

# 硕士学位论文

## 基于双目视觉的四旋翼无人机 定位系统设计与实现

### DESIGN AND IMPLEMENTATION OF QUADROTORS BASED ON STEREO VISION

何芳

哈尔滨工业大学

2017 年 6 月

国内图书分类号: TP242  
国际图书分类号: 621

学校代码: 10213  
密级: 公开

## 工程硕士学位论文

# 基于双目视觉的四旋翼无人机 定位系统设计与实现

硕 士 研 究 生: 何芳

导 师: 马杰 教授

申 请 学 位: 工程硕士

学 科: 控制工程

所 在 单 位: 航天学院

答 辩 日 期: 2017 年 6 月

授予学位单位: 哈尔滨工业大学

Classified Index: TP242  
U.D.C: 621

Dissertation for the Master Degree in Engineering

# **DESIGN AND IMPLEMENTATION OF QUADROTORS BASED ON STEREO VISION**

<b>Candidate:</b>	He Fang
<b>Supervisor:</b>	Prof. Ma Jie
<b>Academic Degree Applied for:</b>	Master of Engineering
<b>Speciality:</b>	Control Engineering
<b>Affiliation:</b>	School of Astronautics
<b>Date of Defence:</b>	June, 2017
<b>Degree-Conferring-Institution:</b>	Harbin Institute of Technology

## 摘 要

本文以自主开发一套能够适用于四旋翼无人机室内自主定位的视觉 SLAM 系统为目标, 设计并编程实现了一种配置灵活、计算量小、性能稳定的四旋翼无人机室内定位系统。本文对部分已有主流开源视觉 SLAM 系统进行测试分析, 归纳总结了这些典型系统实际应用于无人机存在的不足, 针对存在的困难从两个方面进行算法改进。从达到实时处理效率的角度出发, 结合基于特征点法和半直接法完成运动估计的理论基础, 提出了基于半直接法的双目视觉 SLAM 系统的设计方法以及实现方案。从提高系统鲁棒性的角度出发, 采用双目视觉传感器进行三维空间点位置估计, 实现了可估计真实环境尺度的双目匹配算法。设计实验验证本系统的有效性。本文的主要工作内容可归纳如下:

(1) 建立了四旋翼飞行器的位姿模型、视觉系统的基本模型。首先, 采用概率分布模型对 SLAM 系统进行数学描述, 建立了系统相关模型, 并定义了相关坐标系。然后, 结合相机模型, 重点分析了视觉 SLAM 系统的几何原理, 为系统设计及算法改进奠定了基础。

(2) 给出了本课题的整体系统方案以及结构框架。结合前文对现有视觉 SLAM 系统实际应用于四旋翼无人机平台存在困难进行归纳总结, 确定本课题的实际需求。依据运动估计方法, 将主流的视觉 SLAM 系统分为基于滤波器、基于关键帧 BA 以及基于直接法三类视觉 SLAM。然后, 根据分类结果, 对部分已有主流开源的视觉 SLAM 系统进行测试分析, 从实时性以及鲁棒性两个角度分析, 结合这些系统的优缺点, 最终确定本课题的整体系统方案以及结构框架。

(3) 提出了基于半直接法的双目视觉 SLAM 系统。首先, 针对基于特征点法的 ORB-SLAM 系统计算量大、在强烈旋转下鲁棒性差等缺点, 以及基于直接法的 SVO 系统对采用前视摄像机进行运动估计鲁棒性差、无法估计真实环境尺度等缺点, 对算法进行改进。其次, 借鉴半直接法 SVO 系统的思想, 采用半直接法完成双目特征匹配, 在保持其高计算效率的同时, 确定了空间三维点的精确位置, 使其对无人机运动定位更加鲁棒, 并编程实现。最后, 通过相应的实验验证, 充分证明了本系统在无人机平台上应用是可行的。

**关键词:** 四旋翼无人机; 同时定位与地图构建; 立体视觉; 半直接法; SVO

## Abstract

For the design and the implementation of a visual SLAM system which can be applied to the quadrotors, this paper studies and tests the existing open source visual SLAM system in-depth. The disadvantages of the existing typical systems are summed up and improved. For real-time purpose, a stereo vision SLAM system based on semi-direct method is designed and implemented based on the feature point method and the semi-direct method. In order to improve the robustness of the system, a binocular vision sensor is used to estimate the three-dimensional space pose estimation. The estimation of the real environment scale information can thus be realized by stereo matching algorithm. The validity of the algorithm is verified by experiments. The main contributions of this paper can be summarized as follows:

Firstly, the position model of quadrotor and the basic model of the visual system are built. By using probability distribution model, the relevant coordinate system is defined and the system-related model is established. The geometric principles of the visual SLAM system are analyzed using the camera model. The difficulties of the application of the visual SLAM system to quadrotors are summarized and the foundation for the system design and improvement is laid.

Secondly, the overall system framework is proposed. The actual requirements and the indexes of the system are determined by previous research and tests. According to the motion estimation method, the mainstream visual SLAM systems can be classified into the filter based method, the keyframe BA based method and the direct method. The corresponding existing advanced representative systems are tested according to the classification results. The advantages and disadvantages of these typical systems are compared. Taking real-time performance and robustness into account, the overall system framework is ultimately determined.

Finally, the stereo vision matching algorithm for quadrotor based on semi-direct method is proposed. Considering the large computation and poor robustness under strong rotation of the ORB-SLAM system based on the feature point method and non-estimation of real environment scale of the SVO system based on semi-direct method, this system takes both real-time performance and robustness into account. Referring to the idea of SVO system, the semi-direct feature is used for the stereo feature matching. The stereo vision SLAM system based on semi-direct method is realized with high computational efficiency and robustness. Finally, the feasibility of the proposed system is verified by corresponding experiments.

**Keywords:** Quadrotors; SLAM; Stereo Vision; Semi-direct Method; SVO

# 目 录

摘    要 .....	I
Abstract.....	II
第 1 章 绪论 .....	1
1.1 课题研究背景与意义 .....	1
1.2 国内外发展现状 .....	3
1.2.1 国内外无人机自主定位研究现状.....	3
1.2.2 国内外基于视觉的 SLAM 系统研究现状.....	5
1.3 论文主要内容及章节安排 .....	8
第 2 章 基于视觉的 SLAM 基本原理和方法 .....	10
2.1 引言 .....	10
2.2 SLAM 问题的一般定义 .....	10
2.2.1 SLAM 问题的数学描述.....	10
2.2.2 概率 SLAM 问题的描述.....	11
2.2.3 概率 SLAM 的结构.....	13
2.3 基于视觉的 SLAM 系统模型 .....	14
2.3.1 坐标系定义及转换关系.....	14
2.3.2 摄像机模型.....	16
2.4 无人机平台下视觉 SLAM 系统实际应用问题描述 .....	18
2.5 本章小结 .....	20
第 3 章 典型视觉 SLAM 系统的分析对比 .....	21
3.1 引言 .....	21
3.2 典型视觉 SLAM 系统分析 .....	21
3.2.1 基于滤波器的视觉 SLAM 系统.....	21
3.2.2 基于关键帧 BA 的视觉 SLAM 系统 .....	26
3.2.3 基于直接法的视觉 SLAM 系统.....	33
3.3 典型视觉 SLAM 系统性能比较 .....	36
3.3.1 实时性能.....	36
3.3.2 定位精度.....	36
3.3.3 特征缺失鲁棒性.....	37
3.4 本章小结 .....	37
第 4 章 基于半直接法的立体视觉 SLAM 系统设计 .....	38

4.1 引言 .....	38
4.2 基于直接法定位原理 .....	38
4.2.1 Lucas-Kanade 光流跟踪.....	39
4.2.2 基于直接法定位的工作原理.....	40
4.2.3 直接法分类.....	45
4.3 基于半直接法的立体视觉 SLAM 系统设计 .....	45
4.3.1 视觉 SLAM 系统方案选取.....	45
4.3.2 基于半直接的双目匹配算法设计.....	47
4.3.3 SLAM 系统总体方案.....	50
4.4 本章小结 .....	51
<b>第 5 章 实验设计与分析 .....</b>	<b>52</b>
5.1 引言 .....	52
5.2 实验平台搭建 .....	52
5.2.1 硬件系统搭建.....	52
5.2.2 软件系统搭建.....	54
5.2.3 系统定位精度评估指标.....	56
5.3 实验设计及结果 .....	56
5.3.1 EuRoC 无人机数据集.....	56
5.3.2 实验室室内场景图像序列.....	60
5.4 本章小结 .....	62
<b>结    论.....</b>	<b>63</b>
<b>参考文献 .....</b>	<b>65</b>
<b>哈尔滨工业大学学位论文原创性声明和使用权限 .....</b>	<b>70</b>
<b>致    谢 .....</b>	<b>71</b>

# 第 1 章 绪论

## 1.1 课题研究背景与意义

随着无人机(Unmanned Aerial Vehicle, UAV)概念的提出,在世界范围内掀起了无人机技术发展的新一波浪潮。无人机也称为无人驾驶飞行器,顾名思义是一种无人驾驶,可通过机载程序控制以及无线电远程遥控,由动力驱动自动行驶,同时不承载人的航空器<sup>[1]</sup>。自从 1914 年无人机的概念被首次提出,无人机以其体积小、造价低、使用方便、易于隐藏、便于操作、对环境要求低等优点,使该产业得到迅猛发展<sup>[2]</sup>。直到近几十年,无论是在军事领域还是在商业以及民用领域,无人机均备受关注并投入实际应用。尤其是在经过越南战争、中东战争、海湾战争、阿富汗战争等各大战争后,各个国家越来越意识到无人机技术应用的必要性,无人机技术研究得到高度重视。

自 20 世纪 90 年代以来,人们大大简化了无人机的机械机构,使其逐渐向着小型化方向发展。旋翼式无人机不同于直升机,通过尾桨这样复杂的机械结构解决自悬的问题,且具有机动性能好、使用方便、无人员伤亡风险、用途广泛等优势,得到国内外的广泛研究<sup>[3]</sup>。同时,随着军事、农业以及民用等各个应用领域对无人机的大量需求,使旋翼式无人机得到迅猛发展<sup>[4]</sup>,其中四旋翼无人机(Quadrotor Unmanned Aircraft Vehicle, QUAV)是旋翼式飞行器的主要研究对象。

四旋翼无人机的领域十分广泛,如:避障、跟踪、路径规划以及伤员搜索<sup>[5]</sup>等等。四旋翼无人机能够准确地、实时地对自身进行自主定位为实现这些任务奠定了基础。其中,自主定位将成为四旋翼无人机必备的基本功能。目前应用最为广泛的定位方法主要分为两大类:一类是基于机载摄像机估计自身位置和姿态信息,同时实时地重建运动场景,这是即时定位与地图构建<sup>[7]</sup>即 SLAM (Simultaneous Localization and Mapping);另一类是基于外部传感器提供精确的导航参数,如全球定位系统<sup>[10]</sup> (Global Position System, GPS) 运动捕捉系统<sup>[12]</sup>。四旋翼无人机能够垂直起飞和着陆,还可以定点悬停和低速飞行。这些特性使其成为在建筑物内或者其他固定空间内使用的理想选择。然而,无人机在飞行过程中,实现室内自主飞行是特别具有挑战性的。关键问题是在建筑物内或者一些室内固定空间无法接收到有效的 GPS 信号,导致无人机无法确定自身的位置信息。同样,运动捕捉系统只能在特定的场合中使用。因此,室内导航需要其他测量信息,需要融合使用额外的传感器<sup>[13]</sup>。将两类方法进行比较,采用 SLAM 系统获取精确的无人机位置的方法普适性更强。但现如今 SLAM 系统多数应用于在搭载传感器的无人机



上, SLAM 系统同时包含定位与建图两个问题, 被认为是实现无人机自主性的关键问题之一, 对无人机领域的发展有着重要的研究意义<sup>[14]</sup>。但是, 至今仍未出现简单易用的, 具有实际应用价值的 SLAM 系统<sup>[15]</sup>。目前该问题仍处于研究室的开发阶段。为了解决无人机的自主定位问题, SLAM 系统可使用多种类型的传感器, 还可以将两种或多种传感器的数据融合使用, 如: 激光扫描仪<sup>[16]</sup>, 惯性元器件、单目摄像机<sup>[19]</sup>, 双目摄像机<sup>[21]</sup>以及 RGB-D 传感器<sup>[23]</sup>等。使用视觉传感器的 SLAM 系统称为视觉 SLAM 系统。

随着四旋翼无人机自主定位问题解决方案的推广, UAV 可以携带的有效载荷相当低, 可用的机载电源也是如此, 这大大限制了这种自主无人机可以配备传感器的选择。同时, 能够进行精确 3D 定位的传感器(例如 LIDARS)通常太重以至于四旋翼无人机无法携带, 并且功耗很高。这种情况引发了使用摄像机进行自主无人机定位的重要意义, 摄像机不仅具有质量轻, 功耗低, 尺寸小以及成本低等特点, 在此基础上还可以获得丰富的信息量, 同时, 摄像机只需要很少的能量消耗, 可以非常容易得刚性连接在无人机机体上, 同时, 相机可用于未知环境中的三维定位以及地图构建。因此, 相比传统的定位方式, 使用摄像机传感器的定位技术是实现无人机自主飞行的必然趋势。

应用视觉传感器于四旋翼无人机自主定位系统, 需要考虑实时性, 鲁棒性, 适应性能等。由于摄像机提供的信息量过于丰富存在因处理数据的计算量大, 当应用于四旋翼无人机、智能移动等设备, 使用其机载处理器进行数据, 但其处理能力较差, 导致应用于实际系统存在实时性较差的问题, 因此, 达到在线实时处理效率是主要问题之一。同时, 基于视觉系统进行定位需对视觉图像进行处理, 图像处理需要满足一定的假设条件才可以用于后续操作, 在四旋翼出现快速运动或发生大机动的情况下, 导致图像模糊, 跟踪频繁丢失或出现明显漂移, 导致定位系统不可用。对于实际应用中的无人机自主定位应用, 无人机随意运动, 导致上述情况频繁发生。

综上所述, 我们的工作目的主要是在低成本的四旋翼无人机硬件上, 设计并实现一个鲁棒地、可确定环境尺度的基于机载立体视觉的四旋翼定位系统。完成四旋翼无人机在复杂环境中通过自身处理器处理, 可实时获取位姿信息, 在其出现大机动或快速运动的情况下可以高效而鲁棒的实现定位功能, 这将为无人机智能自主导航技术奠定坚实的基础, 具有重要的理论研究意义及应用参考价值。

## 1.2 国内外发展现状

### 1.2.1 国内外无人机自主定位研究现状

将无人机投入实际应用，实现一些复杂任务的基础和前提就是对其准确地、实时地自主定位。而实现自主定位涉及的领域十分广泛，该技术主要将移动机器人、地面车辆、飞行器等作为研究对象。早在 1950 至 1960 年间无人机的自主定位领域得到关注，致力于解决该问题，国内外各领域的研究人员结合其专业知识提出许多解决方案。但是，针对无人机自主定位问题，实现准确实时地定位依旧是一个难点，虽然已经有许多种定位方案，但大多数方案在实际应用中很难得到精确的位姿信息，而环境的复杂性、状态的不确定性等都为无人机的自主定位增加了难度。下面将国内外相关文献作简要的概述。

随着无人机的快速发展，无人机自主定位技术作为自主飞行的基础研究内容，贯穿于无人机飞行技术的整个发展历程。在无人机的研发历史中，随着无人机于二十世纪初的革新到现在，已有四代无人机自主飞行技术。无人机自主飞行的第一个阶段，开始于第一次世界大战，这时的无人机作为训练用的靶机，自主飞行能力十分差。1916 年无人机自主飞行开始了第二阶段，在这个阶段研发的无人机相对第一阶段能为自身提供动力，多数采用遥控的方式操作。第三个阶段是随着 GPS 的普及而开始，此时定位系统得以在无人机上应用，定位系统的出现大大提高了其自主飞行的能力，RQ-4A“全球鹰”<sup>[25]</sup>和 A160T“蜂鸟”<sup>[26]</sup>等均是这一阶段的代表飞行器。在近几年的迅猛发展过程中，双目立体传感器、RGB-D 相机、TOF 深度传感器、激光雷达等新型传感器的普及，这些传感器对环境有了一定的自主感知能力，无人机结合一种或多种新型传感器实现自主定位，迎来了该技术的第四个阶段。

直到现在，无人机自主定位技术还处于第四阶段，且还处于研究室的开发阶段，在实际应用中存在诸多局限，尚未投入实际应用，但是这在四旋翼无人机的发展史上却有举足轻重的意义。在国外多国高校均展开了新一代自主飞行无人机的开发和研制工作，如：美国的耶鲁大学<sup>[27]</sup>、麻省理工学院<sup>[28]</sup>、宾夕法尼亚大学<sup>[29]</sup>、瑞士的苏黎世联邦理工学院<sup>[30]</sup>等，其研究方向大致包括新型传感器的开发，避障、多种传感器数据融合，交互以及轨迹规划等。

接下来，我们对国内外各无人机自主飞行研究团队的定位方法、采用的传感器以及研究成果进行以下总结。早在 1991 年，卡耐基梅隆大学设计实现 Yamaha R50 无人机，这款无人机创新之处在于使用双目视觉传感器，完成机体自身位姿估计，并通过目标跟踪实现了视觉里程计(VO)算法。从而初步实现了无人机应用立

体视觉完成定位导航等任研究。图 1-1 即为 Yamaha R50 无人机。



图 1-1 Yamaha R50 无人机

麻省理工学院设计的四旋翼无人机 Huang A S<sup>[31]</sup>是一款自主性非常高的四轴飞行器，该无人机的感知系统是由 RGB-D 相机和惯性元件组成，同时使用这两种传感器的数据，实现了在 GPS 不可用的环境中定位系统的设计。图 1-2 即为麻省理工的四旋翼无人机。



图 1-2 麻省理工的四旋翼无人机

国外的商业领域也产生了很大的影响。在 2015 年 3D Robotics 公司推出的 Solo 无人机可自主飞行，非常智能，同时 3DR 的开源特性深受用户的喜爱。图 1-3 所示为 3D Robotics 公司的 Solo 无人机。



图 1-3 3D Robotics 公司的 Solo 无人机

无人机自主定位的研究在国内起步比较晚，对四旋翼无人机的定位问题研究还处于初级阶段，从事这方面研究的主要是各科研机构、各个高校以及无人机相关的商业公司。在我国商业领域，大疆创新公司研发的部分无人机均利用机载视觉传感器获取的图像数据，以及惯性元器件提供的数据，在室内飞行时实现自主定位，例如：Phantom 3、Yang S等。大疆Phantom 3无人机如图1-4所示。国内的一些重点院校和科研院所也致力于对四旋翼无人机的研究，其中，中国科学院、上海交通大学、哈尔滨工业大学等高校致力于研究无人机定位问题，并取得一定成果。



图 1-4 大疆公司的 Phantom 3

目前国内针对固定环境、无人机平稳飞行的情况下定位算法的研究成果已经比较丰富，但是针对无人机出现大机动和快速运动等特殊情况下的无人机定位算法的研究，以及定位算法在实际应用时，其处理效率、定位精度以及运动场景鲁棒程度等性能，相对国外的研究成果还较为落后。

### 1.2.2 国内外基于视觉的 SLAM 系统研究现状

现如今，随着计算机视觉、图像处理水平的迅猛发展，许多研究学者将计算机视觉应用到移动终端、移动机器人、旋翼式无人机自主定位中。基于视觉的自主定位是一种在实际应用中结合视觉图像信息，实时地提供机器人位置和姿态信息的技术，它相比传统定位方式可针对未知环境进行自主定位，普适性更强，因此该技术获得了空前的关注。基于视觉的自主定位技术在实现自主避障、路径规划、跟踪控制等复杂任务时需要精确地估计机器人位姿，同时实时地重建场景，这即是同时定位与地图构建技术所需要解决的问题。

SLAM技术的提出是指自主移动的机器人在未知环境中运动，在行进过程中通过估计自身位置、姿态信息以及通过传感器观测得到的信息，对其运动的环境进

行地图构建，同时使用构建的地图推断其自身位姿信息，并在此基础上实现自主定位和导航。因此，SLAM技术被视为实现机器人真正自主定位的前提和基础。

概率SLAM问题起源于1986年在加利福尼亚州旧金山举行的IEEE机器人与自动化大会上。这次大会的成果是将概率方法引入机器人和与人工智能（AI）领域解决问题，同时认识到一致性的概率地图是机器人技术中的一个基本问题，需要解决主要的基本原理和计算问题。一些研究人员一直在研究将估计理论方法应用于建图以及定位问题中，主要包括Peter Cheeseman、Jim Crowley和Hud Durrant-Whyte等。在接下来的几年中，Smith、Cheesman和Durrant-Whyte研究得出描述路标点之间的相互关系和几何不确定性之间符合统计学原理<sup>[32]</sup>。这项工作表明了一个关键问题，地图中不同路标点位置的估计值之间必须存在很高的相关性，同时这些相关性将随着连续观察而提高。

在Smith等人<sup>[34]</sup>的里程碑式论文发表后不久，概率估计理论的SLAM系统就被提出了。Ayache和Faugeras最早投入视觉导航领域的研究工作<sup>[35]</sup>，与此同时，Crowley和Chatila和Laumond将卡尔曼滤波器应用在移动机器人导航中<sup>[37]</sup>。这两项研究有很多共同之处，提出的基于卡尔曼滤波的SLAM系统被广泛研究与应用，到目前为止，依旧是解决移动机器人定位问题的主流方法之一。目前最常见的基于概率SLAM问题的解决方案中主要使用FastSLAM算法、扩展卡尔曼滤波器（EKF）以及Rao-Blackwellized粒子滤波器来解决SLAM问题。近年来，以EKF-SLAM和FastSLAM是这两个最重要的解决方案为基础，研究者们提出了更多的替代方案。

在1995年国际机器人研究学术研讨会上提出的移动机器人测量论文中首先介绍了SLAM问题的结构，收敛结果以及专有名词缩写的SLAM。收敛的基本理论<sup>[38]</sup>和许多初步结果由Csorba开发。几个组织已经开始研究同时定位及建图联合问题，特别是马萨诸塞理工学院<sup>[40]</sup>，萨拉戈萨大学<sup>[41]</sup>，悉尼ACFR<sup>[43]</sup>等组织<sup>[45]</sup>最早开始致力于解决室内、室外、海底环境下的SLAM问题。此时，工作重点是提高计算效率，解决数据关联或闭合回路中的问题。

21世纪初，基于视觉的SLAM系统的研究得到迅猛发展，其中最具代表性的是英国伦敦帝国理工大学的Andrew J. Davison教授，2003年，Davison教授初次采用主动视觉匹配的方法实现对于一个自由运动的摄像头进行了六个自由度的位姿定位，并取得成功。该视觉系统的主要特点是实时性能较好同时对环境有很强的适应能力。同时，近年来概率SLAM系统的实现令人印象深刻，用于解决更具挑战性的无人机在大环境中的定位问题。2001年，Guivant和Nebot在非常大的户外环境中实现了SLAM的应用<sup>[47]</sup>。他们解决了实时操作的计算问题，同时可处理移动机器人在非平坦地形和动态混乱的场景中快速运动等问题。他们提供的维多利亚公园试

验的记录数据可在线获取，并已成为SLAM系统算法验证最受欢迎的基准。2002年，Newman等人的“探索与回归”实验<sup>[48]</sup>中实验场景是一个中等规模的室内环境，机器人在勘探阶段是手动驱动的，尽管没有操作人员的视觉接触，他们完全依赖机器人地图的实时渲染。对于回程，机器人无需人为干预进行路径规划并返回起点。

近年来，基于关键帧的SLAM系统以及基于直接法的SLAM系统被研究者广泛研究及应用。21世纪初直到现在许多最先进的SLAM系统都是采用基于关键帧及基于直接法的框架，比如，2008年，牛津大学的Georg Klein完成了PTAM系统，该SLAM系统采用一种新颖的基于关键帧的并行跟踪和地图创建框架，为了获得实时的性能，将跟踪从地图创建中独立出来；为了保证较高的精度，可以在关键帧之间做集束调整（BA）。ORB-SLAM是2014-2015年西班牙的Raúl Mur-Artal博士提出实现完成的，ORB-SLAM可以看做是PTAM的一个延伸，采用ORB特征来做跟踪、地图创建、重定位以及回路检测，同时通过优化姿态图来使环路闭合。2016年，基于直接法的SLAM系统中一个重量级的系统开源实现，DSO-SLAM（Direct Sparse Odometry）是出自SLAM的重镇慕尼黑工业大学Takob Engel，是当前基于视觉的SLAM系统的代表性算法之一。该方法将最小化光度误差模型和所有模型参数联合优化方法相结合。实验表明，在跟踪精度和鲁棒性方面，所提出的方法显著优于其他的直接法和间接法。

随着国内外研究的开展，关于视觉SLAM系统的研究实现也越来越多。但是在机器人研究领域，受研究基础薄弱等条件限制，无人机、移动机器人技术水平同国外一些工业发达的国家还有较大的差距。国内的SLAM研究主要集中在一些重点院校和科研院所，例如：清华大学、浙江大学、哈尔滨工业大学、上海交通大学以及中国科学院自动化研究所。过去国内的基于视觉的SLAM研究还停留在对国外开源的先进技术进行改进和优化的阶段，但是最近几年国内提出并实现了一些优秀典型的SLAM系统。比如：2013年，浙江大学CAD&CG国家重点实验室章国峰教授带队设计实现了RD-SLAM，该系统可以检测场景中的外观和结构变化，提出了一种改进后的RANSAC方法来实现动态场景中的稳定定位工作。在此基础上，该视觉组成员刘浩敏博士提出RK-SLAM，该SLAM系统是一种鲁棒的基于关键帧的单目摄像机实时跟踪的方法，该方法能够可靠地处理强烈旋转和快速运动。

这些最新的研究成果都推动了基于视觉的SLAM研究的发展，该领域的进步是机器人界在过去十年中取得的显著成就之一。SLAM作为一个理论问题已经以许多不同的形式被定义和解决。同时，SLAM也在室内机器人到室外机器人，水下潜艇和无人机等不同领域得到实现。基于视觉的SLAM系统以其拥有视觉传感器的优

