单位代码： 密 级：



硕 士 学 位 论 文



论文题目：

1220055726

学号

姓名

导 师

学 科 专 业

研 究 方 向

申请学位类别

论文提交日期

X学硕士

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

↑

（请填写论文英文题目，Times New Roman 2号字加粗）

Thesis Submitted to Nanjing University of Posts and Telecommunications for the Degree of

Master of XXXXXXXX学位类别英文



By

XXXXXX作者英文名字

Supervisor: Prof. XXXXXX导师英文名字

XXX XXXX论文提交日期例如：April 2018

注：（论文定稿时，此页直接删除，不要打印）

学位类别英文：

工学硕士：Master of Engineering

理学硕士：Master of Science

管理学硕士：Master of Management

教育学硕士：Master of Education

上一页英文封面中，红色字体的文字打印时直接删除。

南京邮电大学学位论文原创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，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邮电大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

本人学位论文及涉及相关资料若有不实，愿意承担一切相关的法律责任。

研究生学号：\_\_\_\_\_\_\_\_\_\_\_ 研究生签名：\_\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_\_\_\_\_\_\_\_\_\_

南京邮电大学学位论文使用授权声明

本人承诺所呈交的学位论文不涉及任何国家秘密，本人及导师为本论文的涉密责任并列第一责任人。

本人授权南京邮电大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档；允许论文被查阅和借阅；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、汇编本学位论文。本文电子文档的内容和纸质论文的内容相一致。论文的公布（包括刊登）授权南京邮电大学研究生院办理。

非国家秘密类涉密学位论文在解密后适用本授权书。

研究生签名：\_\_\_\_\_\_\_\_\_\_\_\_ 导师签名：\_\_\_\_\_\_\_\_\_\_\_\_ 日期：\_\_\_\_\_\_\_\_\_\_\_\_\_

摘要

《摘要正文》×××××××××××××××××××中文摘要（一至两页），英文摘要，二者应基本对应。它是论文内容的高度概括，应说明研究目的、研究方法、成果和结论，要突出本论文的创造性成果或新的见解、用语简洁、准确。论文摘要的关键词3至8个。关键词应为公知公用的词和学术术语，不可采用自造字词和略写、符号等，词组不宜过长。

英文摘要采用第三人称单数语气介绍该学位论文内容，目的是便于其他文摘摘录，因此在写作英文文摘时不宜用第一人称的语气陈述。叙述的基本时态为一般现在时，确实需要强调过去的事情或者已经完成的行为才使用过去时、完成时等其他时态。

关键词: ， ， ， ，

Abstract

Abstract …………..

Key words:, , ,

目录

[第1章 绪论 1](#_Toc121680303)

[1.1 研究背景与意义 1](#_Toc121680304)

[1.2 国内外研究现状 2](#_Toc121680305)

[1.2.1 多传感器融合研究现状 2](#_Toc121680306)

[1.2.2 三维重建研究现状 4](#_Toc121680307)

[1.2.3 三维环境感知研究现状 6](#_Toc121680308)

[1.3 本文的研究内容 8](#_Toc121680309)

[1.4 本文的章节安排 9](#_Toc121680310)

[第2章 相关背景知识介绍 11](#_Toc121680311)

[2.1 相机成像模型 11](#_Toc121680312)

[2.1.1 针孔相机模型 11](#_Toc121680313)

[2.1.2 相机畸变模型 11](#_Toc121680314)

[2.2 多视图几何模型 11](#_Toc121680315)

[2.3 双目立体视觉模型 11](#_Toc121680316)

[2.4 本章小结 11](#_Toc121680317)

[第3章 多传感器时间一致性表征 12](#_Toc121680318)

[3.1 时间一致性表征系统总体设计 12](#_Toc121680319)

[3.2 激光雷达的时钟同步 12](#_Toc121680320)

[3.2.1 时钟同步PPS+GPRMC原理 12](#_Toc121680321)

[3.2.2 激光雷达时钟同步软硬件实现 12](#_Toc121680322)

[3.3 视觉相机的时钟同步 12](#_Toc121680323)

[3.3.1 相机外触发原理 12](#_Toc121680324)

[3.3.2 视觉相机时钟同步软硬件实现 12](#_Toc121680325)

[3.4 惯性测量单元IMU的时钟同步 13](#_Toc121680326)

[3.4.1 就近时间戳对齐原理 13](#_Toc121680327)

[3.4.2 惯性测量单元IMU时钟同步软件实现 13](#_Toc121680328)

[3.5 本章小结 13](#_Toc121680329)

[第4章 多传感器空间一致性表征 14](#_Toc121680330)

[4.1 空间一致性表征系统总体设计 14](#_Toc121680331)

[4.2 传感器内参标定 14](#_Toc121680332)

[4.2.1 视觉相机内参标定算法原理及实现 14](#_Toc121680333)

[4.2.2 惯性测量单元IMU内参标定算法原理及实现 14](#_Toc121680334)

[4.3 多传感器外参联合标定 14](#_Toc121680335)

[4.3.1 视觉相机与视觉相机之间联合标定 14](#_Toc121680336)

[4.3.2 视觉相机与激光雷达之间联合标定 14](#_Toc121680337)

[4.3.3 视觉相机与IMU之间联合标定 15](#_Toc121680338)

[4.4 实验结果与验证 15](#_Toc121680339)

[4.5 本章小结 15](#_Toc121680340)

[第5章 多源多层地图三维重建 16](#_Toc121680341)

[5.1 重建算法设计方案 16](#_Toc121680342)

[5.2 基于MVE算法的三维地形重建 16](#_Toc121680343)

[5.3 火星场多源多层地图构建 16](#_Toc121680344)

[5.3.1 各坐标系之间的关系 16](#_Toc121680345)

[5.3.2 基于多传感器的多层地图生成 16](#_Toc121680346)

[5.4 实验结果与分析 16](#_Toc121680347)

[5.5 本章小结 16](#_Toc121680348)

[第6章 基于双目视觉的三维环境感知 17](#_Toc121680349)

[6.1 感知系统算法设计方案 17](#_Toc121680350)

[6.2 基于SOLO网络的实例分割算法 17](#_Toc121680351)

[6.3 基于BGNet网络的双目立体匹配算法 17](#_Toc121680352)

[6.4 实验结果与分析 17](#_Toc121680353)

[6.5 本章小结 17](#_Toc121680354)

[第7章 总结与展望 18](#_Toc121680355)

[参考文献 19](#_Toc121680356)

[附录1 攻读硕士学位期间撰写的论文 21](#_Toc121680357)

[附录2 攻读硕士学位期间申请的专利 22](#_Toc121680358)

[附录3 攻读硕士学位期间参加的科研项目 23](#_Toc121680359)

[致谢 24](#_Toc121680360)

# 绪论

* 1. 研究背景与意义

近几年，科技在飞速发展着。随着工业4.0的到来，越来越多的智能制造、智能机器人、智能驾驶等技术在不断的发展和完善，也正逐步的改善着我们生活的方方面面，各类无人设备层出不穷。这些不管是工业设备、机器人、自动驾驶汽车等都对环境的感知尤为依赖。例如工业设备需要根据实际应用场景使用单目相机、双目相机、线扫激光等传感器；室内清洁、服务型机器人多以单线雷达就行地图重建，并使用RGBD相机就能进行场景的感知以及运行中的避障；对于自动驾驶汽车而言，它作为一种特殊类型的移动类机器人，其工作的环境特殊且复杂，所以需要多线激光雷达、视觉相机、惯性测量单元、轮速计、GPS等多种传感器结合的方式进行场景的重建和环境感知。尤其对于复杂、恶劣环境，以及未知环境的探索，将会对多传感器的融合感知技术有更高的需求。例如航天对于火星、月球等环境的探索，都需要高精度的三维地图，高精度的机器人定位和高精度且鲁棒的环境感知系统，这都离不开各种传感器之间的融合与配合。

在2020的7月，由长征五号遥四运载火箭将火星任务车祝融号发射升空。其为天问一号任务火星车，重达240公斤，高度将近2米。火星车主要依赖轨道器预先开展的约三个月的对地观测，以便于对于预选着陆区的详细勘测。之后火星车与轨道器分离，并且利用反推火箭以及降落伞协同的方式在火星表面安全着陆。着陆后，火星车将通过配置的导航相机、多光谱相机、激光雷达、磁场探测仪、气象测量仪、成分探测仪等传感器按计划开展巡视区域的环境感知、火星表面移动和科学探测。同时轨道器也持续运行在轨道中，与火星表面的火星车形成空地一体协同探测，并为火星车的巡视提供稳定的中继通信和保障。

对于早期的室内外无人化设备、机器人等智能体多以一个或两个传感器进行物体的识别以及周围环境感知，甚至最初只能实现二维平面内的一些定位、识别和感知功能。这也导致以前的许多智能化、自动化设备动作机械化，容易受干扰，且不鲁棒。随着多传感器融合技术以及三维重建、三维环境感知等相关技术的发展使得各类智能体，机器人包括智能驾驶汽车逐渐的发展起来，也能实现更多精细化的功能。

早期的室内外重建，大多使用单线激光雷达或单目相机用于室内，多线激光雷达用于室外的地图重建，并在此地图的基础上进行机器人导航。但是由于单一传感器对于大场景地图重建中存在的累积漂移现象，使得基于多传感器融合的即时定位与建图逐渐的成为了主流。基于多传感器融合的定位与建图在没有回环的情况下也能有很好的效果，所重建出来的地图信息也更加丰富，并且基于多传感器的环境感知，也更加精准。由此可见，对于复杂地形环境，尤其是针对像火星、月球等航天领域的探索中，多传感器的配准融合、多维地图的重建和三维环境的感知显得尤为重要。

总而言之，如何对于复杂环境下进行高质量、多信息的三维地形重建与物体感知是本文的主要目标。针对火星、月球等航天领域的环境探索，如何在未知环境下，进行鲁棒性较高的地图高精度、高纬度的重建与感知是目前科技发展与航天领域中需要不断探索与研究的问题。本文将从多传感器的时空一致表征，火星场多层次地图的重建系统，双目视觉环境感知三个方面展开研究，针对航天院火星车多传感器系统以及实验室搭建的多传感器系统，从软硬件方面入手并对现有算法进行改进与应用。

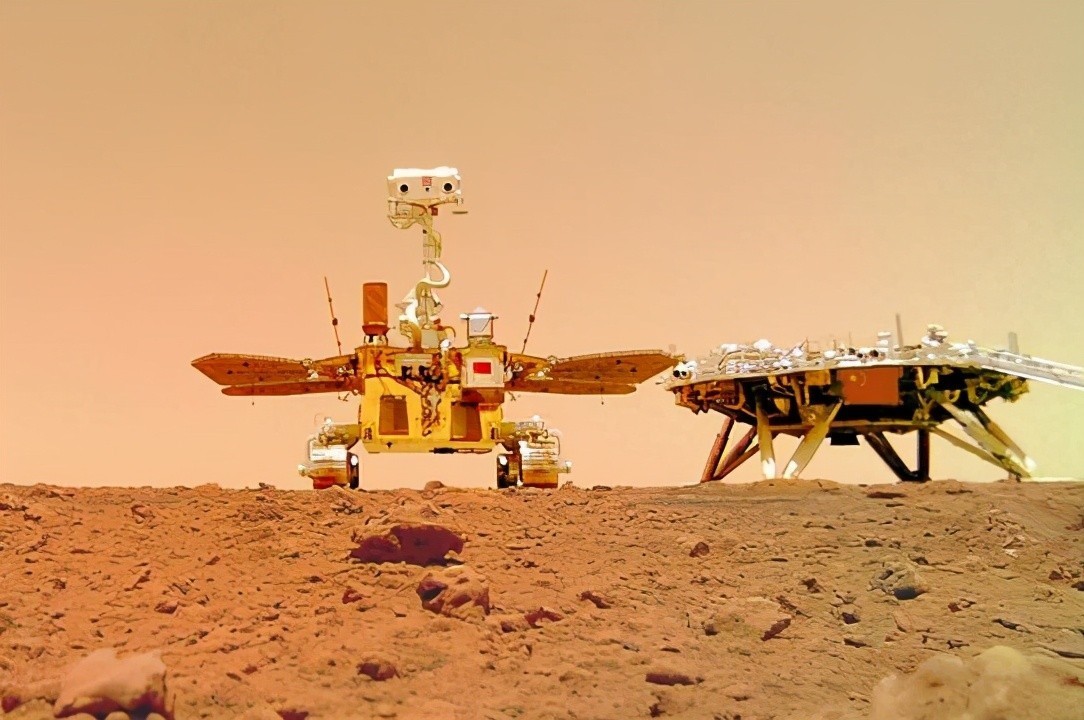


图1.1 火星车祝融号图像

* 1. 国内外研究现状
     1. 多传感器融合研究现状

随着深度学习实用性的显着发展和5G通信技术的超高速信息传输速率将突破数据壁垒车联网传输，智能驾驶技术正在逐渐成为影响交通的关键技术。传感器是智能驾驶系统感知外界环境的关键，其协同性能直接决定了自动驾驶车辆的安全性。

多传感器融合（Multi-sensor Fusion, MSF）技术，是利用现有计算机相关知识与技术，综合利用所获得不同传感器的信息和数据并以一定的准则进行自动分析和融合，避免了感知限制和不确定性，形成对环境更全面的感知和识别或目标，并提高外部感知能力系统。目前，多传感器融合技术已广泛应用于故障检测领域[2], [3], 遥感 [4], 人体健康监测 [5],[6]、机器人系统[7]、人机交互[8]、目标识别和跟踪 [9]、[10]、同时定位和映射 (SLAM) [11] 和高级驾驶员辅助系统（ADAS）[12]等领域。

近年来在智能驾驶中多传感器融合的不同策略也会有不同的融合感知效果，常用的传感器包括毫米波雷达、单线和多线激光雷达、RGB摄像头、深度相机、热成像仪、超声波、GPS、IMU 和轮速计等。各类传感器都有各自的优点与不足，因此在自动驾驶系统中各个传感器通常会有不同任务的划分，也会相互配合完成一项任务，进而达到高效、高精度的效果。目前，几种主要传感器组合形成的用于多传感器融合系统，包括camera-LiDAR(CL)、radar-camera(RC)和radar-camera- LiDAR(RCL)。根据现有数据图1.2表明，最常用的传感器组合是RC，因为这种组合可以在获得周围物体的距离信息的同时获得优异的分辨率。同样，激光雷达和摄像机的组合可以获得具有深度的图像信息，一些研究将激光雷达和毫米波雷达与摄像机相结合，以提高安全冗余度。雷达和摄像机都是全方位的经济技术，激光雷达的性能正在逐步提高，高性能产品的价格仍然很高。虽然相机可以获得目标的轮廓、纹理和颜色分布，但缺点也是显而易见的。双目和深度相机的应用允许图像数据具有深度信息，但在高精度方面还有很长的路要走。与相机和激光雷达相比，毫米波雷达具有更长的波长，可以穿透雨、雪和雾。美中不足的是，雷达更容易受到杂波干扰。激光雷达可以昼夜连续工作，除恶劣天气条件外，还可以提供高分辨率和长距离的三维数据。因此，满足各种工作条件的唯一解决方案是采用多传感器融合技术。

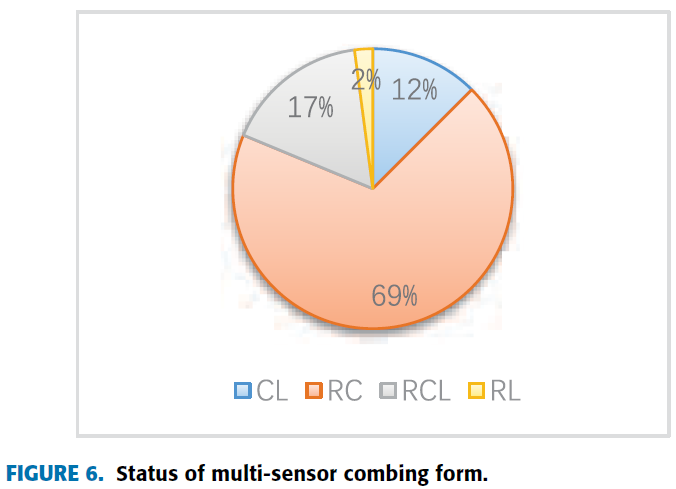


图1.2 常见传感器组合使用比例

目前常见的多传感器融合主要分为两种。一种是如图1.3的前融合算法，另一种是如图1.4的后融合算法。前融合指的是每个传感器都是融合体的一部分，将所有传感器依据他们之间的关联性融合成类似于一个传感器包含所有初始信息的形式，最终输出融合结果；后融合算法又称松耦合算法，它的本质是将所有传感器独立开来，每个传感器进行相应的识别、感知算法，最后将所有感知结果进行一定的汇总，后融合的算法是传感器间相互独立的，不存在传感器之间的约束。



图1.3 前融合算法框图



图1.4 后融合算法框图

* + 1. 三维重建研究现状

在过去几年中，GPU硬件的进步促进了3D重建系统的发展。此前，Pollefeys等人（2008年）、Agarwal等人（2009年）和Furukawa等人（2010年）的大规模努力从非结构化照片集重建了城市场景的部分。最先进的算法所依赖的持续优化理论基础不断加强和拓宽（2011年的Chambolle和Pock；2012年的Goldluecke等人），已成为机器人和计算机视觉应用的基础。硬件和理论结合在一起，使我们能够构建如图1.5所示的创建大规模3D密集重建的系统。然而，许多密集3D重建系统的最新技术很少考虑可扩展性在自动驾驶或检查等地图应用中的实际应用。最普遍的方法是最近的手机和平板电脑的着眼于小规模的重建开发（2015年的Klingesmith等人； 2015年的Engel等人；2015年的Schops等人）。

