算法的快速性。同时,根据上述 16 个像素的计算顺序,得到的算法速度也不同^[32]。

FAST 角点检测算法的速度几乎是 SIFT 中利用高斯差分卷积提取关键点的 30 倍,因此 FAST 角点检测算法具有快速性,可用于实时性要求较高的角点检测中。FAST 的不足之处在于其不具有尺度不变性。

2) BRIEF 算法

SIFT 算法中对特征的检测是基于高斯尺度空间,对特征的描述是通过像素点的梯度方向、模值的处理获得特征描述子。随着对特征提取与匹配的要求在实时性和结果的准确性越来越高,BRIEF(Binary Robust Independent Elementary Features)^[33]应运而生。BRIEF 描述子计算简单、快速。BRIEF 算法原理:在特征点某一邻域内选取 n 对像素点,通过比较每对像素点的灰度值,获得一个 n 位的二进制串,此二进制串便是该特征点的描述子。

与 $s*s$ 尺寸的图像邻域 p 有关的 τ 定义如下:

$$\tau(p; x, y) = \begin{cases} 1 & p(x) < p(y) \\ 0 & \text{其他} \end{cases} \quad (2.46)$$

其中, x 、 y 为一对像素点, $p(x)$ 、 $p(y)$ 分别为其对应的像素灰度值。有 n 个像素点对, n 的值可为 128 或 256 等,则得到 n 的二进制串,可表示为:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x, y) \quad (2.47)$$

其中,对于描述子中的像素对 (x, y) 的空间分布有五种方案^[34]。

BRIEF 描述子对灰度变化具有很强的鲁棒性,但对噪声和旋转敏感。针对 BRIEF 描述子对噪声敏感,需要对图像进行平滑。常见的平滑处理包括均值滤波、中值滤波,高斯滤波。

ORB 是一种基于 FAST 特征检测和将改进的 BRIEF 作为描述子的快速特征点提取与描述的算法^[35]。ORB 算法速度为 SIFT 算法的 100 倍。ORB 算法分为两个步骤:以改进的 FAST 算法用于特征检测;以改进的 BRIEF 特征描述子用于描述特征。如前所述,FAST 算法的缺陷是不具备旋转不变性,因此,ORB 将 FAST 改进为包含方向的 oFAST(FAST Keypoint Orientation)。ORB 算法将特征点坐标到质心构成的向量的方向作为该特征点的方向,其中质心指的是以该特征点为圆心, r 为半径构成的圆的质心。质心可通过矩计算得到,矩的定义如下:

$$m = \sum_{x, y \in r} x^p y^q I(x, y) \quad (2.48)$$

上式中, $I(x, y)$ 是图像的灰度值。矩的质心为:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (2.49)$$

当特征点的坐标为原点时，特征点的方向为 θ ：

$$\theta = \arctan\left(\frac{m_{01}}{m_{00}} / \frac{m_{10}}{m_{00}}\right) \quad (2.50)$$

为使得改进的 BRIEF 具有旋转不变性，ORB 定义了方向引导 BRIEF，称为 Steered BRIEF。

利用 BRIEF 中的 n 对像素点对 (x_i, y_i) 定义 $2*n$ 维矩阵如下：

$$S = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix} \quad (2.51)$$

利用邻域方向 θ 和对应的旋转矩阵 R_θ 获得 S 校正后的 S_θ ： $S_\theta = R_\theta S$ ，其中，

$$R_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (2.52)$$

其中， θ 为特征点检测中求得的 oFAST 特征点的主方向。因此，Steered BRIEF 描述子为：

$$g_n(p, \theta) := f_n(p) | (x_i, y_i) \in S_\theta \quad (2.53)$$

基于 BRIEF 在构造描述子过程中对噪声敏感，ORB 对两像素块之间中各像素块中所有像素的灰度值之和进行比较代替 BRIEF 中两像素间的灰度值进行比较。

2.3.3 特征匹配策略

在获得特征之后，需要完成特征匹配。所谓特征匹配便是依据两个特征点的描述子向量的匹配程度决定是否为一特征点。以欧氏距离作为特征点的匹配策略有：固定阈值、最近邻、最近邻距离比率。

固定阈值匹配策略：两特征点的描述子向量的欧氏距离之差低于某一阈值，判定两特征点匹配成功，表示同一特征点。假设特征点 p 和 q 的特征描述子分别为 D_p 和 D_q ，则欧式距离的定义如下：

$$d = \sqrt{\sum_{i=1}^n (D_p(i) - D_q(i))^2} \quad (2.54)$$

在 SIFT 算法中， n 一般为 128；ORB 算法中， n 可以为 128 或 256。基于阈值的选取的困难程度大，以及容易出现一个特征点存在多个与之匹配的特征点，基于阈值的匹配策略容易出现误匹配。

最近邻匹配策略：顾名思义，在目的特征点集合中选择与待匹配的特征点的描述子向量的欧式距离最小的特征点作为匹配的特征点。因此，此策略获得的特征点具有唯一性。为了减少误匹配，上述的最小欧式距离的低于某一阈值。

最近邻距离比率匹配策略：在获得待匹配的特征点与目的特征点间的描述子向量的欧氏距离时，当最近邻特征点与次最近邻特征点的描述子向量间的欧氏距离之比低于某阈值时，选取该最近邻特征点作为匹配的特征点。此匹配策略比上述两种匹配策略效果更佳。

常见的用于搜索匹配特征点的方法有：穷举法和 $k-d$ 搜索树法。将待匹配的特征点与目的图像的所有特征点一一匹配，这便是穷举法。此方法简单，但是耗费时间，实时性差。 $k-d$ 搜索树是一种二叉树，树中节点为 k 维。 $k-d$ 搜索树法中利用多维搜索树索引结构，可快速搜索特征点，完成匹配。

2.4 特征提取与匹配实验

本文的实验器件:四旋翼无人机、NVIDIA JETSON TK1、ZED 相机、Acer 显示器、USB Hub、键盘、鼠标、HMDI 转 VGA 数据线。操作系统为开源 Linux 操作系统 ubuntu14.04 LTS。室外实验环境为江苏南京航空航天大学江宁校区东区自动化学院 3 号楼旁。室内环境为室外实验环境为江苏南京航空航天大学江宁校区东区自动化学院 3 号楼 403 办公室。本实验的实验平台如图 2.15 所示。



(a) NVIDIA JETSON TK1 开发板

(b) 无人机搭载 ZED 和 TK1

图 2.15 无人机双目视觉 SLAM 实验硬件平台

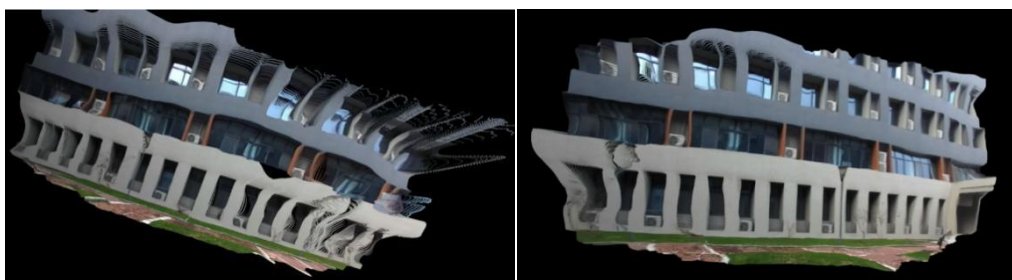
2.4.1 室外 SIFT、ORB 特征提取与匹配

在 k 、 $k+1$ 时刻 ZED 采集而来的两张左相机获得的 RGB 图，分别为 L_k 、 L_{k+1} ，如图 2.16 所示。运用三角测量得到的深度值结合透视投影关系，可得到各 RGB 图对应的三维点云图，如图 2.17 所示。利用 SIFT 进行特征提取，快速最近邻 FLANN (Fast Library for Approximate Nearest Neighbors) 进行特征匹配获得的结果，如图 2.18 所示。



(a) k 时刻左相机采集的室外 RGB 图 (b) $k+1$ 时刻左相机采集的室外 RGB 图

图 2.16 无人机用 ZED 在两个时刻左相机采集到的室外 RGB 图



(a) k 时刻对应的三维点云图

(b) $k+1$ 时刻对应的三维点云图

图 2.17 室外两时刻的三维点云图

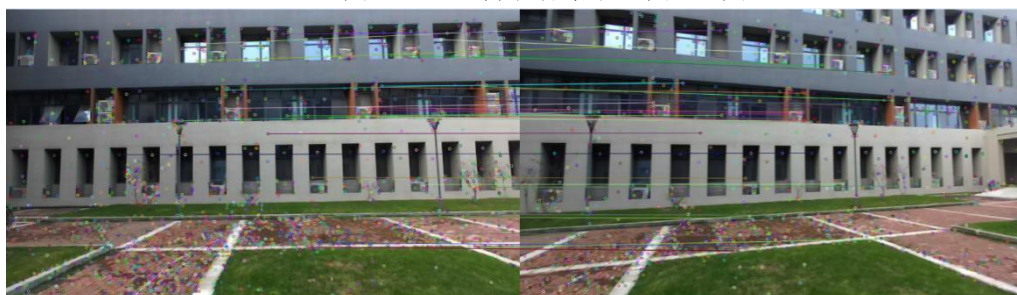


图 2.18 利用 SIFT 算法获得的室外特征匹配图

根据已获得的上述 RGB 图 L_t 、 L_{t+1} , 运用 ORB 进行特征提取, 暴力匹配 (Brute Force match) 进行特征匹配, 获得匹配后的结果, 如图 2.19 所示。



图 2.19 利用 ORB 算法获得的室外特征匹配图

由图 2.18、图 2.19 可直观看知, SIFT 相比 ORB 获得更多的匹配点对, 但是两者还存在误匹配, 且 ORB 算法在室外对光敏感。

对由 ZED 获得的 10、50、100、129 张左相机的 RGB 图进行特征提取, 分别利用上述的

SIFT、ORB 进行特征提取，相应的匹配算法进行特征匹配，所需要的时间如表 2.2 所示：

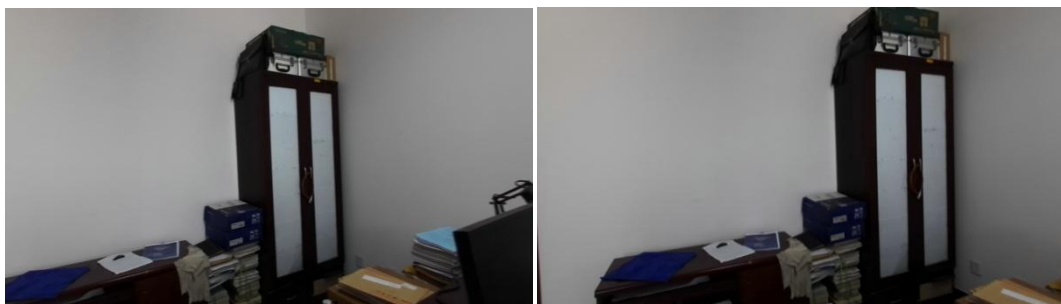
由表 2.2 可知，SIFT 比 ORB 获得更多的特征点，SIFT 特征提取和特征匹配的时间约为 ORB 的特征提取和特征匹配的时间的二十倍，所以 SIFT 比 ORB 获得更多的匹配点，但是 ORB 在处理器计算资源有限且对实时性要求高的 SLAM 系统中更有优势。

表 2.2 室外 ORB 和 SIFT 特征提取与匹配比较

算法	匹配图像 对大小 (张)	特征数量 (个)	特征提取 时间 (ms)	特征匹配 时间 (ms)	总时间 (ms)
SIFT	1	5709	3775	4440	8215
	10	50804	37693	42741	80435
	50	189796	183081	198288	381369
	100	398103	373075	405878	778952
ORB	1	1000	143	183	326
	10	10000	1419	1836	3255
	50	50000	6475	9136	15611
	100	100000	13095	18454	31549

2.4.2 室内 SIFT、ORB 的特征提取与匹配

在室内环境下，无人机进行数据采集如图所示。在 k 时刻和 $k+1$ 时刻 ZED 采集而来的两张左相机获得的 RGB 图，分别为 l_k 、 l_{k+1} ，如图 2.20 所示，各自对应的三维点云图如 2.21 所示。分别利用上述的 SIFT、ORB 进行特征提取，相应的匹配算法进行特征匹配，分别如图 2.22、图 2.23 所示。



(a) k 时刻左相机采集的室内 RGB 图

(b) $k+1$ 时刻左相机采集的室内 RGB 图

图 2.20 无人机用 ZED 在两个时刻左相机采集到的室内 RGB 图



(a) k 时刻左相机的室内三维点云图 (b) $k+1$ 时刻左相机的室内三维点云图
图 2.21 室内两时刻对应的三维点云图

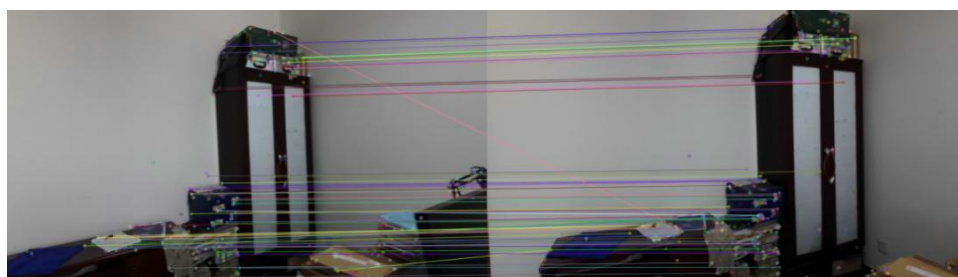


图 2.22 利用 SIFT 算法获得的室内特征匹配图

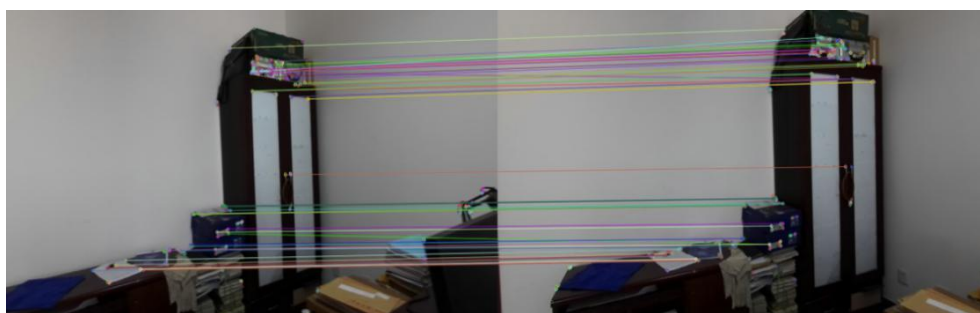


图 2.23 利用 ORB 算法获得的室内特征匹配图

由图 2.22、图 2.23 可知，在室内的情况下，由于光源噪声减少，ORB 算法获得匹配数量与 SIFT 算法获得的匹配数量差距不大。相比于室外，室内的特征点数量相对减少。

2.5 本章小结

本章主要研究了基于 ZED 相机的特征提取与特征匹配算法。首先，研究了 ZED 相机的双目立体成像原理，对 ZED 相机进行标定；其次研究了用于两种特征提取与匹配的算子：SIFT 和 ORB；最后，利用 ZED 获得室外、室内 RGB 图，采用 SIFT 和 ORB 实现了特征提取与匹配。通过室外 ORB 和 SIFT 特征提取与匹配比较得出，SIFT 特征提取和特征匹配的时间约为 ORB 的特征提取和特征匹配的时间的二十倍，SIFT 获得特征点相比 ORB 获得的更多，ORB 在室外对光敏感。在室内的情况下，由于噪声光源的较少，ORB 获得的特征匹配数量和 SIFT 获得的特征匹配数量相差不多。

第三章 无人机的双目视觉里程计研究

3.1 引言

在机器人学与计算机视觉领域,视觉里程计(Visual Odometry, VO)是一个仅仅使用相机获得图像序列来确定机器人位置和姿态的过程。视觉里程计利用连续的图像序列来估计机器人位姿。在 SLAM 中,视觉里程计称为 SLAM 的前端。视觉里程计的领域包括机器人学,可穿戴计算设备,增强现实和自主运动^[36]。术语 VO 一词首先在 2004 年由 Nister 在他的手稿中被提出^[37],被选择的原因是其类似于轮式里程计。通过获得车辆轮胎的转数于周长,轮式里程计增量性地估计车辆的运动。同理,通过视觉获取图像序列,视觉里程计增量性地估计位姿。为了使得视觉里程计良好地工作,必须有良好的光照条件,并且场景拥有足够的纹理可以被提取。视觉里程计相对于轮式里程计的优势在于视觉里程计不会受轮胎打滑的影响,特别是在不平整的地形中。本文涉及的无人机的双目视觉里程计有五个步骤:图像获取、特征提取、特征匹配、运动估计、局部优化。如图 3.1 所示为无人机的双目视觉里程计的流程图。顾名思义,运动估计的目的是估计无人机的位姿,无人机位姿包含三个旋转量和三个平移量,这六个变量组成变换矩阵。在双目视觉里程计中,由于可以直接获取深度值,所以常常采用 2D-3D, 3D-3D 两种方法计算变换矩阵,其中,2D 表示某空间点图像中像素点的坐标,3D 表示该空间点在三维坐标。首先,本章研究了 2D-3D, 3D-3D 两种计算变换矩阵的方法;其次,在 ZED 立体相机获得的图像序列的前提下,利用上章所述的特征提取匹配算法获得图像匹配序列中的 2D 和 3D 信息,以通过上述两种运动估计方法获得一系列位姿变换矩阵。在获得了位姿变换矩阵的基础上获得融合后的三维点云地图。

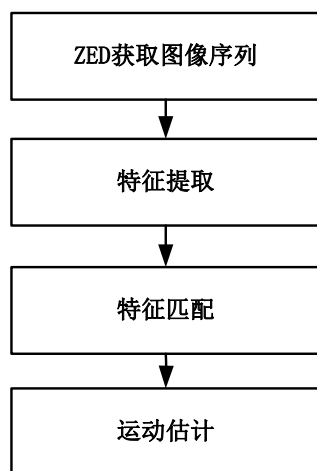


图 3.1 视觉里程计流程图

3.2 双目视觉的运动估计

在双目视觉里程计中,运动估计是核心的部分。在双目视觉里程计中,假设 $k-1$ 时刻, k 时刻获得的图像分别为 I_{k-1} 、 I_k , 相机两时刻间位姿的变换矩阵为 T_k 。在双目视觉里程计中,通过对图像 I_{k-1} 、 I_k 进行特征提取与匹配,实现两者间的数据关联,获得了匹配好的像素点对。在理论上可认为每一对匹配好的像素点对应 3D 空间中的某一共同点。 I_{k-1} 中获得的匹配好的像素坐标集合,可转化为 $k-1$ 时刻相机坐标下获得的 3D 坐标集合。 I_k 中获得的匹配好的像素坐标集合代表 2D 信息。用前一时刻空间点的 3D 位置信息和该空间点在下一时刻的 2D 像素坐标,求解两邻时刻相机位姿间的变换矩阵的方法,称为 3D-2D 运动估计方法。利用 I_{k-1} 、 I_k 中特征提取与匹配后获得的 2D 像素坐标信息,得到在各自时刻相机坐标系下的 3D 位置信息,利用 I_{k-1} 、 I_k 各自获得的 3D 位置信息求解两邻时刻相机位姿间的变换矩阵的方法,称为 3D-3D 运动估计方法。