势，得到迅猛发展并能到广泛应用。

### 1.3 论文主要内容及章节安排

本文将深入研究并以自主开发一套能够适用于四旋翼无人机室内自主定位的 SLAM 系统为目标，解决在无人机机载处理器上很难实时估计自身位姿的问题。本系统设计一种配置灵活、计算量小、性能稳定的四旋翼无人机室内定位系统并编程实现。首先，基于概率分布模型对 SLAM 系统进行数学描述，建立四旋翼无人机的位姿模型以及视觉系统的基本模型。其次，通过对视觉 SLAM 系统实际应用于无人机存在的问题描述，确定系统需求及指标要求，对部分已有主流开源 SLAM 系统进行测试分析，确定这些 SLAM 系统以及采运动估计方法的优缺点。最后，根据不同运动估计方法应用于无人机平台的定位效果，将提出一种将半直接法应用于双目视觉系统的方案，并设计基于半直接法的双目匹配算法，以满足无人机平台运行的实时性要求。论文具体安排如下：

第 1 章论述基于视觉的无人机自主定位技术的研究背景与意义。然后，分别从无人机自主定位以及基于视觉 SLAM 系统两个方面总结分析国内外的研究现状与不足，给出本文的主要研究内容。

第 2 章首先建立基于视觉的 SLAM 系统的数学模型、物理模型，给出基于概率分布模型对 SLAM 系统的问题描述。然后，结合视觉 SLAM 系统的基本原理，确定四旋翼飞行器的位姿模型以及视觉系统的基本模型，在此基础上，建立视觉系统的相关坐标系。最后，为便于确定系统的整体方案，根据配置不同数量的视觉传感器，总结归纳视觉 SLAM 系统的几何原理以及数学模型，为后文算法分析及改进奠定基础。

第 3 章首先针对视觉 SLAM 系统实际应用于无人机平台存在的问题进行描述。同时，结合实际需求给出本系统的指标要求。在此基础上，分类研究近年来视觉 SLAM 领域的主流方法，并对已有代表性地视觉 SLAM 系统进行测试分析。最后，结合本课题的系统需求，通过实验验证对比分析这些典型系统的优缺点，确定本课题的系统方案及框架。

第 4 章提出一种基于半直接法的双目视觉匹配算法。首先，介绍 LK 光流追踪、基于直接法运动估计的基本原理以及直接法的分类，针对典型的基于特征点法的 ORB-SLAM 系统实际应用于无人机存在的问题，确定系统采用半直接法完成运动估计。然后，针对半直接法的 SVO 系统存在的问题，进行算法改进，采用半直接法完成双目特征匹配，完成整个系统及方法设计，并基于 Linux 操作系统以及 ROS 机器人操作系统编程实现。

第 5 章完成系统的实验验证。首先，提出一套视觉 SLAM 系统的测试实验方案。确定实验方案中硬件系统的选型，以及软件环境的搭建。使用的大疆无人机 M100 的视觉传感器提供输入数据，采用 ROS 机器人操作系统对其标定。同时，对图像进行畸变处理并搭建地图绘制环境。最后，分别采用 EuRoC 无人机公共图像数据集和实验室室内无人机实际飞行实验数据，对基于半直接的双目视觉 SLAM 系统进行测试。根据实验结果，从系统计算效率以及定位鲁棒性两各方面验证该系统的有效性。



## 第 2 章 基于视觉的 SLAM 基本原理和方法

### 2.1 引言

在认识到基于视觉 SLAM 系统的普适性和高效性，研究者发现基于视觉的 SLAM 系统都是基于概率估计理论对四旋翼无人机的定位问题进行数学描述。本章将从基本的概率估计理论出发，给出基于视觉 SLAM 系统的数学模型及问题描述。同时，搭建系统的基本模型并定义系统相关坐标系，主要包括四旋翼飞行器的位姿模型、视觉系统模型、世界坐标系、图像坐标系、机体坐标系等。此外，本章还将进一步分析视觉 SLAM 系统的基本原理，为后文中视觉 SLAM 系统的研究和改进提供必要的理论基础。同时，针对视觉 SLAM 系统实际应用于无人机平台存在的问题，从实时性和鲁棒性等两个方面进行描述，并结合实际需求给出本视觉 SLAM 系统的指标要求。

### 2.2 SLAM 问题的一般定义

#### 2.2.1 SLAM 问题的数学描述

移动机器人 SLAM 是指自主移动的机器人在未知环境中运动，在行进过程中通过估计自身位置、姿态信息进行预测，以及通过传感器观测得到的信息进行更新，对其运动的环境进行地图构建，同时使用构建的地图推断其自身位姿信息，并在其基础上实现自主定位和导航<sup>[49]</sup>。在移动机器人 SLAM 问题中，不需要任何先验的位置信息，即可实现在线估计机器人运行的轨迹以及构建地图中所有地标的位置信息。

现在考虑移动机器人在包含丰富路标信息的场景中移动，通过使用刚性连接在移动机器人机体上的传感器获取环境中多个未知地标的相对观测信息，如图 2-1 所示。

在  $k$  时刻，SLAM 问题中相关变量定义如下：

$x_k$ ：表示移动机器人位姿的状态向量， $x_k = (x_k, y_k, \theta_k)^T$ ，其中  $x, y$  是位置坐标， $\theta$  是姿态信息。

$u_k$ ：表示在  $k-1$  时刻到  $k$  时刻对机器人的运动控制向量，使该机器人在  $k$  时刻被驱动到  $x_k$  状态。

$m_i$ ：表示运动环境的地图中第  $i$  个路标特征点的位置坐标， $m_i = (m_{i,x}, m_{i,y})^T$ 。假设地图中路标点的真实位置是不随时间改变的。

$z_{k,i}$ : 表示在  $k$  时刻, 移动机器人对第  $i$  个路标特征点的位置观测量。若任何一个时刻可观测到多个路标特征点, 或者具体的路标点与讨论无关时, 则观测量可简化表示为  $z_k$ 。所有时刻的观测量表示为  $z_{0:k} = (z_0, z_1, \dots, z_k)$ 。

此外, 还定义了以下集合:

$X_{0:k} = (x_0, x_1, \dots, x_k) = (X_{0:k-1}, x_k)$ : 表示机器人在整个运动过程中的位置, 其中初始位置  $x_0$  是已知的, 其他时刻位置未知。

$m = (m_1, m_2, \dots, m_M)$ : 表示路标特征地图。其中,  $M$  为地图中观测路标特征数。

$U_{0:k} = (u_0, u_1, \dots, u_k) = (U_{0:k-1}, u_k)$ : 表示整个控制过程的输入信息。

$Z_{0:k} = (z_0, z_1, \dots, z_k) = (Z_{0:k-1}, z_k)$ : 所有路标特征点观测值的集合。

根据移动机器人 SLAM 问题描述, 结合问题中上述相关变量定义, SLAM 问题的数学描述就是根据已知变量  $Z_{0:k}$ 、 $U_{0:k}$ 、 $x_0$  输出变量  $m$  和  $X_{0:k}$ 。

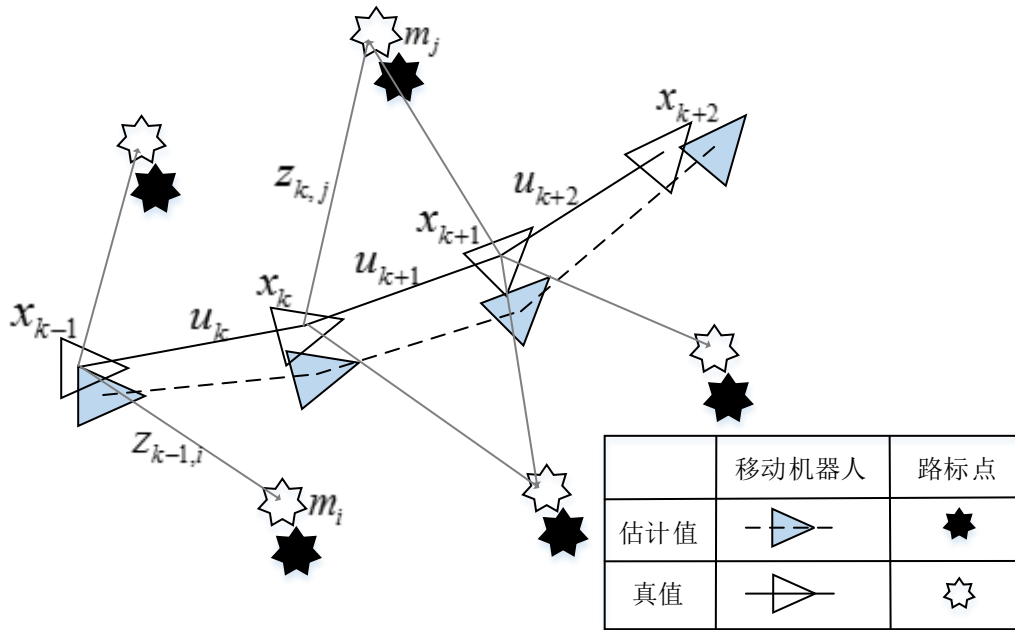


图 2-1 基本 SLAM 问题的数学描述图

### 2.2.2 概率 SLAM 问题的描述

由于环境和噪声的不确定性以及动态运动系统的复杂性, 主要采用概率分布模型来描述 SLAM 问题<sup>[49]</sup>。从概率学的角度出发, 需要计算所有时间  $k$  的概率分布:

$$P(x_k, m | Z_{0:k}, U_{0:k}, x_0) \quad (2-1)$$

如式(2-1)的概率分布描述了给定路标特征点的观测值和控制输入, 直到并包

括  $k$  时刻的路标特征点位置和机器人位姿状态（ $k$  时刻）的联合后验概率密度以及机器人的初始位姿状态。一般来说，对于 SLAM 问题采用递归解决方案是可取的。从  $k-1$  时刻的概率分布  $P(x_{k-1}, m | Z_{0:k-1}, U_{0:k-1})$  估计开始，使用贝叶斯定理计算控制输入  $u_k$  和观测值  $z_k$  之后的联合后验。该计算要求分别定义描述影响控制输入和观测值的状态转换模型和观测模型。

观测模型是指在移动机器人位姿状态以及路标特征点位置均已知的情况下，对观测值  $z_k$  的概率进行描述，通常描述的形式如下：

$$P(z_k | x_k, m) \quad (2-2)$$

我们可以合理地假设，一旦定义了机器人的位姿和环境路标特征地图，给定环境地图和当前机器人位姿的观测值是条件独立的。

运动模型是指可以根据状态转换概率分布的形式对机器人的运动进行描述，通常描述的形式如下：

$$P(x_k | x_{k-1}, u_k) \quad (2-3)$$

也就是说，状态转换过程被假定为马尔可夫过程，其中下一状态  $x_k$  仅仅取决于前一个时刻的状态  $x_{k-1}$  和当前时刻所施加的运动控制向量  $u_k$ ，并且独立于观测值和环境地图。

现如今，SLAM 算法以标准的两步递归的方式实现：预测过程（即：时间更新过程）、校正过程（即：测量更新过程）。时间更新过程如下：

$$P(x_k, m | Z_{0:k-1}, U_{0:k}, x_0) = \int P(x_k | x_{k-1}, u_k) \times P(x_{k-1}, m | Z_{0:k-1}, U_{0:k-1}, x_0) dx_{k-1} \quad (2-4)$$

测量更新过程如下：

$$P(x_k, m | Z_{0:k}, U_{0:k}, x_0) = \frac{P(z_k | x_k, m) P(x_k, m | Z_{0:k-1}, U_{0:k}, x_0)}{P(z_k | Z_{0:k-1}, U_{0:k})} \quad (2-5)$$

等式(2-4)和(2-5)提供了一种基于初始时刻直到并包括  $k$  时刻的所有观测值  $Z_{0:k}$  和所有控制输入  $U_{0:k}$ ，用于计算  $k$  时刻机器人的状态  $x_k$  以及环境地图的联合后验概率  $P(x_k, m | Z_{0:k}, U_{0:k}, x_0)$  的递归过程。该递归过程是关于机器人运动模型  $P(x_k | x_{k-1}, u_k)$  及其观测模型  $P(z_k | x_k, m)$  的函数。

值得注意的是，地图构建问题可以表示为计算条件密度  $P(m | X_{0:k}, Z_{0:k-1}, U_{0:k})$  的问题。假设在已知初始位置的情况下，机器人的位姿状态  $x_k$  始终是已知的（或者至少是确定的）。然后，通过融合来自不同位置的观测数据来构建地图  $m$ 。相反，定位问题可以表示为计算概率分布  $P(x_k | Z_{0:k}, U_{0:k}, m)$  的问题。假定路标特征点的位置是已知的，其目标是为计算机器人位姿相对于这些路标特征点的估计值。

### 2.2.3 概率 SLAM 的结构

为了简化本节的讨论，如上下文许可，我们将去掉公式中历史变量的预处理，并将环境地图和机器人位姿上所需的联合后验概率记为  $P(x_k, m | z_k)$ ，甚至可以记为  $P(x_k, m)$ 。

观测模型  $P(z_k | x_k, m)$  使机器人以及路标位置观测的依赖性更加明确了。因此，联合后验概率不能以明显的方式被分解

$$p(x_k, m | z_k) \neq p(x_k | z_k) p(m | z_k) \quad (2-6)$$

实际上，从关于一致地图的早期论文<sup>错误!未找到引用源。</sup>中已经被人熟知，这样的分解导致不一致的估计。

再次参考图 2-1 可以看出，路标点位置的估计值和真值之间存在的大部分误差在路标点之间是共同的，实际上是来源于同一个误差，在观测路标点时机器人的位姿信息存在误差。相反，这意味着路标点位置估计之间中的误差是高度相关的。实际上，这意味着即使在路标点  $m_i$  绝对位置相当不确定的情况下，也可以高精度地得到任意两个路标点  $m_i - m_j$  之间的相对位置。从概率的角度，这意味着即使当边缘密度  $P(m_i)$  相当分散时，一对路标点  $P(m_i, m_j)$  的联合概率密度也能达到最高点。

SLAM 中最重要的见解是随着越来越多的观察结果，意识到路标点估计之间的相关性单调增加。这些结果仅仅在线性高斯的情况下被证明成立[14]。对于在一般概率情况下的证明仍然是一个待解决的难题。

再次参考图 2-1，考虑机器人在位置  $x_k$  处观察两个路标点  $m_i$  和  $m_j$ ，很显然观测到路标点的相对位置与机器人坐标系无关，并且来自固定位置的连续观测路标点之间的相对关系将更加独立。现在，当机器人移动到  $x_{k+1}$  位置时，它再次观测到路标点  $m_i$ ，这允许机器人以及路标点估计的位置相对于先前位置  $x_k$  得到更新。反过来，尽管机器人在新的位置观测不到这个路标点，但这个过程会继续更新路标点  $m_i$ 。主要因为两个路标点之间的相对关系与之前的测量高度相关（它们的相对位置是固定且已知的）。此外，使用相同的观测数据来更新这两个路标点使得它们更加相关。每一项几乎独立观测，因为通过连续的机器人运动观测误差将相互关联。从图 2-1 中还可以看出，在位置  $x_{k+1}$  机器人除了观测到路标点  $m_j$  还可观测到其他两个新的路标点。因此，这些新的路标点立即与地图的其余部分相关联或相关，这些路标点的后续更新也将更新路标点  $m_j$ ，以及路标点  $m_i$  等等。也就是说，

所有的路标点通过相对位置或相关性最终形成一个网络，其精度随着观测而提高。

上述过程可以被视为一个将所有路标点连接在一起的弹簧网络，邻近的观测就像弹簧系统发生位移一样，随着与其他路标点之间的距离减小，其附近的效果变好，并且取决于局部相关性。在理论极限中，机器人相对位姿精度等于给定地图可实现的定位精度。

## 2.3 基于视觉的 SLAM 系统模型

在视觉 SLAM 系统的设计与实现过程中，准确的系统建模是实现移动机器人避障、路径规划、导航等复杂任务的基础。同时，在视觉测量系统中，利用视觉传感器获取世界空间中图像信息，通过摄像机模型描述三维空间中任意空间点的位置与图像平面像素位置之间的对应关系，及摄像机的成像过程。摄像机模型一般选用最常用的针孔模型（pinhole model），该模型是一种线性的模型。通常，摄像机标定技术可获得相机参数、建立四种坐标系以及它们之间的相互关系来描述摄像机模型<sup>[52]</sup>。

### 2.3.1 坐标系定义及转换关系

(1) **像素坐标系** 假设某一幅图像以一个  $M \times N$  的二维数组的形式存储，图像中每一个像点被称为一个像素。通常在计算机内每个像素对应 0-255 范围的值称为像素点的灰度值。如图 2-2 所示，以一幅图像的左上角为图像原点  $O$ ，建立一个直角坐标系  $O-UV$ 。其中， $(u, v)$  表示像素坐标系下像素点的坐标，单位是像素，该  $O-UV$  坐标系即为图像像素坐标系。

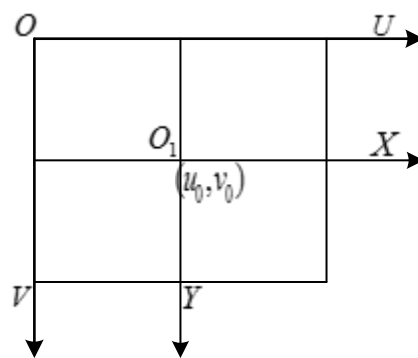


图 2-2 图像像素坐标系和成像平面坐标系

(2) **图像坐标系** 使用像素单位来描述图像二维信息的坐标系称为图像像素坐标系，而对应图像信息中的任意目标在空间中的物理位置是三维信息，通过毫米等物理单位来表示。于是，通过建立图像像平面坐标系来表述图像像素信息与对应的三维空间点之间的相互关系。如图 2-2 所示，一般选取摄像机光轴通过

投影与图像平面的交点为原点  $O_1$ ，建立直角坐标系  $O_1-XY$ ，坐标系中的  $X$  轴、 $Y$  轴分别于与图像像素坐标系中  $U$  轴、 $V$  轴相互平行，该直角坐标系  $O_1-XY$  称为像平面坐标系。同时，该坐标系的原点一般在图像的中心点位置，即主点。如果图像的主点  $O_1$  在  $O-UV$  坐标系中的坐标位置为  $(u_0, v_0)$ ，每个像素对应应在  $X$  和  $Y$  坐标轴上的物理尺寸为  $dx$  和  $dy$ ，则假设图像中的任意像素点在成像平面内的坐标为  $(x, y)$ ，于是得到像平面坐标系到图像像素坐标系的转换关系

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (2-7)$$

采用转换矩阵的形式可以表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-8)$$

上式变换后可得到

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} d_x & 0 & -u_0 d_x \\ 0 & d_y & -v_0 d_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2-9)$$

**(3) 摄像机坐标系** 在已知像平面以及像素坐标系的前提下，将摄像机的光心定为原点  $O_c$ ，建立直角坐标系  $O_c-X_cY_cZ_c$ 。如图 2-3 所示，直角坐标系  $O_c-X_cY_cZ_c$  中的  $X$  轴、 $Y$  轴分别于与图像坐标系中  $X$  轴、 $Y$  轴相互平行，同时沿着摄像机光轴的方向，以摄像机的光心为起点连接图像坐标系中的主点  $O_1$  设为  $Z$  轴， $O_cO_1$  的长度为摄像机的焦距  $f$ ，该轴同时垂直于图像像平面，交于该像平面坐标系的原点，即图像的主点，称  $O_c-X_cY_cZ_c$  为摄像机坐标系。

**(4) 世界坐标系** 任意时刻，摄像机在任意三维空间中的位置都是未知且不确定的，因此建立的摄像机坐标系参考的基准不同，获得的位置信息没有参考，不能直接应用，我们需要将工作中任意时刻建立的摄像机坐标系统一到同一个坐标系下。因此，通常在工作空间中设定一个世界坐标系，将其作为表示摄像机在三维空间中位置信息的基准。如图 2-3 所示，直角坐标系  $O_w-X_wY_wZ_w$  即为世界坐

标系， $O_c - X_c Y_c Z_c$  即为摄像机坐标系，这两个坐标系之间的相对关系可通过齐次变换矩阵表示。

假设三维环境中存在任一个空间点  $P$ ，该点在  $O_w - X_w Y_w Z_w$  和  $O_c - X_c Y_c Z_c$  坐标系下的坐标分别为  $(X_w, Y_w, Z_w)$  和  $(X_c, Y_c, Z_c)$ ，其齐次坐标可分别表示为  $(X_w, Y_w, Z_w, 1)$  和  $(X_c, Y_c, Z_c, 1)$ 。于是，摄像机和世界坐标系之间的相对关系可用变换矩阵的形式表示如下

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_3 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2-10)$$

其中， $R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$ ， $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ ， $R$  是一个  $3 \times 3$  的正交单位矩阵； $t$  是一个  $3 \times 1$  的平

移向量； $O^T = (0, 0, 0)$ ；同时，令  $M_3 = \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix}$ 。

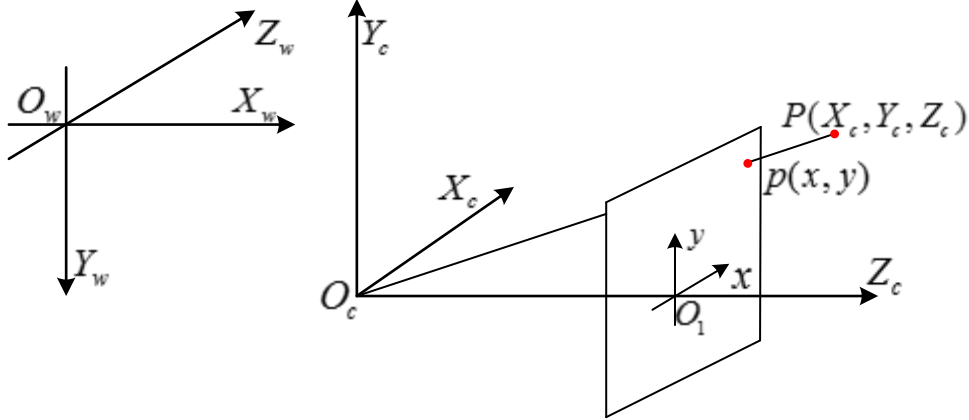


图 2-3 成像平面坐标系、摄像机坐标系与世界坐标系

### 2.3.2 摄像机模型

摄像机成像模型种类很多，通常使用针孔模型，该模型决定三维空间点的物理位置与该点在图像中对应的像素坐标及像素点的灰度值之间的关系，同时将该模型的参数称为摄像机的参数，分为摄像机内参和摄像机外参<sup>[53]</sup>。

假设三维环境中存在任一个空间点  $P$ ， $O_c$  为摄像机的光心， $O_c$  与该点  $P$  的连线交于图像平面于点  $p$ ，该交点  $p$  即为空间点  $P$  在图像中的投影位置，在图像像

平面坐标系下坐标记作  $(x, y)$ ，空间点  $P$  相对于图像像平面坐标系的位姿关系可表示为

$$\begin{cases} x = \frac{fx_c}{z_c} \\ y = \frac{fy_c}{z_c} \end{cases} \quad (2-11)$$

其中， $(x, y)$  为空间点  $P$  的图像像平面坐标； $(x_c, y_c, z_c)$  为空间点  $P$  在摄像机坐标系下的坐标。上述投影关系即针孔模型的透视投影关系，该过程采用转换矩阵以及齐次坐标的形式可描述为

$$z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (2-12)$$

将(2-8)代入(2-12)中可以得到

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} K & O^T \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (2-13)$$

其中，矩阵  $K$  即为摄像机的内参矩阵，可通过摄像机标定技术获得，且摄像机内参在摄像机制造生产后固定不变， $O^T = (0, 0, 0)^T$ 。实际上，摄像机内参矩阵  $K$  的表示形式如下

$$K = \begin{bmatrix} \frac{f}{dx} & \gamma & u_0 \\ 0 & \frac{f}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-14)$$

其中， $\gamma$  参数代表摄像机坐标系下  $X$  轴和  $Y$  轴之间的夹角余弦值。

综上所述，整个视觉系统相关的四个坐标系及其之间的相对关系，可表示为



$$\begin{aligned}
 z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ O^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} K & O^T \end{bmatrix}_{3 \times 4} \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ O^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 M_0 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}
 \end{aligned} \tag{2-15}$$

其中,  $M$  即为  $3 \times 4$  的摄像机投影矩阵;  $M_1$  是图像坐标系与像素坐标系之间的齐次变换矩阵, 只与摄像机内参相关, 可通过相机标定技术确定;  $M_2$  是世界坐标系与摄像机坐标系之间的齐次变换矩阵, 只与两者之间的旋转矩阵  $R$  和平移向量  $t$  有关, 可通过视觉 SLAM 系统获得, 求解出的  $R, t$  即为定位系统所要求解的相机姿态。

## 2.4 无人机平台下视觉 SLAM 系统实际应用问题描述

室内未知环境中实现自主定位是四轴飞行器室内自主飞行的关键技术之一, 也是最近几十年来机器人领域研究的热门问题<sup>[51]</sup>。随着研究的深入, 基于视觉的定位技术成为目前应用最为广泛的定位方法之一。该技术主要是将四旋翼无人机放置在室内未知环境中的任意位置随意飞行, 通过机载摄像机获取环境图像信息, 逐步建立该环境的三维地图, 同时确定其自身在该地图内的位置, 这即是视觉 SLAM 技术所要解决的问题。视觉 SLAM 技术的大致示意图如图 2-4 所示。

基于视觉的摄像机跟踪定位方法具有高效、信息丰富、可扩展等特点。结合本课题的应用背景, 由于视觉传感器提供的信息量过于丰富处理起来计算量大, 同时, 无人机硬件条件存在诸多限制, 其可以携带的有效载荷相当低, 且机载处理器的处理能力相对较差, 导致视觉 SLAM 技术应用于无人机等实际系统时实时性较差。因此, 解决在线实时姿态估计这一问题成为首要目标。

在无人机室内自主定位的多数应用中, 由于其飞行环境是未知的, 且室内真实环境复杂多样, 同时四旋翼无人机具有不同的运动速度以及运动模式 (如悬停等), 导致基于视觉的定位方法在某些复杂场景中鲁棒性较差, 尤其在强烈旋转和快速运动等情况下, 跟踪频繁丢失或出现明显漂移。在无人机基于视觉的自主定位技术投入实际应用的过程中, 实时性、鲁棒性等问题兹待解决。