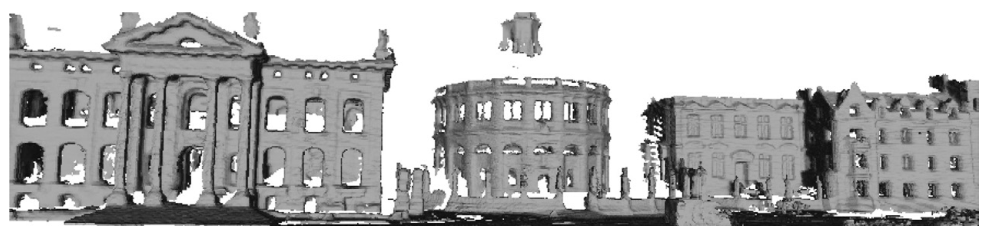


图1.5立体视觉或激光数据高效生成多公里尺度环境的密集模型

一些研究人员研究了更适用于自动驾驶汽车的数据和传感器。最值得注意的是，2010年谷歌发布了一篇学术文章，详细介绍了他们的“树视图”应用程序，该应用程序利用激光和相机数据创建世界各地城市的密集3D重建（Anguelov等人，2010年）。然而，他们的算法过度强调激光数据，并假设所有深度图仅包含分段平面对象。2013年，谷歌提出了一种仅使用相机传感器的运动方法替代结构（Klingner等人，2013年）。Xiao和Furukawa（2014年）提出了一种利用激光和图像输入对大型室内环境进行建模的系统，但他们修改的曼哈顿世界假设限制了重建，无法对除垂直和水平平面以外的任何东西进行建模。Bok等人（2014年）在现有技术的基础上，使用相机和激光创建大规模3D地图。他们的最终重建是稀疏的，仅使用相机进行里程测量和环路闭合，而不是作为额外的深度传感器来改善密集的重建。

在RGB-D相机广泛使用之前，DTAM提出了一种使用单眼相机生成高质量深度图的方法。在相机的焦平面前面构建成本体积，并使用来自连续图像帧的2D正则化深度估计来不断更新成本体积。最终的重建提供了精细的细节，但该系统将单眼相机的范围限制为近场重建。2010年，第一台商品RGB-D相机发布。RGB-D相机为图像中的每个像素提供厘米级的精确深度测量：在第一台设备中，分辨率为640×480，频率为30Hz。Curless和Levoy（1996年）通过 Kinect Fusion 系统扩展了他们的工作，以利用这种高频和高质量的深度图流。利用截断有符号距离函数（TSDF），深度观测值存储在体素网格中，其中每个体素在表面前面时包含一个正数，在表面后面包含一个负数。求解零值水平集会生成原始曲面的密集模型。因此，Kinect Fusion可以为大约7m³的工作空间实时生成前所未有的高质量密集3D重建。

与Kinect Fusion（体素网格固定在空间中的一个位置）不同，Kintinuous试图通过允许体素网格随相机移动来扩展重建场景的大小。先前观察到的区域的连续流作为网格流式传输到磁盘，但如果再次观察到该区域，则可以重新加载到GPU中。该系统理论上无限扩展了重建工作空间的大小。然而，它不能利用距离RGB-D相机3米以外的传感器观测，因为它仍然基本上基于传统的固定尺寸体素网格。

Nießner的哈希体素网格（HVG）也扩展了重建的大小，但只在观察到表面的区域分配体素。这样删除的内存浪费了存储可用空间。当表面远离传感器时，与GPU和硬盘之间的数据流相结合时，重建的大小基本上没有限制。这种实施方式将传感器范围限制在4米，因为这接近Kinect相机的最大有效范围，但该范围可以很小地扩展。Whelan等人（2014年）的解决方案利用滚动循环缓冲作为体积重建数据结构。这是一种有趣的方法，它允许局部体积区域在摄影机在环境中移动时进行虚拟平移。在深度学习环境中，正确选择有效的体积数据结构也获得了关注，因为它会影响3D任务的分辨率，包括3D对象分类、方向估计和点云标记（Riegler等人，2017年）。

三维重建作为计算机视觉领域的一个活跃研究领域，在虚拟现实、逆向工程和机器人视觉等领域有着重要的应用。

* + 1. 三维环境感知研究现状

自从计算机视觉出现[1，2]以来，3D环境重建和物体6D姿态估计一直是一个核心问题。先进的机器人经常出现在科幻电影中，这提高了人类对未来机器人的期望。我们希望机器人能像我们一样感知周围的世界。然而，现实世界的机器无法实现这种智能。我们的人类视觉系统可以快速轻松地推断3D对象的属性，并理解周围环境的组成。相比之下，机器人视觉传感器捕捉环境的图像，但输入设备不具有感知能力或认知能力；它不了解世界。因此，迄今为止，赋予机器这些能力是一项具有挑战性的任务。

计算机视觉系统最重要的目标之一是从复杂环境中估计物体的6D姿态（3D平移和旋转）。精确的6D姿态估计技术可以驱动与机器人操纵、自动驾驶、增强现实等相关的各种新兴技术领域。为了在杂乱场景中感知准确的6D姿势，大量研究致力于建立高效的3D物体感知系统。想法包括利用对象几何信息的基于特征的方法[3，8]和使用深度神经网络的基于学习的方法[13，14，15]。无论使用哪种范式，它们都依赖于高质量的对象模型。换句话说，优越的物体模型是准确的6D姿态估计的先决条件。

在3D计算机视觉的早期，它主要处理多面体物体[16，17]。为了识别细粒度几何形状，一些方法[18，19]直接从3D点云估计物体的6D姿态。他们利用点特征描述来计算对象模型和场景点云之间的一组对应关系，并通过随机样本一致性（RANSAC）算法稳健地估计6D姿态。除了基于点云的方法外，基于图像的方法也取得了重大进展。Hinterstoisser等人[3]提出了一种模板匹配方法Linemod，该方法利用颜色梯度和表面法线来描述对象轮廓和内部信息。在[8]中，通过从CAD 3D模型自动生成模板数据集，对其进行了进一步改进。然而，由于缺乏丰富的纹理，这些作品无法处理杂乱环境中的无纹理对象，以检测可区分的特征进行匹配。

最近，深度卷积网络技术的出现，特别是基于CNN的类别检测器[20，21，22]，已经显示出用于对象检测和对象分割的优异结果。受2D物体检测的显著进步的启发，越来越多的工作将深度学习用于6D姿态估计。对于RGB导出的6D姿态估计，大多数方法遵循类似的范式：首先，它们采用神经网络来检测与目标对象相关的八个3D边界框角点。他们执行透视n点（PnP）算法来计算方向和平移。然而，该范式在无纹理对象的检测精度低和昂贵的后处理步骤方面存在严重缺陷[23]。

与2D图像相比，点云更接近原始的3D几何形状，并具有可靠的深度距离以精确定位对象。Qi等人提出了一种端到端网络PointNet[24]，它直接将点云作为输入，并对输入点云进行分类和分割。在[25]中，提出了PointNet的改进版本，使网络能够学习不同规模的局部结构。虽然这些方法仍处于类别级别，但Li提出了PointRCNN[26]，它可以在实例级别从原始点云实现3D对象检测。然而，一个典型的场景模型包含超过100k个点；训练这样的网络需要高的计算和存储器要求。此外，当前基于点云的学习技术利用公开可用的数据集进行训练和测试。尽管数据集可以加快研究过程，但如何生成适用于真实世界对象的通用系统仍然具有挑战性。

在自动驾驶系统中，环境感知[17]包括物体定位、离线障碍物和道路映射、移动障碍物跟踪以及交通信号检测和识别。在本研究中，我们主要关注自动驾驶系统中的多对象三维（3D）感知问题。通常，环境感知的作用是检测物体，并从城市场景中的传感器数据中获取其定位信息。先前对自动驾驶多目标感知的研究大致可分为手工特征算法和深度学习[11]方法。手工制作的特征算法在很大程度上依赖于专业知识和技能来提取代表性特征；因此，使用该算法很难在复杂环境中获得令人满意的性能。随着深度学习方法的最近出现，已经设计了各种3D感知模型，用于使用来自不同传感器的数据来检测场景中的关注对象。该方法可进一步分为基于相机的、基于激光雷达的和多传感器融合方法。

基于相机的方法直接从二维（2D）RGB图像执行3D对象检测，方法是首先对提出的2D对象进行回归，然后根据其在3D空间中的几何关系或约束预测3D边界框结果[18-21]。Mono3D[18]方法通过语义分割和上下文特征（例如，大小、位置和形状）推断出单眼3D边界框。在参考文献[19]中，几何约束被施加在2D边界框和具有从单个图像生成的对象姿态的3D边界框上。为了模拟来自图像的LiDAR信号，定义了伪LiDAR[20，21]，由此使用深度估计算法从RGB图像生成一组伪点。

基于激光雷达的检测是自动驾驶系统中用于3D物体感知的主流方法。该方法使用PointNet架构直接使用原始点云数据[22-24]，或者将点转换为规则体素表示[25]，并采用卷积运算进行3D对象检测[26，27]。PointNet架构[22，23]将点云数据作为输入，并应用最大池化操作以保持无序点的排列不变性。在此基础上，TANet方法[24]设计了三重注意力模块和用于鲁棒对象定位的粗到细回归策略。参考文献[25]提出了VoxelNet，它使用点体素编码模块，将点云数据转换为用于特征提取的规则网格格式。随后，参考文献[26]报道了3D稀疏卷积运算的发展，以加速体素特征编码和3D对象检测。

多传感器融合方法已成为一种有前途的方法，它使用多模态输入数据（即图像和点云数据），并使用各种融合策略组合其各自的特征图，以实现更准确和鲁棒的检测[28–31]。MV3D[28]使用紧凑的多视图表示对稀疏的3D点云数据进行编码，并结合来自多个视图的区域特征，用于3D空间中的对象检测。截头体PointNet[29]框架利用成熟的2D对象检测器来生成区域建议，并在几何关系的约束下精确估计3D边界框。AVOD[30]在自动驾驶场景中聚集用于对象检测的视图。在所提出的网络中，从点云和RGB图像数据生成共享特征图，然后通过区域建议网络融合多模态特征图以获得可靠的3D对象建议。该方法对于城市场景中的目标检测是有效的。

* 1. 本文的研究内容

在上一节中，主要简述了目前国内外包括多传感器融合、三维重建以及三维环境感知等相关技术的研究现状，以及现有的不足。针对复杂场景，未知场景的多传感器融合感知探索仍需不断研究。随着多传感器融合感知相关技术的发展，以及我国对于航天领域探索的不断深入，对航天领域无人车的多传感器融合感知系统的研究越发的重要。例如我国已经送上火星的祝融号、送上月球的玉兔及玉兔二号，这些无人车对环境的感知与探索都显得尤为重要。所以，本文的主要研究内容也都围绕着基于航天探索背景的二代火星车进行多传感器融合、三维地图重建以及三维环境感知三个方面展开研究。

1. 用于多传感器融合的时空一致表征

在机器人、自动驾驶、复杂环境感知等问题中，常用的算法大多是单一类型传感器进行环境的感知，这也导致了对于复杂场景以及需要高精度的探索感知环境，效果不是很好。随着相关技术的发展，对于多传感器的配合探索、感知也逐渐成熟，多传感器融合技术不论是在三维重建方面还是在环境感知方面都有着不错的效果和精度。但是不管是多传感器的前融合、亦或者是多传感器的后融合，他们的融合前提条件都需要进行多传感器的时间同步和空间同步。与此同时，这些都是需要软硬件结合来实现多传感器的时间同步与空间同步，传感器种类越多，实现起来越困难，同步精度也会越差。现有的多传感器融合大多集中于3类传感器的时空同步以及数据的融合。本文主要围绕着基于航天探索背景的二代火星车，以及智能驾驶相关设备中常见的大于5种的常用传感器进行时间同步与空间同步。分别从多种传感器软硬件着手，进行时间同步和空间同步，对于传感器种类的繁多，进行时间同步和空间同步的难度也随之提高。最终在航天院的探索无人车上的传感器以及实验室的一套传感器上分别进行实验验证，并取得不错的效果。

1. 火星车附近场景的多源多层地图构建

现有的三维重建以及SLAM技术发展越发成熟，基于深度学习的三维重建与SLAM算法也层出不穷。但是针对于航天探索领域来说，火星车的移动速度也十分缓慢，其需要的不是多么新的技术，而是需要十分鲁棒且精度比较高的算法系统，需要的是更多的信息源构成的地图，以便于火星车对自身周围环境的感知与判断。本文主要使用单目MVE三维重建算法进行三维地图的重建，并通过热成像仪、多光谱成像仪与RGB相机的关系，就行多层地图的构建，最终形成一个包含颜色、热、光谱信息的多源多层地图，为不同的场景感知任务做了较好的地图重建工作。

1. 基于双目视觉的三维环境感知

对于室内外三维环境感知，相关算法数不胜数。例如有基于三维点云的室外三维环境感知，基于RGB-D相机的室内三维感知，以及基于单目相机的深度恢复然后进行室内外的三维感知。但是上述相关环境感知算法，大都存在一定的缺陷。例如，基于RGB-D相机的三维感知系统，由于硬件的限制，大多只适用于室内；再例如基于激光雷达的三维点云环境感知，虽然适合室外感知，但其需要大量的对应数据集用于训练，并且目前相关的较好的算法基本都是基于一些城市场景的公开数据集，这些数据集基本只有大型的建筑物、车、树等，对小物体的感知并不鲁棒。但是对于航天领域的探索，通常需要进行一些沙地、岩石、地形的感知与判断，所以多小物体的检测与感知能力较为关键。所以，本文主要围绕基于深度学习的双目立体视觉进行室内外三维环境感知的研究。通过在2D上进行感知进而恢复到三维的方式来实现三维空间的物体感知。

* 1. 本文的章节安排

针对当前航天探索领域的无人车相关环境感知应用的研究。本文主要从多传感器融合；多源多层地图构建；基于深度学习的双目立体视觉感知三个方面着手，基于航天院二代火星车及室内模拟火星场，软硬件结合，进行以上三个方面的实践应用与研究。本文各章节组织结构如图1.6所示。



图1.6 章节组织结构

本文由七个章节组成，各个章节的内容和相关结构安排如下：

第一章绪论部分，主要介绍了目前关于多传感器融合在智能驾驶等无人车领域的应用，以及三维重建和三维环境感知的国内外研究现状。简单介绍了基于航天探索领域的课题应用背景，并且对本文的研究内容与章节安排进行了概述。

第二章将介绍相关的基础知识与原理。首先进行单目相机的小孔成像模型、相机畸变模型的介绍；又对三维重建中较为关键的多视图几何模型，包括刚体的运动模型、对极几何、三角测量等原理进行了介绍；最后介绍了三维环境感知中常见的双目立体视觉模型的相关原理与算法。

第三章主要从多类传感器的软硬件着手，进行多传感器融合的前期准备工作，即多传感器的时间一致性表征。重点是实现多类传感器的时间一致，保障这些传感器能采集到同一时刻的数据。

第四章主要进行的多传感器的空间一致性表征。重点是实现多类传感器的内参标定、畸变校正以及传感器之间的外参标定。将各类传感器统一到一个坐标系下，为后续的多传感器融合、感知、重建等算法提供良好的工作准备。

第五章将利用RGB相机进行基于MVE算法的火星车附近场景的三维重建，再利用第三章与第四章的时空一致性的相关数据，将热成像仪、多光谱成像仪的数据叠加到已经建好的三维地图中，实现一个多源多层的点云地图。

第六章介绍两个深度学习网络，分别实现基于深度学习图像分割和基于深度学习的双目视觉视差图的恢复。最终实现基于双目视觉的对火星场特定目标的三维空间定位，并进行室外场景的测试，实现室外特定目标的测距。

第七章，总结与展望。主要是对本文工作与相关实践内容的一个总结，分析了自身工作中的问题以及相关算法的不足，并对未来工作的相关工作与自身的发展进行了展望。

# 相关背景知识介绍

本章将介绍三维视觉里面的相关的基础知识、数学表达及原理。首先进行单目相机的小孔成像模型、相机畸变模型的介绍；又对三维重建中较为关键的多视图几何模型，包括刚体的运动模型、对极几何等原理进行了介绍；最后介绍了三维环境感知中常见的双目立体视觉模型的相关原理与算法。

* 1. 相机成像模型
     1. 针孔相机模型

最早的照相设备，以及今天很多人依然在使用的照相机。其在成像上无不采用的是针孔成像的原理，只不过对于这个针孔，已经逐渐的由简单的物镜变成了较为复杂且高质量的现代镜头，对于焦点成像平面的位置变成了电子成像原件或者是感光胶片之类的成像载体。

如图2.1为理想的针孔相机的成像模型。其中针孔O与像平面之间的距离我们称之为焦距f。对于理想的针孔相机模型，物体的成像没有畸变，且所成的像是一个倒像，这样对我们计算机处理计算时需要一个负号。因为，我们通常研究的如图2.1中的虚拟像平面。