3.2.1 3D-2D 运动估计方法

由 3D-2D 进行相机运动估计最少需要三组点对,此时的一组点对为一个 3D 点和一个 2D 点。3D-2D 方法中的 3D 信息由双目相机直接获得,3D-2D 方法简洁、有效,是一种常用的运动估计方法。由 3D 空间坐标和 2D 图像像素坐标求得相机位姿变换矩阵 T_k , 这类优化问题被称为 n 点透视问题(Perspective from n Points, PnP)^[38-39], 可以利用直接线性变换^[11]、Bundle Adjustment^[40-41]方法求解。

(1) 直接线性变换

设空间点 P 齐次坐标为 $P = (X, Y, Z, 1)^T$, 在图像 I_1 中的投影点的齐次坐标为 $x_1 = (u_1, v_1, 1)^T$, 三维齐次坐标为 $P_1 = (X_1, Y_1, Z_1, 1)^T$ 。在图像 I_2 中的投影点的齐次坐标为 $x_2 = (u_2, v_2, 1)^T$, 三维齐次坐标为 $P_2 = (X_2, Y_2, Z_2, 1)^T$ 。 I_1 经过 R, t 变换为 I_2 。相机的内参数矩阵 K 已知。设图像 I_1 对应的相机坐标系为世界坐标系,即参考坐标系。根据相机成像原理有:

$$s_2 \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = K [R | t] \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix} \quad (3.1)$$

上式 (3.1) 中:

$$K [R | t] = T_1 = \begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_5 & t_6 & t_7 & t_8 \\ t_9 & t_{10} & t_{11} & t_{12} \end{pmatrix} \quad (3.2)$$

基于式 (3.1)、式 (3.2) 得:

$$\begin{cases} u_2 = \frac{t_1 X_1 + t_2 Y_1 + t_3 Z_1 + t_4}{t_9 X_1 + t_{10} Y_1 + t_{11} Z_1 + t_{12}} \\ v_2 = \frac{t_5 X_1 + t_6 Y_1 + t_7 Z_1 + t_8}{t_9 X_1 + t_{10} Y_1 + t_{11} Z_1 + t_{12}} \end{cases} \quad (3.3)$$

定义矩阵 $\mathbf{T}_1$ 的行向量:

$$\begin{cases} \mathbf{t}_1 = (t_1, t_2, t_3, t_4) \\ \mathbf{t}_2 = (t_5, t_6, t_7, t_8) \\ \mathbf{t}_3 = (t_9, t_{10}, t_{11}, t_{12}) \end{cases} \quad (3.4)$$

整理式 (3.3)、(3.4) 得:

$$\begin{aligned} \mathbf{t}_1^T \mathbf{P} - \mathbf{t}_3^T \mathbf{P} u_1 &= 0 \\ \mathbf{t}_2^T \mathbf{P} - \mathbf{t}_3^T \mathbf{P} v_1 &= 0 \end{aligned} \quad (3.5)$$

由于矩阵 $\mathbf{T}_1$ 具有十二个变量, 所以当存在六对匹配点时可以通过求解线性方程组求解 $\mathbf{T}_1$, 当存在超过六对匹配点时, 对超定方程利用最小二乘法求解 $\mathbf{T}_1$ 。在获得 $\mathbf{T}_1$ 的基础上, 由 $[\mathbf{R} | \mathbf{t}] = \mathbf{K}^{-1} \mathbf{T}_1$ 求解 $\mathbf{R}, \mathbf{t}$ 。因为旋转矩阵 $\mathbf{R}$ 是单位正交矩阵, 而上述方法求解结果未必满足, 因此利用 QR 分解将解投影到旋转矩阵的解空间。

(2) Bundle Adjustment

Bundle Adjustment 描述最小化空间点的重投影误差 (**Reprojection Error**)。重投影误差为空间点 $\mathbf{P}$ 在 $\mathbf{I}_2$ 与 $\mathbf{I}_1$ 像素点匹配的像素点的位置与 $\mathbf{I}_1$ 时刻的空间点 $\mathbf{P}$ 坐标、估计的变换矩阵 $\mathbf{T}$ 投影获得的估计像素点位置的误差。如下图 3.2 所示:

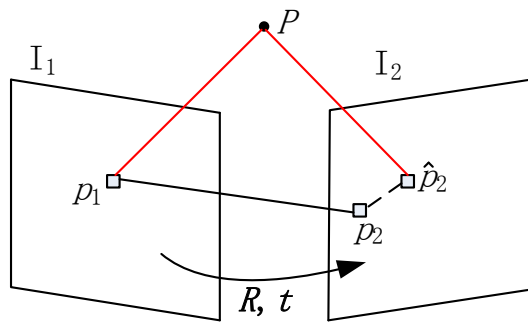


图 3.2 重投影误差图

空间点 $\mathbf{P} = (X, Y, Z, 1)$ 在图像 $\mathbf{I}_1$ 的像素点为 $p_1 = (u_1, v_1, 1)$, p_1 通过特征提取与匹配获得在空间点 $\mathbf{P}$ 在图像 $\mathbf{I}_2$ 的像素点 $p_2 = (u_2, v_2, 1)$, 图像 $\mathbf{I}_1$ 经过旋转变换 $\mathbf{R}$ 和平移变换 $\mathbf{t}$ 到图像 $\mathbf{I}_2$ 。 $\mathbf{P}$ 通过 $\mathbf{R}, \mathbf{t}$ 获得的重投影像素点为 $\hat{p}_2$, 所以根据透视投影关系有:

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R} | \mathbf{t}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.6)$$

由于噪声的影响，式（3.6）等号不成立，所以运用误差 e 构建最小二乘问题，在像素点坐标 $p_i = (u_i, v_i, 1)$ 和参考坐标系下空间点 $P_i = (X_i, Y_i, Z_i, 1)$ 已知情况下，通过调整 $\mathbf{R}$ ， $\mathbf{t}$ 使得误差和最小，优化方程表示如下：

$$(\mathbf{R} | \mathbf{t}) = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} \mathbf{K} [\mathbf{R} | \mathbf{t}] P_i \right\|_2^2 \quad (3.7)$$

为了构建无约束的优化问题，运用李代数^[42]，再使用迭代方法求解。设变换矩阵的李代数为 ξ ，所以变换矩阵可通过李代数的指数运算求得，即：

$$\mathbf{T} = \exp(\xi^\wedge) \quad (3.8)$$

所以，式（3.7）可以转化为：

$$\xi = \arg \min_{\xi} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} \mathbf{K} \exp(\xi^\wedge) P_i \right\|_2^2 \quad (3.9)$$

将误差项进行线性化^[11]：

$$e(x + \Delta x) \approx e(x) + \mathbf{J} \Delta x \quad (3.10)$$

式（3.10）中，

$$\mathbf{J} = - \begin{bmatrix} \frac{f_x}{Z_c} & 0 & -\frac{f_x X_c}{Z_c^2} & -\frac{f_x X_c Y_c}{Z_c^2} & f_x + \frac{f_x X_c^2}{Z_c^2} & -\frac{f_x Y_c}{Z_c} \\ 0 & \frac{f_y}{Z_c} & -\frac{f_y X_c}{Z_c^2} & -f_y - \frac{f_y Y_c^2}{Z_c^2} & \frac{f_y X_c Y_c}{Z_c^2} & \frac{f_y X_c}{Z_c} \end{bmatrix} \quad (3.11)$$

求得雅克比矩阵 $\mathbf{J}$ 后可为迭代优化提供梯度方向，进而求解相机的运动。由于特征提取与匹配过程中存在着错误匹配，采用随机采样一致性(Random sample consensus, RANSAC)算法^[43-44]剔除误匹配。

因此，双目视觉利用 3D-2D 实现视觉里程计步骤如下：

- 1) $k = t$
- 2) 获得两个时刻 $k-1$ ， k 的两个立体图像序列 $I_{l,k-1}$ ， $I_{r,k-1}$ 和 $I_{l,k}$ ， $I_{r,k}$
- 3) 对于 $k-1$ 时刻的图像对 $I_{l,k-1}$ ， $I_{r,k-1}$ 进行立体匹配
- 4) 利用三角测量计算 $k-1$ 时刻 $I_{l,k-1}$ 的深度值，求得 3D 空间点坐标
- 5) 对 k 时刻图像 $I_{l,k}$ 和 $k-1$ 时刻图像 $I_{l,k-1}$ 进行特征提取与匹配，通过 RANSAC 剔除误匹

配

6) 利用 5) 获得的 $k-1$ 时刻图像 $I_{l,k-1}$ 的像素点, 通过 4) 中获得的 3D 空间点坐标查询其 3D 空间点坐标, 再利用 5) 中 k 时刻图像 $I_{l,k}$ 的 2D 像素点坐标, 利用上述方法计算两邻两帧的相机位姿变换矩阵 T_k

7) $t = t+1$ 重复步骤 1)

3.2.2 3D-3D 运动估计方法

双目视觉里程计中, 基于空间三维坐标求相机变换矩阵 T_k 可以转化为求两个三维点集的配准问题, 在获得良好的特征匹配的前提下, 利用迭代最近点算法 (Iterative Closest Point, ICP)^[45-47] 获得相邻两帧之间的变换矩阵。ICP 算法的过程如下: 假设两个时刻 $k-1$, k 对应的相邻两帧分别为 F_1 和 F_2 , 由上述提取的两邻两帧的特征点集合分别为 P , Q 表示为:

$$\begin{cases} P = \{p_1, p_2, \dots, p_n\} \\ Q = \{q_1, q_2, \dots, q_n\} \end{cases} \quad (3.12)$$

假设两帧之间的旋转矩阵 R 和平移向量 t 组成。因此存在下式:

$$p_i = Rq_i + t \quad (3.13)$$

由于误差的存在, 使得左端不等于右端。通过优化下式求得 R , t 。

$$\min_{R, t} \sum_{i=1}^N \|p_i - (Rq_i + t)\|_2 \quad (3.14)$$

将求得 R , t 用于匹配 P , Q 中的特征点, 获得新的对应点关系; 再优化误差项求得新的 R , t 。依次循环直到到达设定的循环次数或者 R , t 变化很小。

以上过程便是经典的 ICP 算法。利用 ICP 算法可以获得相邻两帧之间对应的相机位姿的变换矩阵。综上可得, 利用 3D-3D 的 ICP 求取相机位姿变换矩阵 T_k 可以表示如下:

1) $k=t$

2) 获得两个时刻 $k-1$, k 的两个立体图像序列 $I_{l,k-1}$ 、 $I_{r,k-1}$ 和 $I_{l,k}$ 、 $I_{r,k}$

3) 对 $k-1$, k 两时刻获得的立体图像序列分别进行立体匹配, 再用三角测量原理获得各自时刻的 3D 信息

4) 从两时刻 $k-1$, k 获得的 3D 信息构成集合 P , Q , 利用上述 ICP 算法计算得到相机变换矩阵 T_k

5) $t = t+1$ 重复步骤 1)

3.3 局部优化

在 Bundle Adjustment 中只优化处理相邻两帧之间的变换矩阵, 当优化变量包含空间三维点

坐标, 且不仅仅局限于处理两邻两帧之间的变换矩阵时, 而是处理包含连续 n 帧的窗口时, 称为窗口 Bundle Adjustment(Windowed Bundle Adjustment)。窗口 Bundle Adjustment 的优化模型为:

$$\arg \min_{X^i, C_k} \sum_{i,k} \|p_k^i - g(X^i, C_k)\|^2 \quad (3.15)$$

式 (3.15) 中, p_k^i 表示 3D 路标点 X^i 在第 k 张图像的第 i 个图像点。 $g(X^i, C_k)$ 表示基于当前相机位姿 C_k 的重投影误差。如前所述, 重投影误差是一个非线性函数, 求解方法一般采用列温伯格-马夸而特 (Levenberg-Marquardt, LM) ^[48], 此法需要一个接近最小值的初始解, 相邻两帧间的视觉里程计的解用来做初值。基于窗口 Bundle Adjustment 比两帧间的 Bundle Adjustment 使用了更多的重投影误差项, 所以窗口 Bundle-Adjustment 减少了漂移。图像特征不仅仅与前一帧的图像关联, 而前 $n-1$ 帧图像关联, 其中 n 为窗口大小。窗口大小的选择取决于计算大小。Bundle-Adjustment 计算复杂为 $O((qM + lN)^3)$, 其中 M 为点的数量、 N 为相机位姿的数量, q 为点的参数数量, l 为相机位姿的参数数量。为了实现实时的 Bundle-Adjustment, 需要选定的窗口大小, 参数数量。当 3D 路标点通过立体相机精确的测量获得时, 只优化相机的位姿可以减少计算的复杂度。

3.4 初始位姿与地图融合

通过运动估计及局部优化, 获得了相邻两帧之间相机的位姿变换矩阵 T 。将第一帧图像的相机的位姿设为参考位姿 C_0 , 有:

$$C_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.16)$$

假设双目相机共获得了 N 帧图像, 则有第 0 帧到都 $N-1$ 帧。通过上述运动估计方法求得了第 $i-1$ 帧与第 i 帧的相机位姿变换矩阵为 T_i , 则第 i 帧相机的位姿相对于参考位姿的初始位姿为 C_i , 且 $C_i = T_1 \times T_2 \times \cdots \times T_i$ 。基于此, 可以获得每一帧图像在该时刻相机的位姿, 所有帧的位姿便是相机的位姿, 也是无人机的位姿。这些位姿构成了无人机的位姿曲线。这些由视觉里程计获得的无人机的位姿称为无人机的初始位姿。

在获得了相机在每帧图像对应的位姿时, 将非参考位姿的其余位姿对应的图像中特征点的 3D 点云地图通过各自的位姿矩阵投影到参考位姿的 3D 点云地图中, 这便是点云地图融合。综上所述, 本章的双目视觉 SLAM 方法用流程图表示, 如图 3.3 所示。

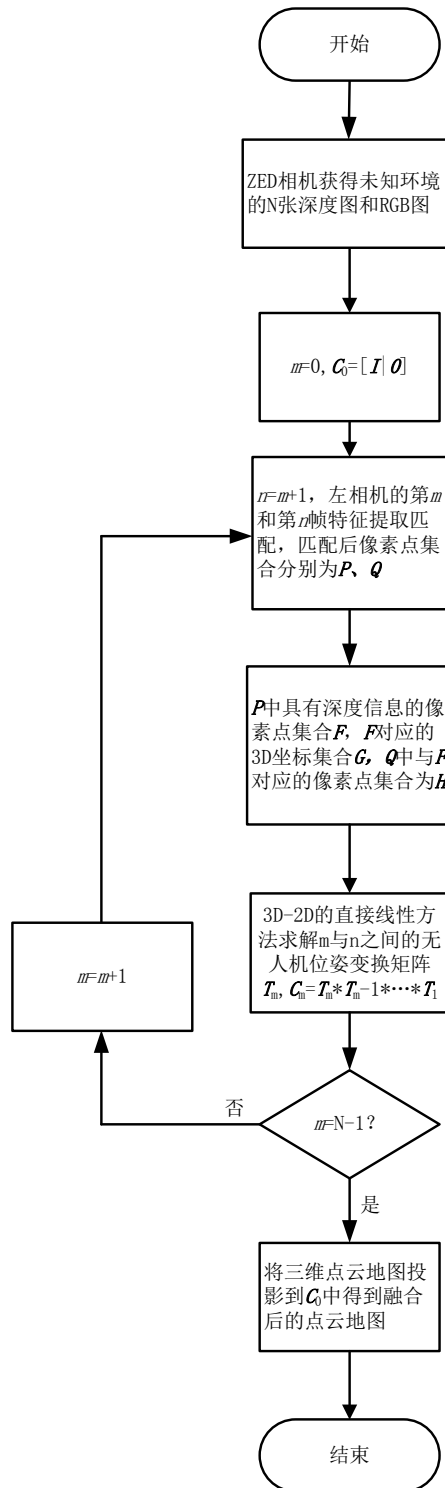
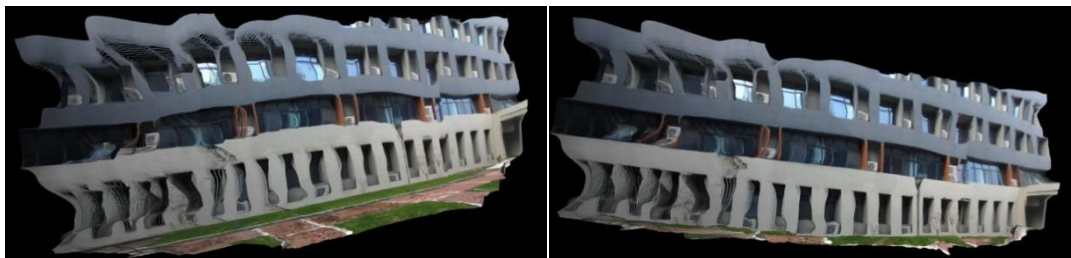


图 3.3 双目视觉里程计流程图

3.5 VO 实验结果及分析

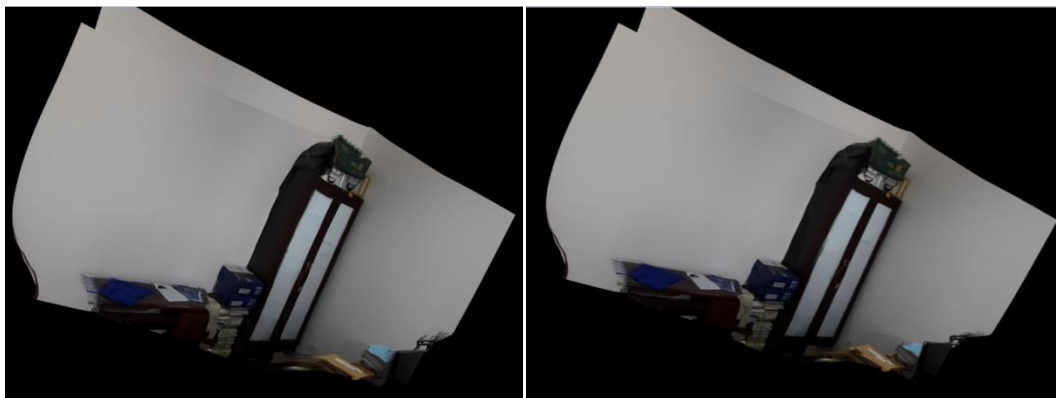
基于本文用的双目 ZED 相机, 用 PnP 方法求解相机运动, 调用 Opencv 中的直接线性变换函数求解相机运动, 获得了第二章中在 t 时刻和 $t+1$ 时刻相机的位姿变换矩阵, 利用位姿变换