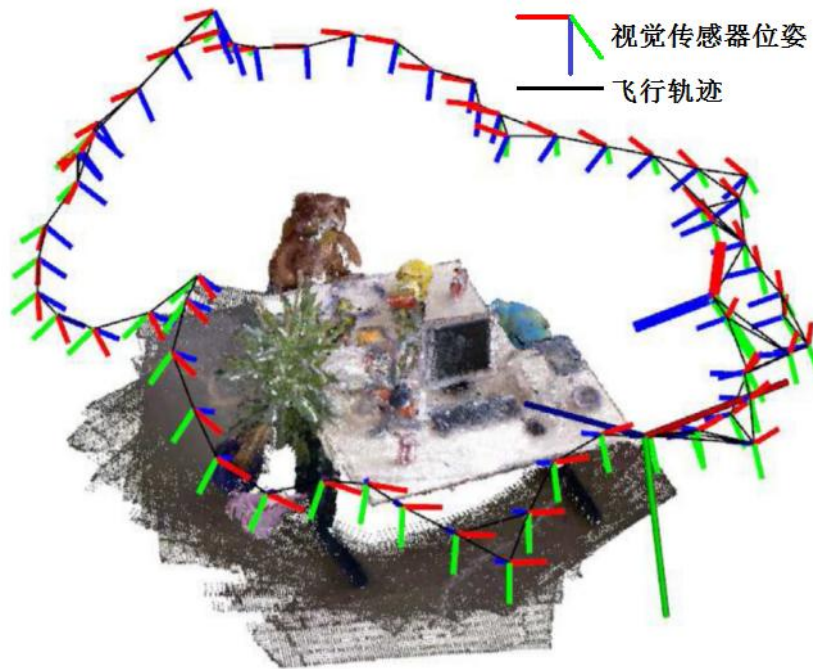


图 2-4 视觉 SLAM 技术的示意图

目前，无人机基于视觉的自主定位技术大多停留于低速水平，在实际应用中，无人机在大多数应用场合都是中高速（3m/s 以上）飞行，导致上述情况频繁发生。同时，无人机在室内环境中随意运动，经常会飞行到走廊、楼梯等纹理重复较为严重的场景中，或者相机获取的是白墙、玻璃等图像，无法根据这些图像提供的数据进行位姿估计。还会存在当无人机飞行过程中出现大机动、快速运动等特殊情况时，摄像机拍摄的图像模糊，图像匹配区域小，可匹配的特征点数量少等问题。这些情况的存在导致视觉 SLAM 系统在实际应用时，准确的位姿估计问题则变得十分尖锐。

这里我们从实时性和鲁棒性两个角度分别研究四旋翼无人机基于视觉自主定位的问题，具体描述如下：

**（1）实时性角度** 由于无人机机载处理器的处理能力相对较差，将视觉 SLAM 系统应用于四旋翼无人机平台计算效率相对较低。针对该问题，设计一个在满足一定精度需求的条件下，尽可能达到实时的计算效率成功跟踪该无人机并完成位姿结算的系统。这部分研究重点在于基于运动估计定位方法的确定以及视觉传感器配置的选取，要求无人机估计的位姿在达到一定精度的同时，解决在线实时姿态估计这一问题。

**（2）鲁棒性角度** 良好地自主飞行体验、精确地避障和路径规划等复杂任务的实现，需要视觉 SLAM 系统能够提供准确的位姿估计。在此基础上，还需要其满足处理无人机飞行过程中出现的复杂情况。并且，在遇到快速运动和严重的

运动模糊的情况下也要保证相机跟踪丢失的频率尽量小。甚至在遇到跟踪失败的情况下，应该要能很快地对相机进行重定位，从而避免长时间的跟踪失败导致系统不可用。

## 2.5 本章小结

本章针对 SLAM 系统复杂、不易理解的特性，详细的给出基于视觉 SLAM 系统的数学描述以及结合概率估计理论对 SLAM 系统的描述。同时，结合视觉传感器的优势，介绍了基于视觉的 SLAM 系统的基本原理。在此基础上，搭建了视觉 SLAM 系统中四旋翼无人机的位姿模型、运动环境的数学模型、视觉系统模型。然后，根据搭建的基本模型以及坐标系之间的变换关系，解算得到了世界坐标系下的四旋翼飞行器的位姿信息。最后，针对视觉 SLAM 系统实际应用于无人机平台存在的问题，从实时性和鲁棒性等两个方面进行描述。本章为后文视觉 SLAM 系统进一步的性能分析和改进奠定了理论基础。

## 第 3 章 典型视觉 SLAM 系统的分析对比

### 3.1 引言

针对第 1 章中本课题的应用背景，同时结合第 2 章中介绍的视觉 SLAM 问题的数学描述及其系统模型，本章将近几年来视觉 SLAM 领域的主流方法进行分类。依据运动估计的方法，将主流的视觉 SLAM 系统分为基于滤波器、基于关键帧 BA 以及基于直接法三类视觉 SLAM。并对部分已有主流开源视觉 SLAM 系统进行测试分析，从实时性、定位精度以及特征缺失鲁棒性等三个方面，通过实验对比分析这些典型系统的优缺点，最终确定本课题的整体系统方案以及结构框架。

### 3.2 典型视觉 SLAM 系统分析

基于视觉的 SLAM 系统的分类方法主要分为两种，第一种是依据视觉传感器配置分类，另一种依据运动估计的方法分类。依据视觉传感器配置分类，通常分为单目视觉 SLAM 系统、立体视觉 SLAM 系统、以及视觉传感器融合其他传感器数据的系统（目前应用最多的是融合 IMU 数据的 VI-SLAM 系统）；另外通常主流的视觉 SLAM 方法依据运动估计的方法分类，可以分为基于滤波器的方法、基于关键帧 BA 的方法以及基于直接法的视觉 SLAM。

针对 2.4 节中从实时性和适应性两个角度出发，分析四旋翼无人机的基于视觉自主定位的问题描述，为确定本课题的系统方案，我们将近年来现有开源主流的视觉 SLAM 系统根据三种运动估计的方法进行分类，并对其进行实际测试，根据测试结果分析这三类运动估计方法以及对比这些开源系统的优缺点。

#### 3.2.1 基于滤波器的视觉 SLAM 系统

视觉 SLAM 系统估计的轨迹均存在严重的漂移，所以在 SLAM 系统中，还会对帧间的相对运动进行处理，消除漂移。早期的 SLAM 后端均使用滤波器方法，经常使用的滤波器有扩展卡尔曼滤波（EKF）、粒子滤波（PF），FastSLAM 等。

基于滤波器的视觉 SLAM 系统基本思想就是：使用高斯概率模型来表示每个时刻  $t$  的系统状态，该高斯概率模型可以写为

$$x_t \sim N(\hat{x}_t, P_t) \quad (3-1)$$

其中， $\hat{x}_t$  为  $t$  时刻系统中被测对象的位姿的估计值， $x_t$  为  $t$  时刻系统更新得到系统状态， $P_t$  为  $\hat{x}_t$  与真值误差的协方差矩阵，使用滤波器不断更新  $x_t$ 。SLAM 系统最早的提出者 R.Smith 等人就把 SLAM 构建成了一个扩展卡尔曼（Extended Kalman

Filter, EKF) 问题。按照 EKF 的形式, 把 SLAM 通过一个运动方程和观测方程来表示预测 (Prediction) 阶段和更新 (Update) 阶段, 基于滤波器的视觉 SLAM 系统中运动方程, 可表示为

$$x_p^{i+1} = f(x_p^i, u_i) + w_i \quad x_p^i, i = 1, \dots, n \quad (3-2)$$

其中,  $x_p$  表示被测对象的位置, 假定由  $n$  个时刻位姿组成运动轨迹,  $w_i$  表示运动过程中的噪声, 假定该噪声服从高斯分布。观测方程如下式所示

$$z_{i,j} = h(x_p^i, x_L^j) + v_{i,j} \quad x_L^j, i = 1, \dots, m \quad (3-3)$$

其中,  $x_L$  表示运动环境中可观测到的路标点位置, 假定环境地图可观测到的  $m$  个路标点。  $v_{i,j}$  表示通过传感器对路标观测引进的噪声, 同样假定该噪声服从高斯分布。

该过程以最小化(3-2)和(3-3)这两个方程中的噪声项  $w_i, v_{i,j}$  为目的, 使用典型的滤波器思路来解决该问题。基于滤波器的视觉 SLAM 系统通过(3-2)和(3-3)这两个方程, 以最小化定位过程中引入的噪声为目的, 对被测对象位姿进行估计, 可以预测出被测对象在当前时刻的位置。同时, 根据以往标记的路标点, 又能计算出一个卡尔曼增益, 以补偿噪声的影响, 更新其均值和协方差。于是, 对当前帧和路标的估计, 即是这个预测和更新的不断迭代的过程。现有主流开源的基于滤波器的视觉 SLAM 系统有 MonoSLA<sup>[54]</sup>系统和 MSCKF<sup>[55]</sup>系统, 本节主要介绍使用扩展卡尔曼滤波的 MonoSLAM 系统。

## 1 MonoSLAM

A.J.Davison 教授是视觉 SLAM 研究领域的先驱, 2007 年, Davison 教授开发的 MonoSLAM 是首个利用一个摄像头做的 SLAM 系统<sup>[32]</sup>, 也叫做 Real-Time Structure From Motion。该系统被人们视为是许多研究的基础。MonoSLAM 的基本流程如下。

(1) **系统初始化** 首先对系统进行初始化操作, MonoSLAM 启动提供放置在相机前面已知目标形状的关于场景先验信息的帮助系统, 该系统提供已知位置和已知外观的几个特征点, 一般选取四个。由于该系统使用单个摄像机, 无法直接测量特征点深度或场景尺度信息, 提供初始化可确定初始目标尺寸, 准确估计地图和位姿精确的尺度。将初始化四个特征放置在地图中, 该帮助系统在启动时设置其不确定度为零。在跟踪第一帧图像时, 相机相对初始目标被保持在某个大致已知的位置, 同时初始相机位置被给予初始不确定性。图 3-1 显示了在第一帧图像中跟踪初始化目标的四个已知特征。

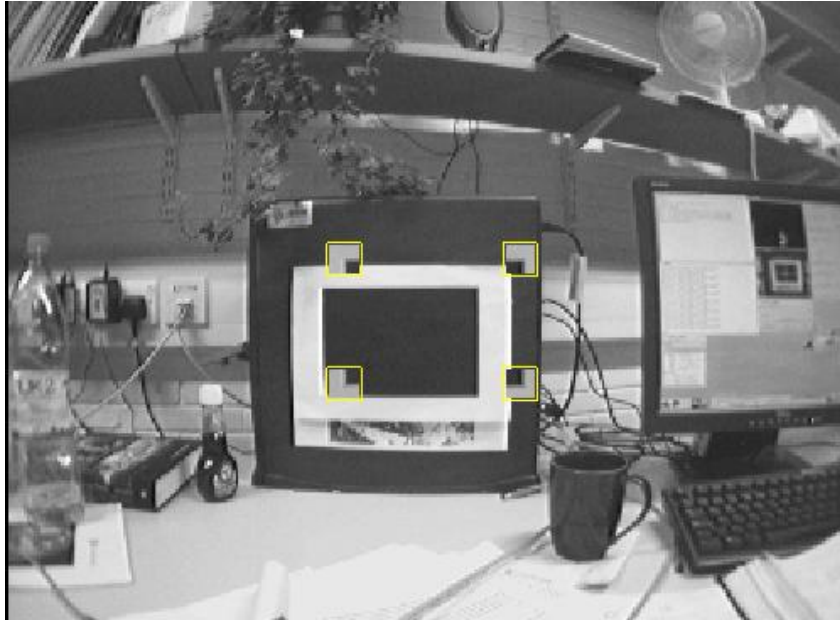


图 3-1 第一帧图像中跟踪初始化目标的四个已知特征

(2) **运动建模和预测更新** 帮助系统启动后，状态向量以两种交替的方式进行更新：(1)预测过程，当相机在图像捕获之间的“盲区”间移动时，通常采用运动模型进行预测；(2)观测更新，在实现了特征测量之后，进行更新。由于本系统采用的时 EKF 方法进行滤波，因此，该系统的预测更新模型与 EKF 的两个方程相对应。预测过程也是先验估计的过程，即已知  $t-1$  时刻的状态与  $t$  时刻的速度，采用匀速、匀角速度运动模型估计  $t$  时刻的状态。该运动模型并不意外着假设相机在所有时间内以恒定的速度移动，但是我们在时间步长中的运动统计模型是平均而言，使用服从高斯分布的不确定性加速度，相机运动模型如图 3-2 所示，且该运动方程可写为

$$C_t = f(C_{t-1}, v) \quad (3-4)$$

其中， $C_t$  表示  $t$  时刻运动估计到的相机运动状态参数， $C_{t-1}$  表示  $t-1$  时刻相机运动状态参数。并获取  $t$  时刻的观测数据，即后验估计过程。由于先验估计一般存在累计误差，不定性大，后验估计相对准确，但获取的时局部信息，为了减少偏差，采用投影方程对观测数据更新。该投影模型可写为

$$\hat{x}_j = h(C_t, X_j) + n_j \quad (3-5)$$

其中， $\hat{x}_j$  表示当前时刻摄像机观测的三维空间点对应图像坐标系中的位置， $X_j$  表示可观测的特征点对应的三维空间点的位置， $n_j \sim N(0, \Sigma_j)$ 。通过对比估计数据协方差和观测数据协方差，调节 EKF 增益完成状态更新。

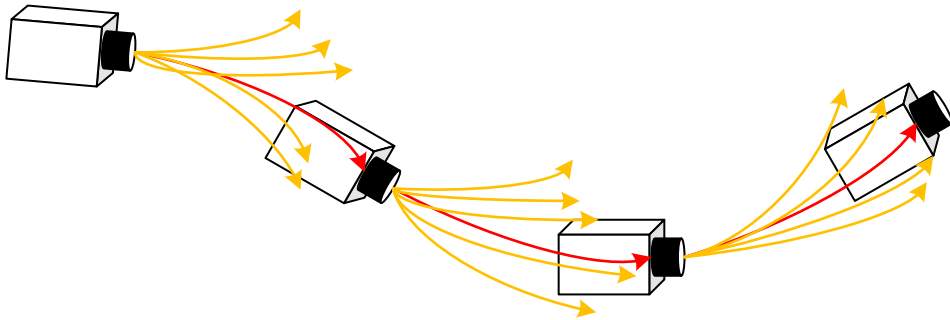


图 3-2 相机运动模型

(3) **特征提取与主动搜索** 为每帧灰度图像提取 Shi-Tomasi 角点，根据该检测算子自动检测显著的图像块，一旦提取作为长期的路标点使用，稀疏的路标点集不断增加并无限存储。使用主动追踪技术追踪其特征点，在 EKF 中假定其符合高斯分布，且每个时刻对应的相机位姿  $C_t$  和这些特征点所对应的三维点的位置  $X_j$  均带有一个概率偏差，如图 3-2 所示。

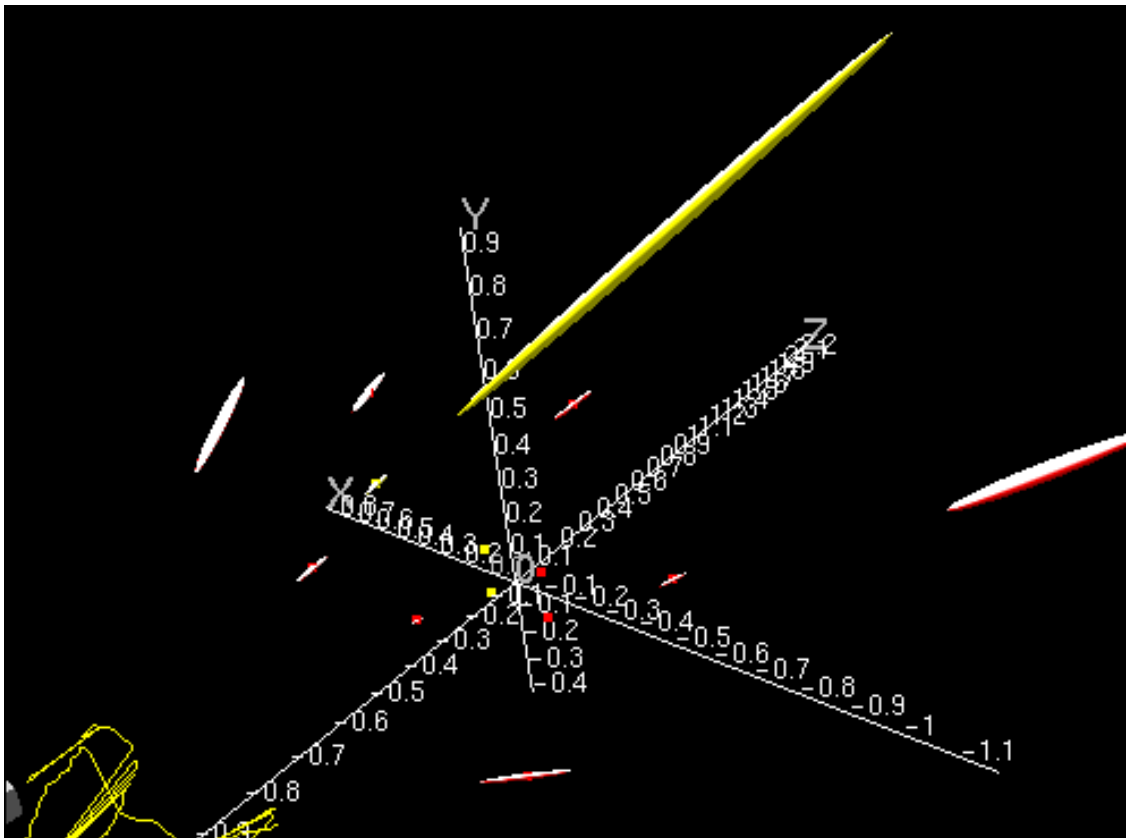


图 3-3 场景点的概率分布

概率偏差包括其均值及不确定性，MonoSLAM 使用一个三维椭球来表示，如图 3-3 所示。每个椭球以其中心为系统状态的估计值，体积代表不确定度。从图



中可以看到，空间中每个椭球的大小不等，在某个方向其长度越长，说明在该方向的位置就越不确定。通过 EKF 滤波更新协方差矩阵使对应特征点从椭球形状缩小至一个小球或点的状态，说明其特征点追踪成功并且处于收敛状态。

(4) **主动地图管理** 根据基于相机和特征的相对位置以及特征被初始化的相机的位置初始化地图，仅当相机观测的区域中可观测到的特征点书小于该阈值时，才将特征添加到地图中。同时在特征应该被观测到的预定次数的检测和匹配过程执行后，超过 50%均是失败的，则从地图中删除该特征。

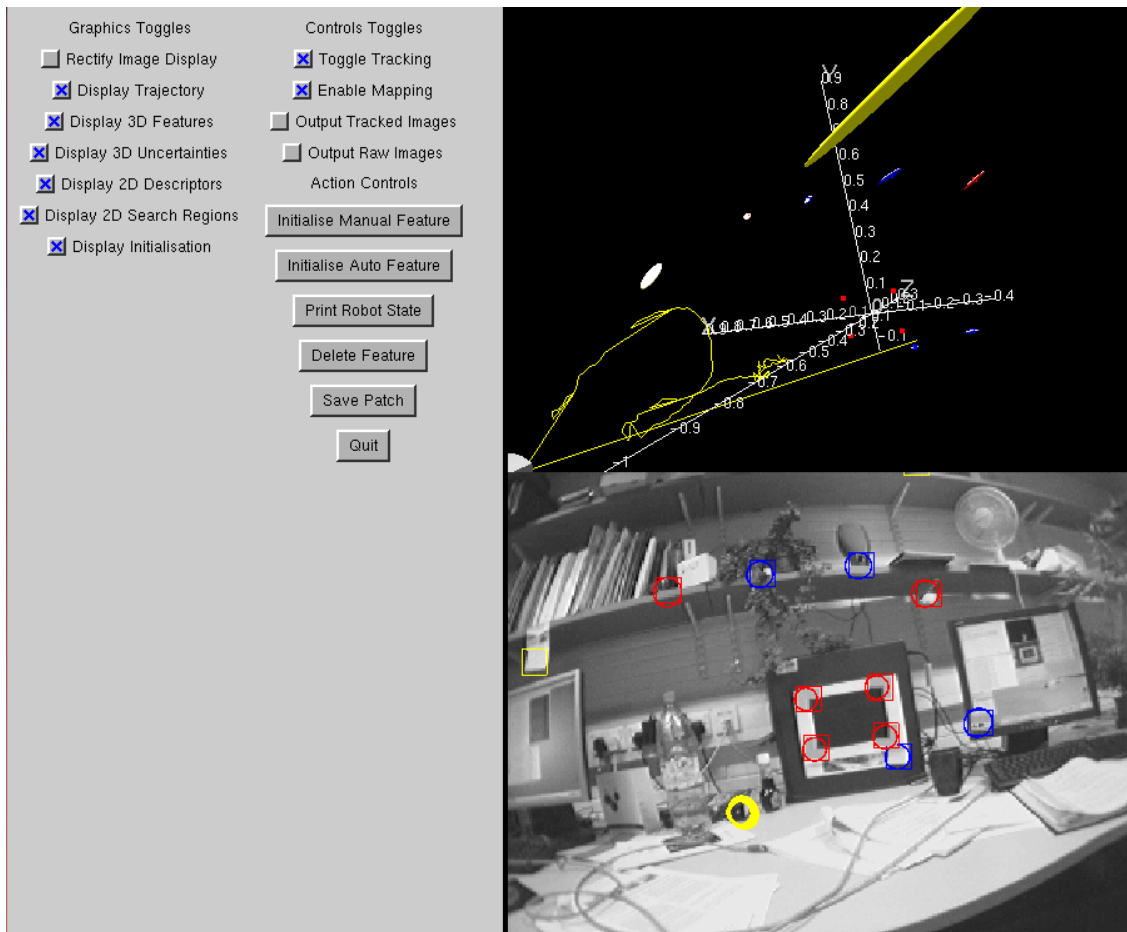


图 3-4 MonoSLAM 在运行过程中的状态

MonoSLAM 在运行过程中状态如图 3-4 所示。追踪特征点在图像中的表示如右下半部分所示。图中使用红色、蓝色、黄色分别表示匹配成功、匹配失败、没有参与使用的特征点。同时特征点对应的三维空间点的表示如右上半部分所示，其中空间点颜色表示与对应特征点形同。同时，图 3-4 中右上半部分黄色线表示视觉传感器的运行轨迹。



### 3.2.2 基于关键帧 BA 的视觉 SLAM 系统

21 世纪之后, SLAM 研究者开始借鉴 SfM (Structure from Motion) 问题中的方法, 把捆集优化 (Bundle Adjustment) 引入到 SLAM 中来。优化方法和滤波器方法有根本上的不同, 它并不是一个迭代的过程, 而是考虑过去所有帧中的信息。通过优化, 把误差平均分到每一次观测当中。在 SLAM 中的 BA 常常以图的形式给出, 亦称之为图优化方法 (Graph Optimization)。图优化可以直观地表示优化问题, 可利用稀疏代数进行快速的求解, 表达回环也十分的方便, 因而成为现今视觉 SLAM 中主流的优化方法。

近年来, 许多最先进的 SLAM 系统都采用基于关键帧的框架, 比如 DTAM[56]、ORB-SLAM<sup>[57]</sup>和 PTAM<sup>[58]</sup>。本节将主要介绍 PTAM 系统和 ORB-SLAM 系统。

#### 1 PTAM

2007 年, PTAM(Parallel Tracking and Mapping)<sup>[58]</sup>系统由 Klein 等人提出, 采用关键帧 BA 的方法完成运动估计, 是第一个基于该方法的视觉 SLAM 系统。不仅如此, PTAM 作者公开其源代码供研究者进一步开发与改进。2009 年该作者将 PTAM 系统移植到智能手机上<sup>[59]</sup>, 引领了后来许多视觉 SLAM 系统的设计。作为视觉 SLAM 领域的又一大突破, PTAM 的主要贡献体现在以下两个方面:

(1) **获得实时性能** PTAM 采用了一种新颖的基于关键帧的并行跟踪和地图创建的框架, 这就是现如今视觉 SLAM 系统中的前端和后端, 即系统设立前后端两个线程并行的处理位姿跟踪与地图构建这两个工作。将相机位姿跟踪从地图创建中独立出来, 相机位姿跟踪任务是通过实时处理图像信息求解相机位姿, 然而地图构建任务是为了优化地图更新矫正相机姿态, 不需要达到实时的计算效率, 就将地图构建任务放在系统后端进行处理, 只需要在必要的时刻进行线程同步即可。

(2) **提高跟踪精度** PTAM 提出关键帧的概念, 同时引入了非线性优化方式, 避免了后端使用基于滤波器优化方式无法实时处理大规模场景的情况。为了保证较高的精度可以在关键帧之间做集束调整 (BA)。集束调整 (BA) 是一种常用的非线性优化方式, 该方法的应用正是传统滤波器优化方式向非线性优化的过渡。

PTAM 系统提出的关键帧概念, 主要是在原来图像流中提取一些符合条件的图像帧, 在构建地图时只处理这些图像帧的信息, 包括: 二维图像中的特征点以及对应的三维空间点, 这样可以避免后端处理所有图像数据。结合第二章中视觉 SLAM 系统的基本原理, 后端 BA 优化的目标函数如下

$$\arg \min_{C_1 \dots C_m, X_1 \dots X_n} \sum_{i=1}^m \sum_{j=1}^n \|h(C_i, X_j) - \hat{x}_{ij}\|_{\Sigma_{ij}} \quad (3-6)$$

对于该优化过程，仅优化关键帧使后端优化满足稀疏性，这样可以大大减少计算量，使其达到实时的计算效率；由于后端 BA 优化可以得到精确地相机姿态以及三维空间点的位置，那么在此基础上，前端跟踪只考虑当前时刻对应相机的姿态即可。

首先，对当前帧进行 FAST 特征点检测<sup>[57]</sup>，通过特征匹配方法将相对应的二维图像点与在地图中三维空间点匹配起来，特征检测及跟踪定位如图 3-5 所示。然后使用匀速运动模型估计得到当前帧对应的相机姿态，并将该姿态作为优化函数的初值，优化求解当前帧位姿。

$$\arg \min_{C_i} \sum_{j=1}^n w_j \|h(C_i, X_j) - \hat{x}_{ij}\|_{\Sigma_j} \quad (3-7)$$