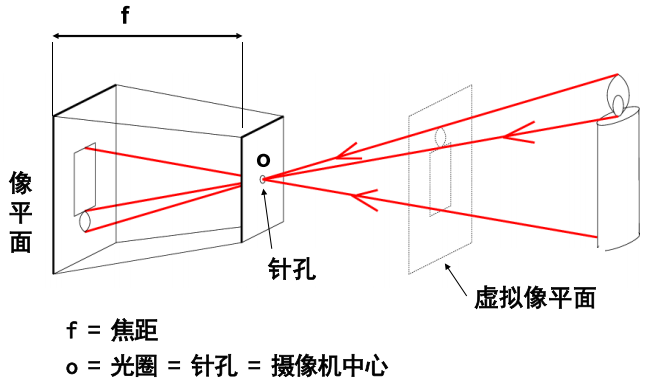


图2.1 针孔相机模型

下面为一个三维点到二维点的映射过程。设O-x-y-z为相机坐标系，其中O为针孔，x-y平面与成像平面平行，z轴指向物体。在此相机坐标系下，设空间中的某一物体点，该三维空间点，经过针孔成像在像平面--上成像点为，其中针孔到像平面的距离为f。所以最终可以通过图2.2所示相似三角形关系可得：

将式子整理可得：

上述式子即根据相似三角形原理，描述了三维点到二维成像平面的映射关系，最终可得到空间点P在相机成像O-u-v平面的成像坐标为

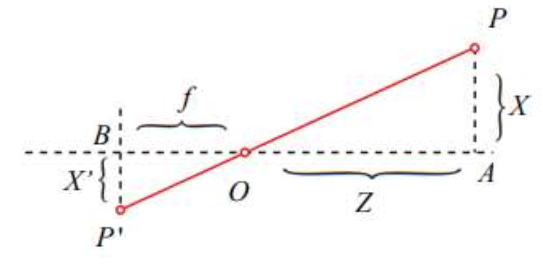


图2.2 相似三角形成像图

* + 1. 相机畸变模型

上述的针孔成像模型是理想状态下的线性成像模型，随着相机镜头的出现，现在也更多的是非线性的透镜成像。但是由于镜头的安装误差以及制作过程中工艺的误差，通常在成像平面上与理想模型计算所得坐标不一致，这也就是透镜畸变。

透镜的畸变主要包括径向畸变和切向畸变。其中径向畸变主要是因为镜头工艺不理想，在成像平面上产生径向失真；切向畸变主要是由于安装的过程中无法使镜头与像平面平行，导致切向失真。

1. 径向畸变

相机的径向畸变主要是由于镜头制作工艺的误差带来的影响，使得像素点无法理想的到达成像平面对应位置，产生径向失真。常见的径向畸变失真有“肥胖”桶型和“瘦小”枕型两种失真。径向畸变的数学模型可由如下公式近似表示：

上述（2.3）中两式分别对应x、y方向的径向畸变分量，其中的为径向畸变系数。

1. 切向畸变

相机的切向畸变主要是由于镜头安装的过程中，我们人为的无法确保镜面与像平面平行，产生切向失真。切向畸变的数学模型可由如下公式近似表示：

上述（2.4）中两式分别对应x、y方向的切向畸变分量，其中为切向畸变系数。式中的r为理想像素点距离畸变中心点的像素长度，即。

所以，对于一个镜头，我们通常需要5个参数来描述畸变以及进行畸变的矫正。

* 1. 多视图几何模型
     1. 三维空间的刚体运动

机器人在三维空间中的位姿变化，一般会描述为一个刚体的运动。在三维的空间里面，通常我们用来描述刚体运动的方式主要是旋转和平移。对于平移来说，相对较好理解和描述，通常用一个平移向量来描述刚体三个自由度的平移量；对于旋转而言，描述的方式就比较多了，例如旋转矩阵、四元数等。对于旋转不同的表示方法来说，他们之间也都有着各自的优势和劣势，通常我们根据实际需要来选择具体的表示方式。下面为几种常见的三维空间中刚体旋转的表示方式：

1. 旋转矩阵

对于一个向量a，其在两个坐标系下的表示是相等的，可知：

我们再将左右两边同乘一个进行等式化简，可得：

由上可得向量经过矩阵R的转换，可以得到向量a，也即是一个向量在两个坐标系下的表示是相等的。上述的矩阵R描述了旋转，所以被称为旋转矩阵。此矩阵又是一个正交矩阵，且行列式为1。最终可将旋转矩阵的集合定义为如下的一个特殊正交群：

1. 旋转向量和欧拉角

对于上面介绍的旋转矩阵而言，其需要9个量，难免会显得有些冗余。旋转向量和欧拉角是更为紧凑的表示方式。对于任意的旋转而言，都是可以由一个旋转轴外加一个旋转角度来进行描述。所以，我们使用一个旋转轴为**n**，旋转角度为的向量来进行描述，这个向量就称为旋转向量。除了旋转向量，欧拉角的描述方式也非常的紧凑，只需要3个量来描述旋转，欧拉角的描述方式在无人机等领域应用较多，一般通常使用roll-pitch-yaw分别表示绕XYZ轴的旋转角度。所以，使用即可描述任意旋转。

旋转向量与旋转矩阵之间还可以还可以通过罗德里格斯公式进行相互的转化，公式如下：

1. 四元数

虽然已经有了紧凑的旋转向量和欧拉角，但是他们存在一个非常严重的问题，那就是万向锁问题，也即是它们存在奇异性。一个三维的向量是不可避免奇异性的，所以才引入了四元数来描述旋转，它即紧凑，又没有奇异性，是一个常用的描述旋转的方式。四元数由一个实部、三个虚部组成，定义如下：

其中***i***，***j***，***k***为三个虚部，他们之间满足如下关系：

四元数到旋转矩阵的转换公式如下：

旋转向量到四元数的转换公式如下（其中设={}，）：

由上述公式可知，在实际的计算中，如果趋近于0，则会导致解的不稳定，此时需要使用其他方法进行旋转矩阵到四元数的求解。

* + 1. 对极几何

对极几何一般指的是现实中的两个相机之间，像平面不平行，并且以两个相机之间的基线作为轴的几何约束关系。这种约束关系只与两个相机的内参以及两个相机之间的相对位置有关，与空间中实际物体无关。对极几何关系图如图2.3所示，表示两个相机拍摄同一场景。其中*O*--P称为极平面，*O、*分别为左右相机的坐标系原点，连线为基线。极平面与两个相机的像平面分别相交于*l*和，P点对应的两个投影点分别为*x*和，基线与像平面的交点称为极点，分别为e和。

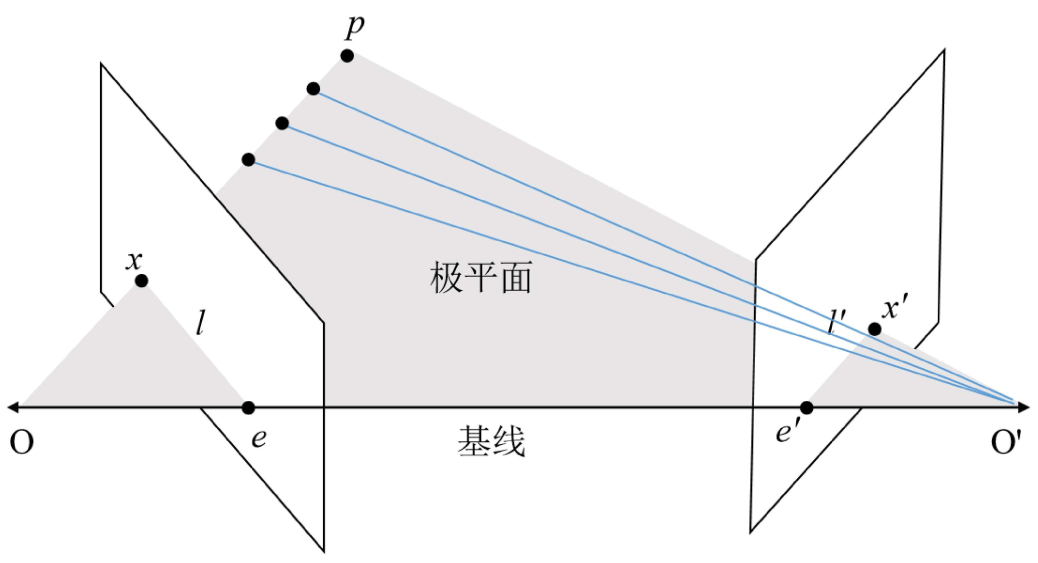


图2.3 对极几何关系图

对于P点在左相机的投影点x，其在右相机上的投影一定在极线上，在P点深度未知的情况下，极线上任意位置都有可能。这就是对极几何的极线约束，通过此几何约束，在进行左右视图之间的匹配时，对于左图的某一点，仅仅只需要在有图的极线上进行检索，大大的减小的搜索范围和计算量，同时也加速的算法的匹配速度。

* 1. 双目立体视觉模型

在立体视觉中，双目相机可以根据两个相机之间的视差进行三维信息的恢复。相较于单目相机，其可以恢复三维信息；相较于深度相机，其又相对简单，成本低。双目视觉常被用来做双目测距、三维重建、深度恢复等工作。

为了简单的描述双目立体视觉模型，减小计算量，通常建立如图2.4所示的两个相机光轴平行的理想双目立体视觉模型。其中P为三维空间中一点，--与--分别为左右相机成像平面，空间点P在两个相机成像点为和。左右相机坐标系分别为---和---，并且将左相机坐标系作为世界坐标---。设P在左相机成像下的像素坐标为，在右相机成像下的像素坐标为。并且如图所示的两个相机的基线间距为b，相机的焦距为f。

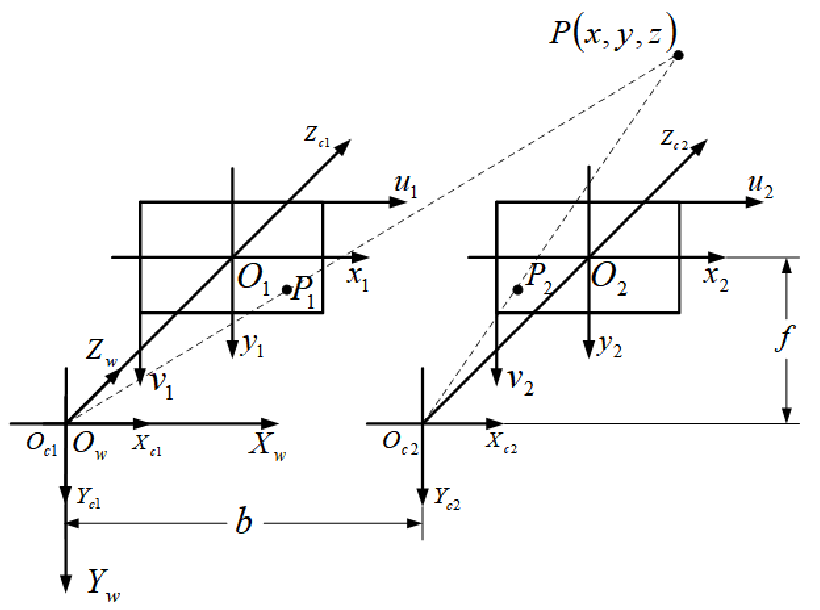


图2.4 双目相机立体成像模型

通过计算可得、之间的距离dis = b – ()。

所以，由图中映射关系，并且结合相似三角形，可建立如下式子：

将式中dis进行替换，并进行整理可得：

公式中的f为相机的焦距，b为双目相机的基线距离，均可通过相机内参标定获得。所以，只需要获得双目相机的视差，即，就可以计算出对应点的深度信息。

* 1. 本章小结

本章首先进行了相机成像模型的介绍，包括对针孔相机模型的介绍以及利用相似三角形进行成像原理的描述，并进行了相机透镜的径向畸变与切向畸变模型；然后介绍了三维重建中十分重要的多视图几何模型，其中先介绍了三维空间中刚体的运动模型，重点介绍了描述旋转的旋转矩阵、旋转向量、欧拉角和四元数，然后通过几何原理介绍了对极几何中的几何约束；最后介绍了三维环境感知中常见的双目立体视觉模型，通过对三维空间中的一点在两个相机中的投影点，并利用相似三角形原理进行了双目测距的数学推导。

# 多传感器时间一致性表征

对于自动驾驶、智能化无人设备中的多传感融合是提高性能与精度的非常实用的办法。在多传感器融合系统中，对于各种各样的传感器，如果无法获取到同一时刻的传感器数据，那么将无法在后续的算法中进行融合，即便个别传感器之间可以融合，也需要复杂的运动矫正。所以在多传感器融合、多传感器感知等系统中，首先需要做的工作就是进行多传感器的时间同步，即确保所有传感器能同一时刻进行数据的采集。

* 1. 时间一致性表征系统总体设计

时间同步的任务是在多传感器的数据进行融合前消除时间上的不同步，各个传感器的初始采集数据会存在一定的时间差，把这个时间差尽量减小，甚至消除，实现各传感器采集的数据在时间上的统一即为时间同步。本章将从软硬件相结合来实现多线激光雷达、彩色相机、鱼眼相机、深度相机、热成像仪、多光谱相机和IMU的时间同步，总体结构示意图如图3.1。



图3.1 时间同步系统总体示意图

该系统主要包含使用激光雷达授时的方式，实现激光雷达与系统UTC时间一致；利用触发信号进行各类相机的硬触发实现相机之间的时间一致；通过测量激光雷达与相机采集数据的时间差调整相机的触发信号达到激光雷达与相机之间时间一致；再通过高频的IMU数据与激光雷达之间进行就近时间戳插值对齐的方式实现IMU与激光雷达时间一致。

* 1. 激光雷达的时间同步

激光雷达内部时钟，默认是以激光雷达上电时刻开始从0走时。所以激光雷达的时间同步主要是采用授时的方式，即将系统的UTC时间写入到激光雷达内部，以此来改变激光雷达内部时钟，实现激光雷达与系统UTC时间同步。激光雷达时间同步方案示意图如下：



图3.2 激光雷达时间同步方案示意图

激光雷达时间同步方案如图3.2所示。其中硬件方面主要由Xavier工控机，通过USB转RS232串口与激光雷达连接，用于发送GPRMC报文，并且通过Xavier工控机的一个IO口实现PPS信号的发送；其中软件方面主要由Xavier工控机端运行模拟GPS数据(GPRMC报文数据)发送模块和PPS信号发送模块。

之所以采用模拟的GPRMC数据，主要原因是本课题的研究背景中不允许使用GPS之类的传感器，但是激光雷达的授时同步又需要此类报文数据，所以只能由工控机通过串口编程，模拟一串GPRMC格式的报文(包含系统UTC时间)，发送到激光雷达端。

对于GPRMC报文数据帧格式包含14个字段，具体如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 | F11 | F12 | F13 |

其中对于每一个字段：

F 0：帧头，$GPRMC

F 1：UTC时间, 格式hhmmss.ssss，代表时分秒.毫秒

F 2：状态 A/V:分别代表定位成功/失败

F 3：纬度，格式ddmm.mmmmmm

F 4：纬度 N/S

F 5：经度，格式dddmm.mmmmmm

F 6：经度 E/W

F 7：速度

F 8：方位角

F 9：UTC日期 DDMMYY 日月年

F 10：磁偏角

F 11：磁偏角方向E/W

F 12：模式，A =自动，D =差分，E =估计，AND =无效数据

F 13：校验和

如下例子，则表示的系统UTC时间为21-03-24日的9:32分17秒：

“$GPRMC,093217.119,A,2237.496474,N,11356.089515,E,0.0,225.5,240322,2.3,W,A\*23”

通过PPS和模拟GPS数据（GPRMC数据）的组合就能够实现对激光雷达时钟同步。激光雷达在接收Xavier工控机发送模拟GPS信号后，以下一个PPS信号的上升沿作为时间同步基准。通过编程实现Xavier发送模拟GPS信号功能，如下图所示，具体要求为：激光雷达串口要求串口波特率：9600；数据位：8；校验：无；停止位: 1；PPS 信号要求为 标准的TTL 信号，脉宽大于 200ns。通过修改激光雷达驱动程序让激光雷达能够接收GPRMC数据帧提供的标准时间信号。使用RS232协议进行串口通信模拟GPRMC数据帧发送如下图所示：

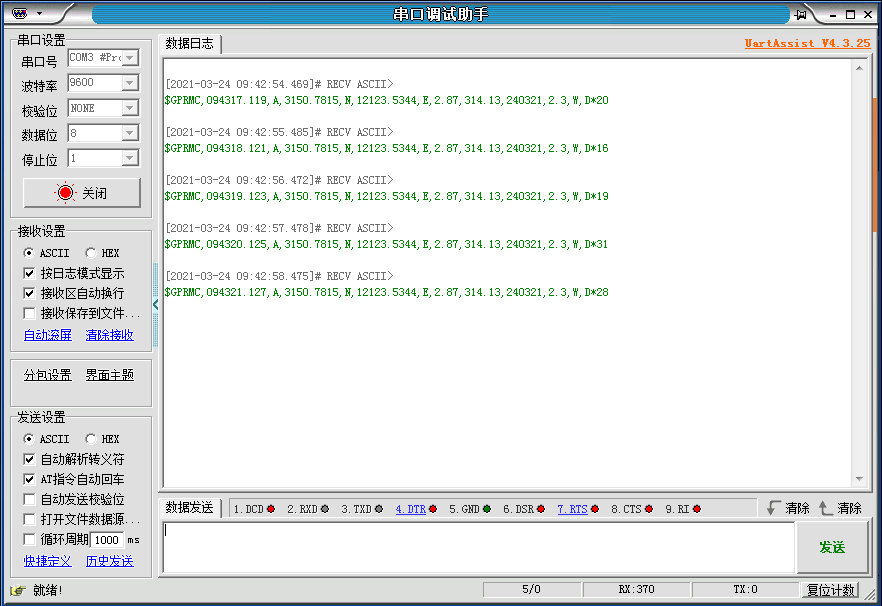


图3.3 串口助手发送GPRMC数据帧

* 1. 视觉相机的时间同步

由于涉及视觉传感器较多，而且每一帧的图像数据量较大，无法采用串行软触发的方式进行串行图像采集。所以本文中使用多相机的硬件触发，以此来实现相机之间的时间同步。视觉相机之间的时间同步示意图如图3.1视觉相机部分。