矩阵将 t 时刻获得的三维点云地图投影到 $t+1$ 时刻的点云地图中，融合得到相邻两帧间的局部三维点云地图。本文第二章的室外的两帧图像，利用 SIFT、ORB 构建 PnP 中的 3D 和 2D 信息，求解后获得的融合点云图，如图 3.4 所示。



(a) 基于 SIFT 的相邻两帧融合后的点云图 (b) 基于 ORB 的相邻两帧融合后的点云图
图 3.4 室外相邻两帧融合后的点云图

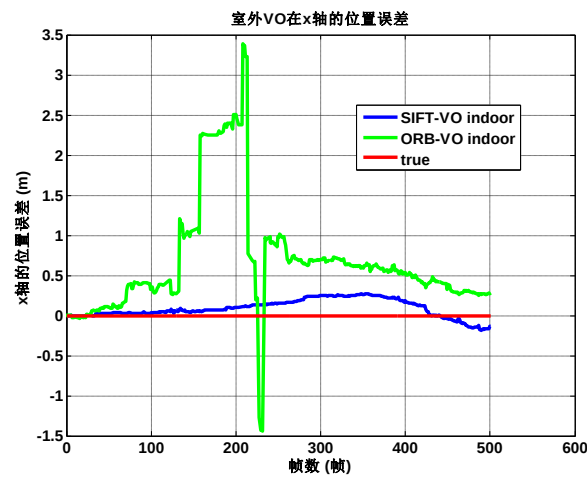
利用 SIFT、ORB 获得的本文第二章的室内的两帧融合后点云图，如图 3.5 所示。



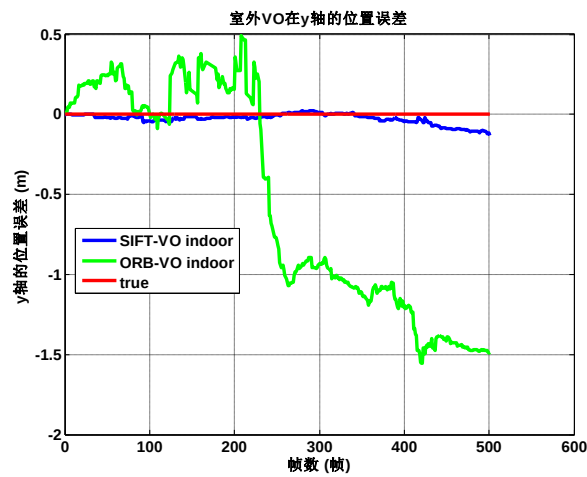
(a) 基于 SIFT 相邻两帧融合后的点云图 (b) 基于 ORB 的相邻两帧融合后的点云图
图 3.5 室内相邻两帧融合后的点云图

从图 2.17、2.21、3.4、3.5 可知，利用上述方法成功融合了该两帧点云图。但从图 3.4、3.5 无法直观地分析两种方法优劣，且两幅融合后的点云图依旧存在噪声，但是室内的点云图效果远远比室外的点云图效果好。点云图未能完全如实际情况那样，与 ZED 采集图像时的位姿、环境中的镜面反射有关。结合第二章所述内容，SIFT 相比于 ORB 更精确，但是更缓慢。

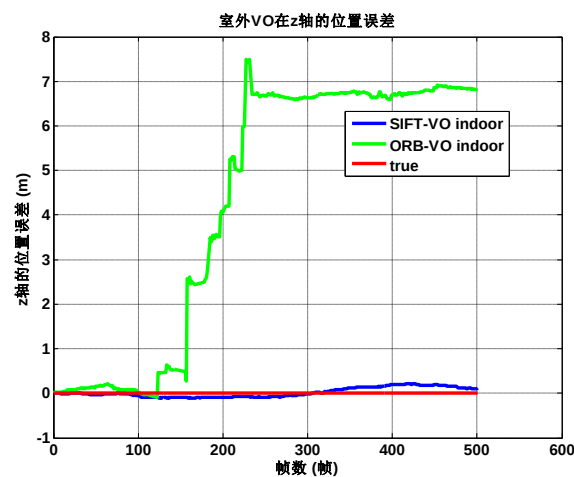
利用 ZED 获取室外的 720HD，30FPS 的视频，获取 502 帧 RGB 图及对应的深度图。为了验证本文提出的双目视觉里程计的精度，在实验过程中，确保相机的位置不变，在该位置水平旋转。在特征提取与匹配的过程中，用 SIFT 及 ORB，对其构建双目视觉里程计，估计的位置结果如图 3.6 所示。



(a) x 轴误差



(b) y 轴误差



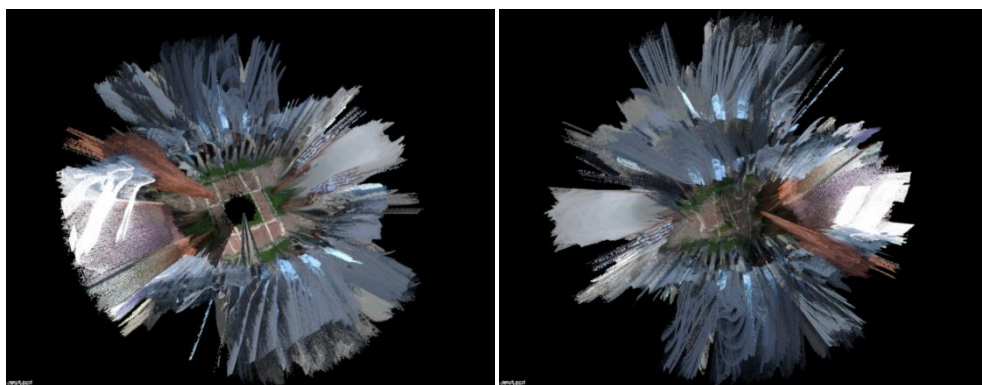
(c) z 轴误差

图 3.6 VO 估计的室外位置误差图

从图 3.6 可知，室外情况下，利用 SIFT 构建 VO 获得的位置估计值远远好于用 ORB 构建

VO 获得的位置估计值。利用 SIFT 构建 VO 获得的位置估计误差值在 0.2m 范围内，而 ORB 构建的 VO 获得的位置估计误差很大。只要是 ORB 对光敏感，室外环境下存在镜面反射、向光场景。

利用获得位姿矩阵，将三维点云图融合，结果如图 3.7 所示。

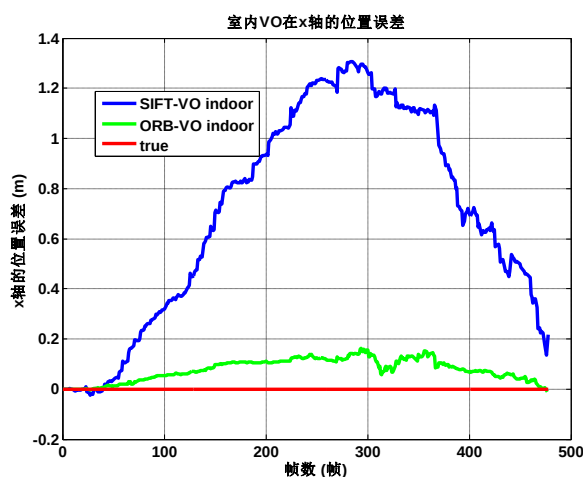


(a) 用 SIFT 构建 VO 获得室外的三维点云图 (b) 用 ORB 构建 VO 获得室外的三维点云图

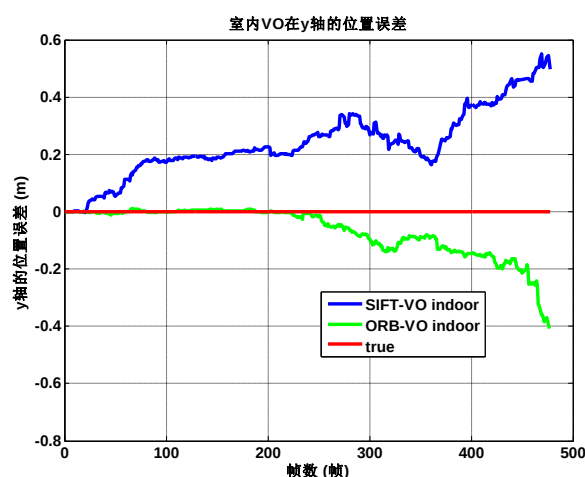
图 3.7 VO 获得的室外三维点云图

从图 3.7 结果进一步说明了室外情况下利用 SIFT 构建 VO 获得的位置估计值远远好于用 ORB 构建 VO 获得的位置估计值。究其原因有以下方面，第一：室外的情况下，由于部分位置向光，室外玻璃反光，导致特征匹配部分效果不佳；第二：而 SIFT 相比于 ORB 鲁棒性更好，所以获得的两帧间的位姿变换矩阵更精确，在 VO 累计误差的情况下，基于 SIFT 的 VO 的位姿结果将明显好于基于 ORB 的 VO；第三，基于 SIFT 的 VO 和基于 ORB 的 VO 都存在误匹配的情况。

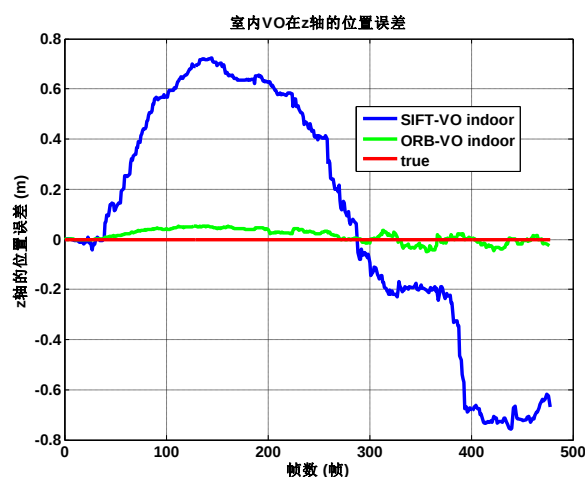
相应地，利用 ZED 获取室内的 720HD, 30FPS 的视频，获取 479 帧 RGB 图及对应的深度图。室内估计的位置结果，如图 3.8 所示。



(a) x 轴误差



(b) y 轴误差

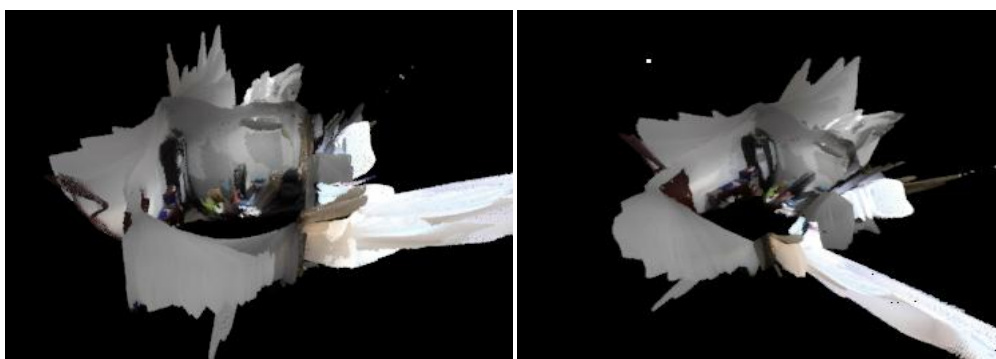


(c) z 轴误差

图 3.8 VO 估计室内位置误差图

从图 3.8 可知：室内情况下，利用 ORB 构建 VO 获得的位置估计值好于 SIFT 构建 VO 获得的位置估计值。SIFT 构建 VO 获得的 x 方向上的位置估计值最大误差约为 1.3m，y 方向上的位置估计值最大误差约为 0.6m，z 方向上的位置估计值最大误差约为 0.7m，因此结果还不算太差。

利用获得位姿矩阵，将室内三维点云图融合，如图 3.9 所示。从图 3.9 可知，在室内环境中，SIFT 构建的 VO 和 ORB 构建的 VO 都能获得较好的三维点云地图，而基于 ORB 获得的三维点云图噪点少于基于 SIFT 获得的三维点云图，且地图更方正，精确。与图 3.7 所示的室外点云图相比，室内的情况远远更好，主要是因为室外存在镜面反射、向光等噪声干扰。



(a) SIFT 构建 VO 获得室内的三维点云图 (b) ORB 构建 VO 获得室内的三维点云图
图 3.9 VO 获得的室内三维点云图

3.6 本章小结

在本章中，主要分析了基于特征的视觉里程计的运动估计部分。依次讲述了运动估计中的 3D-2D、3D-3D 的原理及基于运动估计的视觉里程计实现。实验利用双目视觉对室内、室外构建基于特征的视觉里程计，完成了视觉 SLAM 的前端，为后端优化奠定了基础。对基于 SIFT、ORB 的室内外 VO 的结果进行了分析比较。结果表明，室外情况下基于 SIFT 的 VO 结果远远好于基于 ORB 的 VO 结果，基于 ORB 的 VO 结果不太乐观；室内情况下，基于 ORB 的 VO 结果好于基于 SIFT 的 VO 结果，但是两者的效果都不错。需要注意的是：基于 ORB 的 VO 速度远远小于基于 SIFT 的 VO，在实时性方面具有绝对的优势。从室内室外的 VO 结果可得，在基于特征的视觉里程计中，对于特征的提取与匹配需要更快速、剔除误匹配、更精确地实现特征的匹配是获得良好的视觉里程计的前提，局部优化对于获得正确的位姿、构建视觉里程计也是必不可少。在以上基础上，获得 VO 的结果远远达不到工程需要，所以后端优化是必要的。

第四章 无人机的双目视觉 SLAM 的后端优化

4.1 引言

由上一章构建的视觉里程计还存在着以下的不足。首先，在上一章构建的视觉里程计中，将所有获得的帧都参与了视觉路程计的构建，考虑到某时刻相机的运动缓慢将导致处理大量包含重复信息的图像，这样显然不利于实现在线实时的 VO 或 SLAM，因此本章将引入关键帧以运动大小为衡量删选获得帧，得到的帧便成为关键帧。其次，考虑到光照强度、阴影、相机模型、相机畸变等原因导致获得的位姿矩阵存在误差，由上一章分析得到，由视觉里程计获得的无人机的在某帧的初始位姿的过程是一系列两邻两帧的位姿变换矩阵相乘的过程，因此某帧的相机位姿的误差包含其之前帧的相机位姿误差，即误差具有累加性。这样将导致在大范围的环境下，SLAM 获得的无人机的位姿误差越来越大。视觉里程计没有考虑回环^[49]的情况，以地图的角度严格意义上讲，当再次回到同一位置时称为回环，事实上将无人机经过同一区域，有部分相同的环境信息，便认为是回环。回环将某一帧和之前的多帧联系起来，使得纠正了 VO 的位姿误差。忽略回环检测部分，将导致视觉里程计获得的位姿和地图不是全局一致的，例如：原本是同一位姿，估计获得的位姿离真实位姿相差很大。在获得的地图体现在：环形区域的地图变为不闭合的地图。为处理 VO 中的位姿误差，需要构建视觉 SLAM 的后端。后端的输入包含前端获得的位姿矩阵和回环检测部分，输出为优化后的位姿矩阵；进而获得全局一致的地图。早期的 SLAM 方法在后端中采用滤波方法，主要有扩展卡尔曼滤波^[50-52]、粒子滤波^[53-54]等；现如今用非线性优化的方法处理 SLAM 问题慢慢变成主流。

4.2 SLAM 优化模型

SLAM 问题可以用一个运动方程和一个观测方程^[11]表示如下：

$$\begin{cases} \mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{w}_k \\ \mathbf{z}_{k,j} = h(\mathbf{x}_k, \mathbf{y}_j) + \mathbf{v}_{k,j} \end{cases} \quad (4.1)$$

上式中， $\mathbf{x}_k$ 表示无人机 k 时刻状态变量，在 SLAM 中表示位姿， $\mathbf{u}_k$ 表示 k 时刻运动输入量， $\mathbf{w}_k$ 表示 k 时刻运动方程噪声，函数 f 描述了 $\mathbf{x}_{k-1}$ 与 $\mathbf{x}_k$ 的关系； $\mathbf{y}_j$ 表示第 j 个路标点， $\mathbf{z}_{k,j}$ 表示 k 时刻在第 j 个路标点时的观测量， $\mathbf{v}_{k,j}$ 表示此次观测的噪声。

结合式 (4.1) SLAM 问题表示为：在已知输入 $\mathbf{u}$ 和观测量 $\mathbf{z}$ 的前提下，估计位姿变量 $\mathbf{x}$ 和路标 $\mathbf{y}$ 。求得位姿变量的集合便解决定位问题，相应地，路标集合便构成地图。SLAM 是一个状态估计的问题，在噪声存在的情况下，如何估计状态变量 $\mathbf{x}$ ， $\mathbf{y}$ 。根据观测方程和运动方程

是否为线性，噪声是否符合高斯分布，将此状态估计问题分为线性、非线性和高斯、非高斯系统。线性高斯系统的无偏最优估计可由卡尔曼滤波器求解；非线性非高斯系统用扩展卡尔曼滤波和非线性优化方法求解。以图优化为代表的非线性优化方法已经优于滤波器方法。在视觉 SLAM 中，用变换矩阵 $\mathbf{T}$ 描述两帧间的运动，用针孔相机模型描述观测方程。

在非线形优化中，待估计的位姿变量和路标变量组成状态变量，即： $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)$ 。

在已知 $\mathbf{u}$ 和 $\mathbf{z}$ 的前提下，估计状态 $\mathbf{x}$ 可以表示为求解一个条件概率 $P(\mathbf{x}|\mathbf{z}, \mathbf{u})$ 。在只有视觉传感器，没有运动输入 $\mathbf{u}$ 的情况下，该条件概率为 $P(\mathbf{x}|\mathbf{z})$ 。利用贝叶斯法则得：

$$P(\mathbf{x}|\mathbf{z}) = \frac{P(\mathbf{z}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z})} \quad (4.2)$$

式 (4.2) 中， $P(\mathbf{x}|\mathbf{z})$ 在概率学中称为后验概率， $P(\mathbf{z}|\mathbf{x})$ 称为似然， $P(\mathbf{x})$ 称为先验。由于求解后验概率困难，上述问题转化为在 $\mathbf{z}$ 已知的前提下，求解最优状态估计 $\mathbf{x}$ ，使得后验概率 $P(\mathbf{x}|\mathbf{z})$ 最大(Maximize a Posterior, MAP)^[55]；上述可表达式描述为：

$$\mathbf{x}_{MAP}^* = \arg \max_{\mathbf{x}} P(\mathbf{x}|\mathbf{z}) = \arg \max_{\mathbf{x}} \frac{P(\mathbf{z}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z})} \simeq \arg \max_{\mathbf{x}} P(\mathbf{z}|\mathbf{x})P(\mathbf{x}) \quad (4.3)$$