其中  $w_j$  是 Tukey 函数<sup>[60]</sup>对应的权重，该函数起到减少当特征匹配过程中存在误匹配时对优化的影响。

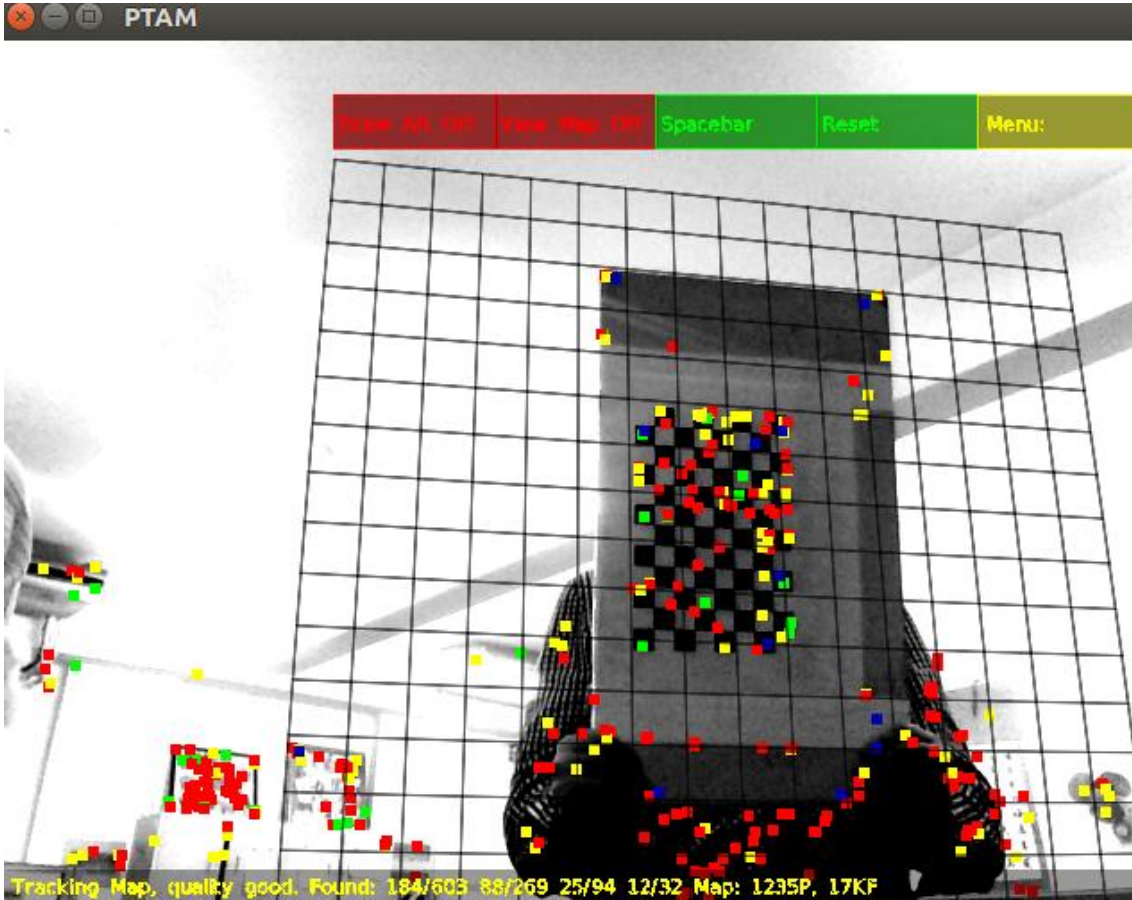


图 3-5 PTAM 跟踪定位过程

在执行跟踪定位线程时，匹配成功的内点数少于一定阈值，则认为跟踪结果不好或者很差，触发重定位程序<sup>[60]</sup>。在关键帧队列中找到与当前帧图像共同视角最大的图像进行对比，通过选定的关键帧判断当前时刻相机的位姿，随后在做一次特征匹配、以及 BA 优化。一旦重定位程序执行成功，判断当前帧来做是否可添加进关键帧的队列，添加成功，则认定新选择的关键帧可以唤醒后端线程执行地图构建功能。通过添加的新关键帧与地图中原有的关键帧进行极线(Epipolar Line)<sup>[62]</sup>匹配，以及局部集束调整后的三角化的三维点来扩张全局地图，构建环境地图中空间点以及关键帧如图 3-4 所示。同时，如图 3-5 所示网格平面作为 AR 应用的虚拟平面用于放置虚拟物体。

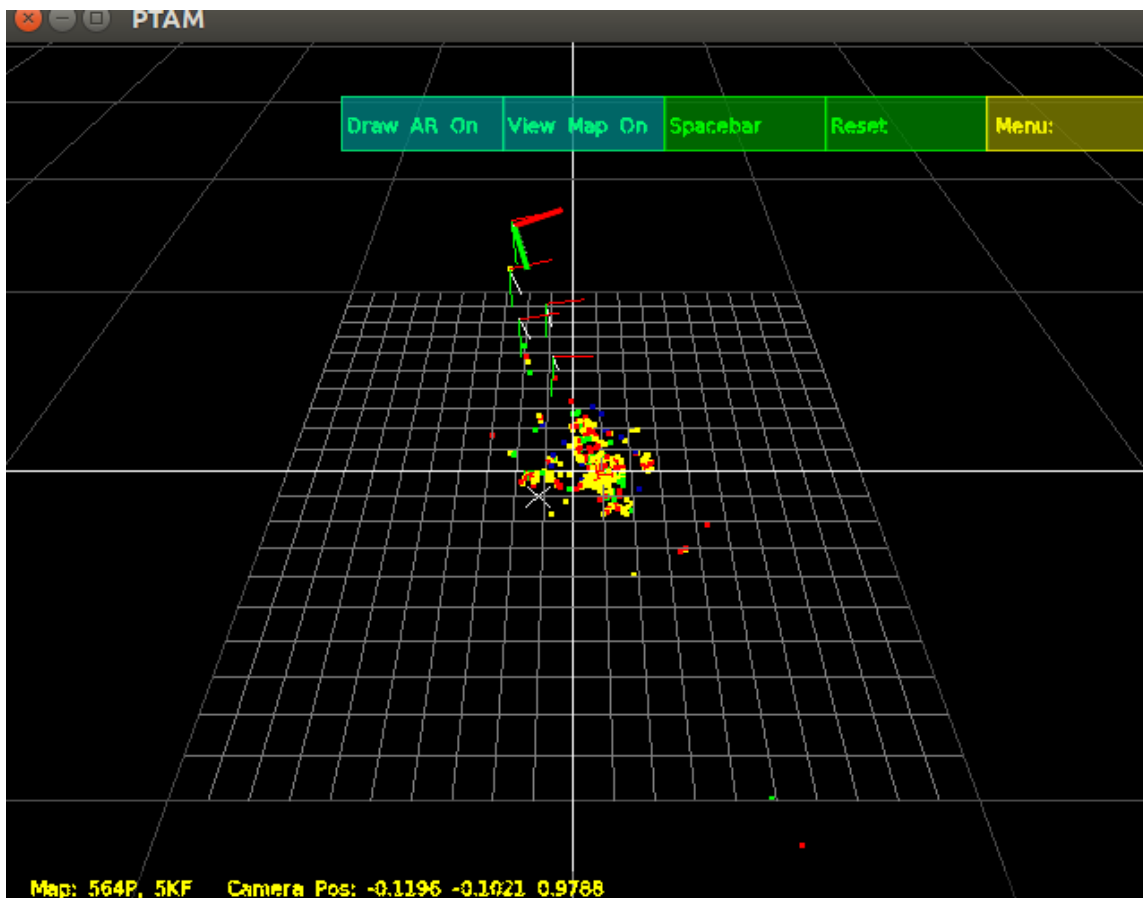


图 3-6 PTAM 构建环境地图及关键帧选择

## 2 ORB-SLAM

基于上述 PTAM 系统的介绍可知，PTAM 作为基于关键帧 BA 的视觉 SLAM 系统的开山之作，其代码开源为对视觉 SLAM 系统的研究奠定了基础。近几年许多视觉 SLAM 系统都是基于 PTAM 的算法框架改进而来的，其中不乏作为目前最适用的、性能最好的单目视觉 SLAM 系统之一的 ORB-SLAM 系统，该系统于 2015 年被 Mur-Artal 等人提出<sup>[63]</sup>。ORB-SLAM 以 PTAM 系统为基础，借鉴 PTAM 思想

对其进行优化并添加新的功能，整个系统的创新处可总结如下：

(1) **泛用性** ORB-SLAM 在其单目系统的基础上，开发可用于立体视觉以及 RGB-D 相机等系统中，使其应用范围大大扩展。

(2) **提取 ORB 特征** ORB-SLAM 系统作为主流的基于特征点的视觉 SLAM 系统，该系统对图像帧提取 ORB 特征，该特征是对 FAST 特征点与 BREIF 特征描述子的一种结合与改进，相比 PTAM 系统提取的 FAST 角点，具有良好的旋转和缩放不变性。ORB 特征是一种高效的具有良好针对视图不变的识别器，同时具有描述子，使系统可以在更宽的基准线上进行特征匹配，在运动距离较大时特征也可以成功匹配，增强了 BA 的精度。同时，整个系统的所有功能都采用 ORB 特征，且该特征是除了 SIFT 与 SURF 外一个很好的选择，而且它具有很高的效率，在 CPU 上可以达到实时的计算效率。图 3-7 显示了 ORB 特征提取的结果。

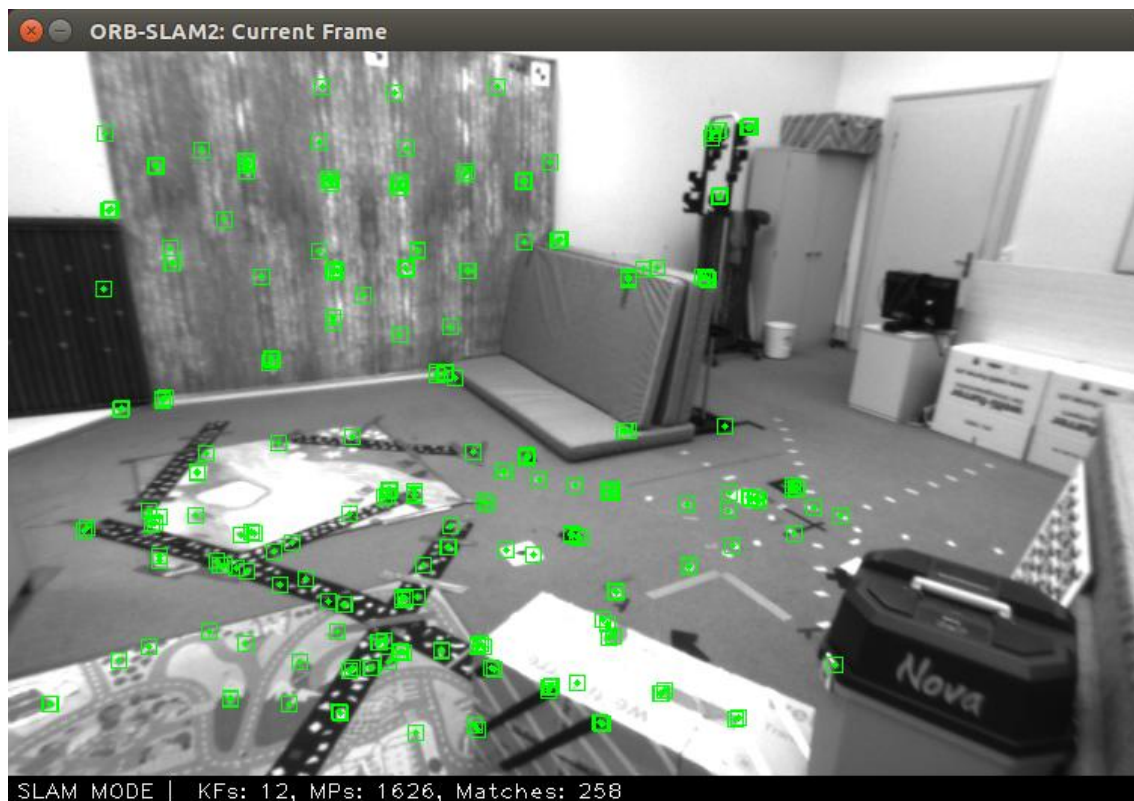


图 3-7 ORB-SLAM 提取 ORB 特征的结果

(3) **优化策略** ORB-SLAM 系统在 PTAM 的基础上引入一些优化策略，使其的鲁棒性得到提升，成为现今主流的系统之一。ORB-SLAM 系统对图像构建金字塔，在使用 OpenCV 提取出特征点，通过四叉树的方式对图像进行划分，根据每层图像的特征数确定四叉树的节点，将每层图像检测到的特征划分到这些节点，保证每个节点里面有一个特征，以确保对检测到的特征尽进行均匀化处理。



从图 3-7 和图 3-8 可看出, ORB-SLAM 系统提取出的 ORB 特征均匀分布于整幅图, 且尽量保证在图像的每个区域均有一个可用的特征点, 而 OpenCV 自带函数库提取出的特征点较为杂乱无序, 且分配不均匀不易于特征匹配, 且易导致特征误匹配。

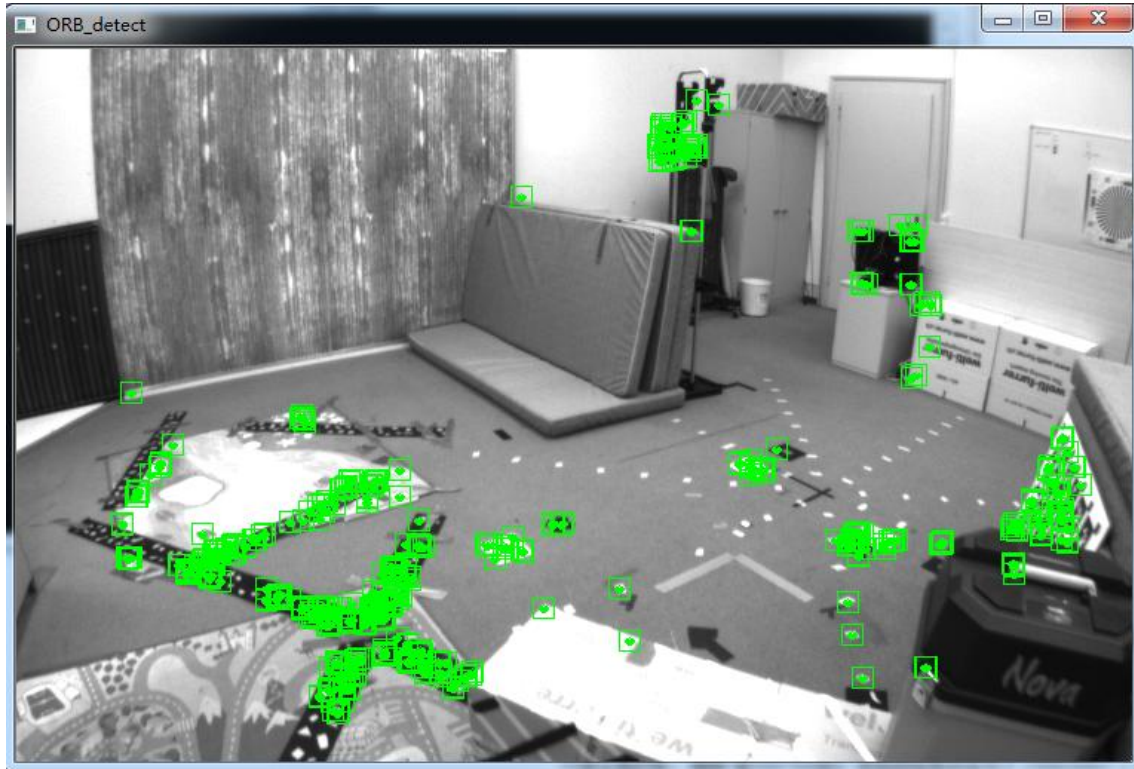


图 3-8 OpenCV 中提取 ORB 特征结果

(4)三个线程并行处理 ORB-SLAM 系统延续 PTAM 多线程工作的框架, 使用三个线程并行工作。在原来多线程基础上, 整合了回环检测线程。该线程通过对闭环进行检测判断新添加的关键帧在地图中是否形成闭环, 计算系统的累积误差, 结合闭合机制, 将该闭环进行闭合控制, 以及共有地三维空间点进行融合处理。该线程可以消除系统累计误差, 确保全局地图的一致性。

(5)提高场景扩展性 ORB-SLAM 提出一个相比 PTAM 更加宽松鲁棒的关键帧选取策略, 将局部地图添加关键帧的条件放宽, 确保新的相机运动更易于估计, 在局部地图构建过程中, 删除一些冗余的关键帧, 可提高系统的场景扩展性以优化的计算效率, 同时地图及时扩展, 可保证系统的平稳跟踪定位。ORB-SLAM 系统选择的关键帧如图 3-9 所示。

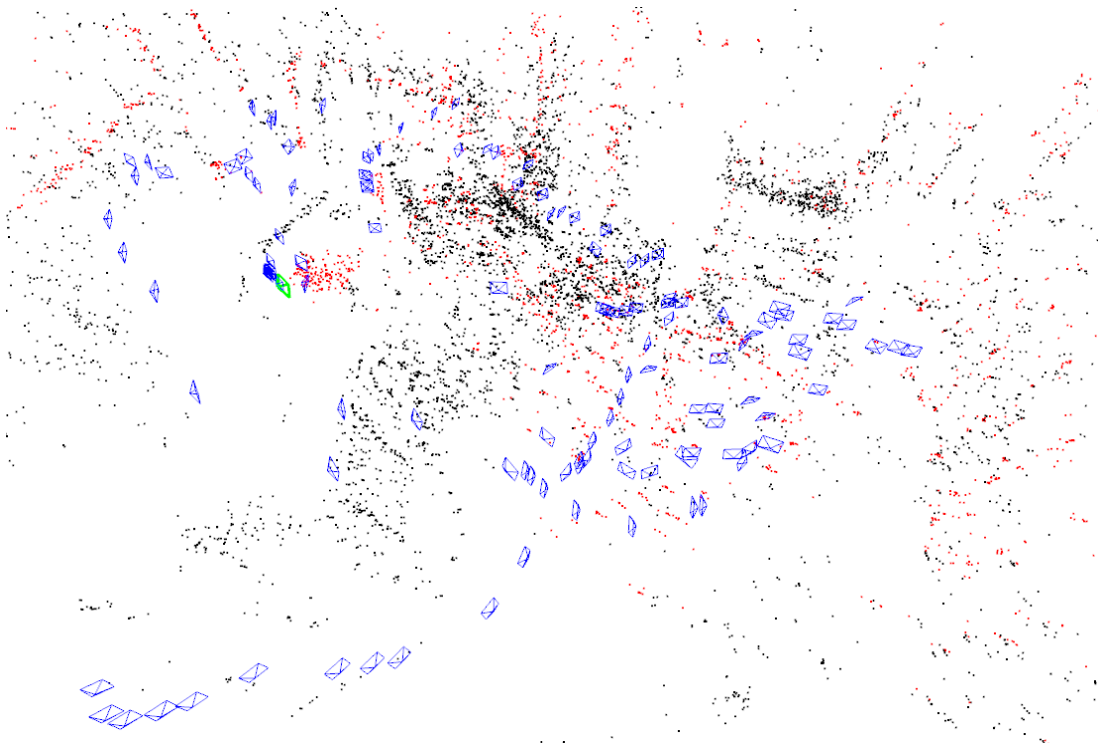


图 3-9 ORB-SLAM 系统选择的关键帧、当前帧

ORB-SLAM 系统创建三个并行的线程：特征跟踪线程（Tracking）、局部地图构建线程（Local Mapping）以及回环检测线程（Loop Closing）。ORB-SLAM 基本延续了 PTAM 的算法框架，ORB-SLAM 系统的大体结构流程如下。

（1）特征跟踪线程（Tracking） 特征跟踪线程主要通过提取每帧图像 ORB 特征点，对前一帧图像使用同样的匀速运动模型估计初始位姿，搜索上一帧观测到的空间三维点进行特征匹配粗略估计其三维位置，通过 BA 优化匹配成功点对之间的对应关系估计相机位姿。如果跟踪丢失，将当前帧转换成词袋模型，寻找最相关的关键帧执行全局重定位。同时，开启局部地图构建线程，创建一个局部地图，即 Co-visibility Graph，如图所示。对每个关键帧执行 RANSAC 迭代计算，用 PnP 算法解算相机位置，并进一步通过当前帧中获得所有的三维空间点进行优化，跟踪线程继续执行。最后判断当前时刻的图像帧是否可标记为关键帧，在该线程快速的插入关键帧，在局部地图构建线程中可以删除冗余的关键帧，这也是 ORB-SLAM 系统跟踪定位相机位姿鲁棒性更好的原因之一。

（2）局部地图构建线程（Local Mapping） 局部地图构建线程主要像 Covisibility Graph,如图 3-10 所示。中添加的关键帧，根据 Covisibility Graph 中相互关联的关键帧，再一次进行特征匹配用于数据关联进行三角化得到新的地图点，但存在误匹配等原因使一些地图点被错误的三角化处理，根据地图点筛选机制去除这些错误的地图点。然后，执行局部 BA 优化重新构建更新当前 Covisibility Graph

中的新添加的关键帧、与其相关联的关键帧和所有被观测到的地图点，同时对冗余的关键帧进行筛选。

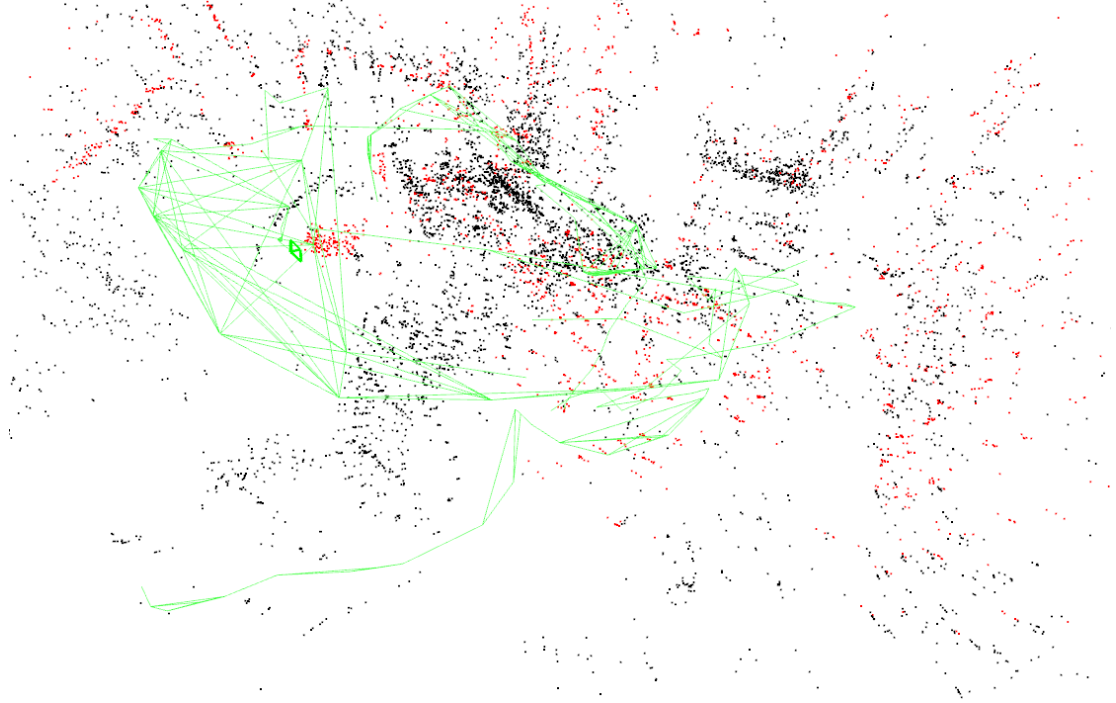


图 3-10 可关联视图

**(3) 回环检测线程 (Loop Closing)** ORB-SLAM 系统最大的亮点在于系统增添了闭合回路检测的线程，用于消除系统引入的累积误差。对新添加的关键帧均进行一次相似度计算，判断是否与其他关键帧构成闭环，如果形成闭环，则通过计算相似变换  $\xi_{ij}$  以获得闭环的累计误差，通过方位图(Pose Graph)优化使回路两端闭合

$$\arg \min_{\xi_1, \dots, \xi_m} \sum_{(\xi_{ij}, \Sigma_{ij})} (\xi_{ij} \circ \xi_i^{-1} \circ \xi_j)^T \Sigma_{ij}^{-1} (\xi_{ij} \circ \xi_i^{-1} \circ \xi_j) \quad (3-8)$$

其中， $\xi_i, \xi_j$  表示每个关键帧的相似变换， $\xi_{ij}$  表示两个关键帧之间的相对变换， $\Sigma_{ij}$  为  $\xi_{ij}$  的协方差矩阵，操作符  $\circ$  按顺序连接 2 个相似变换。最后，相似约束的位姿图优化确保全局地图一致性，相比针对整幅地图执行 BA 优化，大大减少了计算量。本文主要通过关键图像 (Essential Graph) 进行优化，执行全局优化使用 g2o 库中的 Levenverg-Marquardt 算法。因此，ORB-SLAM 能处理大尺度场景。

ORB-SLAM 在运行过程中状态如图 3-11 所示。图 3-11 中，蓝色显示系统选取的关键帧，绿色显示当前帧相机位姿以及黑色和红色显示构建地图的地图点，其中红色代表当前局部的地图点。