相机的采集模式一般分为内触发和外触发两种模式，其中内触发主要靠相机内部给出的信号进行图像的采集，外触发主要由外部信息例如软件信号、硬件信号来触发图像的采集。本文中使用的相机外部触发中的硬件触发，一般是使用外部设备通过相机的I/O接口相连，通过触发脉冲信号给到相机，来触发采集图像，其内部原理是直接对相机内部寄存器进行的读写。本文使用的U3相机背面接口图以及相机I/O接口定义如下：

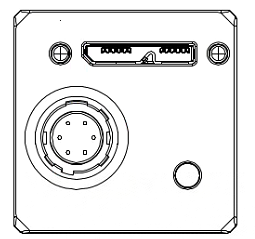


图3.4 U3相机背面接口图

接口中6pin的电源及I/O输入口对应的引脚定义如下表：

表3.1 管脚信号定义

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 管脚 | 信号 | I/O类型 | 说明 |
| 1 | 12V电源 | 输入 | 电源输入 |
| 2 | Opt-Iso In | 输入 | 光耦隔离输入 |
| 3 | GPIO | 输入/输出 | 可配置I/O口 |
| 4 | Opt-Iso Out | 输出 | 光耦隔离输出 |
| 5 | I/O Ground | 输入 | 触发-，信号地 |
| 6 | GND | 输入 | 电源地 |

具体的，硬件方面主要由彩色相机、鱼眼相机、热成像仪、多光谱相机均与Xavier工控机连接，并将各类相机的外部硬件触发引脚统一接到Xavier工控机上的IO口；软件方面需要配置相机采集帧率与激光雷达帧率相同，本文为20HZ，并且设置触发模式为上升沿硬触发，采集帧率20HZ目的是为了与激光雷达帧率相同，便于后续的相机与激光雷达之间的时间同步，另外需要在Xavier工控机上运行触发模块，实现IO口方波信号的输出，并利用上升沿来触发相机采集图像。如图3.5为虚拟逻辑分析仪查看的工控机通过IO输出的方波信号：

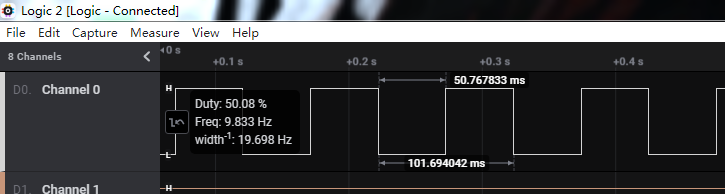


图3.5 工控机IO口输出的方波信号

针对Xavier工控机发出指令到触发信号产生这段时间，由于Xavier是ARM架构，ARM一条指令的执行时间是一个时钟周期，通常小于1纳秒，因此该部分时间误差可以忽略不计。对于相机之间的时间同步效果如下图：

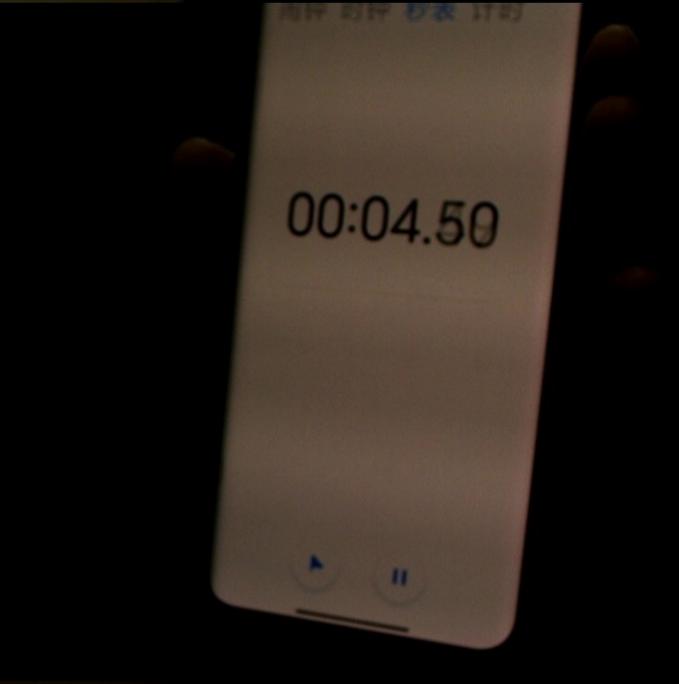
 

图3.6 彩色导航相机时间同步效果图

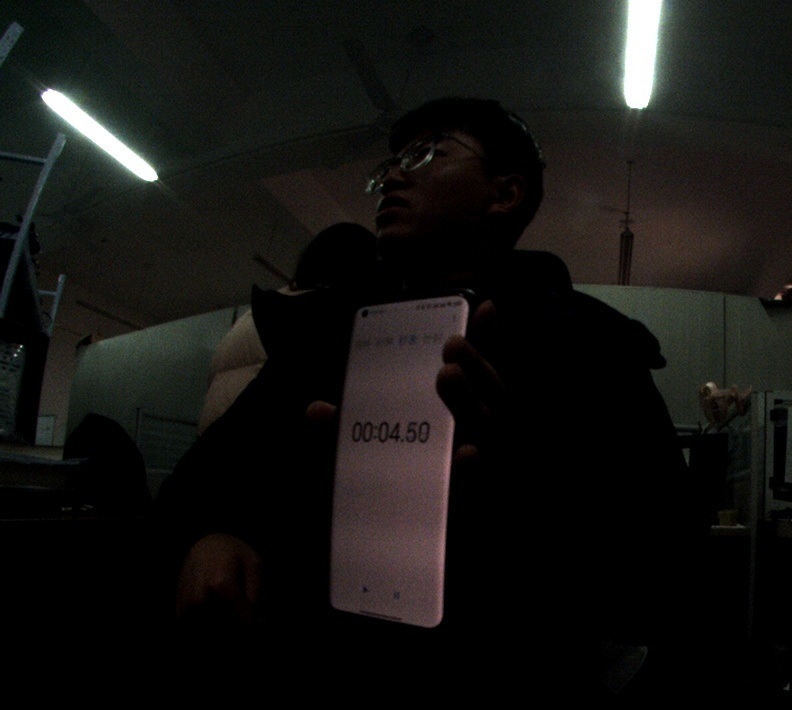
 

图3.7 鱼眼避障相机时间同步效果图

图3.8 多光谱相机时间同步效果图

由于手机秒表最高精度只到10ms级别，为了验证时间同步的精度，本文还使用了高精度的1ms级别的秒表进行相机外触发时的拍摄，以此说明此方法能够达到1ms以内的精度。其中彩色导航相机与鱼眼避障相机的高精度秒表拍摄图如下：

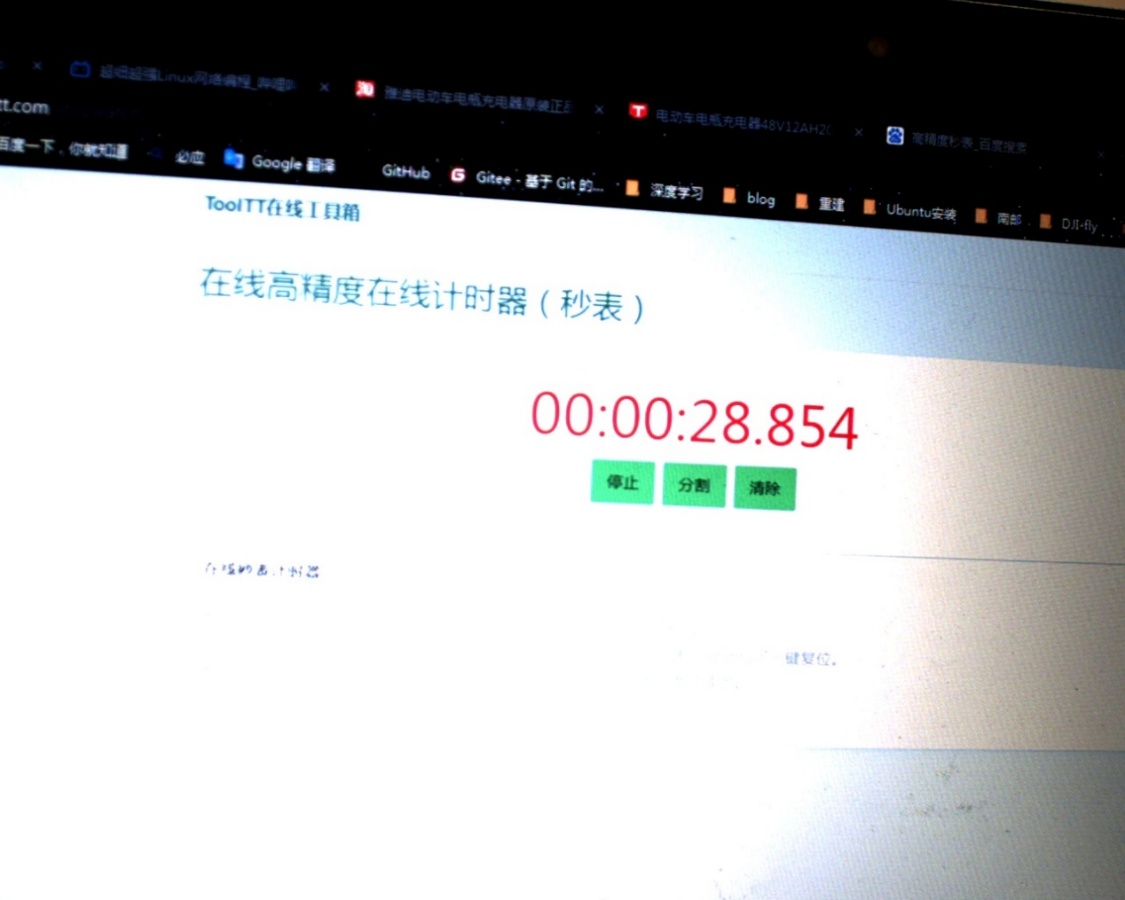
 

图3.9 导航相机与避障相机时间同步效果图

由上述触发模块实现了相机之间的外部硬触发，达到了相机之间的毫秒级时间同步。但是对于相同帧率的相机与激光雷达之间难免会存在一定的相位帧差，如图3.8所示。对于这种情况，我们通过测量t1、t2之间的时间差，进行相机外触发模块的信号就行对应时间的补偿，即可实现激光雷达与各类相机之间的时间同步。

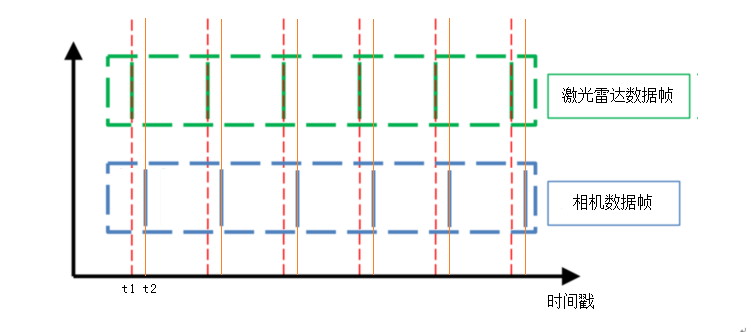


图3.10 雷达与相机时间戳相位帧差

* 1. 惯性测量单元IMU的时间同步

有关IMU的时间同步，同样可以采用3.2小节的与激光雷达相同的授时同步，也可以采用3.3小节中与相机外部硬件触发的同步方式。但是以上两种方式，对于IMU来说，显得较为昂贵。主要原因是，我们常用的IMU大多数是MPU6050或者MPU9250为主的较为便宜的只有6轴数据输出的传感器，且这种IMU已经能满足我们的正常需求，但是对于支持授时同步或者支持外部触发的IMU来说，价格将贵上好几倍甚至几十倍。所以本文中的时间同步方案，出于对成本的考虑，并且考虑到IMU的高频率采集数据的特点，采用的最直接的就近时间戳对齐的方式。

下面以激光雷达与IMU进行举例。如图3.11为雷达与IMU的数据帧示意图：

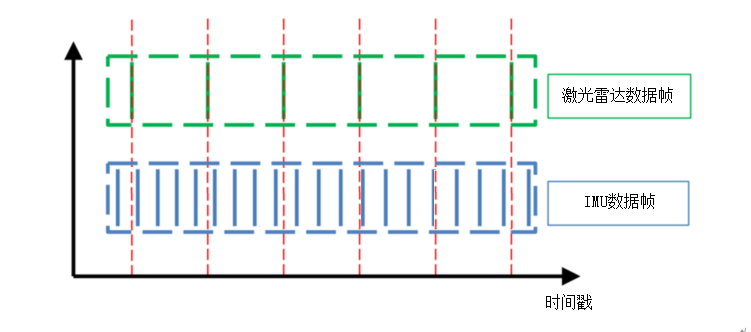


图3.11 激光雷达与IMU数据帧示意图

可以看到IMU的采集频率非常高，一般的IMU采集频率为200HZ，但是一般的激光雷达采集帧率只有20HZ甚至更低。所以，对于IMU的时间同步，采用就近时间戳对齐的方式无疑是最省时省力的。激光雷达与IMU数据帧之间就近时间戳对齐示意图如下：

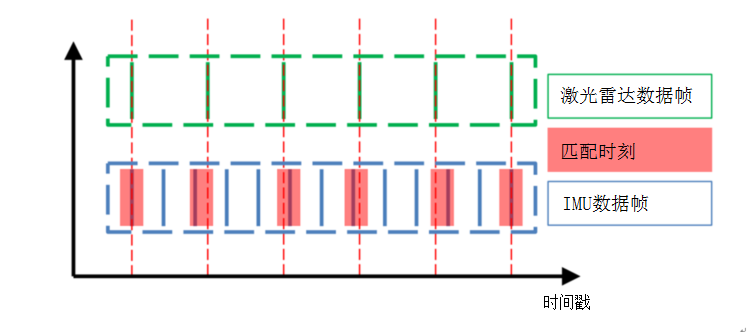


图3.12 激光雷达与IMU数据帧就近时间戳对齐示意图

下图为使用实验室传感器采集的激光雷达数据和IMU数据。数据帧示意图如下：

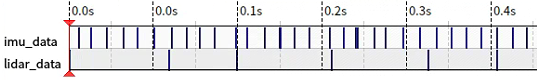


图3.13 雷达与IMU实际采集数据帧

通过就近时间戳对齐后的数据帧示意图如下：

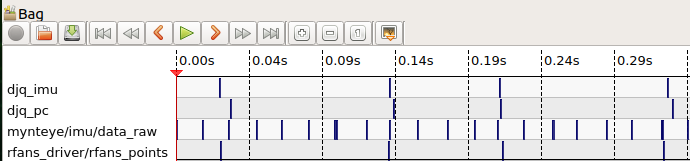


图3.14 雷达与IMU实现就近时间戳对齐

其中由上到下的4个topic分别为，时间戳对齐后的imu数据帧、时间戳对齐后的雷达数据帧、时间戳对齐前的imu数据帧、时间戳对齐前的雷达数据帧。

由图3.14可以看出时间戳对齐后的imu数据和雷达数据已经达到了毫秒级的精度，已经满足本课题中的对应的时间同步指标。但是由上图也可以看到，在经过就近时间戳对齐后的imu与雷达数据之间还是会存在一定的时间相位差，所以，对于更高要求的时间同步系统而言，还可以在本文方案的基础上对高频的imu数据进行插值，以此来提高时间同步的精度。对于本课题来说毫秒级的精度已经满足对应指标，所以没有对imu数据进行插值处理。

以上所述，便实现了IMU与激光雷达的时间同步。在3.2小节中实现了激光雷达与系统时钟的同步，在3.3小节中实现了激光雷达与相机之间的时间同步，在本小节中又实现了激光雷达与IMU之间的时间同步。至此，本文中的多种传感器已经完成了时间同步，且时间同步精度达到了毫秒级。

* 1. 本章小结

本章介绍主要围绕多传感器融合系统中的时间同步方法，基于各类传感器的软硬件特性，从三个方面实现了多种传感器的时间同步。分别为，其一，通过PPS信号+模拟GPRMC信号，使用Xavier工控机实现激光雷达的授时同步，实现激光雷达与系统UTC时间同步；其二，使用Xavier工控机发送相机外触发信号，实现各类相机之间的时间同步，并通过计算相机与激光雷达之间的数据帧的相位差来调整相机的触发信号，实现相机与激光雷达之间时间同步；其三，由于IMU采集频率很高的特性，使用就近时间戳对齐的方式实现IMU与激光雷达之间的时间同步。综上所述，本文中的所有传感器实现了时间同步，且时间同步精度达到了毫秒级。