上式 (4.3) 中，由于 $P(\mathbf{z})$ 与待优化变量 $\mathbf{x}$ 无关，所以分母忽略。由于先验 $P(\mathbf{x})$ 未知，利用 $\mathbf{z}$ 已知的前提，使得似然 $P(\mathbf{z}|\mathbf{x})$ 最大，获得估计值 $\mathbf{x}$ ，认为是该观测 $\mathbf{z}$ 下的状态，所以上述最大后验概率转化为求解最大似然估计 (Maximize Likelihood Estimation, MLE)：

$$\mathbf{x}_{MLE}^* = \arg \max_{\mathbf{x}} P(\mathbf{z}|\mathbf{x}) \quad (4.4)$$

假设观测方程的噪声 $\mathbf{v}_k$ 服从高斯分布，即 $\mathbf{v}_{k,j} \sim N(0, \mathbf{Q}_{k,j})$ ，结合式 (4.1) 的观测方程得，

观测量 $\mathbf{z}_{k,j}$ 的条件概率密度服从均值为 $\mathbf{h}(\mathbf{x}_k, \mathbf{y}_j)$ ，方差为 $\mathbf{Q}_{k,j}$ 的高斯概率密度分布，即：

$$f(\mathbf{z}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{Q}_{k,j})}} \exp\left(-\frac{1}{2}(\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k))^T \mathbf{Q}_{k,j}^{-1} (\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k))\right) \quad (4.5)$$

式 (4.5) 中 N 为 $\mathbf{z}$ 的维数，由概率学知识可得， $f(\mathbf{z}|\mathbf{x})$ 为似然函数 $\mathbf{L}(\mathbf{x})$ ，对其取负对数得：

$$-\ln(\mathbf{L}(\mathbf{x})) = \frac{1}{2} \ln((2\pi)^N \det(\mathbf{Q}_{k,j})) + \frac{1}{2} (\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k))^T \mathbf{Q}_{k,j}^{-1} (\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k)) \quad (4.6)$$

上式 (4.6) 中第一项与 $\mathbf{x}$ 无关，忽略；所以最大似然估计转化为：

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left(\frac{1}{2} (\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k))^T \mathbf{Q}_{k,j}^{-1} (\mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k)) \right) \quad (4.7)$$

上式 (4.7) 为观测方程的误差的平方。定义观测误差如下:

$$\mathbf{e}_{y,j,k} = \mathbf{z}_{k,j} - \mathbf{h}(\mathbf{y}_j, \mathbf{x}_k) \quad (4.8)$$

此误差的平方和对应似然函数, 即:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} F(\mathbf{x}) = \arg \min_{\mathbf{x}} \sum_k \sum_j \mathbf{e}_{y,j,k}^T \mathbf{Q}_{k,j}^{-1} \mathbf{e}_{y,j,k} \quad (4.9)$$

当考虑系统有运动输入时, 运动误差为:

$$\mathbf{e}_{v,k} = \mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_k) \quad (4.10)$$

相应的添加了运动误差的待优化的目标函数为:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} F(\mathbf{x}) = \arg \min_{\mathbf{x}} \sum_k \sum_j \mathbf{e}_{y,j,k}^T \mathbf{Q}_{k,j}^{-1} \mathbf{e}_{y,j,k} + \sum_k \mathbf{e}_{u,k}^T \mathbf{R}_k^{-1} \mathbf{e}_{u,k} \quad (4.11)$$

上式 (4.11) 中 $\mathbf{R}_k$ 为运动方程中的噪声误差 $\mathbf{w}_k$ 的协方差, 上式是一个最小二乘问题。利用迭代方法求解此最小二乘问题, 将目标函数 $F(\mathbf{x})$ 在 $\mathbf{x}$ 附近做一阶泰勒展开, 得:

$$F(\mathbf{x} + \Delta\mathbf{x}) \approx F(\mathbf{x}) + \mathbf{J}(\mathbf{x})\Delta\mathbf{x} \quad (4.12)$$

式 (4.12) 中 $\mathbf{J}(\mathbf{x})$ 为 $F(\mathbf{x})$ 关于变量 $\mathbf{x}$ 的雅克比矩阵。沿着负梯度方向可使得 $F(\mathbf{x})$ 值下降, 所以增量 $\Delta\mathbf{x}^* = -\mathbf{J}^T(\mathbf{x})$, 选择相应的步长, 这便为最速下降法。

牛顿法求解最小二乘问题为将目标函数 $F(\mathbf{x})$ 在 $\mathbf{x}$ 附近做二阶泰勒展开, 如下式 (4.13) 所示:

$$F(\mathbf{x} + \Delta\mathbf{x}) \approx F(\mathbf{x}) + \mathbf{J}(\mathbf{x})\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \hat{\mathbf{H}}\Delta\mathbf{x} \quad (4.13)$$

式 (4.13) 中 $\hat{\mathbf{H}}$ 为 Hessian 矩阵。将上式 (4.13) 右边对 $\Delta\mathbf{x}$ 求导, 并令导数为零, 进一步获得相应的增量方程如下式所示:

$$\hat{\mathbf{H}}\Delta\mathbf{x} = -\mathbf{J}^T \quad (4.14)$$

总结上述两种方法步骤如下:

- 1) 获得初始值 $\mathbf{x}_0$ 。
- 2) 选取增量 $\Delta\mathbf{x}_k$, 使得目标函数 $F(\mathbf{x})$ 下降。
- 3) 如果 $\Delta\mathbf{x}_k \leq \varepsilon$, ε 足够小, 则 $\mathbf{x}_k + \Delta\mathbf{x}_k$ 为最终结果, 结束; 否则, 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}_k$, 返回 2)。

由于 $F(\mathbf{x})$ 描述最小二乘问题, 所以有:

$$F(\mathbf{x}) = (\mathbf{f}(\mathbf{x}))^2 \quad (4.15)$$

将 $\mathbf{f}(\mathbf{x})$ 的一阶泰勒展开代入上式 (4.15), 得:

$$\Delta\mathbf{x}^* = \arg \min_{\Delta\mathbf{x}} \frac{1}{2} \|\mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\Delta\mathbf{x}\|^2 \quad (4.16)$$

将 $\frac{1}{2} \|f(\mathbf{x}) + J(\mathbf{x})\Delta\mathbf{x}\|^2$ 展开并求 $\Delta\mathbf{x}$ 得导数, 令其为零, 得:

$$J(\mathbf{x})^T J(\mathbf{x})\Delta\mathbf{x} = -J(\mathbf{x})^T f(\mathbf{x}) \quad (4.17)$$

上式 (4.17) 称为高斯牛顿方程^[56-57], 相比于上述牛顿迭代法, 高斯牛顿方程不用求解二阶偏导数组成的海瑟矩阵。在求解式 (4.17) 这一线性方程时, 可能出现 $J(\mathbf{x})^T J(\mathbf{x})$ 为奇异矩阵的情况, 此时算法不收敛; 同时, 为了使得局部近似更准确, 步长 $\Delta\mathbf{x}$ 不能太大。

g^2o (general graph optimization)^[58]是用图表示非线性最小二乘问题转化的开源 C++ 库。图由顶点和连接顶点的边组成, 在视觉 SLAM 中, 顶点可以表示无人机的位姿或地标, 边是描述顶点间的关系式, 也可以是位姿间的变换矩阵, 描述位姿与地标间关系的观测方程。如图 4.1 所示, 假设机器人在位姿 C_k 处对空间点 x_k 观测得到 z_k 。

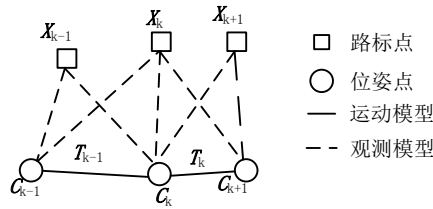


图 4.1 运动观测模型图

由于噪声的存在, 理论值与观测值存在误差, 则这二元边描述为以下最小二乘问题:

$$F_{x_k, C_k, z_k}(x_k, C_k, z_k) = (z_k - K(Rx_k + t))^T Q_k^{-1} (z_k - K(Rx_k + t)) \quad (4.18)$$

无人机在位姿 C_k 到 C_{k+1} 的位姿变换矩阵为 T_k , 由于噪声的存在, 理论值与估计值存在误差, 所以位姿间二元边的描述为以下最小二乘问题:

$$F_k(C_k, T_k, C_{k+1}) = (C_{k+1} - T_k C_k)^T R_k^{-1} (C_{k+1} - T_k C_k) \quad (4.19)$$

因此, 结合式 (4.18) 和式 (4.19)

$$F(\mathbf{x}) = F_{x_k, C_k, z_k}(x_k, C_k, z_k) + F_k(C_k, T_k, C_{k+1}) \quad (4.20)$$

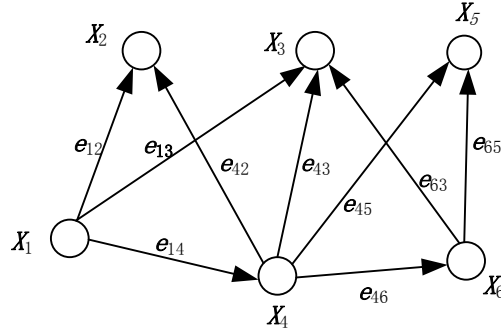
上式 (4.20) 可理解为, 在观测量 z , T 已知的前提下, 调整位姿 C 使得误差最小。将图 4.1 所示的运动观测图转化为 g^2o 中顶点和边的表示, 如图 4.2 所示:

式 (4.20) 描述的最小二乘问题, 用一般的最小二乘形式表示如下:

$$F(\mathbf{x}) = \sum_{(i,j) \in N} e(x_i, x_j, z_{ij})^T \Omega_{ij} e(x_i, x_j, z_{ij}) = \sum_{(i,j) \in N} F_{ij} \quad (4.21)$$

上式中 N 代表自然数集, 此最小二乘的优化变量为:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} F(\mathbf{x}) \quad (4.22)$$


 图 4.2 g^2o 优化图

为了简化推导过程中的表达式，用 $\mathbf{e}_{ij}(\mathbf{x})$ 代替 $\mathbf{e}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij})$ 。基于视觉里程计获得的位姿变换矩阵为位姿间提供初值，而相机测量得到的观测数据为位姿与路标点提供初值，两者构成的初始状态向量为 $\hat{\mathbf{x}}$ 。

采用高斯牛顿法求解时，将误差用一阶泰勒展开近似有：

$$\mathbf{e}_{ij}(\hat{\mathbf{x}} + \Delta\mathbf{x}) \approx \mathbf{e}_{ij}(\hat{\mathbf{x}}) + \frac{d\mathbf{e}_{ij}}{d\mathbf{x}_{ij}} \Delta\mathbf{x} = \mathbf{e}_{ij}(\hat{\mathbf{x}}) + \mathbf{J}_{ij} \Delta\mathbf{x} \quad (4.23)$$

将式 (4.23) 代入式 (4.22) 中，且用 $\mathbf{e}_{ij}$ 代替 $\mathbf{e}_{ij}(\hat{\mathbf{x}})$ 得：

$$\begin{aligned} F_{ij}(\hat{\mathbf{x}} + \Delta\mathbf{x}) &= \mathbf{e}_{ij}(\hat{\mathbf{x}} + \Delta\mathbf{x})^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}(\hat{\mathbf{x}} + \Delta\mathbf{x}) \\ &\approx (\mathbf{e}_{ij} + \mathbf{J}_{ij} \Delta\mathbf{x})^T \boldsymbol{\Omega}_{ij} (\mathbf{e}_{ij} + \mathbf{J}_{ij} \Delta\mathbf{x}) \\ &= \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij} + 2\mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \Delta\mathbf{x} + \Delta\mathbf{x}^T \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \Delta\mathbf{x} \\ &= \mathbf{c}_{ij} + 2\mathbf{b}_{ij} \Delta\mathbf{x} + \Delta\mathbf{x}^T \hat{\mathbf{H}}_{ij} \Delta\mathbf{x} \end{aligned} \quad (4.24)$$

根据式 (4.21) (4.24) 得：

$$\begin{aligned} F(\hat{\mathbf{x}} + \Delta\mathbf{x}) &= \sum_{(i,j) \in N} F_{ij}(\hat{\mathbf{x}} + \Delta\mathbf{x}) \\ &\approx \sum_{(i,j) \in N} \mathbf{c}_{ij} + 2\mathbf{b}_{ij} \Delta\mathbf{x} + \Delta\mathbf{x}^T \hat{\mathbf{H}}_{ij} \Delta\mathbf{x} \\ &= \mathbf{c} + 2\mathbf{b}^T \Delta\mathbf{x} + \Delta\mathbf{x}^T \hat{\mathbf{H}} \Delta\mathbf{x} \end{aligned} \quad (4.25)$$

将式 (4.25) 求导，令导函数为零，得：

$$\hat{\mathbf{H}} \Delta\mathbf{x}^* = -\mathbf{b} \quad (4.26)$$

求解线性方程 (4.26) 出 $\Delta\mathbf{x}^*$ ，使得 $\mathbf{x}^* = \hat{\mathbf{x}} + \Delta\mathbf{x}^*$ 基本不再发生变化，收敛。这便是高斯牛顿迭代法用于 SLAM 过程中。

同理，利用 LM 迭代方法^[48]获得的求解增量的方程如下：

$$(\hat{\mathbf{H}} + \lambda \mathbf{I}) \Delta \mathbf{x}^* = -\mathbf{b} \quad (4.27)$$

变量 $\mathbf{x}$ 由变换矩阵和路标组成，变换矩阵中旋转矩阵对加法不封闭，对乘法封闭，即经过 $\mathbf{x}^* = \hat{\mathbf{x}} + \Delta \mathbf{x}^*$ 后不再表示变换矩阵。因此，重新定义非线性算子 $\boxed{+}$ ，使得：

$$Dom(\mathbf{x}_i) \times Dom(\Delta \mathbf{x}_i) \rightarrow Dom(\mathbf{x}_i) \quad (4.28)$$

因此上述增量方程为：

$$\mathbf{x}^* = \hat{\mathbf{x}} \boxed{+} \Delta \mathbf{x}^* \quad (4.29)$$

当变量 $\mathbf{x}$ 中的变换矩阵采用李代数的形式表示时， $\mathbf{x}^* = \hat{\mathbf{x}} + \Delta \mathbf{x}^*$ 依旧表示变量集中的元素。所以实际中非线性算子 $\boxed{+}$ 表示用李代数的描述变量 $\mathbf{x}$ ，再进行加法操作。

用此非线性算子 $\boxed{+}$ 表示式 (4.23) 表示的误差函数如下：

$$\mathbf{e}_{ij}(\hat{\mathbf{x}}) = \mathbf{e}_{ij}(\hat{\mathbf{x}} \boxed{+} \Delta \mathbf{x}) \approx \mathbf{e}_{ij} + \mathbf{J}_{ij} \Delta \mathbf{x} \quad (4.30)$$

上式 (4.30) 中的雅克比矩阵 $\mathbf{J}_{ij}$ 为：

$$\mathbf{J}_{ij} = \left. \frac{\partial \mathbf{e}_{ij}(\hat{\mathbf{x}} \boxed{+} \Delta \mathbf{x})}{\partial \Delta \mathbf{x}} \right|_{\Delta \mathbf{x} = 0} \quad (4.31)$$

上式 (4.31) 中的求解运算便是李代数中的扰动模型或微分模型。

因为误差函数 $\mathbf{e}_{ij}$ 只与两顶点 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 有关，所以有：

$$\mathbf{J}_{ij} = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{0} & \underbrace{\mathbf{A}_{ij}}_i & \mathbf{0} & \cdots & \mathbf{0} & \underbrace{\mathbf{B}_{ij}}_j & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \quad (4.32)$$

根据式 (4.24)、(4.32) 得：

$$\begin{aligned} \mathbf{b}_{ij}^T &= \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} = \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} (\mathbf{0} \cdots \mathbf{A}_{ij} \cdots \mathbf{B}_{ij} \cdots \mathbf{0}) \\ &= (\mathbf{0} \cdots \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} \cdots \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} \cdots \mathbf{0}) \end{aligned} \quad (4.33)$$

式 (4.33) 中信息矩阵为协方差矩阵的逆，是对称矩阵。得 $\mathbf{b}_{ij}$ ，如下：

$$\mathbf{b}_{ij} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij} \\ \vdots \\ \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij} \\ \vdots \\ \mathbf{0} \end{pmatrix} \quad (4.34)$$

根据式 (4.24)、(4.32) 得：

$$\begin{aligned}
 \hat{\mathbf{H}}_{ij} &= \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \\
 &= \begin{pmatrix} \vdots \\ \mathbf{A}_{ij}^T \\ \vdots \\ \mathbf{B}_{ij}^T \\ \vdots \end{pmatrix} \boldsymbol{\Omega}_{ij} \begin{pmatrix} \cdots & \mathbf{A}_{ij} & \cdots & \mathbf{B}_{ij} & \cdots \end{pmatrix} \\
 &= \begin{pmatrix} \ddots & & & & \\ & \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} & \cdots & \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} & \\ & \vdots & & \vdots & \\ & \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} & \cdots & \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} & \\ & & & & \ddots \end{pmatrix} \quad (4.35)
 \end{aligned}$$

$\hat{\mathbf{H}} = \sum_{(i,j) \in C} \hat{\mathbf{H}}_{ij}$ ，由上图 4.2 可知，任一顶点 x_i 只能与有限个其余顶点 x_j 相连，结合由式

(4.35) 中 $\hat{\mathbf{H}}_{ij}$ 是个稀疏的矩阵，可得 $\hat{\mathbf{H}}$ 是个稀疏矩阵，因此可以快速求解增量方程 (4.26) 或 (4.27)。

4.3 回环检测

在上述过程中，由于相机采集大量图片，为了减少计算量，不能每张图像都进行处理。当运动大小在一定阈值范围内时，选取该张图，此图称为关键帧。阈值的选择在实际实验中获取，一与相机的运动快慢有关，感性的认识便是：相机运动越快，阈值更大。但是相机运动速度不宜太大，否则图像模糊，无法用于获得特征匹配点。在实际过程中，本文衡量运动的变量为旋转矩阵 $\mathbf{R}$ 和平移矩阵 $\mathbf{t}$ 的范数之和。