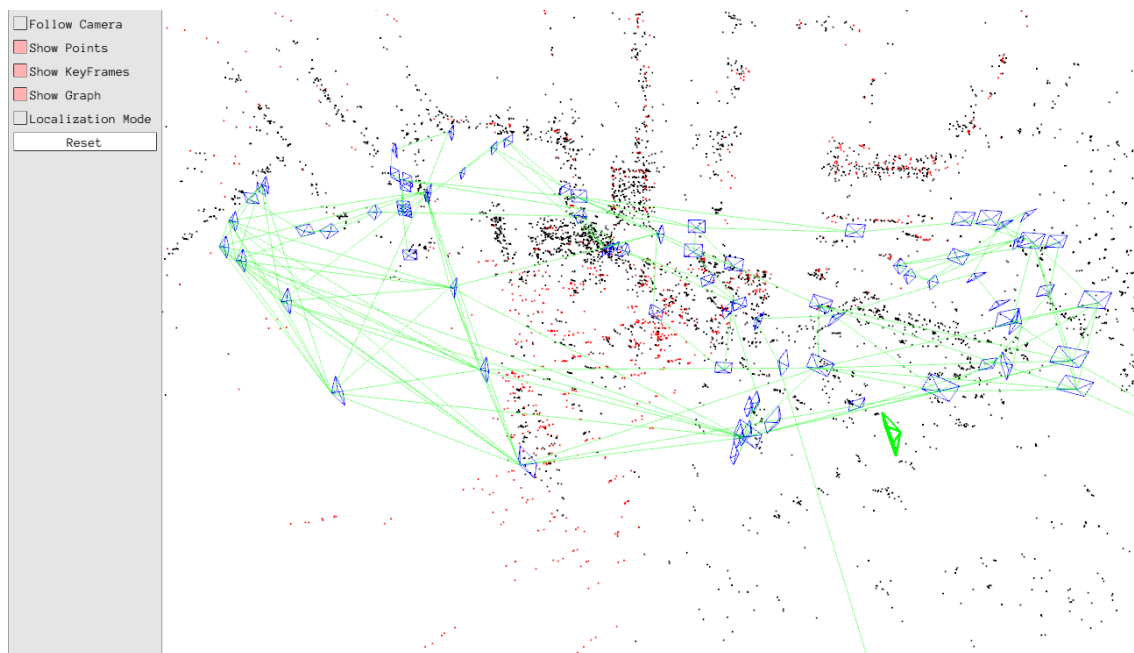


图 3-11 ORB-SLAM 运行状态图。

### 3.2.3 基于直接法的视觉 SLAM 系统

根据前文对其他两种方法的分析可得知，相比于基于直接跟踪的视觉 SLAM 系统，基于滤波器和基于关键帧 BA 的视觉 SLAM 系统通常采用特征点法进行，尽管特征点法在视觉里程计中占据主流地位，但该方法在运动到特征缺失或无明显纹理的地方不能找到足够的匹配点来计算相机运动，且关键点的提取和描述子的计算非常耗时。因此，直接跟踪法(Direct Tracking)是直接根据像素亮度信息估计相机的运动，可以避免特征及描述子的计算时间，也避免了特征缺失的情况下特征法鲁棒性较差。直接法为视觉 SLAM 系统的另一主要分支，近年来，基于直接跟踪法的视觉 SLAM 系统慢慢被研究者熟知，比如 LSD-SLAM<sup>[64]</sup>和 SVO<sup>[65]</sup>。本节将主要介绍 SVO 系统。

#### 1 SVO

2014 年，Forster 等人提出 SVO<sup>[65]</sup> (Semi-direct Visual Odomtry)。该系统是一种基于半直接法的视觉里程计。该方法是指也采用直接法跟踪，但是是采用稀疏的特征点，节省了特征提取的巨大开销，因此这个策略可以实现很高的频率。SVO 的大体流程如下：

(1) **初始化** SVO 首先经过初始化提取 FAST 特征点，然后对后面的图像帧一直用第一帧的特征点做光流跟踪，判断跟踪成功的内点个数是否满足要求，如果满足则求解表示这两帧之间的相对关系的单应矩阵确定前两帧对应相机姿态。图 3-12 显示整个初始化光流跟踪第一帧图像中的 FAST 角点。



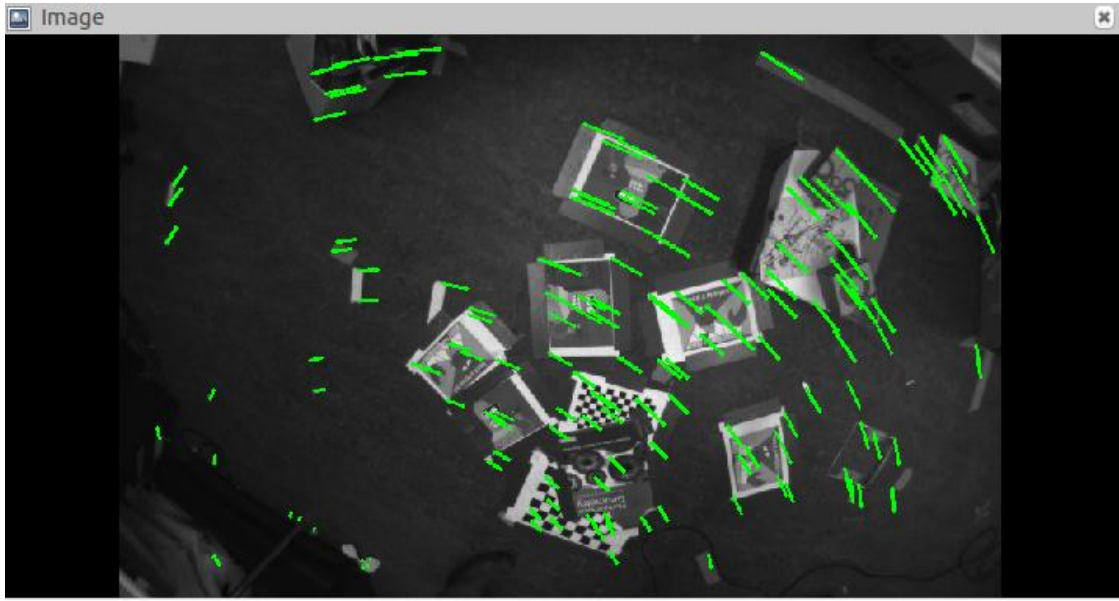


图 3-12 初始化过程。光流跟踪第一帧图像中提取的 FAST 特征点

(2) **图像块对齐** 完成初始化之后，利用追踪成功的这些特征点对应像素周围的  $4 \times 4$  的特征块使用直接法进行配准，通过迭代优化位姿  $T_{k,k-1}$  最小化图像块重投影残差，获取相机位姿。迭代优化的目标函数如下

$$T_{k,k-1} = \arg \min_{T_{k,k-1}} \frac{1}{2} \sum_{i \in \mathcal{R}} \left\| \delta I(T_{k,k-1}, u_i) \right\|^2 \quad (3-9)$$

其中

$$\delta I(T_{k,k-1}, u_i) = I_k \left( \pi \left( T_{k,k-1} \cdot \pi^{-1}(u, d_u) \right) \right) - I_{k-1}(u) \quad (3-10)$$

其中， $T_{k,k-1}$  表示相机位姿， $u_i$  表示特征点像素， $I_k$  表示将三维空间点投影到当前帧图像坐标。

(3) **特征点对齐** 对这些稀疏的特征块进行特征点对齐，通过获取的位姿预测参考帧中的特征块在当前帧中的位置，由于深度估计的不准导致获取的位姿也存在偏差，从而使得预测的特征块位置不准。由于预测的特征块位置和真实位置很近，所以可以使用牛顿迭代法对这个特征块的预测位置进行优化。不过优化变量不再是相机位姿  $T_{k,k-1}$ ，而是像素的位置  $(u', v')$ ，通过光流法跟踪，最小化光度误差对特征块的预测位置进行迭代优化，迭代优化的目标函数如下

$$u'_i = \arg \min_{u'_i} \frac{1}{2} \left\| I_k(u'_i) - A_i \cdot I_r(u_i) \right\|^2 \quad (3-11)$$

其中， $u_i$  表示像素位置，即优化变量； $I_k$  表示将三维空间点投影到当前帧图像坐标； $A_i$  表示参考帧与当前帧之间的仿射变换。

(4) **优化相机位姿及特征点位置** 特征块的预测位置得到优化,说明之前使用直接法预测的有偏差。利用这个优化后得到的特征块预测位置,再次使用直接法,对相机位姿以及特征点位置进行优化。通过迭代优化最小化像素重投影以后位置的差异,迭代优化的目标函数如下

$$T_{k,w} = \arg \min_{T_{k,w}} \frac{1}{2} \sum_{i \in \mathcal{R}} \|u_i - \pi(T_{k,w}, p_i)\|^2 \quad (3-12)$$

其中,  $T_{k,w}$  为相机位姿,也是优化变量。

SVO 作为基于直接法定位系统实现了 SLAM 系统的前端 VO 部分,相比于其他方案, SVO 有两个优势:

(1) **运行速度快** SVO 仅提取特征点,不使用特征描述子进行匹配,同时该系统是半直接法,不必处理像稠密和半稠密那么多的信息,因此即使在低端计算平台上也能达到实时性,而在 PC 平台上则可以达到 100 多帧每秒的速度,在作者后续工作 SVO2.0 中,速度更达到了惊人的 400 帧每秒。这使得 SVO 非常适用于计算平台受限的场合,例如无人机、手持 AR/VR 设备的定位。无人机也是作者开发 SVO 的目标应用平台。

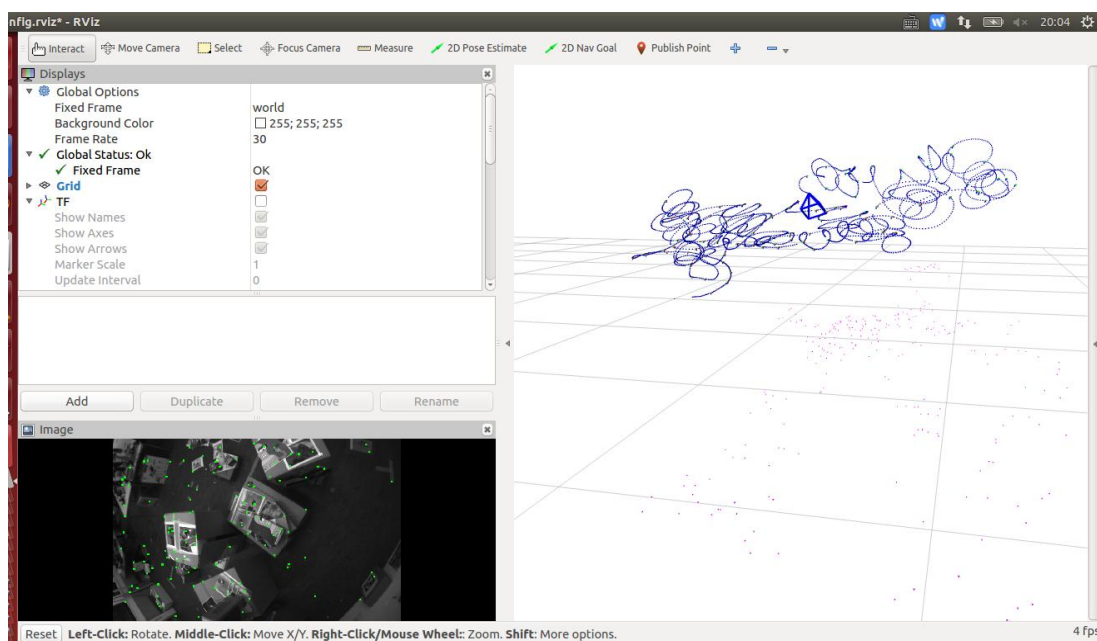


图 3-13 Rviz 显示 SVO 系统运行状态图

(2) **提出深度滤波器** SVO 系统提出了深度滤波器,该系统使用基于均匀-高斯混合分布的深度滤波器。SVO 将这种滤波器用于关键点的位置估计,并使用了逆深度作为参数化形式,使之能够更好的计算特征点位置。

SVO 在运行过程中状态如图 3-13 所示。运动轨迹通过采用 ROS 系统中集成

的 Rviz 可视化工具进行绘制,其中蓝色线显示相机运动轨迹以及选取关键帧位姿、蓝色四棱锥显示当前相机位姿以及红色显示构建的特征点地图。

### 3.3 典型视觉 SLAM 系统性能比较

根据 3.2 节中对四旋翼无人机的基于视觉自主定位的问题的分析及描述,本节结合 3.3 节中对视觉 SLAM 开源系统据实际测试的结果结合理论分析,从实时性、适应性两个角度对比分析 MonoSLAM、PTAM、ORB-SLAM、SVO 这四个典型系统的优缺点。

#### 3.3.1 实时性能

MonoSLAM 的实时性能相对较差,随着系统中处理空间点数量的增长计算量大大增加,计算复杂度为  $O(n^3)$ ,其中  $n$  为处理空间点数量。PTAM 和 ORB-SLAM 作为基于关键帧 BA 的视觉 SLAM 系统的代表,均通过提取特征点,计算特征描述子进行特征匹配完成相机的跟踪定位过程,这个过程十分耗时,导致实时性能相对较差,但这两种系统将地图优化、位姿纠正均放在系统后端进行处理,实时性能及效率相对 MonoSLAM 系统较高。PTAM 和 ORB-SLAM 系统相比较,由于 ORB-SLAM 系统提取 ORB 特征点,且整个系统均根据 ORB 特征进行工作,相比其他基于特征点法的系统实时性相对较高。相比前两种方法,基于直接跟踪的 SVO 系统直接处理图像像素进行定位,避免特征描述子计算的过程,因此该系统的实时性能最高,适用于无人机、移动设备等处理能力较差的平台。

#### 3.3.2 定位精度

由于基于滤波器的视觉 SLAM 系统在轨迹跟踪过程中,使用滤波器根据先前时刻状态对当前状态进行预测更新处理,因此容易导致误差累计,使其轨迹的轨迹存在严重的漂移情况,因此 MonoSLAM 系统的定位精度相对较差。然而基于关键帧 BA 的方法采用图优化的方式进行优化处理,因此其定位精度相比 MonoSLAM 系统较高。同时,由于 PTAM 系统采用运动方程进行预测的方式进行位姿估计,而 ORB-SLAM 系统通过提取 ORB 特征点完成特征匹配的过程,且 ORB 特征点具有视角、尺度不变性。

在此基础上,ORB-SLAM 系统在优化过程中对局部地图进行优化、且增加回环检测控制功能,因此 ORB-SLAM 系统的定位精度相对较高。然后,SVO 作为视觉 VO,后端优化功能较弱,并且该系统采用对像素进行处理的方式,导致其假设条件较为严格,受环境光照影响较大,因此其定位精度相对较差。同样由于这些

原因的存在，导致基于直接法的视觉 SLAM 系统现阶段无法与基于关键帧 BA 的视觉 SLAM 系统进行媲美。

### 3.3.3 特征缺失鲁棒性

由于视觉 SLAM 系统均处理图像信息获取定位，然而在室内环境中，经常遇到白墙、镜面、窗户以及地板等无明显特征或者重复纹理较严重等情况，直接应像视觉 SLAM 系统的定位精度，甚至导致定位系统不可用的情况，尤其对基于特征点法的 ORB-SLAM 系统对精度影响很大。对于 SVO 系统，采用像素灰度信息结算相机位姿的方法，可以一定程度的避免对特征的依赖，却也会受到一定影响。因此，这三种方法对比得出基于直接法的系统性能最好。

## 3.4 本章小结

本章首先针对第二章中对视觉 SLAM 系统实际应用于无人机平台存在问题，将近几年来视觉 SLAM 领域的主流方法进行分类。依据运动估计方法，将主流的视觉 SLAM 系统进行分类。并根据分类结果，实际测试了对应这些方法的典型开源系统，MonoSLAM、ORB-SLAM、PTAM 以及 SVO。在此基础上，通过实际测试结合理论分析，从实时性能、定位精度以及特征缺失鲁棒性三个方面总结了这些典型系统的优缺点。针对本课题的系统需求，最终确定本课题的运动估计方法以及整体系统方案。

## 第 4 章 基于半直接法的立体视觉 SLAM 系统设计

### 4.1 引言

第 3 章针对视觉 SLAM 系统应用于无人机平台存在的问题进行描述，实际测试了近几年来提出的先进系统，并对比分析了这些典型系统的性能及优缺点。虽然这些系统各有优势，但在实际应用中仍存在很大的问题。本章首先针对课题的应用背景，结合现有先进系统的特点，确定使用 SVO 的框架作为本课题的系统方案。虽然 SVO 系统采用半直接法进行运动估计，速度极快，非常适用于四旋翼飞行器这种机载处理器能力差，对实时性要求高的系统，但为了提高系统的实时性能，舍弃了后端优化和回环检测等功能，导致其位姿估计必然存在累计误差，而且丢失后不太容易进行重定位。因此，系统的鲁棒性相对较差，在实际应用中必然存在很大的问题。

在使用 SVO 系统框架下，由于 SVO 是单目视觉定位系统，针对单目视觉在恢复地图场景尺度与真实环境存在偏差的问题，本章首先在 SVO 系统框架下设计使用双目视觉传感器获取场景信息，结合第 2 章中介绍的立体视觉系统可恢复真实地图尺度的特性，提出了一种基于半直接法的双目匹配算法，结合之前确定的系统框架，在满足系统实时性能的条件下，完成整体系统框架搭建，并编程实现整个基于半直接法的立体视觉 SLAM 系统，并通过仿真实验验证整个方案的可行性。

### 4.2 基于直接法定位原理

近年来，随着移动终端、无人机以及可穿戴设备的流行，视觉 SLAM 系统得到更多的应用，由于这些智能设备处理器的处理能力较差，导致系统的实时性能得到空前的关注。因此，一些基于直接法的 SLAM 系统得到更多的研究开发，如 SVO、LSD-SLAM、DSO-SLAM。直接法作为运动估计的方法，直接采用图像中可测量图像的像素的亮度信息恢复未知参数。这与以及基于特征点的方法形成鲜明的对比，该方法首先从每个图像中分别提取出一组稀疏的特征，然后恢复或分析其对应关系，已确定运动和形状。基于特征点的方法最小化几个相应特征之间的基于距离的误差测量，而直接法最小化基于从图像中的所有像素提供的直接图像信息的误差测量，例如图像亮度或基于亮度的相互关系等。直接法对像素进行处理可避免特征提取、特征描述子计算以及特征匹配等操作。因此，可以节省计算特征及匹配的时间，在场景特征不丰富或重复纹理较为严重的环境中鲁棒性更

好。只要场景中存在明暗变化，就可以采用直接法进行定位。根据使用像素的数量，直接法分为稀疏、稠密和半稠密三种。直接法是从按照光流跟踪的思想提出的，并为了克服特征点法的上述缺点而存在的

#### 4.2.1 Lucas-Kanade 光流跟踪

Lucas-Kanade 光流跟踪，来自相机的图像是随时间变化的。图像可以看作时间的函数： $I(t)$ 。那么。一个在 $t$ 时刻，位于 $(x, y)$ 处的像素，它的灰度可以写成

$$I(x, y, t) \quad (4-1)$$

这种方式把图像看成了关于位置与时间的函数，它的值域就是图像中像素的灰度。现在考虑某个固定的空间点，它在 $t$ 时刻的像素坐标为 $(x, y)$ 。由于相机的运动，它的图像坐标将发生变化。我们希望估计这个空间点在其他时刻里图像的位置。怎么估计呢？这里要引入光流法的基本假设：

灰度不变假设：同一个空间点的像素灰度值，在各个图像中是固定不变的。

对于 $t$ 时刻位于 $(x, y)$ 处的像素，我们设 $t + dt$ 时刻，它运动到 $(x + dx, y + dy)$ 处。由于灰度不变，我们有：

$$I(x + dx, y + dy, t + dt) = I(x, y, t) \quad (4-2)$$

灰度不变假设是一个很强的假设，实际当中很可能不成立。事实上，由于物体的材质不同，像素会出现高光和阴影部分；有时，相机会自动调整曝光参数，使得图像整体变亮或变暗。这些时候灰度不变假设都是不成立的，因此光流的结果也不一定可靠。然而，从另一方面来说，所有算法都是在一定假设下工作的。如果我们什么假设都不做，就没法设计实用的算法。我们，暂且让我们认为该假设成立，看看如何计算像素的运动。

对左边进行泰勒展开，保留一阶项，得：

$$I(x + dx, y + dy, t + dt) \approx I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt \quad (4-3)$$

因此我们假设了灰度不变，于是下一个时刻的灰度等于之前的灰度，从而

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (4-4)$$

两边除以，得：

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} = -\frac{\partial I}{\partial t} \quad (4-5)$$

其中 $\frac{dx}{dt}$ 为像素在 $x$ 轴上运动速度，而 $\frac{dy}{dt}$ 为 $y$ 轴速度，把它们记为 $(u, v)$ 。同时 $\frac{\partial I}{\partial x}$ 为

图像在该点处  $x$  方向的梯度，另一项则是在  $y$  方向的梯度，记为  $I_x, I_y$ 。把图像灰度对时间的变化量记为  $I_t$ ，写成矩阵形式，有：

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \quad (4-6)$$

计算的是像素的运动  $u, v$ ，但是该式是带有两个变量的一次方程，仅凭它无法计算出  $u, v$ 。因此，必须引入额外的约束来计算  $u, v$ 。在 LK 光流中，我们假设某一个窗口内的像素具有相同的运动。

考虑一个大小为  $w \times w$  的窗口，它含有  $w^2$  数量的像素。由于该窗口内像素具有同样的运动，因此我们共有个方程：

$$\begin{bmatrix} I_x & I_y \end{bmatrix}_k \begin{bmatrix} u \\ v \end{bmatrix} = -I_{tk} \quad k = 1, \dots, w^2. \quad (4-7)$$

记为

$$A = \begin{bmatrix} \begin{bmatrix} I_x & I_y \end{bmatrix}_1 \\ \vdots \\ \begin{bmatrix} I_x & I_y \end{bmatrix}_k \end{bmatrix}, b = \begin{bmatrix} I_{t1} \\ \vdots \\ I_{tk} \end{bmatrix} \quad (4-8)$$

于是整个方程为：

$$A \begin{bmatrix} u \\ v \end{bmatrix} = -b \quad (4-9)$$

这是一个关于  $u, v$  的超定线性方程，传统解法是求最小二乘解。最小二乘在很多时候都用到过：

$$\begin{bmatrix} u \\ v \end{bmatrix}^* = -(A^T A)^{-1} A^T b \quad (4-10)$$

这样就得到了像素在图像间的运动速度，当取离散的时刻而不是连续时间时，我们可以估计某块像素在若干个图像中出现的位置。由于像素梯度仅在局部有效，所以如果一次迭代不够好的话，我们会多迭代几次这个方程。在 SLAM 中，LK 光流常被用来跟踪角点的运动。

#### 4.2.2 基于直接法定位的工作原理

假设摄像机在未知环境中随意运动，存在一个在两帧图像中均可以观测到的空间点  $P$ ，我们通过这两帧图像信息，求解其对应的相机姿态，以及这两个时刻相机的相对位姿关系，如图 4-1 所示。该三维空间点的世界坐标为  $P = (X, Y, Z)$ ，使用齐次坐标的方式可表示为  $P = (X, Y, Z, 1)$ ，通过相机的成像模型进行投影变换，

与成像平面相交于像素点  $u_i, u_i'$ ，记作  $u_i = (u_i, v_i)$ ,  $u_i' = (u_i', v_i')$ ，根据对应像素点的亮度误差进行位姿估计。图 4-1 显示直接法定位的示意图。

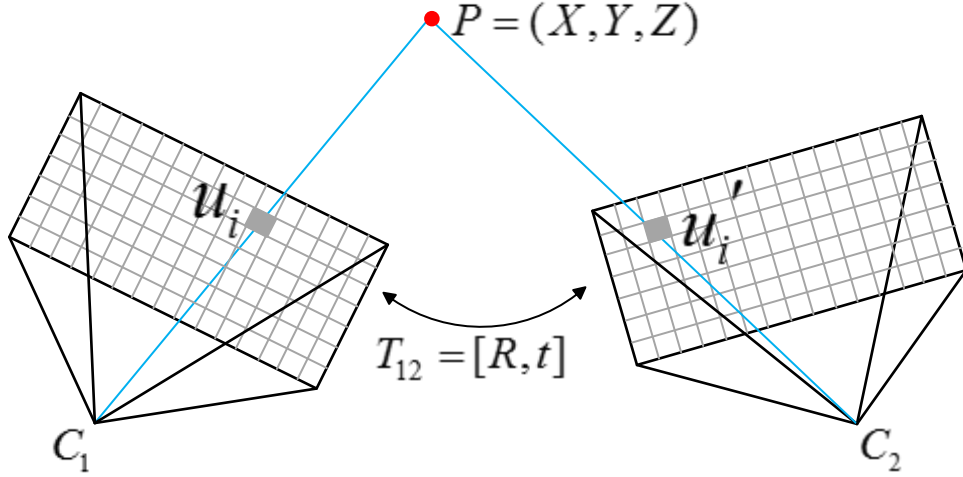


图 4-1 直接法示意图

从图 4-1 中可以看出，视觉定位系统一般假定上一个时刻相机的姿态为单位矩阵，通过估计相邻两帧之间的位姿确定当前时刻的相机姿态，因此相机的姿态通过增量的方式进行估计。由于相机的内参可通过相机标定确定且不会改变，使用  $K$  表示内参矩阵，其中像素点的像素坐标使用齐次坐标  $u_i = (u_i, v_i, 1)$  表示，其两个相邻时刻相机  $C_1, C_2$  之间的相对姿态可用  $3 \times 3$  的旋转矩阵和  $3 \times 1$  平移向量构成的  $T = [R, t]$  表示。因此，投影关系可表示为

$$u_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{Z_i} K(RP + t) = \frac{1}{Z_i} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \quad t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4-11)$$

其中，不妨假设第一帧图像对应相机姿态为单位矩阵  $I$ ，那么第二帧图像对应相机姿态为单位矩阵  $T = [R, t]$ ，那么

$$u_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{Z_i} KP = \frac{1}{Z_i} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (4-12)$$

$$u_i' = \begin{bmatrix} u_i' \\ v_i' \\ 1 \end{bmatrix} = \frac{1}{Z_i'} K(RP + t) = \frac{1}{Z_i'} \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \quad t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$