# 多传感器空间一致性表征

在多传感器融合系统中，除了前面第3章介绍过的多传感器时间一致性之外，还需要进行多传感器的空间一致性。传感器的空间一致性主要是进行传感器之间的空间中相对位置关系的标定，也即是传感器的外参标定。对于不同传感器之间的外参标定来说，还需要提前进行传感器的内参标定，也即是传感器自身参数的标定。对于多传感器融合系统来说，多传感器的内外参的标定精度也会直接影响到后面的融合感知结果。所以，多传感器的内外参标定显得尤为重要。

* 1. 空间一致性表征系统总体设计

多个传感器的空间一致表征是进行数据的高效融合的基本需求，良好的一致表征是多传感器融合的基础，在平台上装了多个/多种传感器，而它们之间的坐标关系是需要确定的。本章节主要用内外参标定来确定它们的固有参数以及它们之间的位置关系，完成多传感器空间一致性表征，总体框架示意图如图4.1所示。

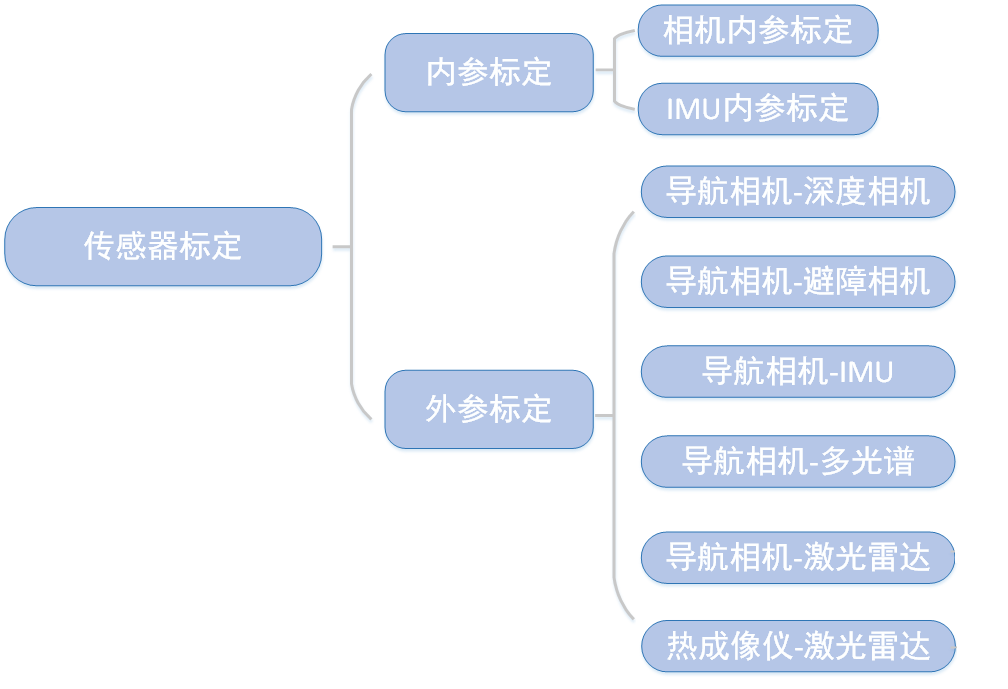


图4.1 传感器标定总体框架示意图

该部分主要分为传感器的内参标定和外参标定。内参标定为相机内参、IMU内参的标定；外参标定为相机与相机之间的标定、相机与IMU之间的标定、相机与激光雷达之间的标定。

* 1. 传感器内参标定

传感器的内参标定在本课题中主要为相机的内参标定和IMU的内参标定。相机的内参标定需要标定出各个光学相机的固有参数，例如焦距、主点偏移、畸变系数等；IMU的内参标定需要标定出其静置状态下的零偏和白噪声。

* + 1. 视觉相机内参标定

视觉相机的内参标定，较为成熟的便是张正友标定法。对于标定工具，可以使用OpenCV标定、使用Halcon标定、使用Matlab标定、使用Kalibr工具箱标定等。张正友标定法的具体步骤流程图如图4.2：

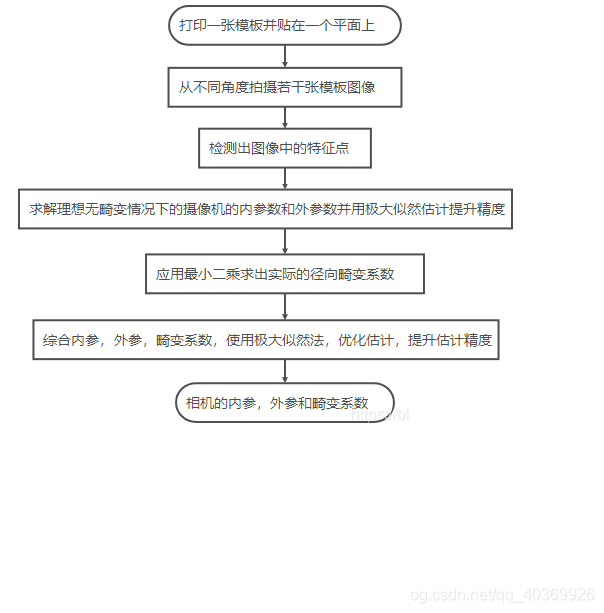


图4.2 张正友标定法步骤

1、打印一张棋盘格，把它贴在一个平面上，作为标定物。

2、通过调整标定物或摄像机的方向，为标定物拍摄一些不同方向的照片。

3、从照片中提取棋盘格角点。

4、利用已知三维世界坐标系下的点坐标和对应的图像坐标，估算解算相机的五个内参和六个外参。（代数法）

5、应用最小二乘法估算实际存在径向畸变下的畸变系数。

6、使用极大似然估计法（光束平差法：bundle adjustment）优化估计重投影误差，提升估计精度。

其中步骤1中可以使用A4纸打印棋盘格贴在平面上，但是此方法标定精度不是特别高，当然也可购买高精度的标定板。特殊的，对于本课题中的热成像仪而言，需要特定的能发热的棋盘格标定板，本课题中，通过导热性较好的铝板和隔热材料进行自制的简易发热标定板，也达到了不错的标定精度。

其中步骤3中的提取棋盘格角点，具体算法步骤为：先进行图像的灰度化，其次进行角点检测，最后为了提高精度在角点位置进行亚像素角点检测。

对于普通光学相机和热成像仪拍摄标定板图像分别如图4.3和图4.4：

|  |  |  |
| --- | --- | --- |
| E:\Hyper\公司项目\张正友标定\Calib\1617700762.622108.png | E:\Hyper\公司项目\张正友标定\Calib\1617700770.376519.png | E:\Hyper\公司项目\张正友标定\Calib\1617700771.720387.png |
| E:\Hyper\公司项目\张正友标定\Calib\1617700782.146869.png | E:\Hyper\公司项目\张正友标定\Calib\1617700787.504745.png | E:\Hyper\公司项目\张正友标定\Calib\1617700789.643445.png |
| E:\Hyper\公司项目\张正友标定\Calib\1617700790.981100.png | E:\Hyper\公司项目\张正友标定\Calib\1617700796.868442.png | E:\Hyper\公司项目\张正友标定\Calib\1617700801.948426.png |
| E:\Hyper\公司项目\张正友标定\Calib\1617700804.890375.png | E:\Hyper\公司项目\张正友标定\Calib\1617700806.487640.png | E:\Hyper\公司项目\张正友标定\Calib\1617700778.407103.png |

图4.3 光学相机拍摄标定板图像

|  |  |  |
| --- | --- | --- |
| F:\My_Halcon_Pro\07\1.jpg | F:\My_Halcon_Pro\07\3.jpg | F:\My_Halcon_Pro\07\5.jpg |
| F:\My_Halcon_Pro\07\7.jpg | F:\My_Halcon_Pro\07\9.jpg | F:\My_Halcon_Pro\07\11.jpg |
| F:\My_Halcon_Pro\07\13.jpg | F:\My_Halcon_Pro\07\16.jpg | F:\My_Halcon_Pro\07\17.jpg |
| F:\My_Halcon_Pro\07\18.jpg | F:\My_Halcon_Pro\07\20.jpg | F:\My_Halcon_Pro\07\22.jpg |

图4.4 热成像仪拍摄自制标定板图像

下面是本课题中用到的视觉传感器内参具体标定流程介绍：

关于相机内参标定，棋盘格的拍摄有一定的要求。首先是需要保证拍摄的图像数量大于4张，但也不是越多越好，经验所得20张以内较好；其次需要保证拍摄到的棋盘格所占图像空间的1/4至3/4最好，不易太小也不易太大；另外还需要保证多张图像的棋盘格覆盖过整个图像视野，且拍摄的棋盘格具有角度倾斜。

对于普通光学相机，可以直接进行图像的二值化，然后进行棋盘格角点提取并进行标定。但是对于热成像仪拍摄的棋盘格需要先进行图像预处理，即将棋盘格部分提取出来，在进行角点的提取和内参的标定。如下图所示即为热成像仪拍摄棋盘格的图像预处理结果：

|  |  |
| --- | --- |
| F:\My_Halcon_Pro\07\1.jpg | F:\My_Halcon_Pro\save\1.bmp |
| F:\My_Halcon_Pro\07\32.jpg | F:\My_Halcon_Pro\save\32.bmp |

图4.5 热成像仪拍摄棋盘格预处理结果

各类光学相机内参标定结果以及畸变系数分别如表4.1和4.2：

表4.1 相机内参标定结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 相机种类 | fx | fy | cx | cy |
| 彩色相机 | 610.9 | 610.2 | 325.9 | 228.5 |
| 深度相机 | 384.6 | 384.6 | 319.5 | 240.3 |
| 热成像仪 | 548.6 | 541.3 | 321.5 | 268.9 |
| 多光谱相机 | 1449.2 | 1448.6 | 638.9 | 490.3 |

表4.2 相机畸变系数

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 相机种类 | k1 | k2 | p1 | p2 | k3 |
| 彩色相机 | 0.137251 | -0.268334 | 0.001417 | -0.000494 | 0 |
| 深度相机 | 0.000287 | 8.327e-05 | -0.000458 | 0.000675 | 0 |
| 热成像仪 | -3.564e-01 | 1.660e-01 | -5.323e-03 | 2.431e-03 | 0 |
| 多光谱相机 | -0.102196 | 0.167307 | 0.000881 | 3.774e-05 | 0 |

* + 1. 惯性测量单元IMU内参标定

惯性测量单元IMU是由陀螺仪和加速度计构成的，其中陀螺仪有3个轴，加速度计也有3个轴，最理想的情况下，陀螺仪和加速度计各自的3个轴是正交的。但是对于我们常见的一般IMU来说，这种情况是不存在的。

准确的IMU数据对目前较为普遍的多传感器融合VIO，例如VINS、ORB3等算法起到非常关键的作用，能使定位的精度更高，视觉里程计算法更鲁棒。所以我们在使用IMU之前，都需要对IMU进行内参标定，也即是标定出IMU自身的一些参数，也叫参数辨识。这些参数主要包括加速度计和陀螺仪各自的零偏、标度因数和安装误差。

下面是对于标定参数的分析：

1. 零偏：零偏就是IMU的输出比输入多了一个常量误差。

加速度计的零偏一般可以表示为：

陀螺仪的零偏一般可以表示为：

（2）标度因数误差：标定因数误差也称为刻度误差，也即标准输出和输入的比值不是1时，就需要标定。

加速度计的标度因数一般可以表示为：

陀螺仪的标度因数一般可以表示为：

对于常见的VIO、LIO算法中的IMU来说，需要标定的噪声一般分为两种：零偏（Bias）和白噪声（White Noise），它们都有连续时间模型和离散时间模型两种形式，两种方式也都可以进行相互转换。有的VIO算法需要的是连续型的噪声参数、有的需要的则是离散型的噪声参数。对于我们常用的标定工具，例如imu\_utils和kalibr\_allan，其中imu\_utils工具包输出的是连续型的，kalibr\_allan输出的是离散型的。

由于后续需进行IMU与相机的联合标定，所以此处需要介绍一下Kalibr工具所需的IMU噪声参数的形式。Kalibr标定工具需要的噪声参数形式属于连续时间模型，具体参数如下表：

表4.3 IMU噪声参数格式

|  |  |  |  |
| --- | --- | --- | --- |
| 参数 | YAML文件元素 | 符号 | 单位 |
| 陀螺仪“White Noise” | gyroscope\_noise\_density |  |  |
| 加速度计“White Noise” | accelerometer\_noise\_density |  |  |
| 陀螺仪“random walk” | gyroscope\_random\_walk |  |  |
| 加速度计“random walk” | accelerometer\_random\_walk |  |  |
| IMU采集频率 | update\_rate |  | Hz |

关于IMU内参的标定较为简单，本课题中使用港科大开源的imu\_utils工具包进行的内参标定。标定步骤操作简答，只需要录制imu在静置状态下的2个小时以上的数据即可进行标定。使用imu\_utils工具包进行IMU内参标定的结果如下图所示：

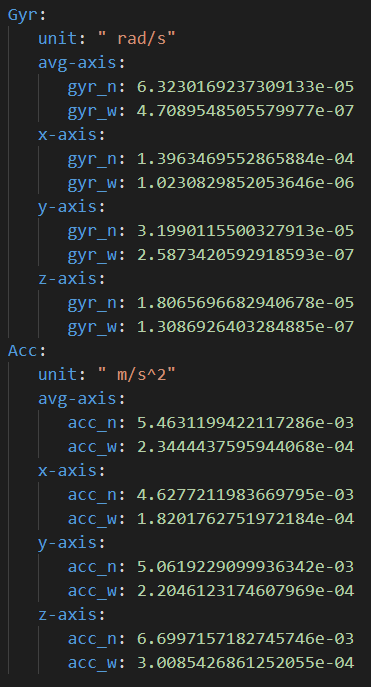


图4.6 IMU内参标定结果

* 1. 多传感器外参联合标定

上述4.2小节主要介绍了视觉相机的内参标定和惯性测量单元IMU的内参标定，其目的是为了标定出对应传感器的自身参数、噪声等。本小节主要在前面内参标定完成的基础上进行的传感器之间的外参标定，也即是求解出传感器之间的相对位置关系，以便后续的融合感知等系统的使用。

* + 1. 视觉相机与视觉相机之间联合标定

视觉相机之间的外参标定，主要应用在三维重建、双目立体视觉匹配等系统中。本课题中的多类视觉相机，包括彩色相机、灰度相机、热成像仪和多光谱相机之间的标定都可以看做是双目相机的标定，也即是分别进行彩色相机-灰度相机之间外参的标定、彩色相机-热成像仪之间外参的标定、彩色相机-多光谱相机之间外参的标定。

双目相机标定的主要目的是确定两个相机之间的相对位置关系，以便用于后续的三维重建，双目测距等系统。通过图4.5所示，一般的两个不同的相机或是双目相机的成像平面是不平行的，我们需要进行双目相机标定来将如图4.5所示的双目相机模型图转换成图2.4所示的，两个相机光轴平行的双目视觉模型。

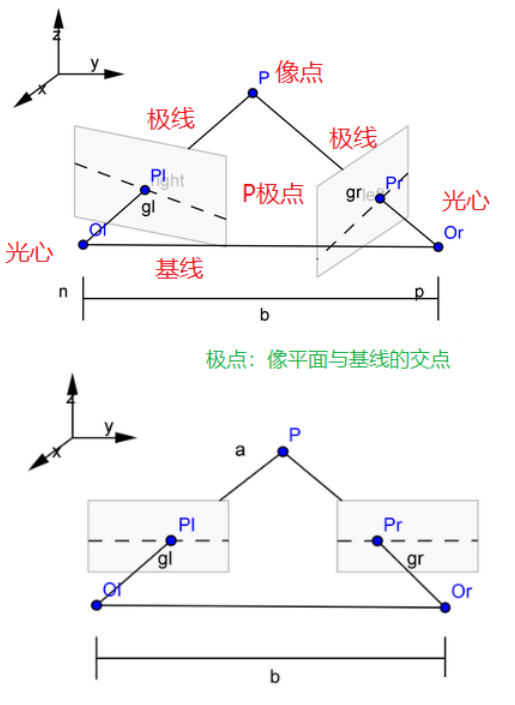


图4.5 光轴不平行的双目模型

下面将介绍如图4.6所示的双目相机标定的算法原理：

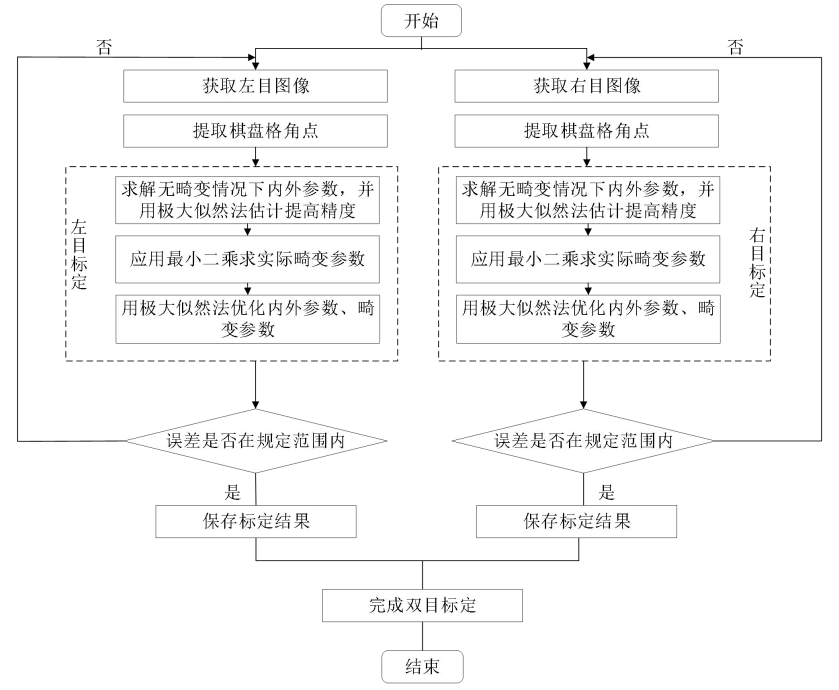


图4.6 双目标定算法原理

对于双目相机的标定，主要是需要求解出两个相机之间的相对位置关系，即相对旋转和相对平移。假设为所拍摄的标定板上某一点在空间中的坐标；、分别为左右两个相机成像的世界坐标系下的坐标；、为点相对左目相机的旋转和平移；、为点相对右目相机的旋转和平移。所以，有如下公式：