位姿图为环形是位姿图优化的重要约束，VO 获得的无人机的初始位姿形成了链状图，时间越靠后的位姿，累积误差越大，VO 获得的无人机的初始位姿无法进行图优化，因为没有误差项的存在。当非相邻的位姿之间有约束时，即该两帧间有共同的数据，使得两帧间足够获得变换矩阵时，意味着该两帧对应的位姿之间有边的约束，此时将产生误差项。通过这种方式，可以减少位姿的累积误差，获得更精确的位姿，这也就是以上讲述的位姿图优化原理。一般的，当无人机经过一段时间观测再次观测到某路标，或者重新回到之前的位置区域处，称为回环^[59]。回环检测的功能便是将回环找出来，体现在位姿图中便是增加顶点间边的关系。通过比较两帧的视觉相似性可以判断回环，利用全局图像描述子^[59-60]或局部图像描述子^[61]计算视觉相似性。视觉词典^[62-65]是一种受到研究人员大量关注的基于局部图像描述子的回环检测方法。本文采用简单回环检测^[11]，将旋转矩阵 $\mathbf{R}$ 和平移矩阵 $\mathbf{t}$ 的范数之和在某阈值范围内，认为检测到回环。回环采用近距离回环和随机回环。顾名思义，近距离回环指对当前帧之前的 m 帧进行回环检测；

随机回环指对之前任意 n 帧进行回环检测。将检测到的回环对应的两帧的边添加到位姿图中，进行优化，如此便能获得更精确的位姿，在基于更精确的位姿获得更精确的地图。

4.4 位姿图优化

在上述待优化目标函数只考虑无人机位姿间的误差时，视觉里程计得到的位姿矩阵作为顶点，位姿变换矩阵作为边，这便是位姿图优化^{[44][66]}。将此问题描述为最小二乘问题，如上式(4.19)所示。位姿图优化可以描述为：在图为非链状时，由视觉里程计获得的位姿矩阵作为初始值，通过高斯牛顿迭代法或 Levenberg-Marquardt 求解位姿，使得求得的位姿使得该误差最小。

综上所述，本文双目视觉 SLAM 方法的流程图如图 4.3 所示。

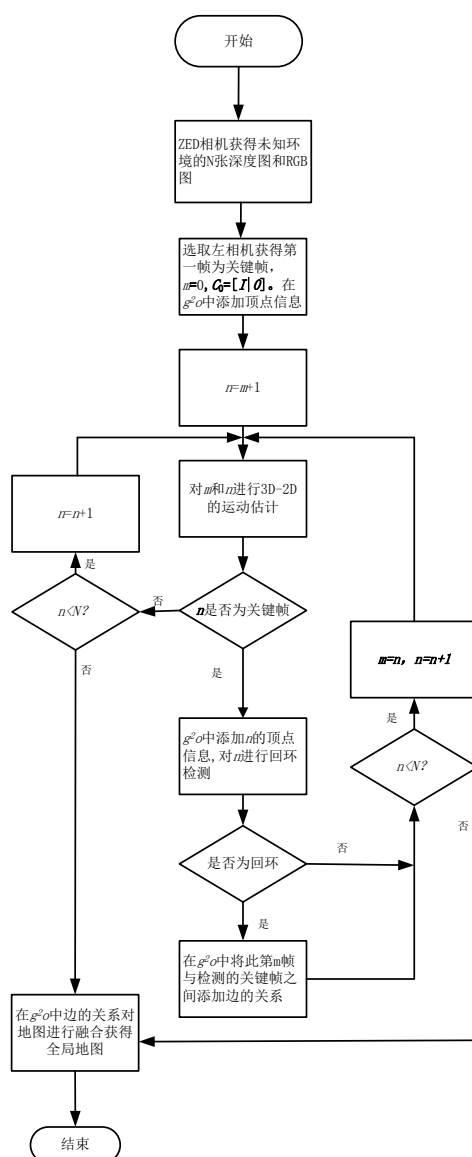
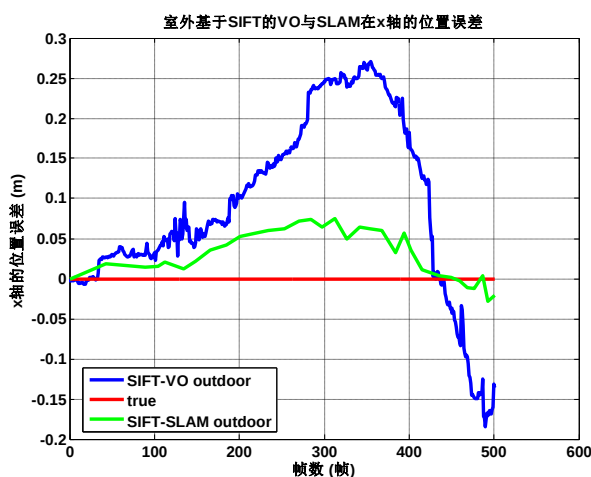


图 4.3 基于无人机的双目视觉 SLAM 流程图

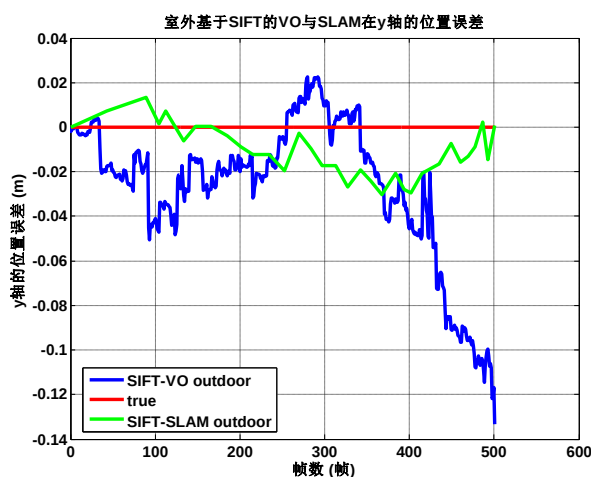
4.5 SLAM 实验结果及分析

为了验证本文提出的双目视觉里程计的精度，在实验过程中，确保相机的位置不变，在该位置水平旋转。在室外的情况下，基于 SIFT 构建的双目视觉 SLAM 方法，利用位姿图优化方法，通过 Levenberg-Marquardt 迭代求解。获得优化后的位置在三轴上的位置误差与本文第三章基于 SIFT 构建的 VO 获得的位置在三轴上的位置误差对比图，如图 4.4 所示。

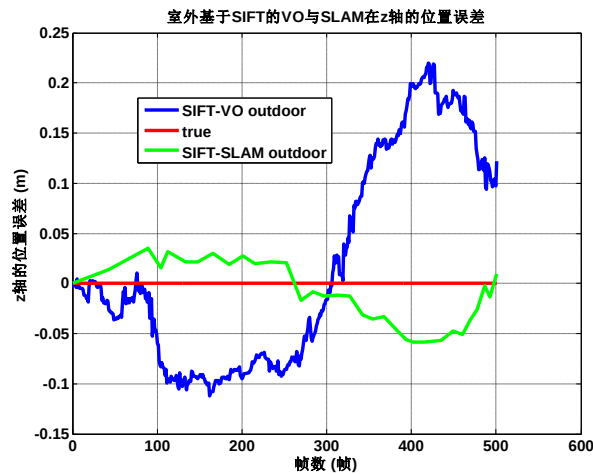
从图 4.4 可知，在室外情况下，添加回环检测、后端优化的基于 SIFT 的 SLAM 方法获得的位置误差相比基于 SIFT 的 VO 小，即获得了更精确的估计位置。



(a) x 轴误差



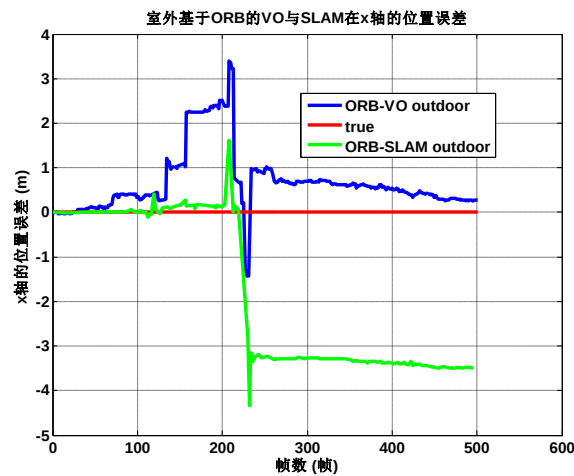
(b) y 轴误差



(c) z 轴误差

图 4.4 室外基于 SIFT 的 VO 与 SLAM 的位置误差对比图

基于 ORB 构建的双目视觉 SLAM 方法获得无人机的位置, 与基于 ORB 构建的 VO 获得的无人机的位置, 构成的位置误差对比图, 如图 4.5 所示。



(a) x 轴误差

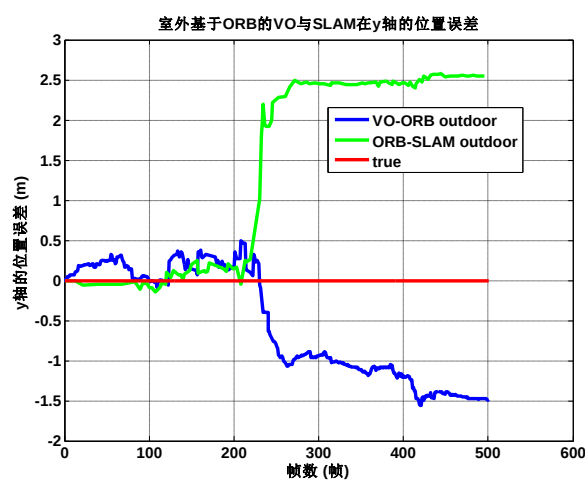
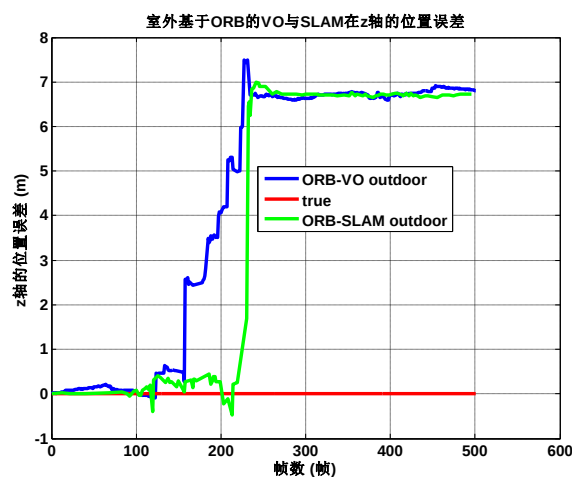
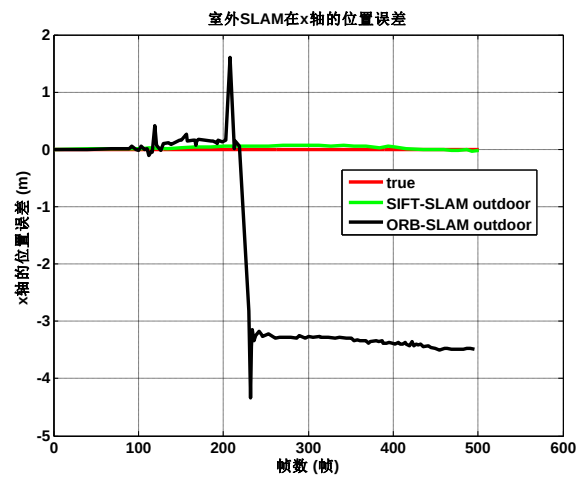
(b) y 轴误差(c) z 轴误差

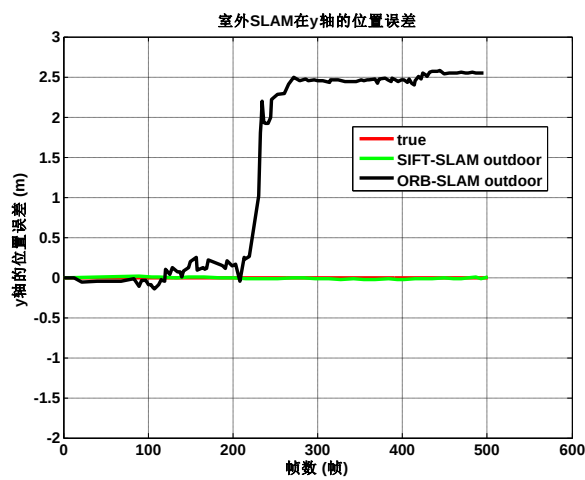
图 4.5 室外基于 ORB 的 VO 与 SLAM 的位置误差对比图

由本文第三章及图 4.5 中可知, 室外基于 ORB 的 SLAM 在前 200 帧获得了更精确的位置估计, 而对于后约 300 帧的位置没有很大的改善。主要是因为室外情况下 ORB 对光敏感, 由 ORB 构建的 VO 获得的初始位姿矩阵误差太大。

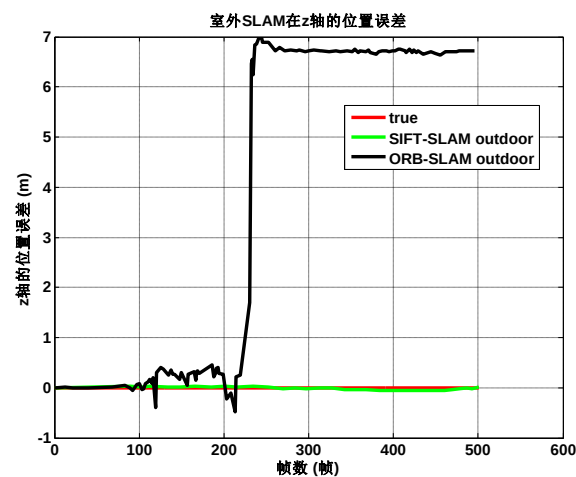
在室外的情况下, 基于 SIFT 和 ORB 的双目视觉 SLAM 方法估计的位置误差对比图, 如图 4.6 所示。



(a) x 轴误差



(b) y 轴的误差



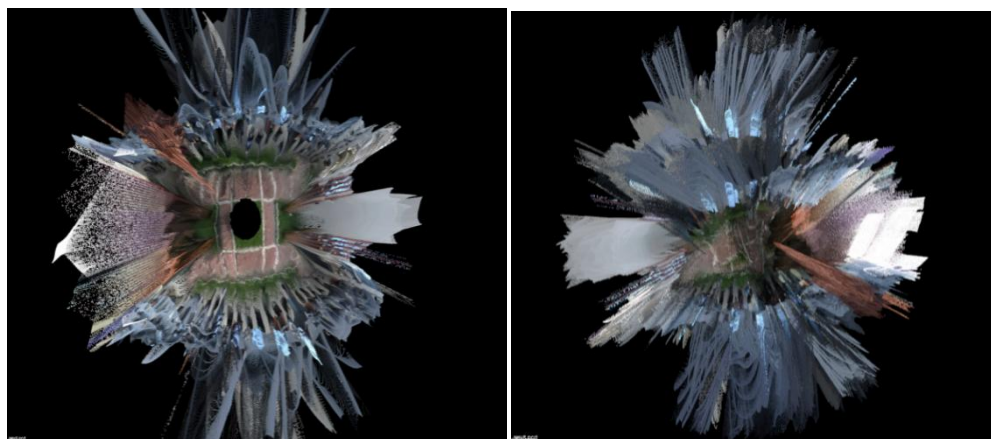
(c) z 轴的误差

图 4.6 室外 SIFT 和 ORB 的双目视觉 SLAM 系统估计的位置误差对比图

从图 4.6 可知：在本实验室外数据下，基于 SIFT 的双目视觉 SLAM 方法优于基于 ORB 的

双目视觉 SLAM 方法。

在上述过程中，ZED 获得的 RGB 图有 502 帧，基于 SIFT 的 SLAM 系统中，选取的关键帧数量为 33 帧。基于 ORB 的 SLAM 系统中，选取的关键帧数量为 111 帧。在获得优化后的位姿矩阵后，将位姿矩阵用于地图融合，获得的全局地图，如图 4.7 所示。

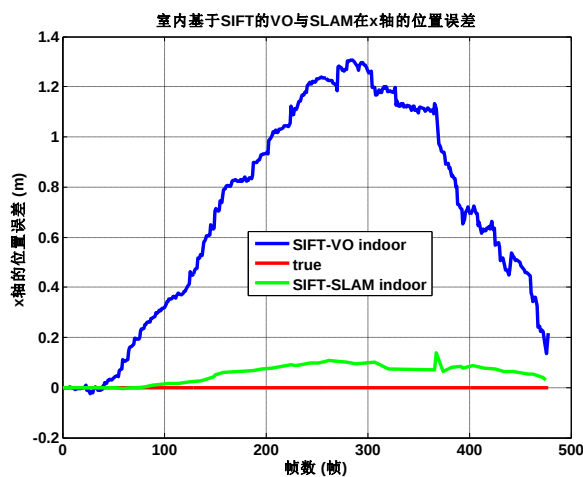


(a)室外基于 SIFT 的 SLAM 获得的地图 (b)室外基于 ORB 的 SLAM 获得的地图

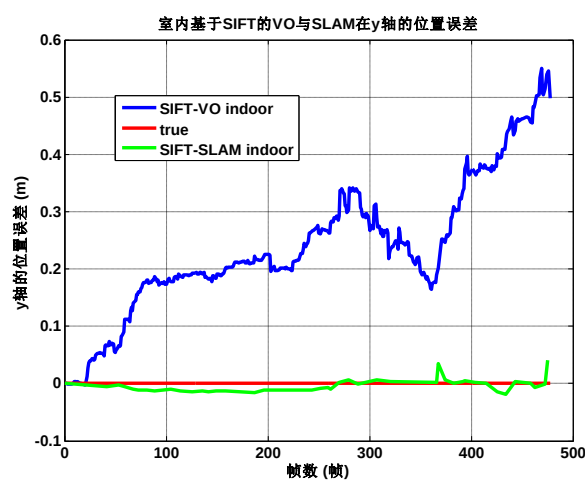
图 4.7 室外双目视觉 SLAM 获得的三维点云地图

从图 4.7 中可得：室外基于 SIFT 的双目视觉 SLAM 获得了较精确的地图，而基于 ORB 的双目视觉 SLAM 获得的地图存在着一定的误差。

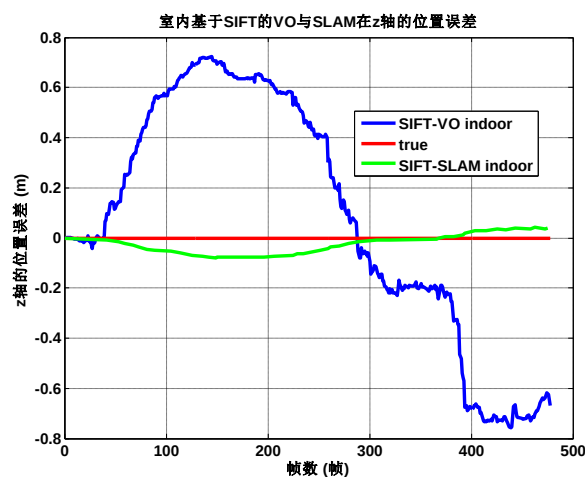
在室内的情况下，基于 SIFT 构建的双目视觉 SLAM 方法，获得的位置误差与相同情况下 VO 获得的位置误差对比图，如图 4.8 所示。