其中,  $Z_i, Z_i'$  分别是空间点  $P$  在两个相机坐标系下的深度信息。

由于特征点法通过提取出的特征点进行特征匹配, 寻找对应关系估计位姿信息。然而相比特征点法, 直接法直接对像素进行处理, 缺少空间点对应相同像素点的先验信息。直接法与光流跟踪法十分相似, 他们都是基于灰度不变假设的, 即假定同一空间点的像素灰度在各个视角下是固定不变的。因此, 在直接法中, 同样是求解一个优化问题, 通过优化相邻两个时刻的相机姿态  $T$ , 确定相对应的两个像素点, 其优化目标是最小化光度误差, 即两个时刻对应同一空间点  $P$  的像素点  $u_i, u_i'$  的亮度误差达到最小

$$e = I_1(u_i, v_i) - I_2(u_i', v_i') \quad (4-13)$$

其中,  $I_1, I_2$  分别代表相邻两个时刻获取的图像,  $(u_i, v_i), (u_i', v_i')$  分别代表两个时刻图像对应的像素点的坐标,  $e$  代表对应像素点的灰度误差, 是一个标量。其优化的目标函数形式可表示成下

$$\min J = \|I_1(KP) - I_2(K(RP+t))\|_2^2 = e^T e = \|e\|^2 \quad (4-14)$$

由于该方法是基于灰度不变的假设, 因此该优化函数确实成立。在相机运动估计中, 会提取多个空间点进行目标函数优化, 假设前后两个时刻可共同观测到多个空间点  $P$ , 那么, 整个相机位姿估计问题可表示为

$$\min J = \sum_{n=1}^N e_n^T e_n, \quad e_n = I_1(u_{i,n}) - I_2(u_{i,n}') \quad (4-15)$$

其中, 该优化函数的优化变量是相机位姿变换关系  $T = [R, t]$ , 且该目标函数的取值只与  $R, t$  有关。

求解这个优化函数, 就是解决其灰度误差  $e$  是如何随着相机位姿变换关系  $T$  变化的, 因此需要对其求导, 分析他们的导数关系。由于  $T$  在李群中, 无加法运算, 不好求导, 因此, 使李代数  $\zeta$  表示相机位姿, 因此使用李代数的形式表示投影关系

$$\begin{aligned} u_i &= \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{Z_i} KP, \\ u_i' &= \begin{bmatrix} u_i' \\ v_i' \\ 1 \end{bmatrix} = \frac{1}{Z_i'} K(RP+t) = \frac{1}{Z_i'} K(\exp(\zeta^\wedge)P) \end{aligned} \quad (4-16)$$

其中,  $T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} = \exp(\zeta^\wedge)$  表示相机位姿。

通过使用李代数上的扰动模型, 给  $\exp(\zeta)$  左乘一个小扰动  $\exp(\delta\zeta)$  可表示为

$$\begin{aligned} \exp(\zeta \oplus \delta\zeta) &= I_1 \left( \frac{1}{Z_i} KP \right) - I_2 \left( \frac{1}{Z_i'} K \exp(\delta\zeta^\wedge) \exp(\zeta^\wedge) P \right) \\ &\approx I_1 \left( \frac{1}{Z_i} KP \right) - I_2 \left( \frac{1}{Z_i'} K (1 + \delta\zeta^\wedge) \exp(\zeta^\wedge) P \right) \\ &= I_1 \left( \frac{1}{Z_i} KP \right) - I_2 \left( \frac{1}{Z_i'} K \exp(\zeta^\wedge) P + \frac{1}{Z_i'} K \delta\zeta^\wedge \exp(\zeta^\wedge) P \right) \end{aligned} \quad (4-17)$$

还可简略表示为

$$\begin{aligned} q &= \delta\zeta^\wedge \exp(\zeta^\wedge) P \\ u &= \frac{1}{Z_i'} Kq \end{aligned} \quad (4-18)$$

其中,  $q$  表示空间点  $P$  在给予扰动之后, 在当前时刻相机坐标系下的三维坐标;  $u$  表示在发生扰动后对应的图像像素坐标。于是, 对式(4-17)进行一阶泰勒展开处理, 得到

$$\begin{aligned} \exp(\zeta \oplus \delta\zeta) &= I_1 \left( \frac{1}{Z_i} KP \right) - I_2 \left( \frac{1}{Z_i'} K \exp(\zeta^\wedge) P + u \right) \\ &\approx I_1 \left( \frac{1}{Z_i} KP \right) - I_2 \left( \frac{1}{Z_i'} K \exp(\zeta^\wedge) P \right) - \frac{\partial I_2}{\partial u} \frac{\partial u}{\partial q} \frac{\partial q}{\partial \delta\zeta} \delta\zeta \\ &= e(\zeta) - \frac{\partial I_2}{\partial u} \frac{\partial u}{\partial q} \frac{\partial q}{\partial \delta\zeta} \delta\zeta \end{aligned} \quad (4-19)$$

由式(4-19)可得知, 图像残差对位姿的迭代增量进行求导, 可以通过链式求导法则得到

$$J = - \frac{\partial I_2}{\partial u} \frac{\partial u}{\partial q} \frac{\partial q}{\partial \delta\zeta} \quad (4-20)$$

其中, 式(4-20)中的中间变量可表示为

$$\begin{aligned} q &= \exp(\zeta^\wedge) P, \\ u &= Kq \end{aligned} \quad (4-21)$$

式(4-20)中链式求导的第一项  $\frac{\partial I_2}{\partial u}$  表示像素点  $u$  处的像素梯度, 为  $1 \times 2$  的矩阵, 经过

简单的处理，得到

$$\frac{\partial I_2}{\partial u} = [I_2(u+1, v) - I_2(u, v) \quad I_2(u, v+1) - I_2(u, v)] \quad (4-22)$$

由于  $u, v$  通常为浮点数，需要对图像进行插值，该过程描述了像素点位置变化时，图像的读数的变化情况。第三项  $\frac{\partial q}{\partial \delta \zeta}$  表示添加扰动后的三维空间点对扰动过程求

导，于是  $\frac{\partial q}{\partial \delta \zeta}$  可表示为

$$\frac{\partial q}{\partial \delta \zeta} = \frac{\exp(\delta \zeta^\wedge) \exp(\zeta^\wedge) P - \exp(\zeta^\wedge) P}{\delta \zeta} = (\exp(\zeta^\wedge) P)^\square \quad (4-23)$$

$\square$  运算符把  $4 \times 4$  的向量变成  $4 \times 6$  的矩阵，得到

$$q^\square = \begin{bmatrix} q_3 \\ 1 \end{bmatrix}^\square = \begin{bmatrix} I_{3 \times 3} & -q_3^\wedge \\ 0_{1 \times 3}^T & 0_{1 \times 3}^T \end{bmatrix} \quad (4-24)$$

$\frac{\partial u}{\partial q}$  表示投影方程关于相机坐标系下的空间点  $q = [X, Y, Z]^T$  的导数。将投影方程展开，得到

$$\begin{aligned} u &= \frac{f_x X + c_x}{Z} \\ v &= \frac{f_y Y + c_y}{Z} \end{aligned} \quad (4-25)$$

于是  $\frac{\partial u}{\partial q}$  可表示为

$$\frac{\partial u}{\partial q} = \begin{bmatrix} \frac{\partial u}{\partial X} & \frac{\partial u}{\partial Y} & \frac{\partial u}{\partial Z} \end{bmatrix} = \begin{bmatrix} \frac{f_x}{Z} & 0 & -\frac{f_x X}{Z^2} \\ 0 & \frac{f_y}{Z} & -\frac{f_y Y}{Z^2} \end{bmatrix} \quad (4-26)$$

通过上述对链式法则分解得到三项导数的描述，可知，由于除了第一部分  $\frac{\partial I_2}{\partial u}$  只受与扰动变换后的空间点  $q$  位置影响，并不受图像信息的影响。因此，后两部分可表示为

$$\frac{\partial u}{\partial \delta \zeta} = \begin{bmatrix} \frac{f_x}{Z} & 0 & -\frac{f_x X}{Z^2} & -\frac{f_x XY}{Z^2} & f_x + \frac{f_x X^2}{Z^2} & -\frac{f_x Y}{Z} \\ 0 & \frac{f_y}{Z} & -\frac{f_y Y}{Z^2} & -f_y - \frac{f_y Y^2}{Z^2} & \frac{f_y XY}{Z^2} & \frac{f_y X}{Z} \end{bmatrix} \quad (4-27)$$

其中,  $\frac{\partial u}{\partial \delta \zeta}$  是一个  $2 \times 6$  的矩阵。这两部分描述了当相机发生微小运动后, 投影在第二个图像的像素点变化过程, 因此, 称  $J$  为雅克比矩阵, 表示图像残差对位姿的迭代增量的求导。上面的运算过程为非线性最小化二乘问题, 可以用高斯牛顿迭代法以及 Levenberg-Marquardt 算法求解位姿的迭代增量。

### 4.2.3 直接法分类

基于直接法的定位系统, 通常根据处理的像素点的个数不同, 可分为三类: 稀疏直接法, 半稠密直接法, 稠密直接法。其中:

稀疏直接法, 也称作半直接法 (semi-direct method), 该方法的代表系统在第三章中介绍过的 SVO 系统。顾名思义, 稀疏直接法对图像提取稀疏的特征点, 同时假定提取出的像素点附近的像素点的亮度信息均保持不变。由于处理的数据量少, 且避免计算特征描述子以及特征匹配等处理过程, 因此算法的实时性能非常好, 适用于对实时性能要求较高, 并且计算资源有限、处理能力较差的平台。也因此导致该方法只能构建稀疏地图, 在利用构建的地图做后续路径规划及避障任务时难度加大。

半稠密直接法 (semi-dense method), 正如上述有关直接法工作原理的介绍, 直接影响直接法对运动估计结果的是采用存在梯度的像素信息进行处理, 因此, 该方法只采用存在梯度且梯度较为明显的像素点进行运动估计。

稠密直接法 (dense method), 稠密直接法采用图像的所有像素进行处理, 因此该方法处理的数据量过大, 导致实时性能非常差, 但其构建的地图相对完整。正如半稠密直接法中介绍, 使用所有信息并非全部是有效信息, 因此, 冗余的信息导致很难执行地图重建及位姿优化过程。

综上所述, 对比分析三种直接法, 根据处理数据量进行分类, 算法的实时性能与处理的数量成正比关系, 构建地图的完整性正好相反。结合本课题的应用平台及系统需求, 为满足实时性能, 本课题选用半直接法完成运动估计。

## 4.3 基于半直接法的立体视觉 SLAM 系统设计

### 4.3.1 视觉 SLAM 系统方案选取

对现有的 SLAM 系统结合论文以及开源出的代码进行实际测试, 综合分析, ORB-SLAM2 系统中基本延续了 PTAM 的算法框架, 但对框架中的大部分组件都做了改进, 在工程角度也更加适合应用、平台移植。ORB-SLAM2 采用 ORB 特征来做跟踪、地图创建、重定位和回环检测。同时通过优化姿态图来使环路闭合。

测试对比着几种先进的 SLAM 系统得出结论, ORB-SLAM2 可以实时运动在 PC 上, 适用于各种场合, 室内的或者室外的, 大场景或者小场景。同时该系统采用了所有 SLAM 相同的功能: 追踪, 地图构建, 重定位和闭环检测、甚至全自动位置初始化。选用了比较适合策略, 地图重构的方法采用云点和关键帧技术, 具有很好的鲁棒性, 生成了精简的、可追踪的地图, 当场景的内容改变时, 地图构建可持续工作。

对 ORBSLAM2 系统进行测试发现, 该系统仍有所有基于特征点方法的 SLAM 系统的共性问题: 移动端应用的实时性以及图像模糊跟踪丢失等问题。尽管 ORBSLAM 系统在 PC 机以 30ms/帧的速度进行实时计算, 但在嵌入式平台上表现不佳, 针对应用平台是四旋翼无人机, 仍不能实现实时处理。由于无人机在发生大机动的情况下会导致获取的图像模糊, 在这种情况下, 定位系统的鲁棒性较差、经常出现跟踪丢失的情况。

ORB-SLAM2 系统, 该系统是一个基于特征识别的 SLAM 系统, 正如传统的基于特征点的方法有一个缺点, 该方法需要提取特征点、对其计算描述子、以及特征匹配等步骤, 这些处理是很耗时的, 在移动端、处理能力较差的嵌入式平台不满足实时性。同时针对鲁棒性也就是图像模糊跟踪丢失无法继续定位的问题, 是由于获取的图像模糊很难提取到特征点, 找不到足够的匹配点计算相机运动导致的。课题现在的主要目的就是解决针对无人机应用平台使用 ORBSLAM2 系统存在的问题进行改进。

同时, 使用摄像机个数不同对系统实际应用存在困难也不同。单目摄像机的缺点是环境的真实尺度信息不能单独通过单目相机来确定, 想获得真实的环境信息需要额外的传感器辅助完成(例如 IMU、双目摄像机等)。针对单目相机存在的限制, 双目摄像机的应用, 不仅可以确定场景中的三维深度信息, 还可以估算出精确的轨迹以及环境地图, 在解决四旋翼定位问题上起到关键性的作用。因此, 立体视觉技术广泛的应用于未知场景下的导航信息获取, 这使得立体视觉定位系统在室内和室外环境均可以使用。

综上所述, 针对应用在无人机平台使用 ORBSLAM2 系统存在的问题, 根据上述对直接法(direct-method)和特征法(feature-method)的对比分析, 可以发现直接法可以完全不用计算关键点和描述子, 既避免了特征的计算时间, 也避免了特征缺失的情况。由于计算关键点和描述子以及特征匹配占用整个系统的一大半时间, 所以直接法可以很大程度提升实时性。同时只要满足场景中存在明暗变化, 直接法就能工作, 可以一定程度上改善特征点法由于图像模糊导致的特征缺失无法准确跟踪的问题。同时前面介绍过直接法分为稀疏、半稠、稠密三种, 对于应用在

实时性较高而且计算资源有限的无人机平台上，稀疏直接法(semi-direct)是最好的选择。稀疏直接法方法另一大优势是它利用特征块的配准来对直接法估计的位姿进行优化。因此速度快，鲁棒性好，更适用于该课题的需求。同时，双目视觉系统提供更为准确的环境尺度信息，以及摄像机深度估计值，因此本课题采用基于半直接法的视觉 SLAM 系统—SVO 框架为基础，并将其改进优化应用于双目视觉中。

#### 4.3.2 基于半直接的双目匹配算法设计

经过研究测试，得出结论，利用直接法进行运动估计的方案处理速度更快、效率更高，满足系统实时性的需求，该方法更适合应用于四旋翼无人机等机载控制器处理能力较差的平台上。近年来，随着 SVO, LSD-SLAM、DSO-SLAM 这些基于直接法的 SLAM 系统的流程，直接法本身也得到越来越多的关注。正如 4.3.1 节中介绍，本课题采用双目立体视觉进行位姿估计，但现有的整幅图像的直接法(direct-method)常见于 RGB-D 相机，稀疏直接法(semi-direct)方法常应用于单目视觉中。结合已看过的文献和综述，直接法还没有应用在双目系统上。但是，对于现有的双目系统获取空间点 3D 位置信息常采用的还是特征法(feature-method)，根据提取出的特征点进行左右图像匹配，获取视差图以及深度信息，为提升整个系统的实时性，尽量避免使用特征点进行位姿估计，由于本系统确定采用双目摄像头，所以在不进行特征匹配的前提下怎么进行双目匹配估计深度，是待解决的主要问题之一。

本文为了解决半直接法应用于双目视觉系统存在的问题，使得本课题设计的定位系统可以准确估计出地图相对真实环境的尺度信息，提出了一种基于半直接法的双目视觉系统估计 3 维点深度的方法，本方法的关键是一种基于半直接法的双目匹配方法设计，本算法主要分为以下几步：

(1) **初始化过程** 使用第一时刻左图像为基准，构建金字塔，提取 FAST 特征点，如果第一帧图像提取的特征点数满足一定阈值，然后根据第一时刻右帧图像用左帧图像的特征点做光流跟踪，解算该时刻左右两帧图像之间的单应矩阵，同时求解出当前跟踪到的特征点的内点数，如果内点数满足一定阈值，则标记设置该时刻的左帧图像为关键帧，添加到地图中，用于地图扩展。同时对该时刻左右两帧图像进行最小化图像重投影残差计算，估计初始时刻左右两帧之间的内点对应的空间点的深度。否则，使用后面时刻的图像重新进行初始化。图 4-2 中显示初始化过程的大致流程。

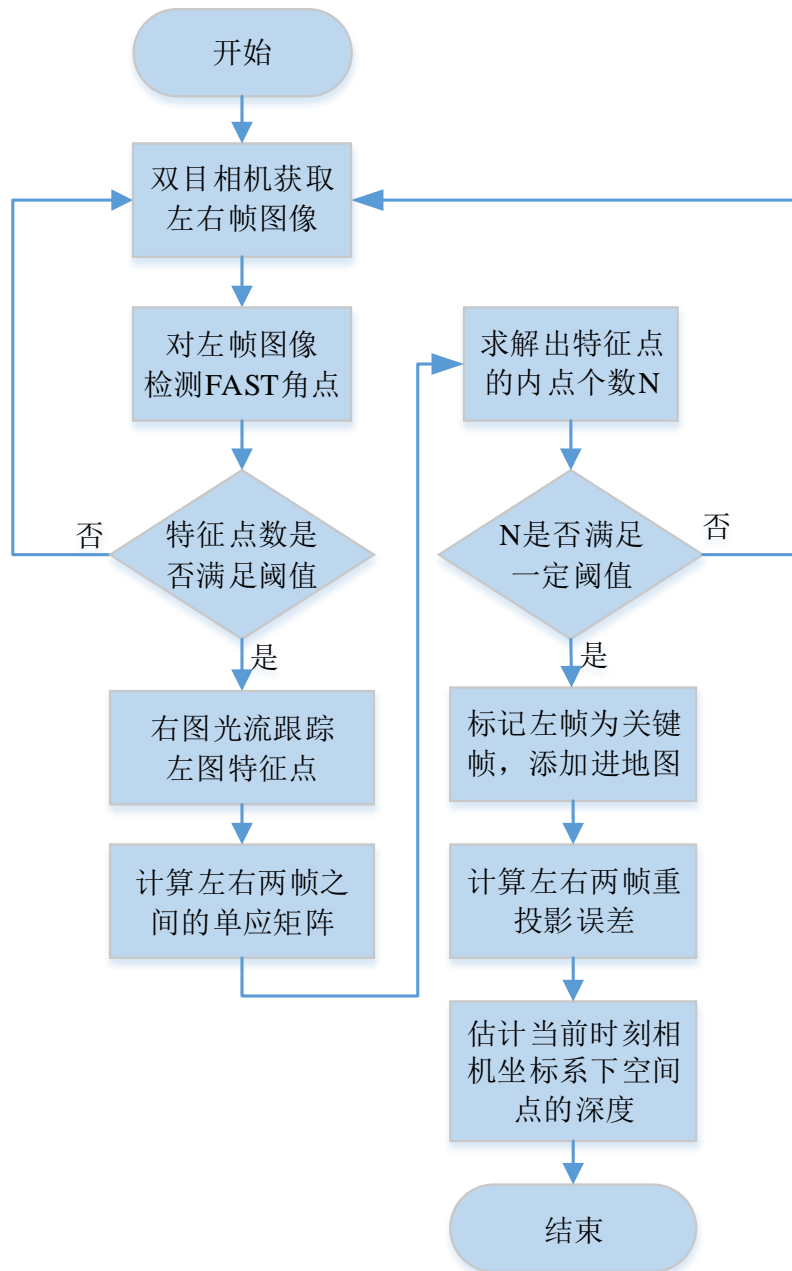


图 4-2 基于半直接法的双目匹配算法具体流程

(2) **初始位姿估计** 初始位置通过相机标定，平行双目视觉中左右相机之间的相对位姿  $T_{l,r}$  已经确定，一般初始化为当前时刻  $k$  左帧图像的位姿或者假设为单位矩阵。通过之前多帧之间的特征检测以及深度估计，可得知当前时刻  $k$  左帧中特征点位置以及他们的深度，还可估计出当前时刻  $k$  右帧中特征块的初始位置。该过程的基本原理为多视图几何原理<sup>[65]</sup>，其原理示意图如图 4-3 所示。

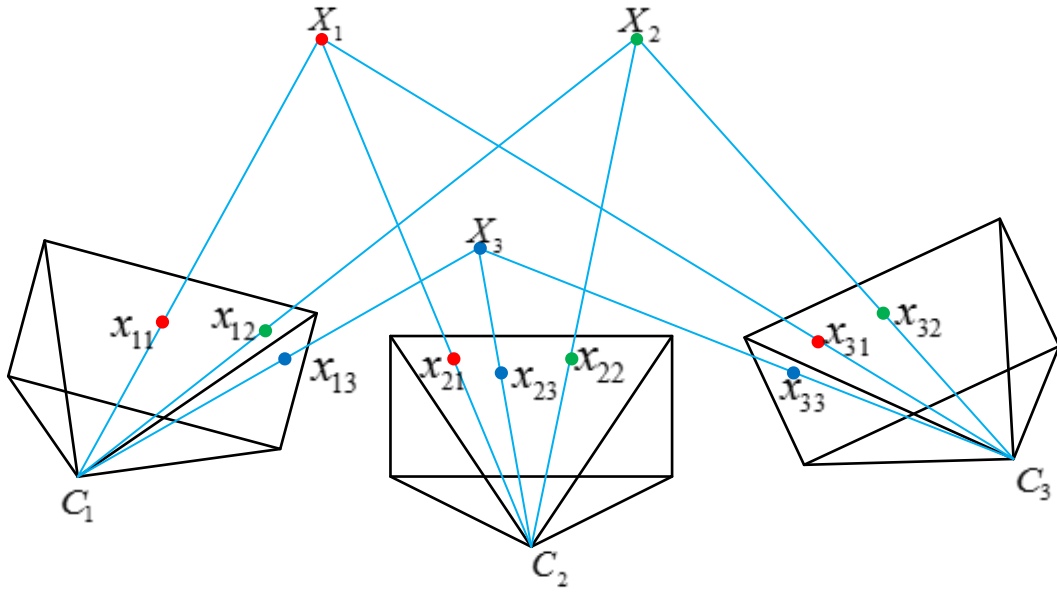


图 4-3 多视图几何原理

(3) **重定位** 知道当前时刻左帧图像  $I_{k,l}$  中的某个特征在图像平面的位置  $(u,v)$ ，需其深度  $d$ ，能够将该特征投影到三维空间  $p_{k,r}$ ，该三维空间的坐标系是定义在当前时刻左摄像机  $I_{k,l}$  坐标系的。所以，我们要将它投影到当前时刻右摄像机  $I_{k,r}$  图像中，需要进行位姿转换  $T_{l,r}$ ，得到该点在当前时刻右摄像机坐标系中的三维坐标  $p_{k,r}$ 。最后通过摄像机内参数，投影到当前时刻右帧图像  $I_{k,r}$  的图像平面  $(u',v')$ ，完成重投影。该过程原理示意图如图 4-4 所示。

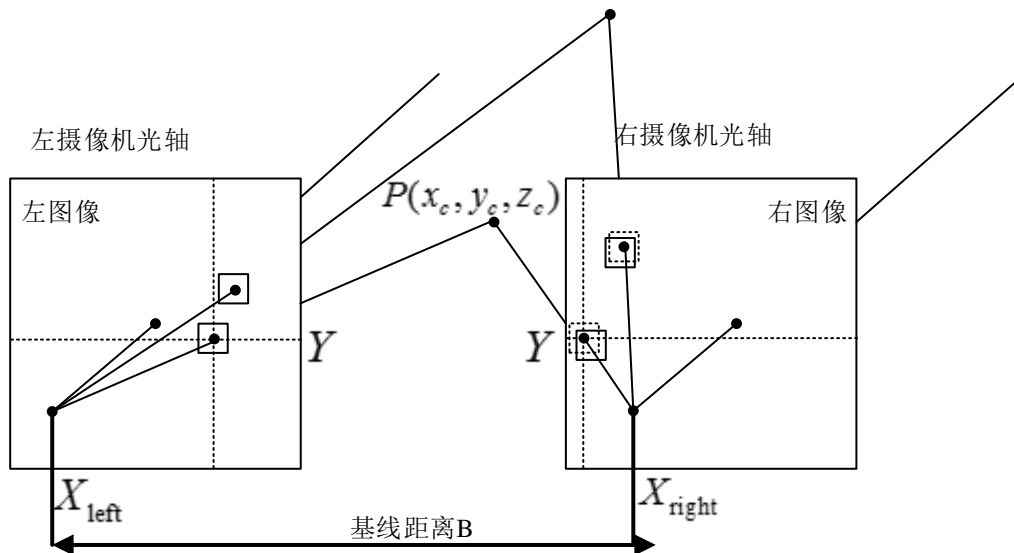


图 4-4 3D 点被投影到图像中对应 2D 点特征位置的估计原理图

(4) **迭代优化更新 3D 点的深度** 按理来说对于空间中同一个点，被极短



时间内的相邻两帧拍到，它的亮度值应该不发生变化。但由于深度是假设的一个值，所以重投影的点不准确，导致投影前后的亮度值是不相等的。不断优化位姿使得残差最小，就能得到优化后的深度信息  $d$ 。

将上述基于半直接法的双目匹配过程公式化如下：通过不断优化空间点深度最小化残差损失函数。

$$\min_d \frac{1}{2} \sum_{i \in \mathcal{R}} \|\delta I(T_{k_l, k_r}, u_i)\|^2 \quad (4-28)$$

其中：

$$\delta I(T_{k_l, k_r}, u_i) = I_{k_r}(\pi(T_{k_l, k_r} \cdot \pi^{-1}(u, d_u))) - I_{k_l}(u) \quad (4-29)$$