将4.1式进行化简可得：

对4.2式进行两式相减，消去可得：

对4.3式两侧同乘并进行化简可得：

对于4.4式中，令、，则有：

所以，上式就是右目相机相对于左目的空间变换，R为旋转矩阵，T为平移向量。又因为旋转矩阵是正交矩阵，所以。所以对于4.5式中的R、T可以表示为：

其中，、、、这几个参数在进行单目相机标定的时候即可获得。所以，带入式4.5，因为拍摄了许多组双目标定板图像，即可利用最小二乘法，也可进行奇异值分解的方法，进行最小化误差，最终可得到最佳的旋转矩阵。

下面是本课题中用到的视觉传感器外参具体标定流程介绍：

关于相机外参标定，棋盘格的拍摄有一定的要求。采集数据为棋盘格图像数据，采集图像数据要保证棋盘格角点的清晰，最好保持相机的采集帧率不要过高，防止图像模糊，不易于提取角点。采集图像总数量最好为20张左右，并且需要所有数据中的棋盘格尽量覆盖图像所有区域，且拍摄的棋盘格具有角度倾斜。其中双目相机外参标定靶标位置示意图如下：

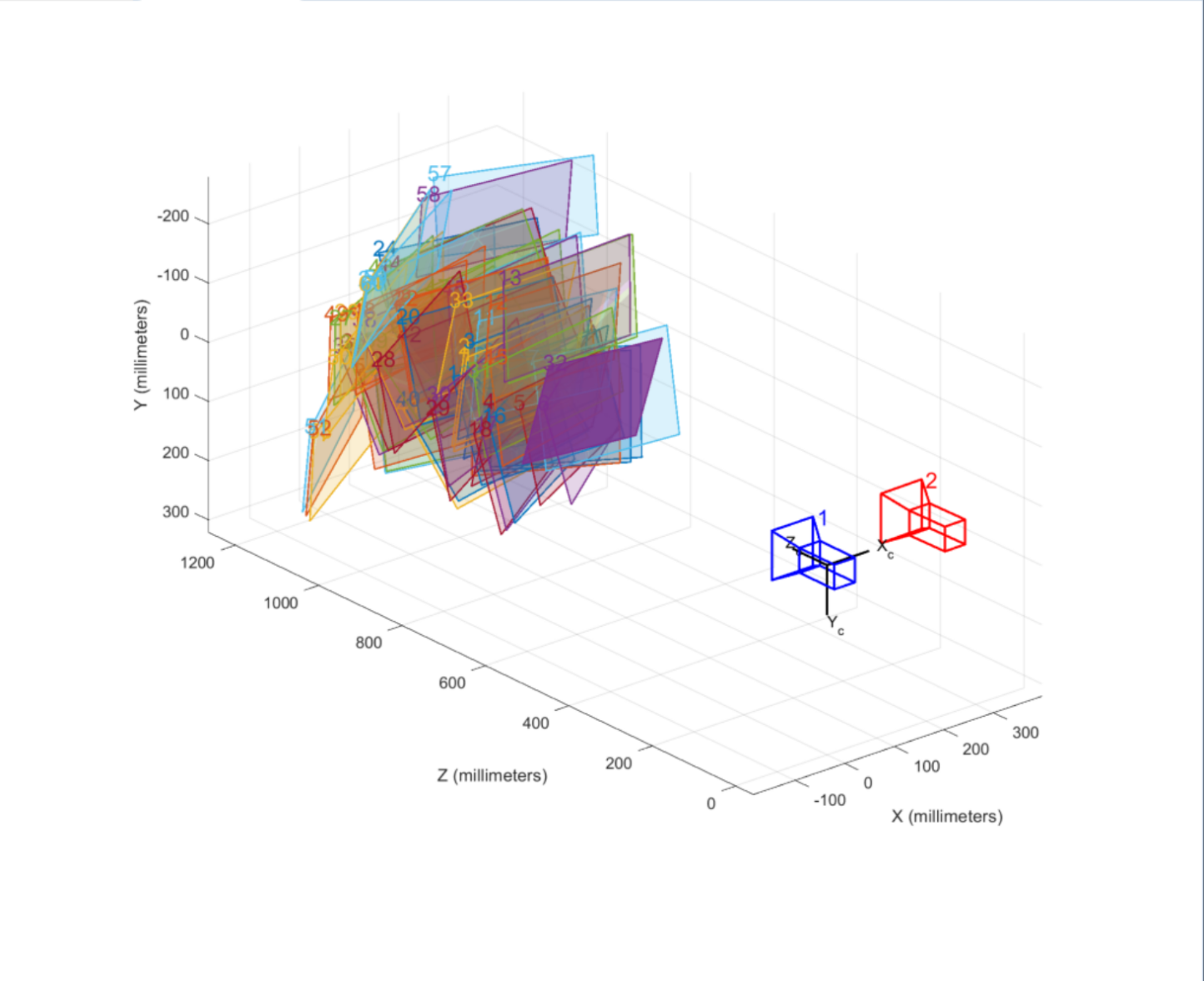


图4.7相机外参标定靶标位置示意图

最终进行两两相机之间外参标定，彩色相机与灰度相机、分别有彩色相机与深度相机、彩色相机与热成像仪、彩色相机与多光谱相机。最终将标定结果分别如下图所示：

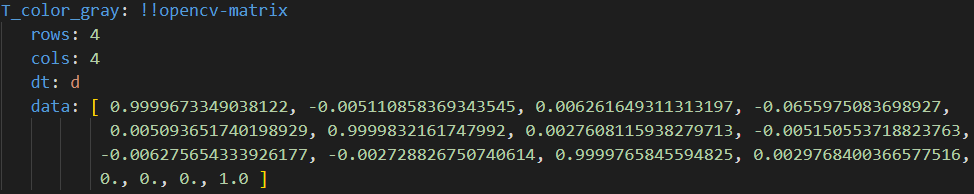


图4.8 彩色相机与灰度相机外参标定结果

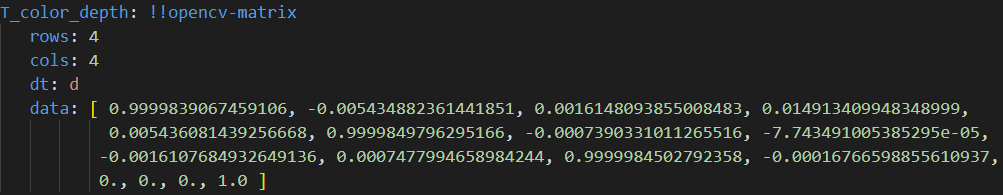


图4.9 彩色相机与深度相机外参标定结果

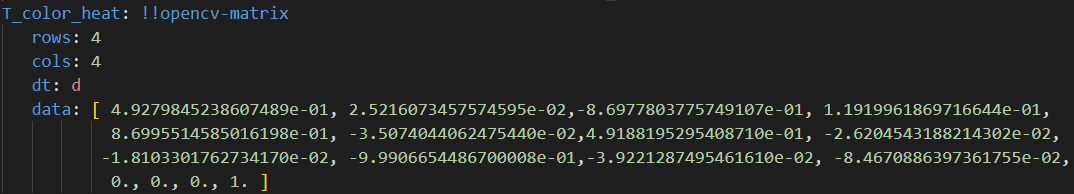


图4.10 彩色相机与热成像仪外参标定结果

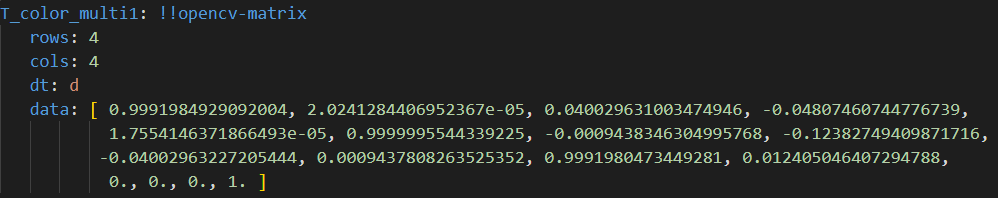


图4.11 彩色相机与多光谱相机外参标定结果

* + 1. 视觉相机与激光雷达之间联合标定

激光雷达与相机之间的联合标定在常见的多传感器融合感知系统中应用非常多，所以使用一个精度较高的标定方法是取得高精度结果不必可少的一步。常见的激光雷达与相机之间联合标定方法有手动法和自动法，手动法即为人工手动的选取点云和图像中的对应的特征点，然后进行求解计算；自动法则为算法自身进行点云和图像特征点的提取，但这种方法精度一般，也有进行线面特征等进行关联，然后3D配准进而增加标定精度的方案。

常用的标定工具有Autoware和Apollo工具箱。其中Autoware是一个由日本开源的一个自驾相关的平台工具，其包含了相机的标定，相机与雷达的联合标定，所需的数据较为简单，但是需要在标定过程中由人工来选取激光点云中的3D特征点与图像中的2D特征点进行匹配，进而求解出外参；Apollo是百度公司推出的阿波罗自动驾驶平台，其也包含相机与雷达之间的联合标定，但是其标定对于数据的需求较高，不仅需要普通的数据，有时还需要IMU以及里程计等传感器的相关数据，并且其还需要一个粗略的外参才能进行外参的求解与优化。所以我们最常见的还是使用Autoware来进行相机与激光雷达之间的外参标定。

下面将介绍如图4.12所示的lidar-cam联合标定的算法原理：

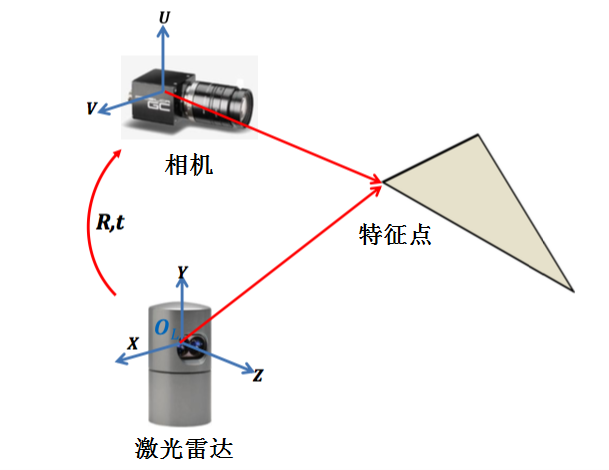


图4.12 lidar-cam联合标定示意图

在如图4.7所示的lidar-cam联合标定示意图中，相机坐标系为-U-V，激光雷达坐标系为-X-Y-Z，他们之间相差一个旋转变换对应的旋转矩阵R和平移变换对应的平移向量t，这也就是我们需要标定的量。假设某一特征点的三维坐标为,此点映射到二维像素平面点为，则由映射关系可得如下式子：

其中、、、均为相机参数，也即是相机的内参，、分别为相机XY方向的焦距，、是像平面的主点坐标。其中R为旋转矩阵、t为平移向量。其中的M为变换矩阵，具体可表示为：

对4.7进行矩阵展开可得：

将4.8式进行变量提取，再转换为矩阵乘积形式可得：

对于4.9式，我们可以通过不同姿态下标定板平面，或者通过激光雷达的三维点云特征点与图像中对应的二维图像特征点，可以得到一系列的线性方程，进而可求解出M矩阵，也即是我们所求的外参标定参数。

下面是本课题中用到的视觉传感器与激光雷达之间外参标定流程介绍：

本课题中，使用的视觉与激光雷达之间外参标定的算法为Autoware工具箱中的标定算法，其原理同前面所介绍。但需要手动的进行激光雷达中3d特征点点的选取，以及图像中与其对应的2d特征点的选取。标定场景选取如下图：

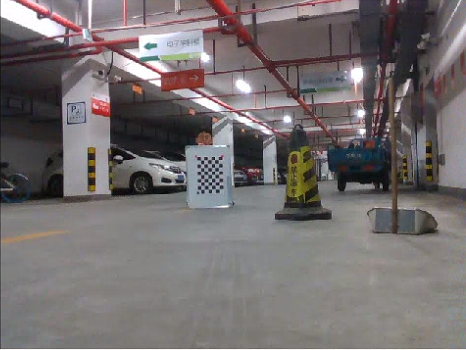
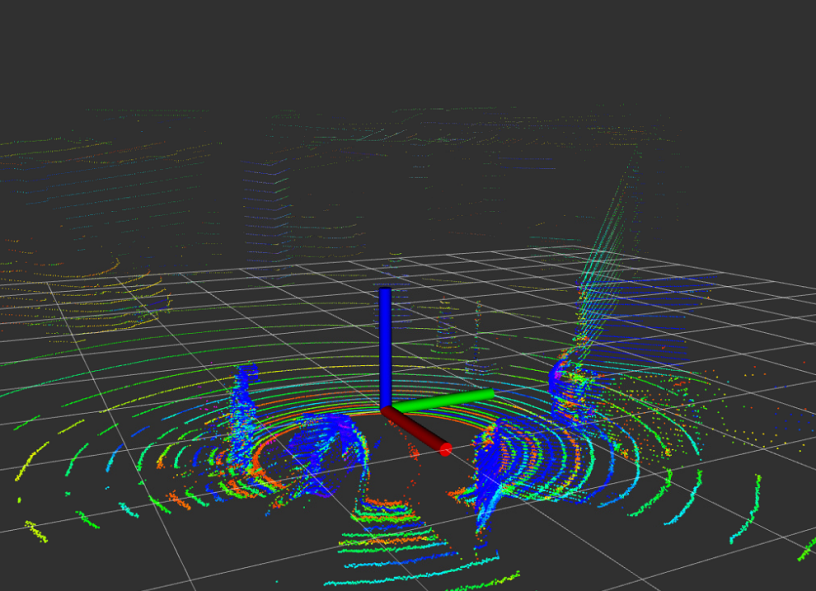
 

图4.13 彩色相机原始图像与激光雷达原始点云

使用Autoware进行相机与激光雷达的外参标定，需要首先在rviz界面中进行手动选择3d点，然后在图像中选取对应的2d像素点，共要9对数据即可进行标定。结果如下图所示：



图4.15 彩色相机与激光雷达外参标定结果

相机与激光雷达标定后可以进行点云到图像的投影可视化，如下图所示即为滤除噪点后的激光雷达点云投影到彩色图像可视化结果：

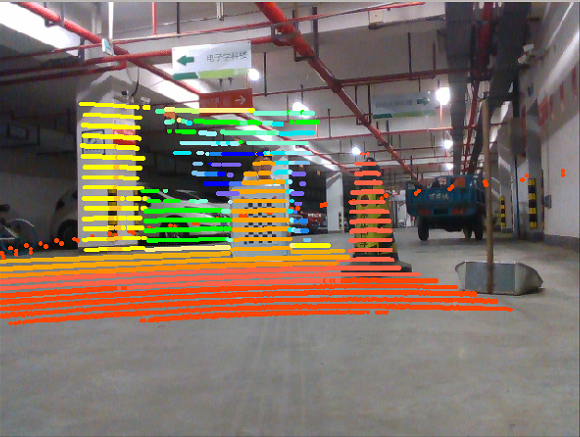


图4.16 激光点云到彩色图像的投影可视化

* + 1. 视觉相机与IMU之间联合标定

在现有的多传感器融合SLAM方案中，视觉与惯导的组合一直十分被认可，主要也因其成本低、精度高等优点。对于这两种传感器的组合，之所以有比较好的融合效果，这是因为他们之间可以做到信息的互补：例如相机在快速运动时会产生模糊，在光照环境不好的条件下视觉算法会失效，但有了IMU的加入，由于IMU的高频数据采集的特性，且不受光照等环境的影响，可以大大的弥补视觉相机的不足。相机与IMU配合如图4.8，可以获得相机较为丰富的图像数据，又能通过视觉回环检测来有效的修正IMU长时间工作带来的累积误差。