(a) x 轴误差



(b) y 轴误差



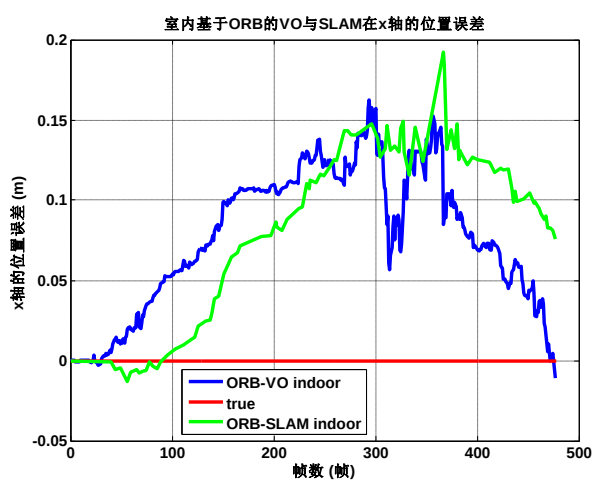
(c) z 轴误差

图 4.8 室内基于 SIFT 的 VO 与 SLAM 的位置误差对比图

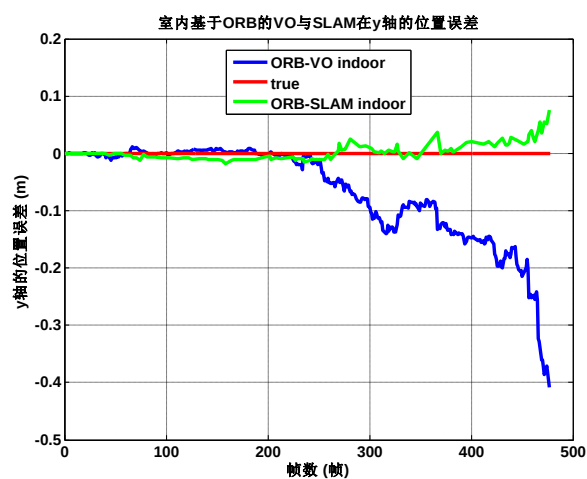
从图 4.8 可知,室内情况下,基于 SIFT 的双目视觉 SLAM 获得的 x 轴、z 轴位置误差在 0.1m 以内,而 y 轴误差在 0.03m 左右;相比于 SIFT 构建的 VO 估计的位置误差小太多。

在室内的情况下,基于 ORB 构建的双目视觉 SLAM 方法,获得的位置误差与相同情况下 VO 获得的位置误差对比图,如图 4.9 所示。

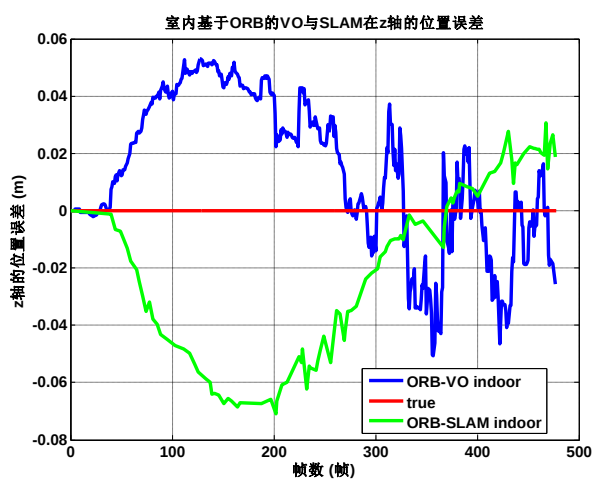
从图 4.9 可知,室内环境下,基于 ORB 的 SLAM 估计的位置在 y 轴方向的误差远远小于基于 ORB 的 VO 估计的位置,而在 x 轴、z 轴方向,两者差距不大。



(a) x 轴误差



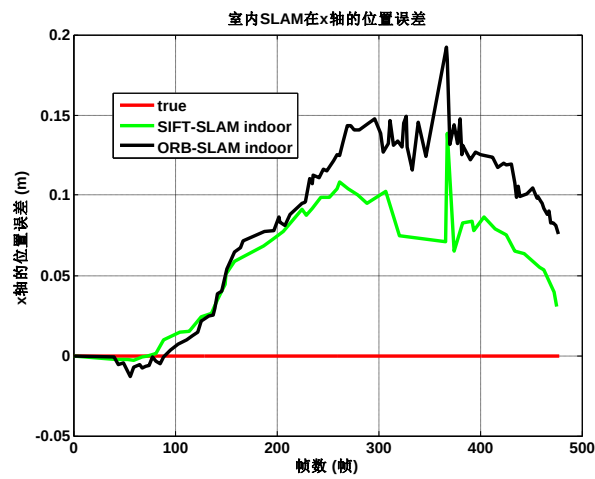
(b) y 轴误差



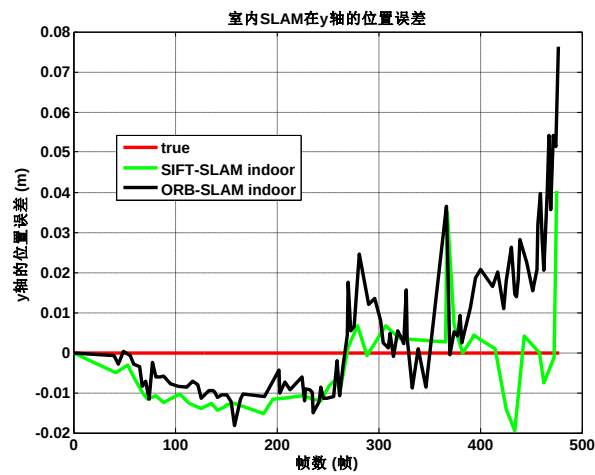
(c) z 轴误差

图 4.9 室内基于 ORB 的 VO 与 SLAM 的位置误差对比图

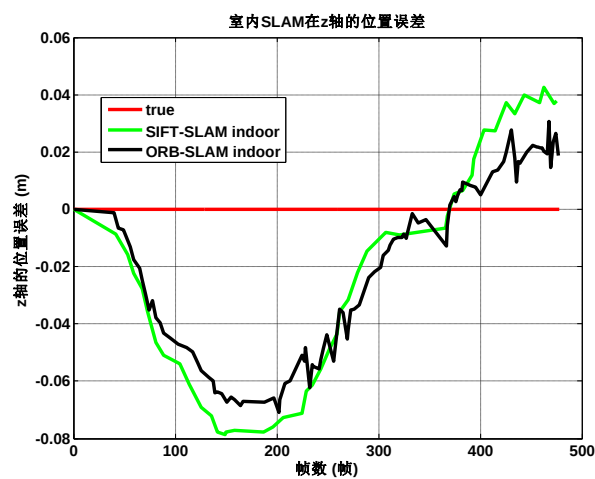
在室内，基于 SIFT 和 ORB 的双目视觉 SLAM 方法估计的位置误差图，如图 4.10 所示。



(a) x 轴误差



(b) y 轴误差

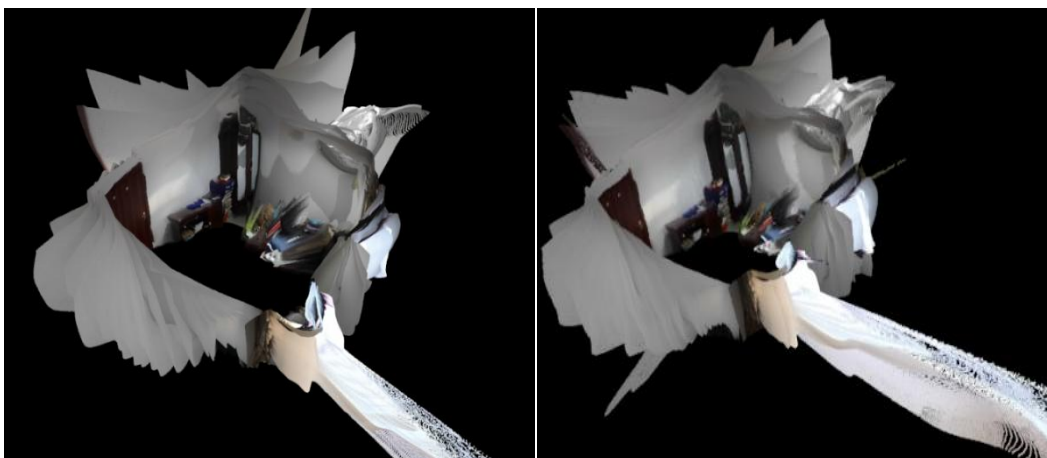


(c) z 轴误差

图 4.10 室内 SIFT 和 ORB 的双目视觉 SLAM 系统估计的位置误差对比图

从图 4.10 可得：在本文室内环境中，基于 SIFT 的双目视觉 SLAM 方法比基于 ORB 的双目视觉 SLAM 方法获得的 x 、 y 轴方向的位置估计更精确，而在 z 轴的位置估计较差一点，所以在精确性方面，基于 SIFT 的双目视觉 SLAM 方法占优；但是由于基于 SIFT 的双目视觉 SLAM 方法在特征提取阶段耗费大量时间，故基于 ORB 的双目视觉 SLAM 系统在快速性方面有足够的优势。

在上述双目视觉 SLAM 系统中，ZED 获得的 RGB 图有 478 帧，基于 SIFT 的 SLAM 系统中，选取的关键帧数量为 46 帧。基于 ORB 的 SLAM 系统中，选取的关键帧数量为 104 帧。



(a) 室内基于 SIFT 的 SLAM 获得的地图 (b) 室内基于 ORB 的 SLAM 获得的地图

图 4.11 室内双目视觉 SLAM 获得的三维点云地图

如图 4.11 所示，室内情况下，基于 SIFT 和 ORB 的双目视觉 SLAM 可获得较精确的三维点云地图，相比于基于 SIFT 和 ORB 的 VO 获得的三维点云地图，更精确、帧数更少、耗费时间更少。

4.6 本章小结

本章研究了视觉 SLAM 的数学优化模型，对双目视觉 SLAM 优化模型进行了参数化。将 SLAM 优化问题用 g^2o 描述求解，进行回环检测，用主流的图优化方法获得更精确的相机位姿，同时也获得更精确的融合后的三维点云图。在理论的基础上，构建了基于 SIFT 和 ORB 的双目视觉 SLAM 系统，在室内、室外两种环境下，构建了全局连续性三维点云地图，获得无人机的位姿，对无人机的位置误差进行了分析。比较、分析了基于 SIFT 和 ORB 的双目视觉 SLAM 系统，比较、分析了室内外双目 SLAM 系统。

第五章 多无人机双目视觉 SLAM 方法研究

5.1 引言

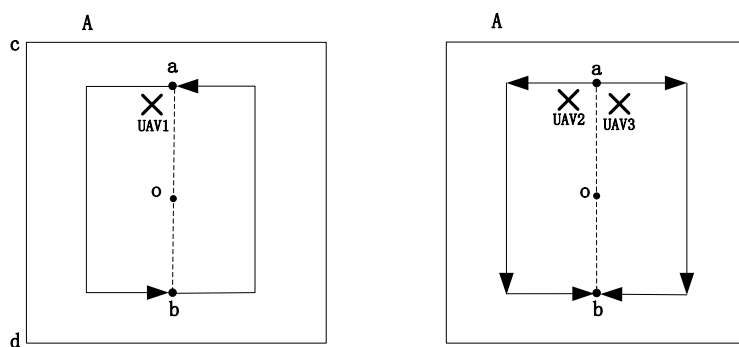
单无人机视觉 SLAM 系统面临着大范围的未知环境、有限任务时间的考验；当未知环境范围大时，任务时间增加，位姿估计的累积误差增加，单无人机视觉 SLAM 获得的位姿和地图误差增加；再者，考虑到单无人机可能出现故障，因此，本章对多无人机双目视觉 SLAM 进行研究，着重对两无人机双目视觉 SLAM 进行研究。

5.2 多无人机双目视觉 SLAM 系统

5.2.1 两架无人机双目视觉 SLAM 方法

如图 5.1(a)所示， A 代表未知环境的俯视图， o 点为俯视图的中心， c 点、 d 点为边界点。过 o 点作与 cd 平行线，选取平行线中两点 a 、 b 。无人机 UAV1 从 a 逆时针到 b ，再从 b 逆时针到 a ，对区域 A 进行视觉 SLAM。

如图 5.1(b)所示，两无人机 UAV2、UAV3 对区域 A 进行视觉 SLAM，无人机 UAV2 逆时针从 a 到 b 进行 SLAM，与此同时，无人机 UAV3 顺时针从 a 到 b 进行 SLAM。



(a) 单无人机双目视觉 SLAM 区域图

(b) 两无人机双目视觉 SLAM 区域图

图 5.1 无人机双目视觉 SLAM 区域图

两架无人机 SLAM 后获得两区域的位姿轨迹和地图，对于两区域的位姿轨迹的融合分为以下两种方案。

- 1) 方案一：直接将两位姿轨迹拼接。此种方法相比于单无人机视觉 SLAM 的优点在于：首先，在忽略一定误差的情况下，UAV2、UAV3 在 a 有公共位姿点，且利用两无人机 SLAM，使得单架无人机 SLAM 的路径减半，VO 中的位姿累积误差减小，因此在 VO 过程中，由 UAV2、UAV3 获得的初始位姿轨迹相对于 UAV3 在该区域的位

姿轨迹更精确。其次，两无人机视觉 SLAM 比单无人机 SLAM 所用时间更短。相比于单无人机的缺点为：由 VO 获得位姿矩阵作为位姿图优化中初值时，单无人机视觉 SLAM 将整个区域的位姿进行处理；而本方案中两无人机视觉 SLAM 方法对一个区域分为独立的两部分处理，自然两区域本应该存在数据关联的部分没有描述出来；即在 a 、 b 两点附近，UAV2 进行 SLAM 获得的关键帧本应该与 UAV3 进行 SLAM 获得的关键帧存在回环检测，增加位姿图中边的关系，但是因为两部分独立处理，所以没有考虑。

- 2) 方案二：将两位姿轨迹拼接后，再进行回环检测，进行位姿图优化。相比于上述方法，此法不仅仅拥有其所有优点，而且根据上述提到的部分数据关联信息丢失的情况，将 UAV2 和 UAV3 前端 VO 及回环检测后，再进一步对获得的 a 、 b 两点附近的关键帧进行回环检测，再进行一次位姿图优化。相对于上述方法，此种方法将带来更多的时间消耗，用于回环检测和后端优化；但是，相比于对单无人机视觉 SLAM，所耗费时间依然更少。

两架无人机进行视觉 SLAM 获得两张点云地图，因为两张点云地图的参考位姿点是相同的，所以对应的地图融合方法为：直接将两张点云地图中的三维点放在同一三维点集中，该三维点集用点云图描述，便是全局三维点云地图。

5.2.2 多无人机视觉 SLAM 方法

对于利用两架以上无人机进行视觉 SLAM 的情况，假设无人机数量为 m ，分别为 $n_1, n_2, \dots, n_i, \dots, n_m$ 。将待执行 SLAM 的区域均等划分为 n 块，每区域利用一架无人机进行 SLAM。每个区域 n_i 选取了一个参考位姿点 p_i ，其中 $i \in \{1, 2, \dots, n\}$ 。选取区域 n_j 的参考位姿点 p_j ，作为全局参考位姿点，其中 $j \in \{1, 2, \dots, n\}$ 。各区域的无人机进行 SLAM 后，获得相对于该区域参考位姿点的位姿轨迹。利用特征提取与匹配获得 p_j 与 $p_{i, i \neq j}$ 的变换矩阵，因此，将区域 n_j 之外的其他 $n-1$ 区域的参考位姿点转化为相对于全局位姿点的值，即统一了参考位姿点。将各位姿点的变换矩阵乘积的形式级联起来，便获得了多无人机在整个区域的位姿轨迹。由于获得了统一连续的位姿轨迹，将所有的局部地图通过当前的位姿变换矩阵融合到全局参考位姿点的全局地图中，这便是多无人机视觉 SLAM 的地图融合。此外，在没有获得连续一致性的位姿轨迹时，可以利用迭代最近点方法获得各局部地图相对于全局地图的变换矩阵，再进行地图融合。

5.3 Octomap 地图研究

在无人机的应用领域中，要求一个概率表示，对空白区域（可自由移动区域），占有区域（障碍物）和未知区域建模，在时间和存储空间具有较高的效率的地图。当无人机通过双目相机感知环境，测量值受不确定性因素影响，例如：动态物体。当要求从嘈杂的环境中创建环境的精

确模型时，必须概率地考虑潜在的不确定。多个不确定性的测量结果融合后可以鲁棒性地估计正确的结果。除此之外，此种概率的表示还可以融合来自于其他传感器的数据。对未知区域建模：在自主导航任务中，无人机可以由相机确定占有区域和空闲区域，规划无碰撞路径。未知区域认为是占有区域，需要避免出现在规划的路线上，因此需要对未知区域进行建模。无人机可以路径规划和执行任务的关键是地图，同时，地图需要在时间和内存两方面具有高效率。在 3D 建图方法，内存消耗是一个大问题。因此，在大规模的环境中此地图必须有效地、节省内存的。

Octomap^[67]满足上述三种要求，octomap 形式紧凑，占内存较小，可以用于导航，方便压缩，更新且分辨率可调。因此采用 octomap 地图用于无人机的自主导航是很好地选择。

Octomap 是通过八叉树以概率的形式表示空间，八叉树的每个节点表示空间的立体空间，称为一个体元。如图 5.2(a)^[67]所示，将某空间区域用一个立方体表示，不断将该立方体均匀划分为八等份，直到达到最小的立方体未知，这个最小的立方体对应 octomap 的分辨率；图 5.2(b)表示将划分的体元用存储概率的八叉树表示，其中白色表示空白区域，黑色表示占有区域。