结合半直接法通过迭代求解式(4-28)，优化图像中特征点对应的 3 维空间中的深度变量。

#### 4.3.3 SLAM 系统总体方案

该基于半直接法的双目定位系统前端的流程图，如图 4-5 所示。

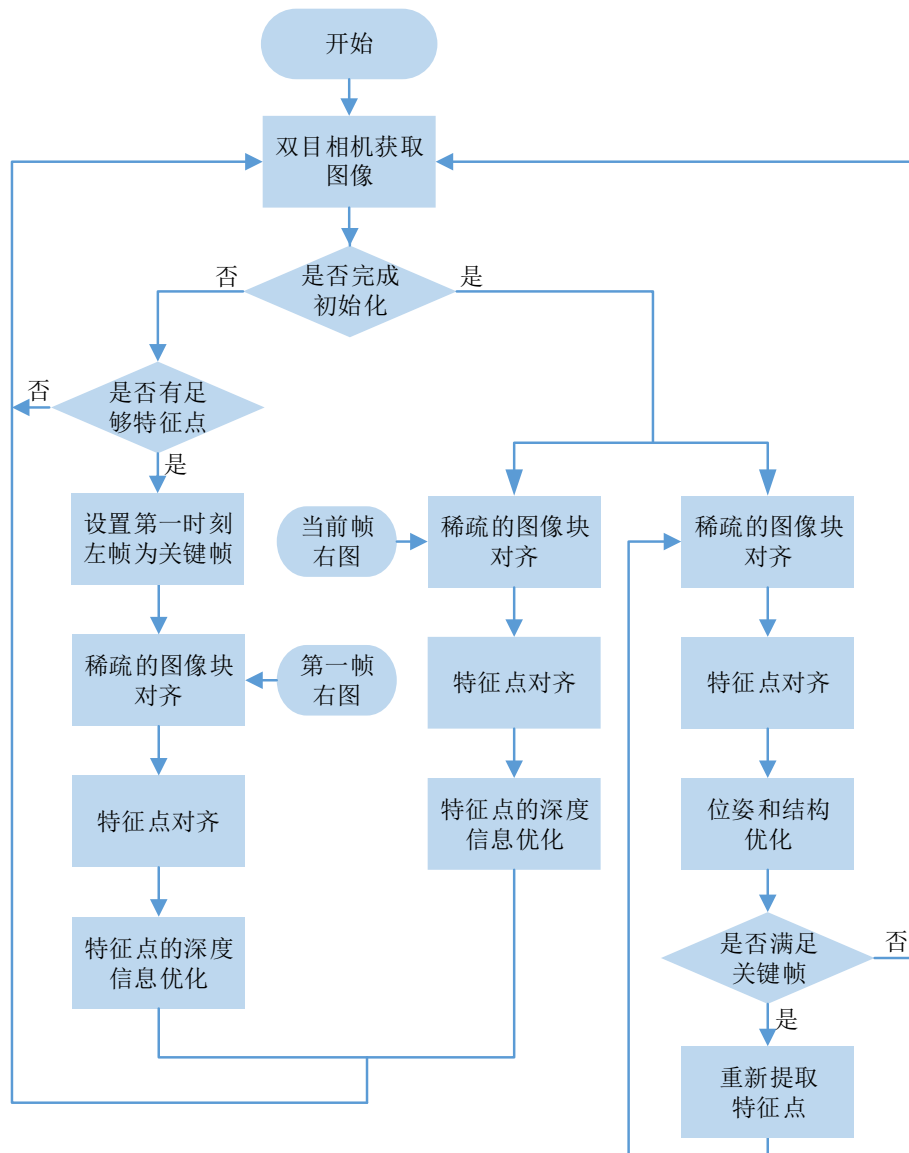


图 4-5 基于半直接法的双目定位系统前端的流程图

## 4.4 本章小结

本章首先介绍了 Lucas-Kanade 光流跟踪以及直接法运动估计的基本原理，然后针对现有典型的 ORB-SLAM 系统和 SVO 系统的优缺点分析，确定本课题以 SVO 系统框架为基础，对算法进行一定改进。设计了基于半直接法的双目视觉匹配算法迭代优化三维点的深度值，准确估计构建地图与真实环境的尺度因子。然后，完成在单目 SVO 系统框架的基础上，设计并实现成为基于双目视觉的 SLAM 系统，并在 Linux 操作系统以及 ROS 机器人操作系统中编程实现。

## 第 5 章 实验设计与分析

### 5.1 引言

为了验证第四章中提出算法的有效性和正确性，本章在实际实验室室内环境中对基于半直接法的四旋翼无人机双目视觉 SLAM 系统进行相关实验研究。首先，主要介绍本系统软硬件平台的搭建，包括硬件系统选型以及软件环境搭建。其次，基于真实的室内场景，在两个不同的实验场景中，进行四旋翼无人机视觉 SLAM 系统的分析实验，实现无人机的准确位姿估计以及特征点地图创建功能。依据这两组实验运行结果，本实验从实时性以及鲁棒性等两个角度验证本系统的性能。最后，结合实验数据中提供的真实路径，采用 RGBD-SLAM 系统中提出的测量轨迹局部精度的相对姿态误差（Relative pose error, RPE）指标，计算所有时间内相对姿态误差平移分量的均方根误差（RMSE）。依据实验结果分析，进一步验证本系统实现了机载 SLAM 的在线处理，且对运动估计鲁棒性更高，对飞行环境的适应能力更强。

### 5.2 实验平台搭建

为验证本课题提出的基于半直接法的双目视觉 SLAM 系统在四旋翼无人机上运行的可行性，所搭建的硬件平台以及软件环境需满足系统整体的需求。系统主要由硬件系统选型和软件基本框架组成。硬件系统主要包括四旋翼无人机、视觉传感器以及数据处理平台的选型。软件系统主要包括系统开发环境、辅助依赖库、以及地图绘制工具。

#### 5.2.1 硬件系统搭建

##### (1) 四旋翼无人机机体选型

本课题由于采用平行式双目立体视觉定位的模型，所以为满足平行式模型的要求，选取的两个摄像头要满足相互平行的连接方式。同时，本课题的另一个主要贡献是为了解决无人机出现大机动、快速运动情况下的准确定位问题，这种情况会导致在四旋翼平台上固定摄像机相对位置发生变化，导致模型不匹配，所以选择内置双目摄像头的大疆 Matrice100 飞行器。图 5-1 为大疆 M100 飞行器实物图。



图 5-1 Matrice100 飞行器实物图

该飞行器的参数如下：最大外形尺寸（含保护架） $1208 \times 1208 \times 230\text{mm}$ ，最大俯仰角度为 $35^\circ$ ，最大俯仰角速度为 $300^\circ/\text{s}$ ，最大偏航角速度为 $150^\circ/\text{s}$ ，最大上升速度为 $5\text{m/s}$ ，最大下降速度为 $4\text{m/s}$ ，最大水平速度为 $15\text{m/s}$ 。同时，飞行器速度控制的精度可达到 $0.2\text{m/s}$ ，相机云台角速度控制的精度可达到 $0.1^\circ$ ，程序指令的频率为 $50\text{Hz}$ 。当开放程序接口时，飞行器默认以 $50\text{Hz}$ 的频率读取各状态参数，如速度，加速度，姿态角等，并以 $50\text{Hz}$ 的频率给飞控发送指令信号，飞控接收指令后调用内部封闭的控制算法去执行指令并保持飞行器的平稳飞行。

## (2) 数据处理平台

本系统的工作机制是对无人机机载视觉传感器输入的图像数据，经过图像处理并结合计算机视觉的方法完成无人机的运动估计。正如本文第三章所述，本系统从实时性、鲁棒性等两个角度出发，设计并实现一个满足无人机机载处理器处理能力的视觉 SLAM 系统。因此，本实验为验证系统的计算效率，需确定数据处理平台完成整个系统的数据处理过程。其中数据处理平台负责几乎整个系统的数据计算，如传感器数据融合、位姿解算。

由于本课题试验的设计，涉及到多线程处理的 ORB-SLAM 系统与本文提出系统的对比实验，因此，为充分验证各个系统的准确性并体现其优势，本系统采用笔记本电脑 ThinkPad S420 作为算法实现及数据处理平台。该款笔记本电脑的配置是 Inter(R) Core(TM) i5-2450M CPU @ 2.50GHz，并且配备四核处理器满足 ORB-SLAM 系统的多线程工作机理。

### (3) 双目视觉传感器

本系统采用大疆 Matrice100 飞行器作为实验平台，且搭建了大疆首个量产的无人机视觉传感器导航系统—Guidance。Guidance 模块配备了视觉超声波组合传感器以及五组立体视觉系统，可感知 20 米范围内的物体，实时监测多个方向的环境信息，并帮助飞行器避免碰撞。图 5-2 中显示的即是 Guidance 模块及其配备的其中一组双目摄像机。这五组双目视觉系统实时输出双目摄像头采集的 8 比特灰度图，分辨率为  $320 \times 240$ ，视角在  $60^\circ$  左右，其图像采集频率为 20Hz。由于本系统确定采用 SVO 系统的基本框架，结合本文第三章对 SVO 系统的理论分析，可确定本系统更适合使用下视相机采集图像，完成运动估计。因此本系统选取五组中的下视相机录制的视频信息进行实验。



图 5-2 M100 机载 Guidance 模块及其双目摄像机

## 5.2.2 软件系统搭建

### (1) 开发环境及三方库依赖

本实验的软件环境是在笔记本电脑 ThinkPad S420 的 Ubuntu14.04 操作系统中完成搭建。同时，本文设计的视觉 SLAM 系统是在机器人操作系统 ROS(Robot Operating System)的 Indigo 版本下开发实现。ROS 操作系统对无人机导航定位提供非常强大的支持，这个可以在没有详细先验知识的前提下完成对无人机导航、定位、路径规划等任务。

为了方便本系统的开发以及系统后续维护和二次开发，本系统在实现过程中调用已开源的三方库。例如，提供相关视觉方法、以及图像处理的基础计算机视

觉库 OpenCV; 主要用于 BA 的优化的 g2o 工具, 提供用于优化基于图形的非线性误差函数的开源 C++ 框架; 提供矩阵相关运算的 Eigen 库; 实现李群李代数中平移、旋转、投影变换的 Sophus 库, 以及特征提取的 Fast 库, 该库对 fast 角点检测是基于 `see` 指令实现, 速度快, 利于提升系统实时性。

## (2) 地图绘制

本课题的系统是在机器人操作系统 ROS 下开发实现的, 由于 ROS 支持机器人定位与导航, 其完成的关键也是第一步就是要建立一个封闭环境的地图。因此, ROS 提供了非常强大的图形化模拟环境 Rviz, 可利用 ROS 中的 Rviz 工具实现对机器人在路径探测定位过程中对地图动态观测, 实时显示其位置和姿态信息, 并绘制环境地图。

本课题以设计一个实时性能好、鲁棒性强的定位系统为目的, 因此, 本实验环境仅用于测试算法执行效率以及定位精度。由于无人机硬件条件存在过多的限制, 为满足系统可在处理能力较差的无人机机载处理器上实时执行的需求, 所以本系统选择构建简单的稀疏特征点地图。稀疏地图绘制过程中, 只需标记显示提取的路标点、四旋翼无人机的运动路径以及当前时刻对应视觉传感器的姿态信息。该系统绘制的稀疏地图如图 5-3 所示, 其中蓝色线显示相机运动轨迹以及选取关键帧的位姿、蓝色四棱锥显示当前时刻相机位姿以及红色显示用于构建地图中的路标点。

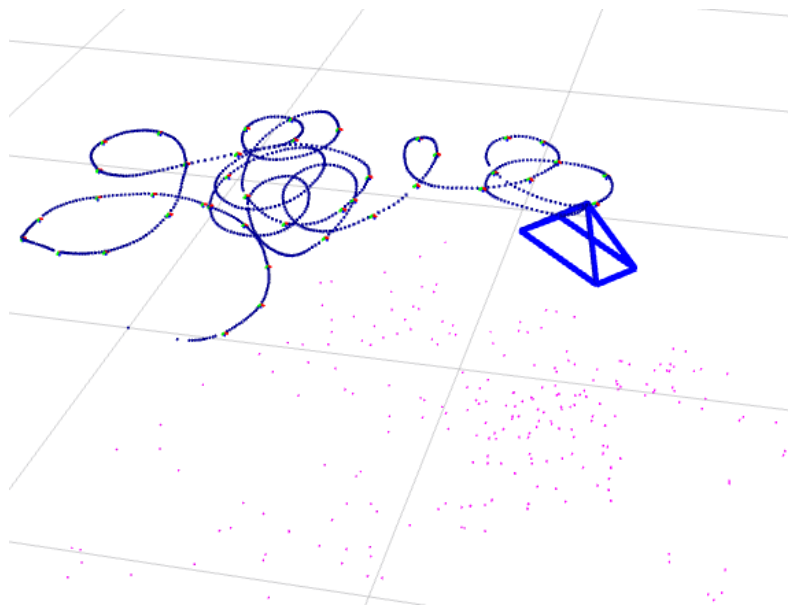


图 5-3 本系统绘制的稀疏地图

### 5.2.3 系统定位精度评估指标

视觉 SLAM 系统通常输出摄像机的估计轨迹以及所构建的运行环境地图。虽然原则上可以评估所得地图的质量，但准确的地面实况图很难获得。因此，本实验采用从图像给定的输入序列中评估估计轨迹的质量，这种方法大大简化了评估过程。在对系统定位精度指标进行测试时，采用 RGBD-SLAM 系统中提出的概念，相对姿态误差（Relative pose error, RPE），RPE 表示在固定时间间隔  $\Delta t$  上测量轨迹的局部精度。相对姿态误差对应于轨迹的漂移，其对于评价视觉定位系统特别有用。对于评估，我们假设从估计的轨迹中给出一系列相机姿态  $P_1, \dots, P_n \in SE(3)$ ，以及地面的真实轨迹  $Q_1, \dots, Q_n \in SE(3)$ 。简化表示方式，假设序列是时间同步的，并对其均匀采样，且对应时间步长相同的图像序列为  $n$  个。定义每个时间步长  $i$  之间的相对姿态误差为

$$E_i := (Q_i^{-1} Q_{i+\Delta t})^{-1} (P_i^{-1} P_{i+\Delta t}) \quad (5-1)$$

一般采用从所有时刻的相对姿态误差，计算所有时间内其真实轨迹和估计轨迹之间在平移分量上的均方根误差（RMSE）。RMSE 误差可表示为

$$\text{RMSE}(E_{1:m}, \Delta t) := \left( \frac{1}{m} \sum_{i=1}^m \|\text{trans}(E_i)\|^2 \right)^{\frac{1}{2}} \quad (5-2)$$

其中， $\text{trans}(E_i)$  表示相对姿态误差  $E_i$  的平移分量。

本实验采用该方法，计算以 m/s 为单位的位姿平移分量漂移的均方根误差以及位姿平移分量漂移的均值。

## 5.3 实验设计及结果

为了体现本课题提出的基于半直接法的双目视觉 SLAM 系统在实时性和鲁棒性两方面的优势，设计了在两种不同的实验环境下对整个系统进行性能评估实验。由于本系统应用于四旋翼无人机平台，分别使用具有代表性的 EuRoC 数据集，其提供四旋翼飞行器上采集的视觉惯性数据集，评估系统的执行效率、重定位和程序运行能力；使用用在实验室室内遥控大疆 M100 无人机任意飞行获取图像序列，评估整个系统的性能。

### 5.3.1 EuRoC 无人机数据集

EuRoC 是一个通过刚性连接在无人机机体上的平视双目摄像机记录的数据集，该数据集包含同步的双目图像、IMU 测量数据以及通过动态捕捉系统获得的地面实际轨迹数据。EuRoC 提供了 ETH 机器大厅数据以及室内环境的 11 个数据集，



本实验选取是一个室内房间的 V1\_01\_easy 数据集，该视频中无人机的飞行路径长 58.6m，平均在地面以上 0.8m 处飞行，平均飞行速度达到 0.4m/s，该图像的分辨率为 752×480。该数据集包含几个回环和大角度的旋转，这对视觉定位技术非常具有挑战性。实验运行界面如图 5-4 所示。

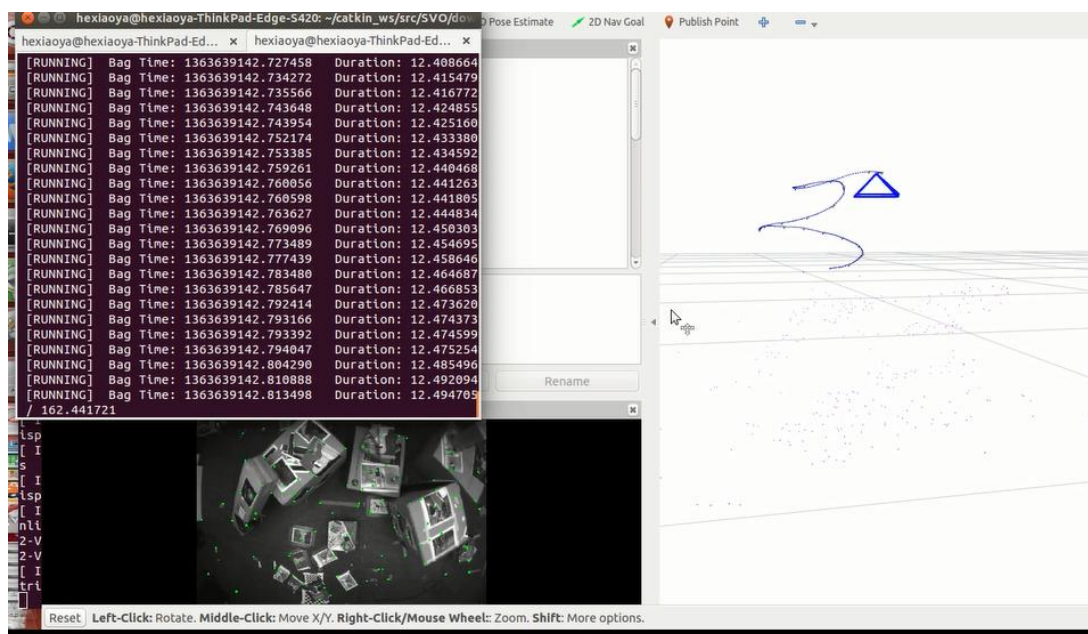


图 5-4 Stereo-SVO 运行界面

本实验选取 SVO 系统以及改进后基于半直接法的双目视觉 SLAM 系统（简称 Stereo-SVO）在笔记本电脑上执行。同时借鉴 SVO 系统的思想，采用两种不同的参数设置方式执行，将本课题提出的系统也按照两种不同参数设置进行实验，一个针对速度优化，一个针对精度优化。当实际应用于嵌入式平台时，仅使用快速参数设置。实验具体参数设置如表 5-1 所示。

表 5-1 Stereo-SVO 实验测试系统参数设置

	速度优化 Fast	精度优化 Accurate
每帧图像提取特征点	120	200
最多的关键帧数	10	50
开启局部 BA 优化	否	是

该实验将 SVO 系统的性能与本系统修改后的半直接法双目视觉系统进行对比，同时针对两种不同的参数设置，分别运行 SVO 与改进后的 Stereo-SVO 系统，并将其运行结果从实时性和鲁棒性两个方面进行对比。

#### (1) 运行时间评估

首先，采用该组实验数据运行，图 5-5 给出了分别利用两种参数设置的改进后的 Stereo-SVO 系统以及 ORB-SLAM 系统整个处理过程中每帧图像对应处理时间



的对比。表 5-2 示出这些系统处理每帧图像所用的时间以及每秒处理图像的帧数。

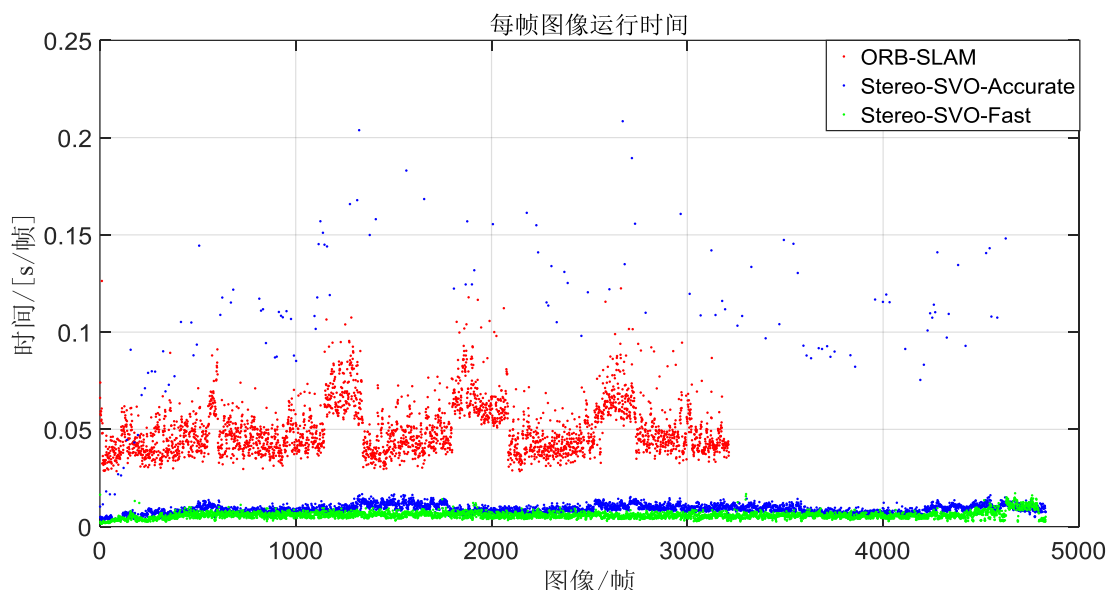


图 5-5 两种参数设置的 Stereo-SVO 及 ORB-SLAM 系统运行时间对比

表 5-2 两种参数设置的 Stereo-SVO 及 ORB-SLAM 系统处理每帧的时间均值

	双目 SVO-Fast	双目 SVO- Accurate	ORB-SLAM
处理每帧的时间均值 (帧/s)	0. 00585	0. 011500	0.04994
每秒处理的图像帧数 (fps)	140	80	20

采用笔记本电脑测试系统的实时性如图所示，图中红色的实点代表 ORB-SLAM 系统处理每帧图像所用的时间，相应处理频率为 20fps。蓝色的点代表 Stereo-SVO 系统在采用精度优化参数设置情况下处理每帧图像所用的时间，能够以大约每秒 100 帧 (fps) 的速度处理帧。绿色的点代表 Stereo-SVO 系统在采用速度优化参数设置情况下处理每帧图像所用的时间，能够以大约每秒 170 帧 (fps) 的速度处理帧。主要差别在于基于半直接法的 SVO 系统不需要在运动估计期间提取特征，当该处理过程耗时构成 ORB-SLAM 系统的大部分时间（相同笔记本电脑上处理每帧图像耗时 25ms）。从图 5-5 中可以明显看出，基于半直接法的 Stereo-SVO 系统在处理时间方面明显优于基于特征点法的 ORB-SLAM 系统。同时速度优化参数设置的 Stereo-SVO 系统在一定程度上提升了实时性。在我们应用于嵌入式操作平台时，选择采用速度优化的参数设置方式。

该系统可以使用较少的特征可靠地跟踪摄像机的原因是使用深度滤波器，这确保了被跟踪的特征位置信息是可靠的。运动估计准确的参数设置在笔记本电脑上平均需要 17ms。运行时间的增加主要是由于局部 BA 优化，其在每个关键帧运

行并且完成优化过程仅花费 18ms。地图构建线程用新的图像帧来更新所有深度过滤器所需的时间高度依赖于过滤器的数量。在选择关键帧之后，滤波器的数目较多，并且在滤波器收敛时迅速减小。平均来说，地图构建线程比运动估计线程处理速度更快，因此它不是影响系统实时性的限制因素。因此，本系统的实时性相比其他系统更好，计算效率更高，满足本课题的需求，可实际在处理能力较差的无人机机载处理器上进行处理。

## (2) 鲁棒性

该实验采用的视频中,无人机飞行的真实轨迹是通过动捕系统获得的。该测试视频无人机的飞行路径长 58.6m，平均在地面以上 0.8m 处飞行。图 5-6 给出了四旋翼无人机随时间变化的真实路径、两种参数设置的基于半直接法的 Stereo-SVO 系统以及 ORB-SLAM 系统得到的飞行轨迹。

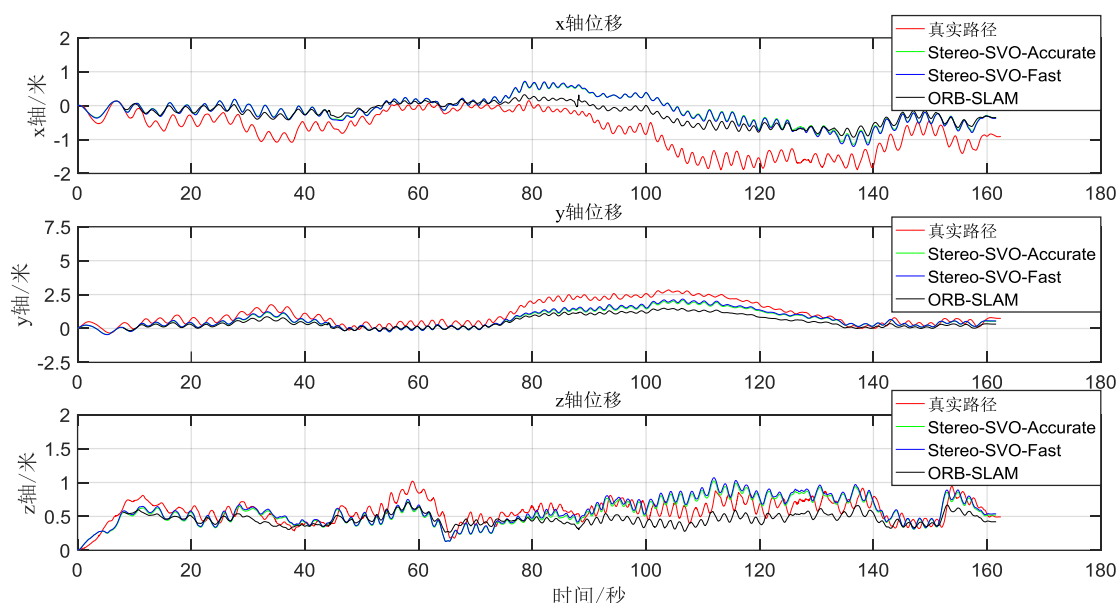


图 5-6 两种参数设置的 SVO 及 ORB-SLAM 系统定位精度对比

总体而言，改进后的 Stereo-SVO 的两个版本比 ORB-SLAM 系统的鲁棒性更强。从图 5-6 中可以看出，速度和精度的参数设置之间的定位精度差异并不显著，但系统在速度参数设置条件下运行，每次迭代时分别优化姿态和观察到的三维空间点的位置信息，对于无人机运动估计是相对准确的。

同时，本实验采用 5.2.3 节中介绍的评价指标对本系统进行定位精度评估。本实验采用该方法，计算以 m/s 为单位的位姿平移分量漂移的均方根误差以及位姿平移分量漂移的均值，如表 5-3 所示。从表中数据可知，改进后的 Stereo-SVO 鲁棒性有所提高，对环境的适应能力较强。