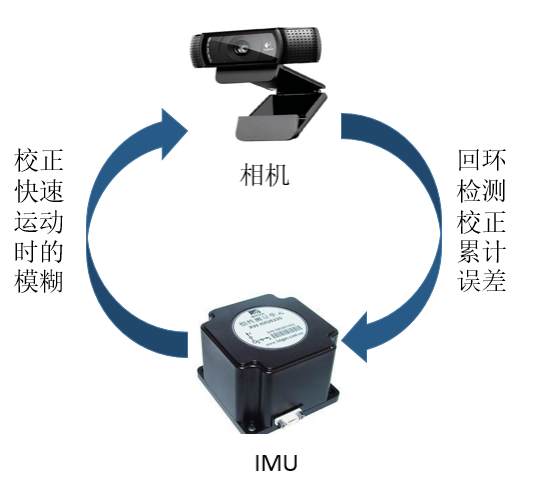


图4.17 相机与IMU融合示意图

使用Kalibr进行相机-IMU之间的外参标定，一般分为三个步骤：

1. 粗略进行相机与imu之间的时延估计：

由于传感器的传输延时和触发延时等情况的存在，导致如下图所示的传感器的时间戳存在时间差，即：

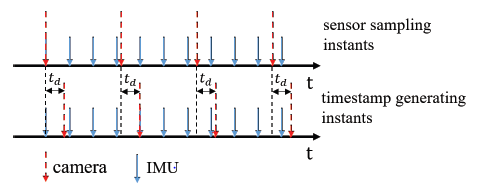


图4.18 相机与IMU之间时间差

在前面的4.2.1小节，以及进行过相机内参的标定。现在已知了图像每一帧的3d与2d之间的对应，所以就可以计算出相机每一帧的离散位姿。通过这些离散的位姿就可以进行连续的B-spline构造，进而获得任意时刻的相机位姿。

这样就可以利用相机的样条曲线获得任意时刻下相机对应的旋转角速度，同时对于imu来说，其陀螺仪又可以测量出角速度。所以先忽略噪声的影响，相机与imu之间只相差一个旋转。所以利用相机与imu数据的两个曲线的相关性，可以粗略估计相机与imu之间的时延，并且利用此相关性，可以将时延的误差缩小至2个imu数据周期范围内。

1. 获得cam-imu之间的初始旋转，以及重力加速度、陀螺仪偏置等初值：

同样的，还可以利用角速度测量的关系，再进行一个优化问题的构造：

通过上述问题的优化，就可以获得相机与IMU之间的旋转，并且也能够获得陀螺仪偏置初始值。

假设在整个标定的过程中，imu的加速度计平均加速度为零，并且在忽略加速度计的偏置与噪声的同事，可以获得重力加速度在参考坐标系下的表示如下：

1. 进行大优化，具体包括相机的重投影线误差和imu的测量误差与噪声：

该步骤的大优化，主要是调整所有的待优化的量，使得其观测误差最小。前面两步也为该步骤提供了一个不错的初始值。待优化的误差项主要包括相机拍摄标定板的重投影误差、imu自身的测量误差和噪声。

下面是本课题中用到的视觉传感器与IMU之间外参具体标定流程介绍：

本课题中使用的cam-imu外参标定的工具为kalibr工具箱。首先需要进行bag包的录制。其中需要相机一直能拍摄到棋盘格，以便优化时进行重投影误差；并且相机的拍摄频率不能太快；最重要的是需要给imu足够激励，即使得imu在xyz各个方向都有移动和旋转。

对于kalibr工具箱的输入需要有上述录制的bag包以及相机与imu各自的内参文件。通过kalibr工具箱进行cam-imu之间的外参标定结果如下：

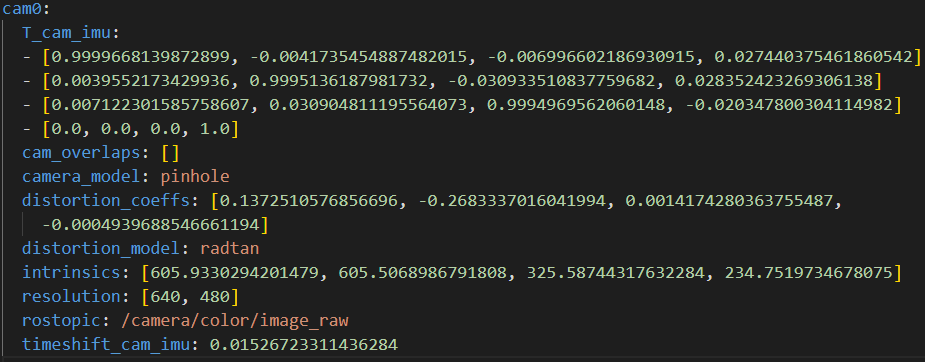


图4.19 相机与IMU外参标定结果

* 1. 本章小结

本章主要围绕多传感器融合系统中的空间同步方法，从两个方面实现了多种传感器的空间同步。分别为传感器各自的内参标定和传感器之间的外参标定。其一，对于本课题中的传感器各自的内参标定，主要包括所有视觉相机的内参以及IMU的内参；其二，对于本课题中的传感器之间的外参标定，主要包括视觉相机之间的外参标定，相机与激光雷达之间的外参标定，相机与IMU之间的外参标定。通过测量传感器的安装位置与标定结果的对比，实现了多传感器的空间一致，且传感器之间外参标定的精度为毫米级。

# 多源多层地图三维重建

三维重建发展至今，相关技术取得了巨大的成就。不管是在游戏开发、遗迹保存、医学三维CT等领域，其都有着及其广泛的应用。在进十几年的发展过程中，不论是基于单目相机的三维重建，还是基于RGB-D相机的三维重建算法层出不穷，并且基于深度学习的许多端到端三维重建网络在这几年发展的也越来越好。各种算法都有其自身的优势和缺点，例如基于深度学习的端到端的三维重建网络虽然发展火热，但是此类算法需要相机的已知位姿，无法做到直接由一系列图像到三维点云的重建，并且针对火星此类场景，由于火星车的慢速移动且无回环，很难通过传统的SLAM算法给出精确的位姿。所以，基于航天院火星场的研究背景，本课题使用了基于MVE三维重建算法，并且进行了热和光谱信息在三维地图的叠加，实现多层地图的构建，以便后续的感知和路径规划等工作。

* 1. 重建算法设计方案

本章节基于航天院火星场的场景进行三维地图重建，并在三维地图上加入热、光谱信息，形成一个高维的多层地图。主要使用的传感器有RGB相机、热成像仪和多光谱成像仪。具体算法框图如图5.1所示，首先，利用RGB图像序列通过SFM进行相机位姿的计算；然后，利用已知的相机参数通过MVS进行各个视角下的深度图恢复；最后通过FSSR算法进行物体表面的恢复；在完成了上述的重建算法之后，利用已知的RGB相机与热成像仪、多光谱之间的外参，可以进行热和光谱信息到三维彩色地图的映射，最终形成多层地图。



图5.1 多源多层地图三维重建算法框图

* 1. 基于MVE算法的多层地图重建

本课题中的多源多层地图重建，主要基于航天院火星场的场景进行实验，传感器主要包括RGB相机、热成像仪和多光谱相机。本章节采用MVE算法，实现一个端到端的三维地图重建，最后再将热和光谱信息向地图中映射，生成多层地图。MVE算法，即多视图环境，它的输入是场景的图像，最终输出的是三角形网格，该算法主要由从运动恢复结构、多视图立体重建、密集点云生成和曲面重建构成。具体算法流程示意图如图5.2所示。

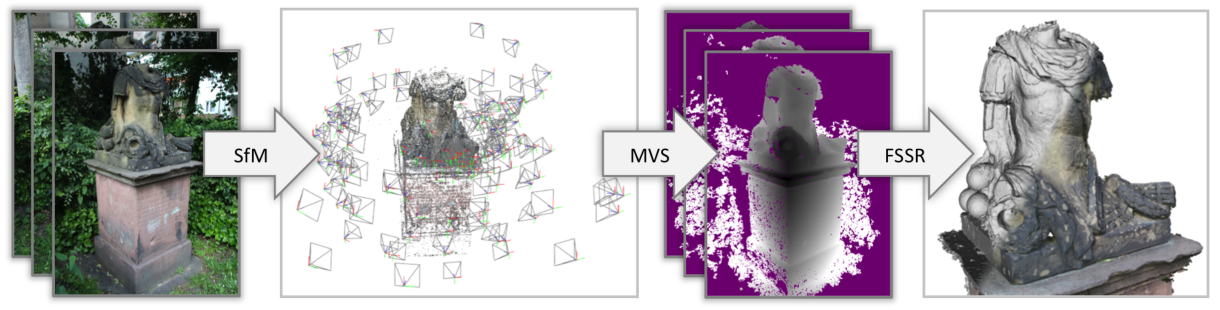


图5.2 多视图重建MVE算法流程示意图

* + 1. 运动恢复结构SFM

从运动恢复结构的SFM算法可以说是计算机视觉和三维重建中非常高的成就之一。本质上来说，SFM是从非结构化的一组图像中的稀疏对应关系重建相机的参数和稀疏点云。其恢复的相机参数主要包括相机位姿。对于常见的增量式SFM算法可以分为如下几步：

1. 特征点提取与匹配：

特征点提取：对输入图像进行特征提取需要考虑特征对于旋转、噪声、光照以及图像尺寸等易变的信息的不变性。常用是具有尺度不变性和旋转不变性的SIFT描述子，该描述子对于上述提到的一些易变信息具有较强的鲁棒性，非常适合用来提取具有尺度变换或是带有旋转的图像特征点信息。

特征点匹配：通过上述特征点提取，可以获得两幅图像的SIFT特征点描述子，又由于两幅图像之间的对应点收到相机模型之间的极线约束，所以通过这些约束可以去除掉许多的错误匹配。对于两两匹配的图像，常以欧氏距离进行匹配成功与否的判断，如果距离小于某一阈值则认为匹配成功，但是此种匹配方式依然会存在较多的误匹配，还需要进行误匹配的剔除。针对误匹配点的剔除常见方法有两种，第一，剔除掉最近距离与次近距离的比值较大的；第二，对匹配点采用RANSAC算法计算其对应的基础矩阵F，若不满足基础矩阵，也需要进行剔除。

1. 相机位姿估计，并重建三维坐标点；

相机相对位姿求解：在进行了第一步的特征点匹配后，即可获得两张图像中的多对2D-2D匹配特征点。可以通过8点法估计本质矩阵E。有了本质矩阵E，即可将其进行SVD分解出两个相机之间的旋转矩阵R和平移向量t，但是SVD分解会得到四组可能的解，需要通过检查旋转矩阵R的正确性，并找到能尽可能把2D点投影到相机坐标系下的3D点的R和t，对应的解则为最可能的解。

重建三维坐标点：在已知了两个相机之间的变换关系以及成对的匹配点坐标之后，便可以通过三角测量来恢复出匹配点在三维空间中的坐标。最后再将这些恢复出的三维空间点向两个视角进行重投影，计算与原来2D坐标的距离，以此来说明三角化误差。

计算第三个视角到世界坐标系的变换矩阵：由前面的两个视角的特征点匹配计算出的视角二相对于视角一，也即是世界坐标系的变换矩阵。在加入第三个视角后，使其与第二帧进行特征点匹配，在这些匹配点中，有一部分是第一帧与第二帧已经匹配成功且已恢复三维坐标的点，所以这种情况下就可以根据PnP算法进行求解第三个相机到世界坐标系的变换矩阵。对于加入的更多图像帧，同上方法即可求得新的一帧可恢复出的三维点云，与全局点云进行融合即可。具体流程如图5.3所示。



图5.3 SFM恢复三维点云流程图

1. 进行BA优化

在随着上述不断的新图像加入到算法中的时候，累积误差会逐渐增大，可能会导致最终的重建误差较大甚至是完全偏离重建目标。为了重建过程中相机位姿以及重建的三维点云的准确，需要进行点云和相机位姿的优化。常用的方法为光束平差法（Bundle Adjustment），其本质上是一个非线性的优化算法，对于BA的求解，Ceres Sover库具有很高的效率。

* + 1. 多视图立体视觉MVS

《正文》×××××

* + 1. 表面重建FSSR

《正文》×××××

* + 1. 多层地图映射

《正文》×××××

* 1. 实验结果与分析

实验部分

* 1. 本章小结

《正文》×××××

# 基于双目视觉的三维环境感知

在三维环境感知领域，常见的算法大多是基于激光雷达进行的端到端的3D点云目标检测。现有的3D点云检测算法和分割算法大都停留在对公开数据集或者车、人、建筑物等大物体的检测。但是对于一些小物体，例如果园的树枝、水果、石头等，以及类似于航天探索领域的砂石等物体便难以鲁棒的进行检测和分割。初次之外，现有的算法大都依赖激光雷达的点云数据，这也就导致了高昂的成本，并且对于3D点云的网络训练数据的标注费时费力。因此，本课题采用了成本较低的双目视觉系统来进行三维环境的感知和特定目标的空间定位。

* 1. 感知系统算法设计方案

本课题主要研究背景为航天探索相关的火星场场景，使用成本较低的双目视觉系统进行目标的三维空间定位。具体算法分为两步，如图6.1所示。首先，通过双目相机获取的图像，使用BGNet立体匹配网络进行训练，并进行视差图估计，再将视差图通过相机内参转换成深度图进而恢复出3D点云。其次，使用SOLO分割网络实现对特定目标2D位置信息的获取，从而估计目标物体的在左目相机中的空间位置信息，即物体在左目相机中的三维空间坐标。因为无人车在行进过程中触碰到物体即会发生危险，目标物体的空间位置信息定义为：距离左目相机光心最近的物体表面三维点的空间坐标。

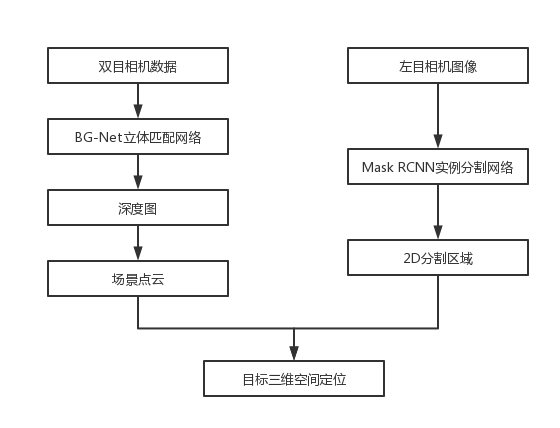


图6.1 双目感知系统算法

* 1. 基于双目视觉的目标空间定位

本课题中的基于双目视觉的三维环境感知系统主要以火星场、仿真软件中的模拟场景进行实验。通过双目立体匹配网络进行左目相机下的深度图恢复，并使用左目图像进行对应物体的分割，最终在多个3D物体点云中再进行特定物体的剔除与筛选。

* + 1. 基于SOLO网络的实例分割算法

实例分割算法在这些年发展的越来越好，但是同时也存在了很大的挑战性。实例分割相较于语义分割更加细致，一个语义类别包含的实例个数可能是多个。在SOLO算法之前，现有的方法基本要么遵循Mask R-CNN一类的算法先检测后分割测量，要么先进行类别的预测，再进行聚类、关系亲合场等方式来进行实例的划分。SOLO算法的效果和优势包括：端到端算法，无需任何的后处理；在COCO数据集上达到了Mask R-CNN的效果；不需要bbox标注信息，只需要mask的标注信息。SOLO与Mask R-CNN算法相比，不同之处如图6.2所示：

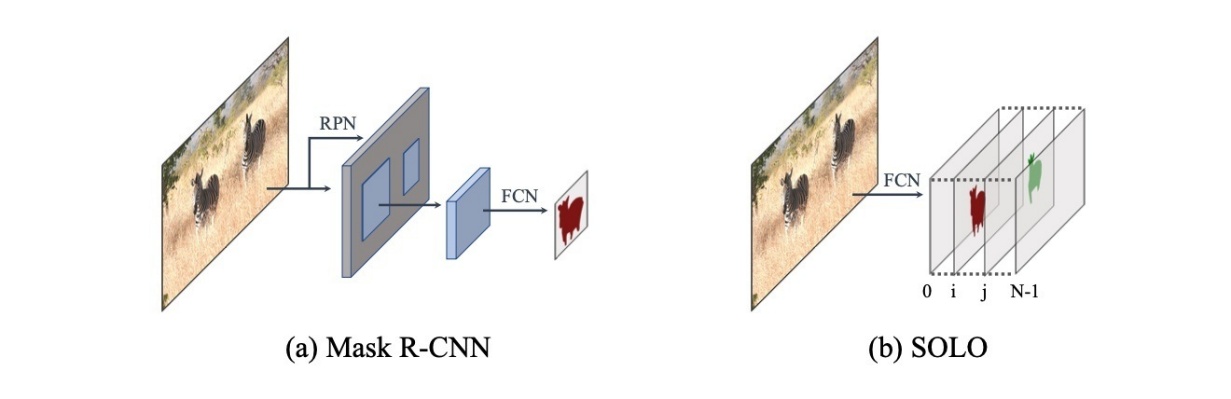


图6.2 SOLO与Mask R-CNN算法对比图

SOLO实例分割算法的目的是确定图像中是否有对应的实例，有的话就返回实例的mask。该算法的核心思路是将实例分割的问题转化成类别预测和实例 mask 生成的两个问题，具体的：首先将图像分成S×S个格子，然后分别通过分类分支和mask分支，如果对应目标的中心在某个格子里，则对应格子需要输出实例类别和分割实例。具体算法如图6.3所示：