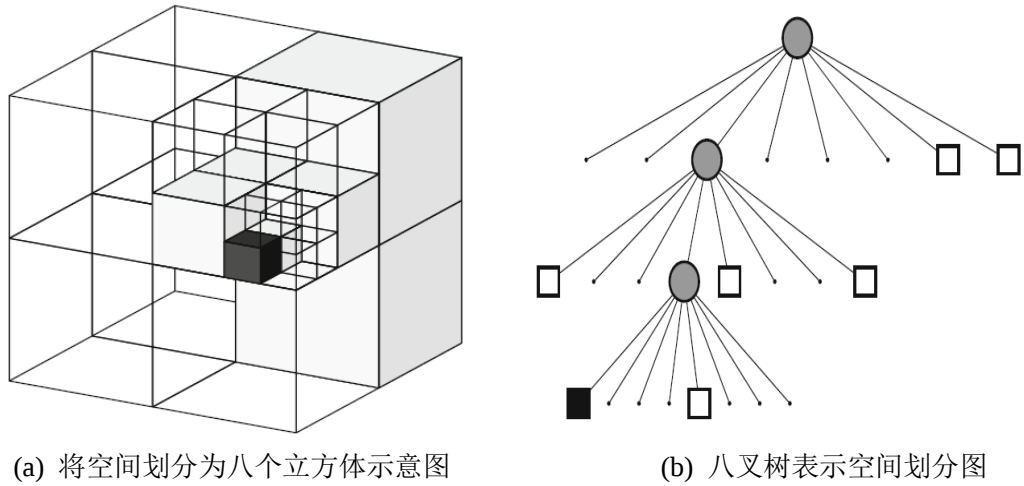


图 5.2 八叉树表示空间示意图

节点 m 不断地被观测，随着观测过程不断被观测为占有区域，概率值不断增加，随着被观测为空白区域而概率值不断减小。假设某节点 m 的初始占有概率 $P(m)$ 为 0.5，节点 m 在观测数据 $z_{1:t}$ 的前提下的概率 $P(m|z_{1:t})$ 表示如下：

$$P(m|z_{1:t}) = \left[1 + \frac{1 - P(m|z_t)}{P(m|z_t)} \frac{1 - P(m|z_{1:t-1})}{P(m|z_{1:t-1})} \frac{P(m)}{1 - P(m)} \right]^{-1} \quad (5.1)$$

在式 (5.1) 中， $z_{1:t}$ 表示从开始时刻 1 到 t 时刻节点 m 的观测数据， $P(m|z_t)$ 表示 t 时刻时观测到节点 m 获得的概率， $P(m|z_{1:t-1})$ 表示从开始时刻到 t 时刻节点 m 的观测数据。

在将 octomap 用于导航时，基于反传感器模型，将从障碍物到相机之间的空间认为是空白区域，障碍物处的表示的空间则为占有区域。在 t 时刻，节点 m 的观测 z_t 为空白区域时，概率

$P(m|z_t)=0.4$ ； z_t 为占有区域时， $P(m|z_t)=0.7$ 。

为了使得式 (5.1) 中乘法运算变为加法运算，利用概率对数值表示概率 $P(m)$ ，如下：

$$\hat{L}(m) = \log \left[\frac{P(m)}{1-P(m)} \right] \quad (5.2)$$

将式 (5.1) 用 (5.2) 表示的概率对数表示为：

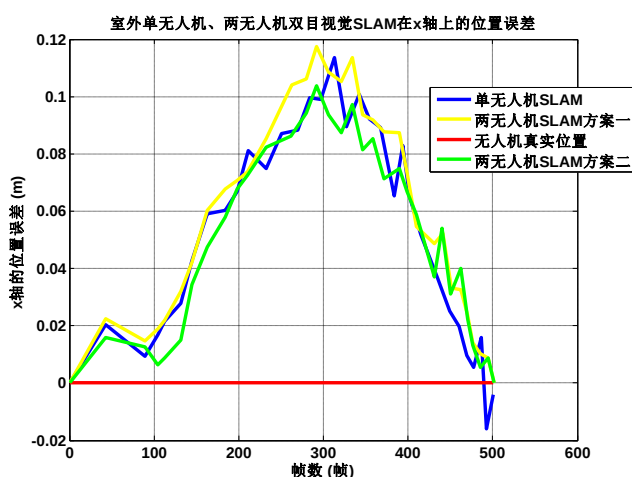
$$\hat{L}(m|z_{t'}) = \hat{L}(m|z_{t'-1}) + \hat{L}(m|z_t) \quad (5.3)$$

利用概率对数值表示存储概率，所以空白区域 $P(m|z_t)=0.4$ 对应的概率对数值 $\hat{L}(m|z_t)=-0.4$ ，占有区域 $P(m|z_t)=0.7$ 对应 $\hat{L}(m|z_t)=0.85$ 。

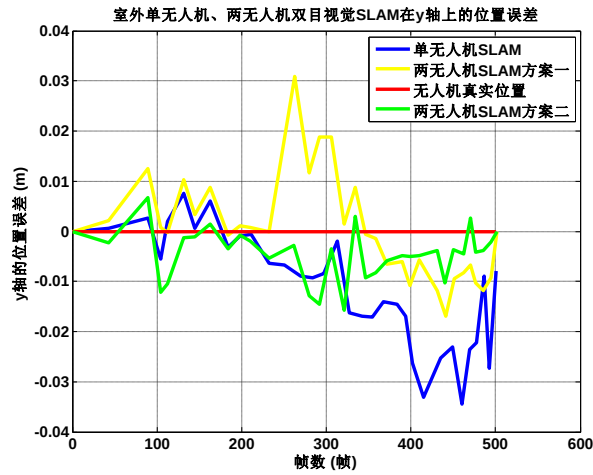
Octomap 相比于点云地图的优势体现在：节省空间、实时更新。从图 5.2 可知，octomap 中部分体元的体积大小不是叶子节点对应的体元的体积大小，即空白区域不会一直划分为八个子节点，被完全占据的空间不会再一直划分为八个子节点，这样有效地节省了空间。从式 (5.2)、式 (5.3) 可知 octomap 中体元的概率一直更新，与之前的时刻的时刻有关，这使得 octomap 可以表示动态物体。

5.4 仿真实验

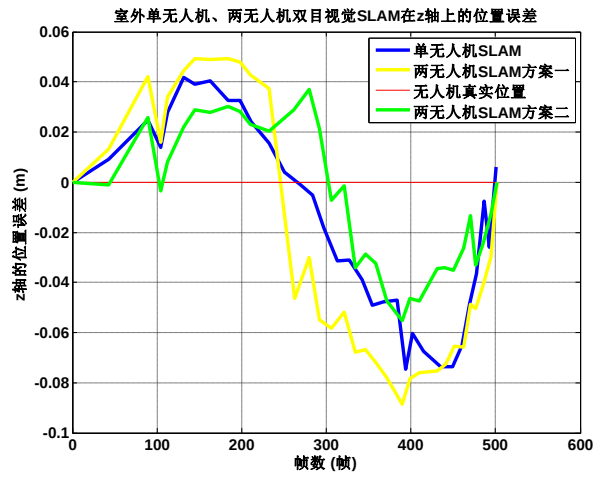
本章实验平台和实验环境如前三章所述一致，为了得到一个可靠有效的无人机的参考位置来验证本章所提的方案的有效性，在实际实验中，将无人机在待进行 SLAM 的区域的某一空间位置固定，绕相机的轴心旋转。考虑到 SIFT 特征提取过程耗费时间，本章的特征提取部分为 FAST 角点特征提取，采用的描述子为 SIFT。在室外的情况下，基于 SIFT 的单无人机双目视觉 SLAM 和两种多无人机视觉 SLAM 方法仿真实验得到的位置曲线如下图 5.3 所示。



(a) x 轴误差



(b) y 轴误差

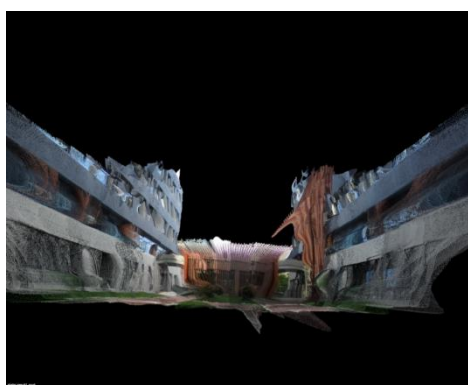


(c) z 轴误差

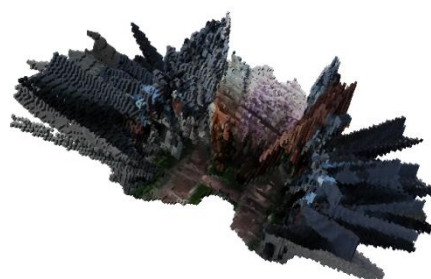
图 5.3 室外单无人机、两无人机双目视觉 SLAM 在三轴的位置误差对比图

从图 5.3 可得, 在室外情况下, 本章提出的两无人机双目视觉 SLAM 方案二相比于单无人机双目视觉 SLAM 在三轴的位置估计更精确, 因此地图也更加精确; 而方案一与单无人机双目视觉 SLAM 相比, 对三轴的位置估计相差不大。

方案一中, UAV2 获得 304 帧, 关键帧为 16, SLAM 所需时间为 678s。获得的点云地图及将点云地图用 octomap 地图表示, 如图 5.4 所示。



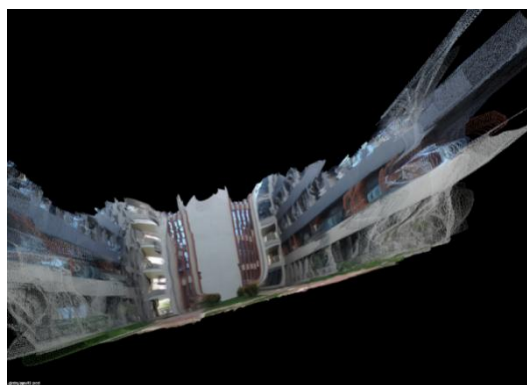
(a) 室外点云地图



(b) 室外 octomap 地图

图 5.4 UAV2 获得的室外地图模型

方案一中，UAV3 获得 200 帧，关键帧为 18，SLAM 所需时间为 492s。获得的点云地图及 octomap 地图，如图 5.5 所示。



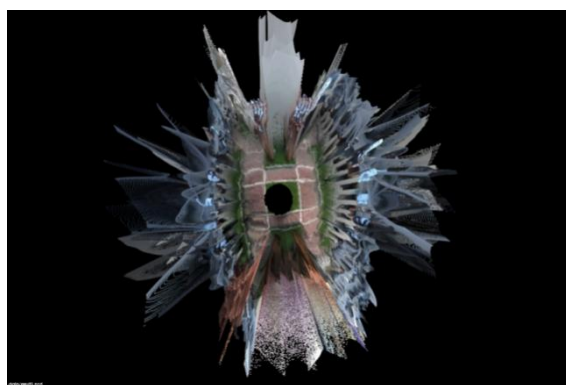
(a) 室外点云地图



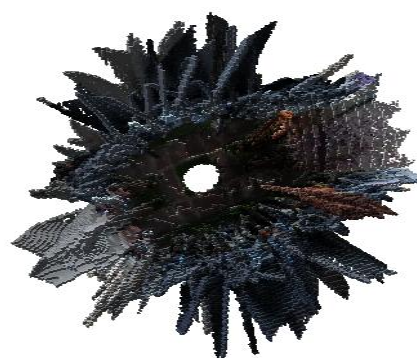
(b) 室外 octomap 地图

图 5.5 UAV3 获得的室外地图模型

利用上述介绍的方法，将点云地图融合，且将其表示为 octomap，如图 5.6 所示。



(a) 融合后点云地图

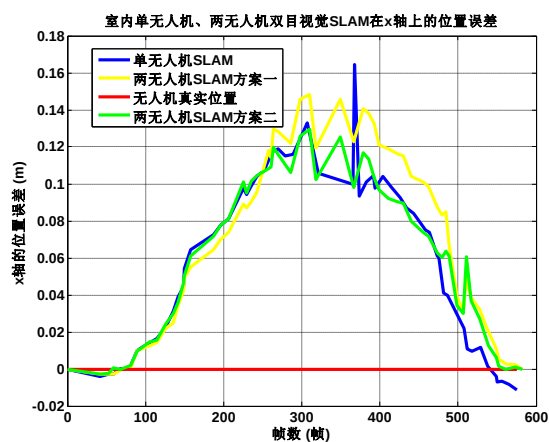


(b) 融合后 octomap 地图

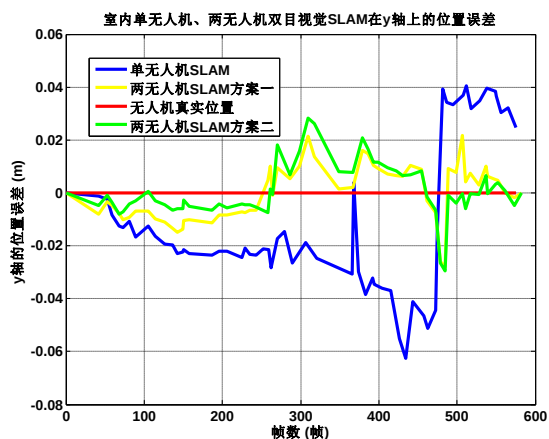
图 5.6 室外地图融合后的地图模型

在以上仿真实验中，单无人机双目视觉 SLAM 所用时间为 1112s，方案一所用时间为 678s，

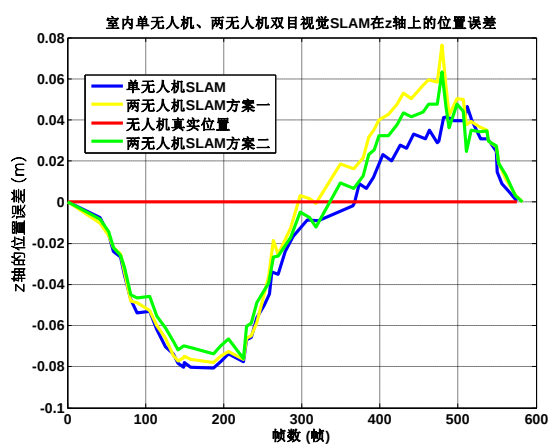
方案二所用时间为 746s。上述方法，在室内情况下获得的位置曲线，如图 5.7 所示。从图 5.7 可得，在室内情况下得到的结论和室外情况下式吻合的。



(a) x 轴误差



(b) y 轴误差



(c) z 轴误差

图 5.7 室内单无人机、两无人机双目视觉 SLAM 在三轴的位置误差对比图

在室内情况下的方案一中，UAV2 获得 267 帧，关键帧为 26，SLAM 所耗时间为 526s。

获得的点云地图及 octomap 地图，如图 5.8 所示。

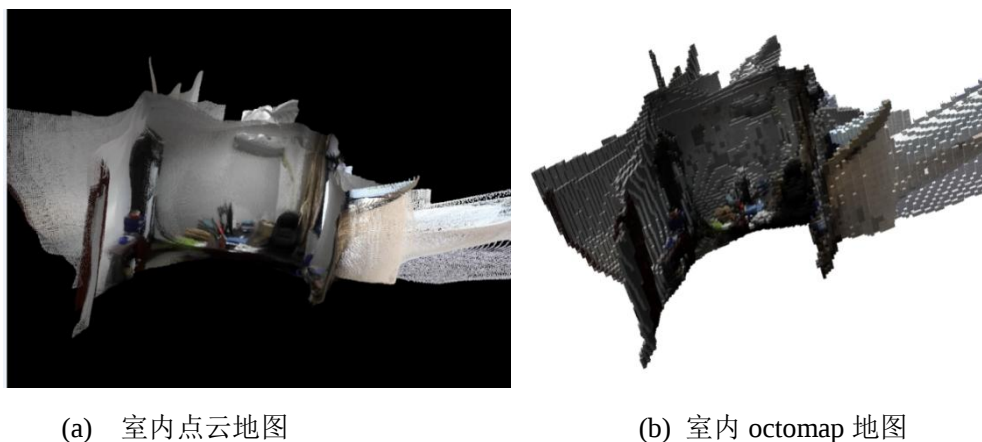


图 5.8 UAV2 获得的室内地图模型

UAV3 获得 315 帧，关键帧为 33，SLAM 所需时间为 596s。获得的点云地图如图 5.9 所示。将两无人机地图融合后，得到点云地图及 octomap 地图，如图 5.10 所示。

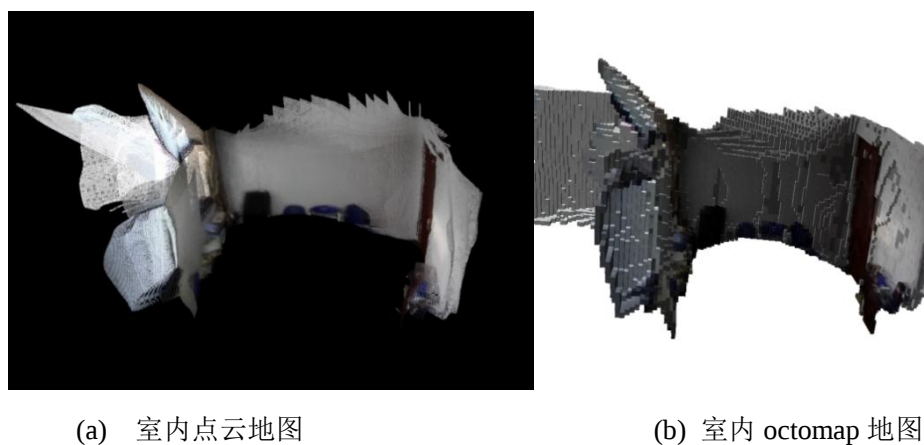


图 5.9 UAV3 获得的室内地图模型

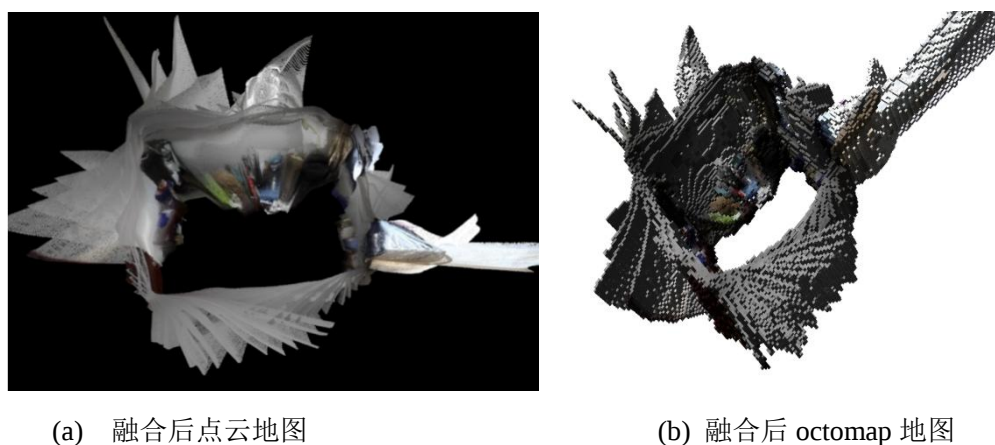


图 5.10 室内融合后的地图模型

在室内仿真实验中，单无人机双目视觉 SLAM 所用时间为 1144s，方案一所用时间为 596s，方

案二所用时间为 651s。

综上可得：在室内外情况下，方案二相比于单无人机双目视觉 SLAM 在位置估计、地图的精确性和时间效率上具有优势；方案一相比单无人机双目视觉 SLAM 在时间效率上具有优势，但在位置估计和地图精确性上相差无几。