表 5-3 两种参数设置的 SVO 及 ORB-SLAM 系统定位误差

	位置的均方根误差[m/s]	位置误差的中值[m/s]
Stereo-SVO-Fast	0.0059	0.0047
Stereo-SVO-Accurate	0.0051	0.0038
ORB-SLAM	0.0164	0.0142

### 5.3.2 实验室室内场景图像序列

为验证本课题设计的视觉 SLAM 系统实际应用于四旋翼无人机平台的有效性,本实验通过遥控无人机在实验室室内环境中任意飞行,采用大疆 M100 无人机机载 Guidance 模块内置的双目摄像机拍摄实验室室内场景,录制成 ROS 的 bag 包形式,该数据包包含同步的双目图像、IMU 测量数据以及通过 Guidance 模块获得的无人机轨迹数据。该实验主要采用 Guidance 模块的下视双目摄像头拍摄的一组视频流。

本实验录制的是一个实验室室内房间的数据集,该视频流中无人机的飞行路径长 28m,平均在地面以上 1.5m 处飞行,平均飞行速度达到 0.5m/s,该图像的分辨率为 320\*240。该视频包含几个回环运动以及出现无人机发生机动的情况,无人机发生机动,导致录制的视频出现模糊情况,这对视觉定位技术非常具有挑战性。实验运行界面如图 5-7 所示。

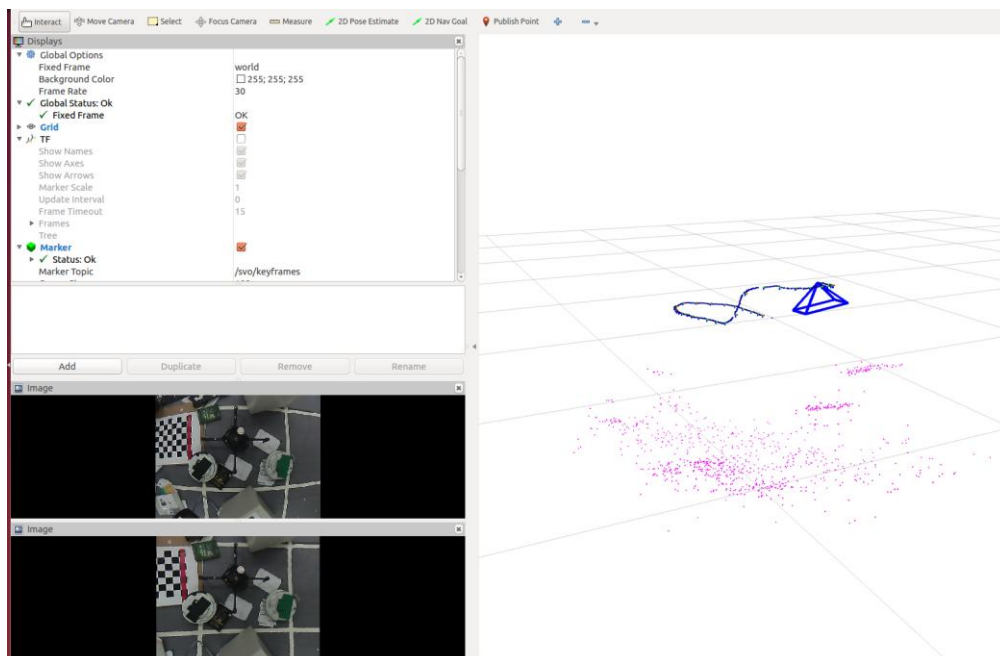


图 5-7 实验室室内场景图像运行界面

#### (1) 鲁棒性

该实验采用的视频中,无人机飞行的真实轨迹是通过 Guidance 模块获得的。

图 5-8 给出了在实验室室内环境中，四旋翼无人机随时间变化的真实路径、基于半直接法的 Stereo-SVO 系统以及原始 SVO 系统得到的飞行轨迹。

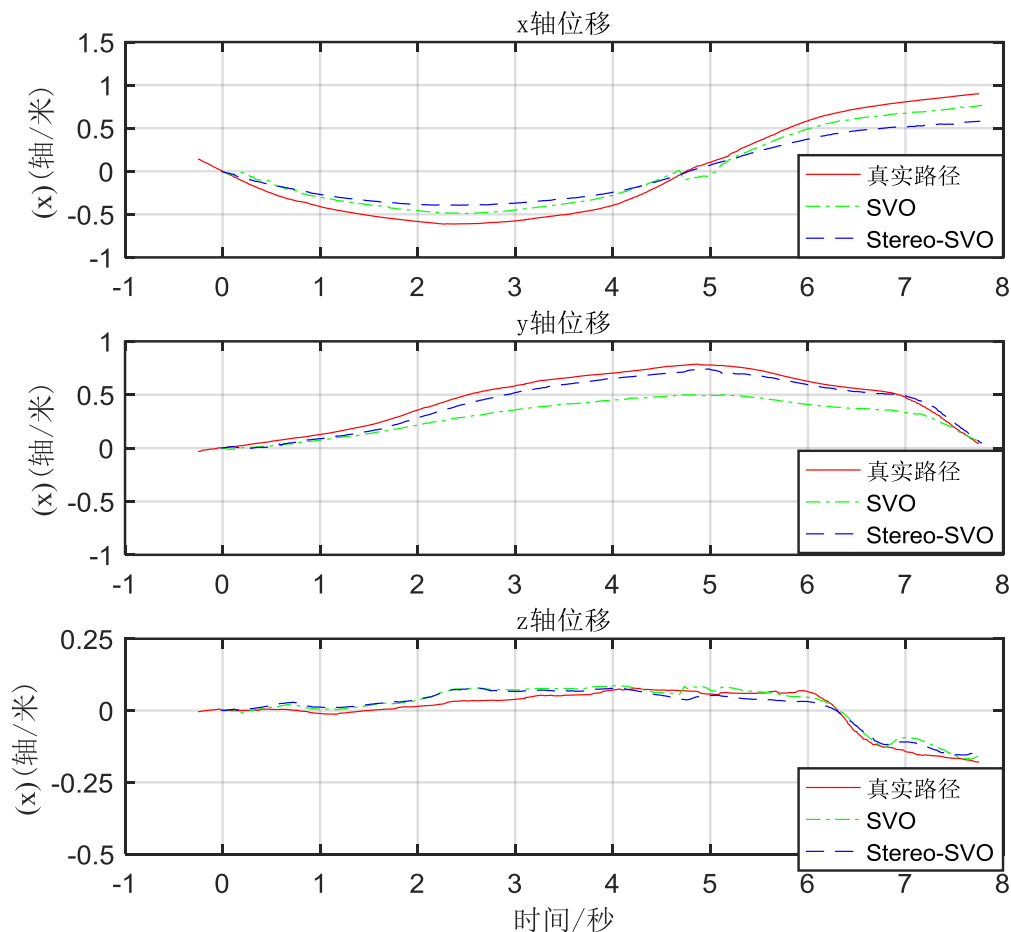


图 5-8 Stereo-SVO 以及 SVO 系统定位精度对比

从图 5-8 可以看出，本课题提出的 Stereo-SVO 相比 SVO 系统定位精度较高，图中蓝色的线表示改进的 Stereo-SVO 系统估计的无人机位置信息，绿色的线表示 SVO 系统估计的无人机位置信息。相比原始 SVO 系统，本系统采用双目视觉进行定位，可以恢复出场景的真实尺度，且得到特征点精度相对更高，因此，在程序对特征位置进行校正部分，不需要过于苛刻，视场内满足条件的特征数量不至于过少，导致不满足重定位条件，因此，改进的 Stereo-SVO 系统重定位能力更强，系统对于无人机快速运动的鲁棒性更好。

在实验过程中改进的 Stereo-SVO 系统与原始版 SVO 的版本相比还有一点优势，即原始 SVO 系统对采用平视相机进行运动估计的鲁棒性较差，由于 SVO 系统采用单目摄像机，无法确定环境真实尺度，估计出的三维空间点深度值不完全准确，存在一定尺度差异，且前视相机场景深度的影响较大，致使其原始 SVO 系

统在基于前视相机完成运动估计的鲁棒性较差，而本系统提出的 Stereo-SVO 采用立体视觉技术，解决了深度估计不准确的问题，并可以完整并确定的恢复场景尺度，因此其对环境的适应性更强。

结合实验结果可知，本系统提出的 Stereo-SVO 系统的实时性更强，满足在无人机等处理能力较差的处理器上运行，相比原始 SVO 系统的环境适应性，无人机快速运动的鲁棒性更强。同时，对相机安装方式要求更为灵活。综上分析，基于半直接法的双目视觉定位系统可实时应用于无人机平台，符合本文要求，在一定程度上解决了视觉 SLAM 系统实际应用的困难。

## 5.4 本章小结

本章主要完成基于半直接法的双目视觉 SLAM 系统的实验验证。首先，通过合理的无人机硬件选型以及软件系统搭建，搭建整个实验平台。其次，通过两组不同的真实室内场景的图像序列，完成无人机视觉定位及建图的过程。由实验结果可得知，第四章中提出算法的可在机载处理器上实时处理，解决了 SLAM 系统在处理能力较差的处理器上实时性较差的问题。同时，将原始 SVO 系统应用于双目视觉中，解决了其深度估计不准确，且单目视觉无法估计真实环境尺度的困难，使该系统在前视摄像头工作的情况下，系统更容易重定位成功，提升了整个系统的鲁棒性以及适应性。

## 结 论

本文针对在未知环境中随意运动的四旋翼无人机，较为深入的分析了现有的视觉 SLAM 系统应用于无人机完成定位及地图构建等任务存在的困难，然后结合视觉 SLAM 系统根据不同的运动估计方法进行分类的结果，完成了对现有典型的开源视觉 SLAM 系统的实际测试，分别从实时性以及鲁棒性两个不同的角度对比分析了测试结果，总结了现有典型系统的不足，并针对存在的不足进行了改进，为实现一个满足嵌入式平台适应性的视觉 SLAM 系统提供基础。基于以上分析及研究，从达到实时处理效率的角度出发，确定采用半直接法完成运动估计，提出了基于半直接法的双目视觉定位系统的设计方法以及实现方案，并编程实现。从提高鲁棒性的角度出发，采用双目视觉传感器进行三维空间点位姿估计的方法，实现了可估计真实环境尺度信息的双目匹配算法。并通过实验验证了本系统的有效性。基于本文的工作内容，可归纳如下：

(1) 本文利用概率分布模型对 SLAM 系统进行数学描述，建立了四旋翼飞行器的位姿模型、视觉系统的基本模型及其相关坐标系。结合相机模型，重点分析了视觉 SLAM 系统的几何原理，针对视觉 SLAM 系统应用于四旋翼无人机平台存在的难点进行归纳总结。

(2) 结合前文对系统实际应用于无人机平台存在困难的描述，确定本课题的实际需求，给出系统的指标要求。依据运动估计方法，将主流的视觉 SLAM 系统分为基于滤波器、基于关键帧 BA 以及基于直接法的视觉 SLAM。根据分类结果，实际测试现有先进地具有代表性地系统，通过实验验证了这些典型系统的优缺点，兼顾系统实时性能和鲁棒性能，最终确定本课题的整体系统方案以及框架。

(3) 提出了一种基于半直接法的双目视觉匹配算法。首先，针对现有典型的基于特征点法的 ORB-SLAM 系统计算量大和在强烈旋转下鲁棒性差等缺点，以及基于直接法的 SVO 系统对采用前视摄像机进行运动估计鲁棒性差和无法估计真实环境尺度等缺点，从实时性以及鲁棒性两个角度考虑，借鉴半直接法 SVO 系统的思想，采用半直接法完成双目特征匹配，在保持其高计算效率的同时，确定了空间三维点的精确位姿以及真实环境的尺度。然后，编程实现了基于半直接法的双目视觉 SLAM 系统。最后，通过相应的实验验证，充分说明了本课题设计的系统在无人机平台上应用的可行性。

综上所述，本论文对视觉 SLAM 系统的建模和方法做了较为深入的研究，同时围绕视觉 SLAM 系统应用于四旋翼无人机平台存在的问题，提出了一些

创新想法，并实际编程实现了整个视觉 SLAM 系统。本系统的实现为无人机视觉定位技术提供了一些参考和借鉴价值。但是，由于个人研究深度有限和时间不足等原因，课题研究仍可进一步深入，系统实现可进一步完善，许多问题可进一步研究，具体如下：

（1）本文只针对视觉 SLAM 系统在无人机平台上实际应用其实时性能较差的问题进行研究，对无人机在快速运动或纯旋转运动的情况下，其鲁棒性能较差的问题还没有研究解决，可进一步研究。

（2）本文只研究了基于视觉传感器数据的定位系统，对融合其他传感器数据，如无人机提供的 IMU 数据还未涉及，进一步考虑实现从图像数据、IMU 数据到位姿输出的 VIO 系统。

（3）对实现的视觉 SLAM 系统还未进行嵌入式移植，仅停留在 PC 平台处理阶段，进一步可考虑完成实现使用无人机机载处理器进行实验。

## 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]. 实验力学, 2013(03): 326-332.
- [2] 黄亮, 刘忠, 石章松, 等. 无人机系统纯方位定位技术及应用[M]. 北京: 国防工业出版社, 2015: 1-8.
- [3] 林鹏宏. 四轴无人机多约束条件下的跟踪控制和轨迹规划方法研究[D]. 哈尔滨: 哈尔滨工业大学学位论文, 2016: 2-4.
- [4] Tuna G, Gulez K. Aided Navigation Techniques for Indoor and Outdoor Unmanned Vehicles[C]// International Conference on New Technologies, Mobility and Security. IEEE, 2012:1-4.
- [5] Ghadiok V, Goldin J, Ren W. Autonomous indoor aerial gripping using a quadrotor[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2011:4645-4651.
- [6] Mellinger D, Shomin M, Michael N, et al. Cooperative Grasping and Transport Using Multiple Quadrotors[M]// Distributed Autonomous Robotic Systems. Springer Berlin Heidelberg, 2013:545-558.
- [7] Michael N, Fink J, Kumar V. Cooperative manipulation and transportation with aerial robots[J]. Autonomous Robots, 2011, 30(1):73-86.
- [8] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(6):1052-67.
- [9] Choi H, Yang K W, Kim E. Simultaneous Global Localization and Mapping[J]. IEEE/ASME Transactions on Mechatronics, 2014, 19(4):1160-1170.
- [10] Weiss S, Achtelik M W, Chli M, et al. Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV[C]// IEEE International Conference on Robotics and Automation. IEEE, 2012:31-38.
- [11] Kushleyev A, Mellinger D, Powers C, et al. Towards a swarm of agile micro quadrotors[C]// Robotics: Science and Systems. 2012:287-300.
- [12] Durrant-Whyte H, Roy N, Abbeel P. Construction of Cubic Structures with Quadrotor Teams[C]// Robotics: Science & Systems VII, University of Southern California, Los Angeles, Ca, Usa, June. MIT Press, 2011:177-184.
- [13] 宋昱慧. 基于单目视觉的四旋翼无人机位姿估计与控制[D]. 深圳: 哈尔滨工业大学深圳研究生院学位论文, 2015: 1-2.
- [14] Zhenhe Chen, Jagath Samarabandu, Ranga Rodrigo. Recent advances in



- simultaneous localization and map-building using computer vision[J]. *Advanced Robotics*, 2007, 21(3-4):233-265.
- [15]Salasmoreno R F, Newcombe R A, Strasdat H, et al. SLAM++: Simultaneous Localisation and Mapping at the Level of Objects[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2013:1352-1359.
- [16]Grzonka S, Grisetti G, Burgard W. Towards a navigation system for autonomous indoor flying[C]// *IEEE International Conference on Robotics and Automation*. IEEE, 2009:2878-2883.
- [17]Du Z, Wu H. Research of Mobile Robot SLAM Based on Laser Range Finder and Binocular Vision Data Fusion[J]. *Computer Measurement & Control*, 2013.
- [18]Genevois T, Zielinska T. A simple and efficient implementation of EKF-based SLAM relying on laser scanner in complex indoor environment[J]. *Industrial Research Institute for Automation & Measurements Piap*, 2014, 8(2):58-67.
- [19]Blösch, M, Weiss, S, Scaramuzza, D, et al. Vision based MAV navigation in unknown and unstructured environments[C]// *IEEE International Conference on Robotics and Automation*. IEEE, 2010:21-28.
- [20]Jeong W Y, Lee K M. CV-SLAM: a new ceiling vision-based SLAM technique[C]// *Ieee/rsj International Conference on Intelligent Robots and Systems*. IEEE, 2005:3195-3200.
- [21]Roy N. Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments[J]. *Proceedings of SPIE - The International Society for Optical Engineering*, 2009, 7332(1):373-375.
- [22]Fraundorfer F, Heng L, Honegger D, et al. Vision-based autonomous mapping and exploration using a quadrotor MAV[C]// *Ieee/rsj International Conference on Intelligent Robots and Systems*. IEEE, 2012:4557-4564.
- [23]Huang A S, Bachrach A, Henry P, et al. Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera[M]// *Robotics Research*. Springer International Publishing, 2017.
- [24]Bylow E, Sturm J, Kerl C, et al. Real-Time Camera Tracking and 3D Reconstruction Using Signed Distance Functions[C]// *Robotics: Science and Systems*. 2013.
- [25]Pike J. Rq-4a Global Hawk (Tier I+ Hae Uav)[J]. *Global Security. ORG*, 2006.
- [26]Jane's I H S. Boeing A160/A160T Hummingbird[J]. *Jane's Unmanned Aerial Vehicles and Targets*, 2013.
- [27]Pounds P, Mahony R, Corke P. Modelling and control of a large quadrotor robot[J]. *Control Engineering Practice*, 2010, 18(7):691-699.

- 
- [28] Valenti M, Bethke B, Fiore G, et al. Indoor Multi-Vehicle Flight Testbed for Fault Detection, Isolation, and Recovery[C]// AIAA Guidance, Navigation, and Control Conference and Exhibit. 2006.
- [29] Michael N, Mellinger D, Lindsey Q, et al. The GRASP Multiple Micro-UAV Testbed[J]. IEEE Robotics & Automation Magazine, 2010, 17(3):56-65.
- [30] Müller M, Lupashin S, D'Andrea R. Quadcopter ball juggling[C]//2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2011: 5113-5120.
- [31] Huang A S, Bachrach A, Henry P, et al. Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera[M]// Robotics Research. Springer International Publishing, 2017.
- [32] Smith R, Cheeseman P. On the Estimation and Representations of Spatial Uncertainty[C]// International Journal of Robotics Research. 1987:1-22.
- [33] Durrant-Whyte H F. Uncertain geometry in robotics[C]// IEEE International Conference on Robotics and Automation. Proceedings. IEEE, 1987:851-856.
- [34] Smith R, Self M, Cheeseman P. Estimating uncertain spatial relationships in robotics. in Proceedings. 1987 IEEE International Conference on Robotics and Automation 4, 850–850.
- [35] Ayache N, Faugeras O D. Building, registering, and fusing noisy visual maps[J]. International Journal of Robotics Research, 1988, 7(7):45-65.
- [36] Crowley J L. World modeling and position estimation for a mobile robot using ultrasonic ranging[C]// IEEE International Conference on Robotics and Automation, 1989. Proceedings. IEEE, 1989:674-680 vol.2.
- [37] Chatila R, Laumond J. Position referencing and consistent world modeling for mobile robots[C]// IEEE International Conference on Robotics and Automation. Proceedings. IEEE, 1985:138-145.
- [38] Csorba M. Simultaneous localisation and map building /[J]. University of Oxford, 1997, 1(3):699 - 704.
- [39] M. Csorba and H.F. Durrant-Whyte, “A new approach to simultaneous localisation and map building,” in Proc. SPIE Aerosense, Orlando, FL, 1996.
- [40] Leonard J J, Feder H J S. A Computationally Efficient Method for Large-Scale Concurrent Mapping and Localization[M]// Robotics Research. Springer London, 2000:169--176.
- [41] Tard'Os J D. Experiments In Multisensor Mobile Robot Localization And Map Building[J]. Ifac Symposium on Intelligent Autonomous Vehicles Machid, 1998.
- [42] Castellanos J A, Tardos J D, Schmidt G. Building a global map of the environment of a mobile robot: the importance of correlations[C]// IEEE International

- Conference on Robotics and Automation, 1997. Proceedings. IEEE, 1997:1053-1059 vol.2.
- [43]J. Guivant, E.M. Nebot, and S. Baiker, “Localization and map building using laser range sensors in outdoor applications,” J. Robot. Syst., vol. 17, no. 10, pp. 565–583, 2000.
- [44]Williams S B, Newman P, Dissanayake G, et al. Autonomous underwater simultaneous localisation and map building[C]// IEEE International Conference on Robotics and Automation, 2000. Proceedings. ICRA. IEEE, 2000:1793-1798 vol.2.
- [45]Chong T K S, Chong K S, Kleeman L. Feature-based Mapping in Real, Large Scale Environments using an Ultrasonic Array[J]. International Journal of Robotics Research, 1999, 18(1):3-19.
- [46]Deans M, Hebert M. Experimental comparison of techniques for localization and mapping using a bearing-only sensor[M]// Experimental Robotics VII. Springer Berlin Heidelberg, 2001:395-404.
- [47]Guivant J E, Nebot E M. Optimization of the simultaneous localization and map-building algorithm for real-time implementation[J]. IEEE Transactions on Robotics & Automation, 2002, 17(3):242-257.
- [48]Newman P, Leonard J, Tardos J D, et al. Explore and return: experimental validation of real-time concurrent mapping and localization[C]// IEEE International Conference on Robotics and Automation, 2002. Proceedings. ICRA. IEEE, 2002:1802-1809 vol.2.
- [49]Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: Part I[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
- [50]刘浩敏, 章国峰, 鲍虎军. 基于单目视觉的同时定位与地图构建方法综述[J]. 计算机辅助设计与图形学学报, 2016(06): 855-868.
- [51]Triggs B, Mclauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis[C] //Proceedings of International Workshop on Vision Algorithms: Theory and Practice. Heidelberg: Springer, 1999: 298-372.
- [52]吴哲岑. 基于双目立体视觉的特征点提取与定位方法研究[D]. 吉林: 吉林大学学位论文, 2015.
- [53]庄瞳. 单目视觉、惯性室内无人机自主导航算法研究[D]. 南京: 南京航空航天大学学位论文, 2012.
- [54]Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(6):1052-67.
- [55]Mourikis A I, Roumeliotis S I. A Multi-State Constraint Kalman Filter for

- Vision-Aided Inertial Navigation[J]. 2007, 22:3565-3572.
- [56]Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C]// IEEE International Conference on Computer Vision. IEEE, 2012:2320-2327.
- [57]Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// IEEE International Conference on Computer Vision. IEEE, 2012:2564-2571.
- [58]Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]// IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE, 2008:1-10.
- [59]Klein G, Murray D. Parallel Tracking and Mapping on a camera phone[C]// IEEE International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2009:83-86.
- [60]Huber P J. Robust statistics[M]. Hoboken: Wiley, 2009.
- [61]Klein G, Murray D. Improving the Agility of Keyframe-Based SLAM[C]// European Conference on Computer Vision. Springer-Verlag, 2008:802-815.
- [62]Zisserman A. Multiple view geometry in computer vision[J]. Kybernetes, 2003, 30(9/10):1865 - 1872.
- [63]Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5):1147-1163.
- [64]Engel J, Schöps T, Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM[M]// Computer Vision – ECCV 2014. Springer International Publishing, 2014:834-849.
- [65]Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]// IEEE International Conference on Robotics and Automation. IEEE, 2014:15-22.

## 哈尔滨工业大学学位论文原创性声明和使用权限

### 学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于双目视觉的四旋翼无人机定位系统设计与实现》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：何莉 日期：2017年6月19日

### 学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：何莉 日期：2017年6月19日

导师签名：马杰 日期：2017年6月19日

## 致 谢

弹指一挥间，两年的研究生学习生活已经接近尾声。从成为哈尔滨工业大学的一员直到毕业设计即将完成之际，现在回想起这两年的时光，感激之情，铭记于心！

首先，在此谨向我的导师陈松林教授表达最诚挚的谢意。本课题是在陈老师的悉心指导下完成的。从论文选题到具体研究细节以及最终论文的撰写，始终得到陈老师的关心与帮助。在课题研究期间，我经常陷入问题不能找到很好的解决办法浪费了很多时间，也经常出现考虑问题过于片面等问题。陈老师总是在我陷入问题的时候为我指引道路，深刻记得他对我说过的一句话，“埋头苦干的时候，记得抬头看路！”此话铭记于心。从大四毕设到研究生近三年的时间里，每周一次组会，两次课题交流从不间断，即使出去实习老师也会和我电话、邮件交流给予我意见帮助我分析解决问题。最为遗憾是我在课题完成期间和老师交流不够主动，没能更好的像陈老师请教。作为导师，他指导我学习研究；作为长辈，他伴随我成长，教会我许多做人的道理。他对工作的认真与严谨的态度，让我受益匪浅，能师从陈老师，是我人生一笔宝贵的财富，我深感荣幸！

然后，还要感谢实验室的马杰老师。我多次向马老师请教问题，每次她都不厌其烦的对我悉心指导，给予我宝贵的建议，在开题和中期提出的宝贵指导意见。马老师丰富的学识，扎实的作风，认真的态度，让我敬佩不已。

此外，还要感谢实验室其他老师、博士师兄师姐们在学习和生活上给与我的帮助。感谢实验室二楼无人机组的郝宁师弟、王宁远师弟、侯谊师弟、许中研师弟和王思强师弟一次又一次在课题上给予我帮助。感谢张双琳、张红、姚晓艳、刘建军、邢宝祥、王鑫、闫伟、张震等同学，一起学习、做课题、聚餐、玩耍的时光终身难忘，有你们的陪伴与帮助真好！感谢朝夕相处的 953 寝室的舍友们以及实验室的同学们，和你们一起渡过了难忘的两年研究生生活。感谢师弟师妹们带来的欢声笑语。

感谢家人对我一如既往的支持与付出。他们的理解和支持，让我能够专心投入学习与课题研究，克服期间遇到的困难，使得我能够顺利完成研究生学业。在以后的日子，我会继续努力，绝不辜负他们寄予的厚望。

最后，感谢所有关心和支持过我的人，谢谢大家！没有最好只有更好，我将继续努力！