5.5 本章小结

本章研究多无人机双目视觉 SLAM 方法，主要对两无人机的双目视觉 SLAM 方法进行了研究，对具有导航功能的 octomap 的地图进行了研究，进行了针对基于 SIFT 特征的两无人机双目视觉 SLAM，提出了两种方案。在室内外的实验环境下，对两方案开展了仿真实验。在室内外的情况下，方案二与单无人机双目视觉 SLAM 相比，获得了更精确的位置估计和地图，且有效地节省了时间；方案一与单无人机双目视觉 SLAM 相比，在保证位置估计不受太大影响的前提下，节省了时间，提高了效率。

第六章 总结和展望

6.1 本文的主要工作

在缺失 GPS 信号的室内、特殊室外环境下, SLAM 是无人机导航的关键。本文研究了在未知环境下无人机的双目视觉 SLAM, 研究了双目视觉 SLAM 的理论方法, 搭建了无人机双目视觉 SLAM 平台, 开展了室内外双目 SLAM 的实验, 并对多无人机 SLAM 进行了理论分析。本文的研究对复杂未知环境下提高无人机的自主导航性能具有重要的理论意义。

本文主要研究内容总结如下:

(1) 研究了双目相机的立体成像原理, 完成了 ZED 立体相机的标定, 对 SIFT 和 ORB 两种特征提取与匹配算法进行了研究, 搭建了无人机双目视觉 SLAM 的平台。在室内外环境下, 利用无人机双目视觉 SLAM 平台采集数据, 通过 SIFT 和 ORB 算法对所采集数据进行特征提取与匹配, 实验对比了两种算法优缺点。

(2) 构建基于特征的双目视觉 SLAM 的前端, 即视觉里程计。针对视觉里程计中运动估计问题, 研究了 3D-2D、3D-3D 两种运动估计方法, 实现了室内外基于 SIFT 和 ORB 的双目视觉里程计, 获得了无人机的初始位姿轨迹和融合后的三维点云地图。对基于 SIFT 和 ORB 的室内外视觉里程计的实验结果进行了分析比较。

(3) 构建基于特征的双目视觉 SLAM 的后端。在视觉里程计、关键帧、回环检测的研究前提下, 采用非线性优化的方法, 利用位姿图优化的方法完成对 SLAM 的优化。建立了视觉 SLAM 的后端优化模型, 基于 SIFT 和 ORB 实现了室内外的无人机的双目视觉 SLAM, 获得了优化后的精度更高的位置轨迹及全局一致地图。

(4) 研究了多无人机双目视觉 SLAM 方法, 提出了两种基于 SIFT 特征的两无人机双目视觉 SLAM 方案。在室内外的实验环境下, 对上述两方案开展了仿真实验。

6.2 前景展望

本文研究的单无人机双目视觉 SLAM 的算法, 搭建了单无人机双目视觉 SLAM 实验平台, 在室内外的环境下, 实现了单无人机基于 SIFT 和 ORB 的视觉 SLAM; 并对两架无人机双目视觉 SLAM 进行了仿真实验, 对多无人机视觉 SLAM 进行了理论分析。但是由于双目视觉 SLAM 问题是个多学科交叉的课题, 理论性强, 加上时间有限, 在理论研究方面还存在不足; 同时, 双目视觉 SLAM 是个工程性很强的课题, 在实验方面还有待进一步开展。

首先, 改善前端。本文对基于特征的双目视觉 SLAM 系统进行了研究及搭建了实验平台, 但是对于视觉里程计中只是利用特征点进行构建, 构建的地图是稀疏的地图。其一, 在某些研究

中, 研究人员充分利用图像的 RGB 信息与特征信息进行融合, 以获得两帧间更精确匹配, 获得提供给后端更精确的初值, 对于获得精确的无人机位姿轨迹和地图具有重要意义。其二, 在视觉里程计中的局部优化还需进一步研究与改进, 局部的位姿调整对于后期获得精确的位姿和地图同样重要。其三, 在视觉里程计中, 本文只是利用双目立体相机 ZED 构建里程计, 这将导致无人机在做视觉 SLAM 中, 无法处理动态的障碍物, 结合 IMU 可以减轻动态物体的影响, 这便是视觉惯性里程计。以上三点都是今后值得研究的方向。

其次, 改善后端。本文的双目视觉里程计以较简单的方式实现的关键帧检测和回环检测在容错性方向还存在欠缺。回环检测部分以 DBOW2 为主流方法, 随着技术的发展, 慢慢地结合深度学习等方法, 优质的回环检测对于获得矫正无人机的位姿误差具有重要意义。同时, 在非线性优化中, 本文考虑算法的复杂性和有效性, 没有将位姿点与路标点之间的误差添加到优化模型中。

然后, 本文只是研究了两架无人机的视觉 SLAM, 对两架无人机视觉 SLAM 进行了仿真实验, 并未搭建实验平台。今后, 搭建多无人机的实验平台是一项重要任务, 因为视觉 SLAM 是一项工程性很强的课题, 如果仅仅停留在理论的研究, 在现如今 SLAM 的研究已经进入最后一个智能感知阶段的背景下, 这显然是不够的。

最后, 无人机的双目视觉 SLAM 设计的最终目的是用于导航。虽然本文介绍了可用于导航的 octomap, 但是距离 SLAM 用于实时自主导航还有很长一段路程, 这决定了 SLAM 的研究意义, 这是最终目标。

参考文献

- [1] 陶于金, 李沛峰. 无人机系统发展与关键技术综述[J]. 航空制造技术, 2014, (20): 34~39.
- [2] Cesar Cadena. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309~1332.
- [3] P Newman, J Leonard, J Tardos, J Neira. Explore and Return: Experimental validation of Real-Time concurrent mapping and localization[J]. IEEE International Conference on Robotics and Automation, 2002, 2(2): 1802~1809.
- [4] C Forster, L Carlone, F Dellaert D. Scaramuzza. Onmanifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics, 2017, 33(1): 1~21.
- [5] H Durrant Whyte, T Bailey. Simultaneous localization and mapping: part I[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99~110.
- [6] T Bailey, Hugh Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II[J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108~117.
- [7] Hauke Strasdat, Jos é MM Montiel, Andrew J Davison. Visual slam: why filter?[J]. Image and Vision Computing, 2012, 30(2): 65~77.
- [8] A Davison, I Reid, N Molton, O Stasse. Monoslam: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052~1067.
- [9] G Klein, D Murray. Parallel tracking and mapping for small ar workspaces[J]. 6th IEEE and ACM International Symposium, 2007, 225~234.
- [10] R Mur Artal, J Montiel, J D Tardos. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147~1163.
- [11] 高翔, 张涛. 视觉SLAM十四讲: 从理论到实践[M]. 北京: 电子工业出版社, 2017, 3.
- [12] J Engel, V Koltun, D Cremers. Semi-dense visual odometry for a monocular[J]. IEEE International Conference on Computer Vision, 2013: 1449~1456.
- [13] J Engel, T Schops, D Cremers. Lsd-slam: Large-scale direct monocular slam[J]. Computer Vision-ECCV 2014, 2014: 834~849.
- [14] Jiaxin Li, Yingcai Bi. Real-time Simultaneous Localization and Mapping for UAV: A Survey[J]. International micro air vehicle competition and conference, 2016: 237~242.
- [15] Stefan Kohlbrecher, Oskar Von Stryk. A Flexible and scalable slam system with full 3d motion estimation[J]. IEEE International Symposium on Safety, Security, and Rescue Robotics, 2011:

155~160.

- [16] Albert S Huang, Abraham Bachrach. Visual odometry and mapping for autonomous flight using an rgb-d camera[J]. International Symposium on Robotics Research(ISRR), 2011, 2.
- [17] Nesrine Mahdoui, Enrico Natalizio, Vincent Fremont. Multi-UAVs Network Communication Study for Distributed Visual Simultaneous Localization and Mapping[J]. International Conference on Computing, Networking and Mapping, 2016.
- [18] A Nemra, N Aouf. Robust Cooperative UAV Visual SLAM[J]. Cybernetic Intelligent Systems, 2010 IEEE 9th International Conference on, 2010.
- [19] 贾云得. 机器视觉[M]. 北京: 科学出版社, 2000.
- [20] Bradski G, Kaehler, A 著; 于仕琪,刘瑞祯译. 学习 Opencv(中文版)[M]. 北京: 清华大学出版社, 2009.10(2016.5 重印).
- [21] Zhengyou Zhang. A Flexible New Technique for Camera Calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, 22(11): 1330~1334.
- [22] Fusiello A, Trucco E, Verri A. A Compact Algorithm for Rectification of Stereo Pairs[J]. Machine Vision and Applications, 2000, 12(1): 16~22.
- [23] 白明, 庄严, 王伟. 双目立体匹配算法的研究和进展[J]. 控制与决策, 2008, 7: 721~729.
- [24] 于乃功, 秦永钢, 阮晓钢. 立体匹配算法进展[J]. 计算机测量与控, 2009, 5: 817~819+840.
- [25] Heiko Hirschmuller. Stereo Processing by Semiglobal Matching and Mutual Information[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328~341.
- [26] Sara Martull, Martin Peris, Kazuhiro Fukui. Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps[J]. ICPR, 2012, 1038~1042.
- [27] David G Lowe. Object recognition from Local Scale-Invariant Features[J]. International Conference on Computer Vision, 1999, 1150~1157.
- [28] Junseong Bang, Dongchun Lee, Youngjun Kim, Hunhoo Lee. Camera Pose Estimation Using Optical Flow and ORB Descriptor in SLAM-Based Mobile AR Game[J]. 2017 International Conference on Platform Technology and Service, 2017: 1~4.
- [29] David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2), 91~110.
- [30] Edward Rosten, Tom Drummond. Machine learning for high-speed corner detection[J]. 9th Euproean Conference on Computer Vison, 2006, 1:430~443.
- [31] Rosten E, Drummond T. Machine Learning for High-Speed Corner Detection[M]. Springer Berlin Heidelberg, 2006: 430~443.

- [32] Rosten E, Porter R, Drummond T. Faster and better: A Machine Learning Approach to Corner Detection[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(1): 105~119.
- [33] Calonder M, Lepetit V, Strecha C, et al. Brief: Binary Robust Independent Elementary Features[M]. Springer Berlin Heidelberg, 2010: 118~792.
- [34] 章杰. 基于 ORB 特征和粒子滤波的目标跟踪算法研究[D]. 吉林: 吉林大学, 2014.
- [35] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski. ORB: An efficient alternative to SIFT or SURF[J]. 2011 International Conference on Computer Vision. 2011: 2564~2571.
- [36] Scaramuzza Davide, Fraundorfer Friedrich. Visual Odometry Part I: The First 30 Years and Fundamentals[J]. IEEE Robotics & Automation Magazine, 2011, 18(4): 80~92.
- [37] D Nister, O Naroditsky, J Bergen. Visual odometry[J]. Computer Vision and Pattern Recognition, IEEE Computer Society Conference, 2004, 652~659.
- [38] F Moreno Noguera, V. Lepetit, P Fua. Accurate non-iterative O(n) solution to the PnP problem[J]. IEEE International Conference, Computer Vision, 2007: 1~8.
- [39] X S Gao, X R Hou, J Tang, H F Cheng. Complete solution classification for the perspective-three-point problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25: 930~943.
- [40] B Triggs, P McLauchlan, R Hartley, A Fitzgibbon. Bundle adjustment a modern synthesis[J]. Workshop Vision Algorithms: Theory and Practice (ICCV'99). 2000: 298~372.
- [41] Chris Engels, Henrik Stewenius, David Nister. Bundle Adjustment Rules[J]. Photogrammetric Computer Vision, 2006.
- [42] N. Bourbaki. Lie Groups and Lie Algebras[M]. Berlin: Springer, 2002.
- [43] Lebeda K, Matas J, Chum O. Fixing The Locally Optimized RANSAC[J]. 23rd British Machine Vision Conference, 2012.
- [44] Friedrich Fraundorfer, Davide Scaramuzza. Visual Odometry: Part II: Matching, Robustness, Optimization, and Applications[J]. IEEE Robotics & Automation Magazine, 2012, 19(2): 78~90.
- [45] P Besl, N McKay. A method for registration of 3-d shapes[J]. IEEE Trans. Pattern. Anal. Machine Intell. 1992, 14(2): 239~256.
- [46] 祝继华, 周颐等. 基于图像配准的栅格地图的拼接方法[J]. 自动化学报, 2015, 41(2): 285~294.
- [47] F Pomerleau, F Colas, R Siegwart. A review of point cloud registration algorithms for mobile robotics[J]. Foundations and Trends in Robotics(FnTROB), 2015, 4(1): 1~104.

- [48] Marquardt Donald. An Algorithm for Least-Squares Estimation of Nonlinear Parameters[J]. SIAM Journal on Applied Mathematics, 1963, 11(2): 431~441.
- [49] T Bailey, H Durrant Whyte. Simultaneous localisation and mapping (SLAM): Part II State of the art[J]. IEEE Robot Automation Magazine, 2006 , 13(3): 108~117.
- [50] Blanco J L, Fernández-Madriral J A, Gonzalez J. A New Approach for Large-Scale Localization and Mapping: Hybrid Metric-Topological SLAM[J]. IEEE International Conference on in Robotics & Automation. 2007: 2061~2067.
- [51] 丛楚滢. 未知环境下无人机自主导航的SLAM方法研究[D]. 南京: 南京航空航天大学, 2016.
- [52] 陈含欣. 无人机视觉特征提取及同步定位与地图的构建方法研究[D]. 南京: 南京航空航天大学.
- [53] Montemerlo M, Thrun S. Simultaneous localization and mapping with unknown data association using FastSLAM[J]. Robotics and Automation, IEEE International Conference, 2003, 2: 1985~1991.
- [54] Hahnel D, Burgard W, Fox D, et al. An efficient fastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[J]. Intelligent Robots and Systems, Proceedings, IEEE/RSJ International Conference on. IEEE, 2003:206 ~211.
- [55] Timothy D Barfoot. State Estimation for Robotics[M]. Cambridge University Press, 2017.
- [56] Miguel Arg áez, Horacio Florez, Osvaldo Méndez. A model reduction for highly non-linear problems using wavelets and the Gauss-Newton method[J]. Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS), 2016: 1~7.
- [57] Jorge Nocedal, Stephen J Wright. Numerical Optimization[M]. Verlag New York: Springer, 1999.
- [58] Rainer Kümmerle, Giorgio Grisetti, et al. G2o: A general framework for graph optimization[J]. IEEE International Conference on Robotics and Automation, 2011: 3607~3613.
- [59] I Ulrich, I Nourbakhsh. Appearance-based place recognition for topological localization[J]. IEEE Int. Conf., Robotics and Automation, 2000: 1023~1029.
- [60] M Jogan, A Leonardis. Robust localization using panoramic view-based recognition[J]. ICPR, 2000, 4: 136~139.
- [61] K Mikolajczyk, T Tuytelaars, C Schmid, A Zisserman, J Matas, F Schaffalitzky, T Kadir, L Van Gool. A comparison of affine region detectors[J]. Compute Vision, 2005, 65: 43~72.
- [62] P Newman, D Cole, K Ho. Outdoor SLAM using visual appearance and laser ranging[J]. IEEE Int. Conf. Robotics and Automation, 2006: 1180~1187.

- [63] M Cummins, P Newman. FAB-MAP: Probabilistic localization mapping in the space of appearance[J]. The International Journal of Robotics Research, 2008: 647~665.
- [64] F Fraundorfer, C Engels, D Nister. Topological mapping, localization and navigation using image collections[J]. IEEE/RSJ Conf. Intelligent Robots and Systems, 2007: 3872~3877.
- [65] F Fraundorfer, C Wu, J M Frahm, M. Pollefeys. Visual word based location recognition in 3d models using distance augmented weighting[J]. 4th Int. Symp. 3D Data Processing, Visualization, and Transmission, 2008:1~8.
- [66] E Olson, J Leonard, S Teller. Fast iterative optimization of pose graphs with poor initial estimates[J]. ICRA, 2006: 2262~2269.
- [67] A Hornung, K M Wurm, M Bennewitz, C Stachniss, W Burgard. Octomap: An efficient probabilistic 3d mapping framework based on octrees[J]. Autonomous Robots. 2013, 34(3): 189~206.

致谢

当我要开始回忆你们对我的帮助时，这意味着这段旅程告一段落，于我而言，学生生涯也将彻底结束。为此，我虽不曾取得多大的成绩，但我想对你们表达我内心的谢意。

首先，我要对我的导师王从庆教授表示最衷心的感谢。感谢老师在研究生期间给予我的帮助，在这期间了解到老师的平易近人，知识渊博，严谨的学习态度，被老师的魅力深深折服，同样希望自己在将来的道路上和老师一样保持对科学的热爱和追求。再次对老师表示感谢，谢谢老师的指导和教诲！

其次，感谢实验室的贾峰、谢勇、吴林峰、古嘉樾、于鸿达、左超等人在课题研究过程中的帮助和配合！谢谢大家共同营造了一个良好的科研环境，一个温暖的大家庭！祝愿你们前程似锦，有情人终成眷属，在尘世获得幸福。

然后，感谢我的好友。谢谢你们关怀和问候，谢谢你们的包容与无私的付出，因为有你们，多少秋风萧瑟处，都付笑谈中。

再者，感谢我的家人和爱人。谢谢你们，谢谢你们的包容与无条件的付出，此等厚重，费劲纸墨也无法言尽。我爱你们，谢谢！

此外，感谢给予我美好的所有人，哪怕是一个真诚的微笑，于我而言，却如春风细雨，给我的心田带来温润。

最后，向百忙之中审阅此文的专家、教授们表示深深的谢意，感谢您对我提出批评指正意见。

钟家跃

2017 年 5 月

在学期间的研究成果及发表的学术论文

攻读硕士学位期间发表（录用）论文情况

1. 钟家跃, 王从庆, 丛楚滢. 一种未知环境下多无人机主从合作 FastSLAM 方法[C]. 2016 IEEE 中国制导、导航与控制学术会议, 2016.

攻读硕士期间申请专利情况

1. 钟家跃, 王从庆, 贾峰. 城市复杂环境下多无人机视觉 SLAM 的地图融合方法, 公开号: 106595659A. 公布日期: 2017 年 4 月 26 日.
2. 钟家跃, 王从庆, 贾峰. 一种可适应不同运动场景的改进 AODV 路由协议, 申请号: 201611122427.4

攻读硕士学位期间参加科研项目情况

2014 年江苏省科技支撑项目 (SBE2014070836) 面向复杂城市环境的多无人机应急监测技术研究

2015 年南京航空航天大学研究生创新基地 (实验室) 开放基金资助(kfjj20150321) 面向城市复杂环境的多无人机协同监测通信技术