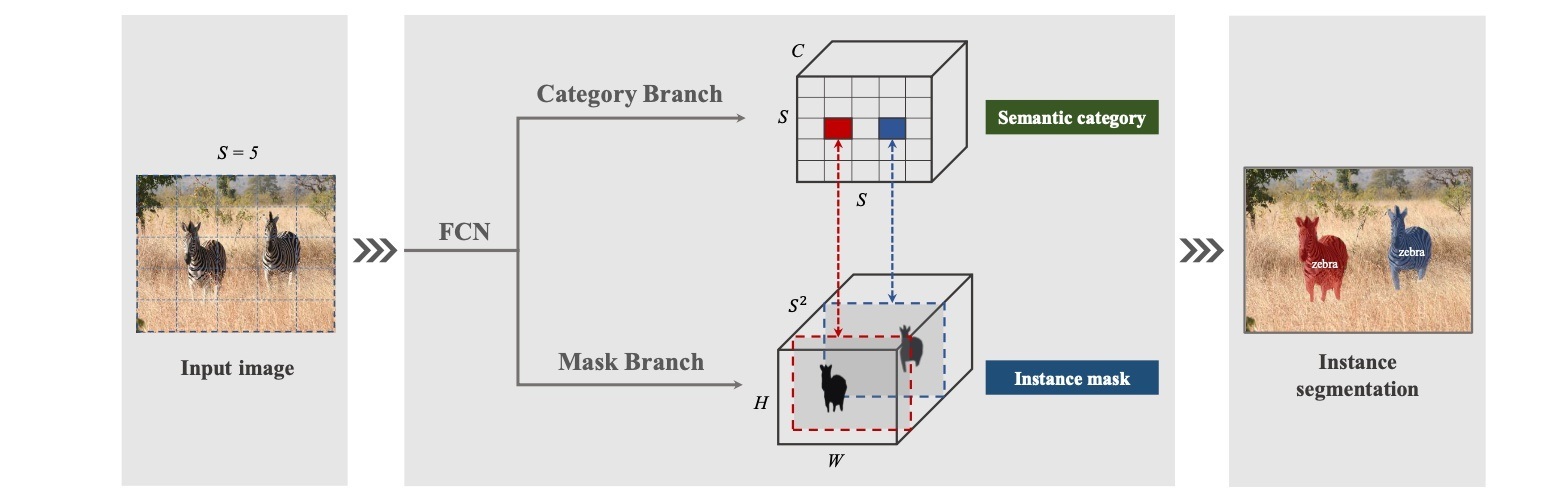


图6.3 SOLO算法网络框图

实例类别：如图6.3所示中的类别分支，对于输入图像中的每一个grid网格，该算法都会进行C维的预测输出，这里的C为类别个数，每一维代表一个类别的置信度得分。所以类别分支的输入是S×S个格子的数据，输出则为S×S×C的多维数据，这样每个格子就只会属于一个实例。

实例掩码：如图6.3所示中的mask分支，对于输入图像I中的S×S个格子，最多会进行次预测，将这些mask以一个3D的tensor进行输出，也即是有3个通道。则mask分支的输出维度为：。并且类别分支和mask分支之间也具有一对一的对应关系，例如某一个cell位置为(i,j)，那么对应的mask分支中的第i\*S+j个mask。

网络架构：该算法的backbone使用FPN，FPN可以在每一个通道上产生大小不同的特征图（通道数通常为256的倍数），FPN输出的特征图作为类别分支head和mask分支head的输入。同一个head的参数都是共享的，但是最后的1×1卷积参数的没有共享的。

关于SOLO网络的学习：网络对于每一个格子都需要给定预测的目标的概率，如果某一格子内落入了任一的gt mask的中心，就会被分为正例。对于每一个gt mask的正样本，都会设定一个二值分割mask，由于mask分支的输出是，所以此处共有个grid，对每个正样本，其对应的binary mask都会被标记。其中，训练过程中的损失函数公式定义如下：

上式中的是类别分支的分类损失，是用于预测mask的损失函数：

如果网格对应的类别索引是从左到右、从上到下的，则上式中i=|k/S|，j=k%S，则为正样本的个数，以及分别对应的类别和对应的mask。代表指数函数，如其内部式子成立则为1，否则为0。

关于SOLO网络的预测过程：（1）图像经过FPN的backbone网络以及分类分支和mask分支，得到(i,j)位置的类别得分以及对应位置的mask；（2）使用特定的阈值来进行类别得分的过滤；（3）对前得分前500对应的mask进行NMS操作；（4）最后将目标类别的mask进行二值mask输出。

* + 1. 基于BGNet网络的双目立体匹配算法

三维重建和三维感知中立体匹配是十分关键的步骤。给定双目图像数据，立体匹配的目的是找到左目图像和右目图像的像素之间建立密集的对应关系。现有的传统的双目立体匹配方法针对火星表面地形无纹理区域和反射表面弱的缺陷，很难找到准确的对应点，并且计算效率很低。所以，本课题采用比较新的基于深度学习的立体匹配网络BGNet。通过双目相机和TOF相机制作的火星场双目数据集以及现有公开数据集，利用端到端的BGNet网络进行训练，计算出视差图，再将视差图利用相机内参进行转换成深度图，以便进行后续的空间定位。现有方法针对火星表面地形无纹理或纹理差区域和反射表面弱的缺陷，很难找到准确的对应点。双目图像通过BGNet网络转化为深度图的流程如下图：

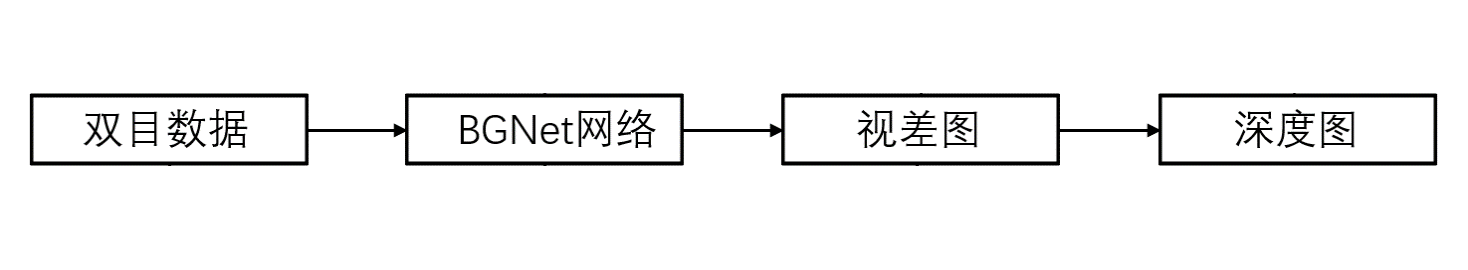


图6.4 基于BGNet的双目立体匹配

实时的、高精度的立体匹配算法是目前三维感知和三维重建中研究的热点，在例如机器人，自动驾驶，AR等领域都有很广泛的应用。虽然相关的立体匹配网络的发展也在逐步的发展，但是在速度和精度上依然存在比较大的发展空间。

BGNet算法使用了一种基于可学习的双边网格的代价空间上采样模块（Cost volume Upsampling in the learned Bilateral Grid, CUBG），该模块结构如图6.5所示。基于这个具有边缘保持特性的上采样模块，再通过没有参数的切片层（slicing layer），网络就可以从低分辨率的代价空间获得高质量的高分辨率代价空间。使用这种方式，可以将代价聚合在低分辨率中执行，大大的降低了时间消耗。该算法中的此模块也可以应用到许多现有的其他立体匹配网络当中（如GCNet，GANet等），并且其在保持不错的精度条件下依然可以取得最高将近30倍的加速，这是可微双边网格首次在立体匹配网络中的应用。此外BGNet网络在一些公开数据集，例如KITTI数据集下能够达到30-40fps，并且精度非常高，超过了现有的大部分实时立体匹配网络。

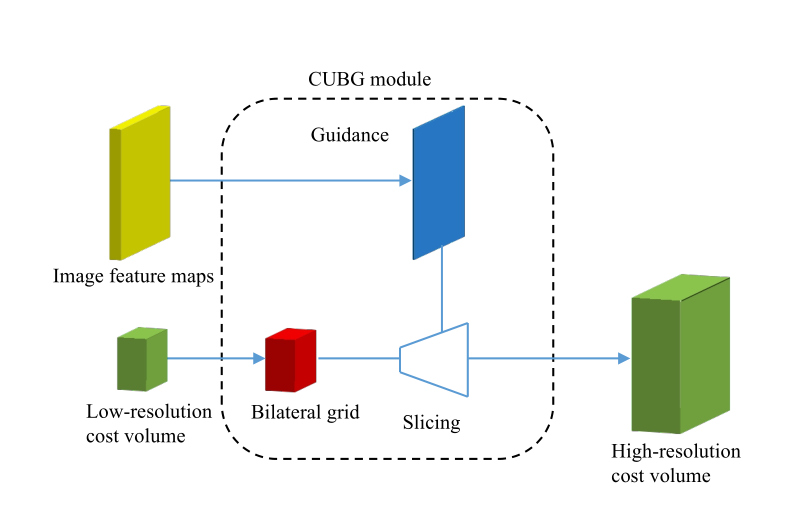


图6.5 CUBG模块结构

BGNet网络就是基于上述的CUBG模块实现的一个高效的端到端的立体匹配网络，其在KITTI数据集下可以实时运行。该网络主要有特征提取、成本聚合、CUBG和视差优化几个模块组成。具体的BGNet网络结构如下图：

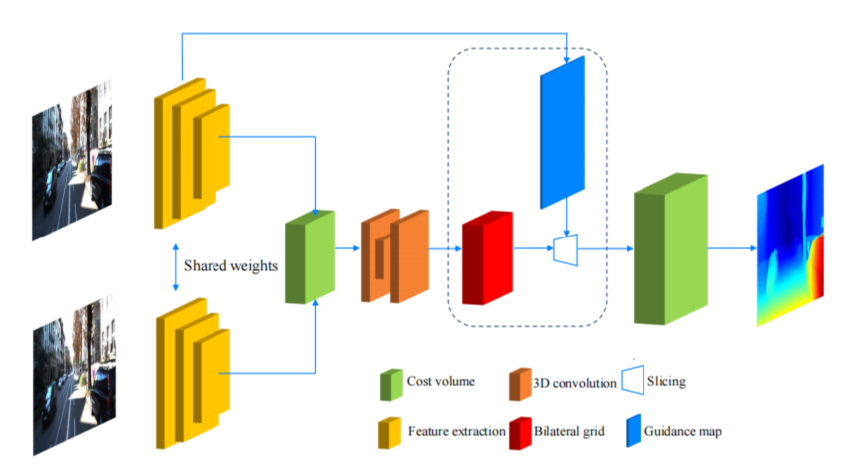


图6.6 BGNet网络结构

对于算法中的特征提取模块：本文采用类ResNet的架构来进行图像特征的匹配。首先，对于开始的三层，分别使用的步长为2、1、1且kernel的大小的3×3的卷积核进行图像的下采样。其次，使用步长分别为1、2、2、1的连续的四个残差层，并且使用1/8的分辨率生成一元特征。其中遵循两个沙漏特征的网络用以扩大感受野。最后，把所有的1/8分辨率的特征图都进行连接，最终生成一个352通道的特征图，并用来生成成本量。

对于算法中的成本聚合模块：在经过了上述的特征提取后，就为成本聚合构建了一个分组相关成本量，该成本量结合了相关联以及串联量的有点。并且，考虑到效率，算法中只使用了一个沙漏的架构来进行成本量的过滤。具体来说，首先使用两个3D卷积进行代价量通道数的降低，再使用类U-Net结构的3D卷积网络进行代价聚合以降低计算成本。

对于算法中的视差回归的定义公式如下：

对于算法中使用的损失函数的公式定义如下：

其中：

经过上述端到端的BGNet深度学习网络，可以得到左目相机坐标系下的视差图。通过相机内参可以将视差图转换为左目相机坐标系下的深度图，进而可以恢复出3D点云来进行相关的环境感知以及目标空间定位。视差图转化为深度图的公式较为简单，同2.3小节中介绍的双目立体视觉模型中的式(2.14)。

为了验证本课题中的方案在三维空间中定位的准确性，本文还采用了TOF相机作为真值来进行对比分析，即将TOF相机获得的深度图转换到左目相机坐标系，并与双目生成的左目相机坐标系下的深度图进行定位误差分析。设TOF相机内参为，其坐标系下的一点坐标为[x,y,z]，那么该点转换到左目相机坐标系的转换，在已知左目相机内参矩阵为K的情况下，可以根据如下公式获得左目相机下对应的深度图：

其中u、v对应左目相机坐标系下的x、y坐标，d对应深度。

* 1. 实验结果与分析

经过上述小节的算法介绍，为了验证本课题中的双目视觉目标空间定位系统的效果，对许多场景进行了对应的分割算法与立体匹配算法，并进行了空间中特定目标的三维空间定位，并以TOF的数据为真值进行了定位的精度分析。实验场景主要包括航天院火星场，以及仿真软件内的模拟火星环境。

* + 1. 实例分割算法实验结果

本小节主要进行分割算法的实验，针对的场景包括航天院火星场、仿真软件里的模拟环境以及NASA的部分航天数据。具体的软硬件配置表如下：

表6.1 分割算法软硬件配置表

|  |  |
| --- | --- |
| 类别 | 配置 |
| 系统 | Ubuntu18.04 |
| CPU | Intel i7-12700 |
| 显卡 | Nvidia 2080ti \* 2 |
| 搭载相机 | 工业相机，分辨率2048\*2048 |
| 实验场景 | 航天院火星场、仿真场景、NASA航天数据 |

对于航天院火星场一期、二期的场地进行分割检测，主要检测的对象包括较大的岩石rock以及较为扁平的基岩bedrock。

|  |  |
| --- | --- |
|  |  |
| （a）效果图1 | （b）效果图2 |
|  |  |
| （c）效果图3 | （d）效果图4 |

图6.7 航天院火星场一期场地分割结果

对于航天院火星场一期的场地，场景主要为平整的沙地，以及少部分的岩石和基岩组成，作为最初的算法实验场地。但是为了更加模拟出火星的场景和后续的三维重建等工作，进行了场地的更换。二期的场景包括岩石、基岩、细沙、沙坑、斜坡、土堆、山脊等多种环境。

图 6.8 航天院火星场二期场地分割结果

另外还在仿真软件中进行了场景的搭建，进行了更多环境的设置，包括有沙坑bunker，矿石ore，岩石rock和突起protuberance。

|  |  |
| --- | --- |
|  |  |
| （a）效果图1 | （b）效果图2 |
|  |  |
| （c）效果图3 | （d）效果图4 |

图6.9 仿真环境地形分割结果

针对NASA的数据集，由于数据量较小，所以分割结果可能存在一定的误差。其中rock为岩石，bedrock为基岩，soil为土壤，shadow为阴影。

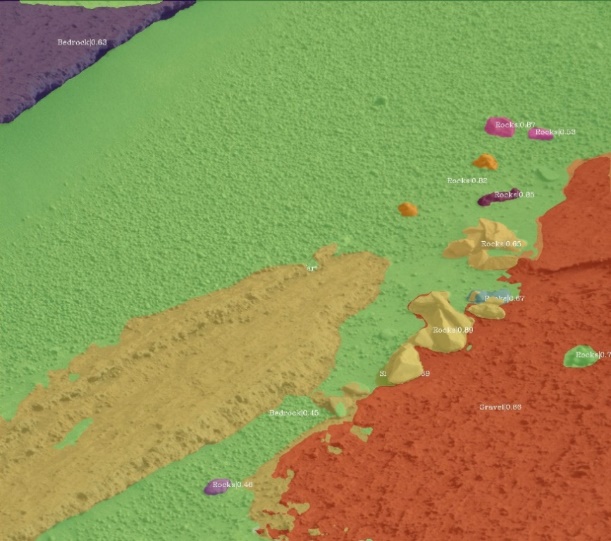
 

图6.10 NASA数据分割结果

通过上述不同场景的实验，训练数据只用了几百张图像，可以看出该算法的分割效果是非常好的。这也为后续的利用双目恢复出的三维信息进行空间定位提供了良好的条件。

* + 1. 双目空间定位算法实验结果与分析

本小节主要进行双目立体匹配与物体的空间定位算法的实验，针对的场景主要为航天院火星场的数据。具体的软硬件配置表如下：

表6.2 双目定位算法软硬件配置表

|  |  |
| --- | --- |
| 类别 | 配置 |
| 系统 | Ubuntu18.04 |
| CPU | Intel i7-12700 |
| 显卡 | Nvidia 2080ti \* 2 |
| 搭载相机 | 工业相机\*2，分辨率2048\*2048；TOF相机，分辨率640\*480 |
| 实验场景 | 航天院火星场 |

基于航天院火星场的场景，利用6.2.2中介绍的BGNet双目立体匹配网络进行左目相机坐标系的视差图恢复，再进行视差图到深度图的转换。下图所展示的即为网络输出并转化得到的深度图：

|  |  |  |
| --- | --- | --- |
| C:/Users/绿油油如你/AppData/Local/Temp/picturecompress_20211123193124/output_38.pngoutput_38 | C:/Users/绿油油如你/AppData/Local/Temp/picturecompress_20211123193124/output_39.pngoutput_39 | C:/Users/绿油油如你/AppData/Local/Temp/picturecompress_20211123193124/output_40.pngoutput_40 |
| （a）左目相机图像 | （b）左目视差图 | （c）左目深度图 |

图6.11 立体匹配网络输出及转化结果

通过上述算法生成的深度图，结合左目相机内参即可生成3D场景点云，再结合分割算法检测出的特定目标，进行3D点云与2D目标的映射，便可以计算出特定目标的空间坐标。例如在火星车行进过程中导致无法移动或者会对火星车造成损害的地形和物体便属于本课题中需要进行目标空间定位的，例如岩石、陨石坑等。

要保证火星车安全的行进，则需要周围障碍物的三维位姿信息，而在图片中的二维位姿不带深度信息，我们无法从2D图片中获得物体的绝对尺度和位置，没有办法有效避免碰撞，具有一定的安全隐患。因此，场景的3D感知在火星车的移动，路径规划等方面都发挥着极其重要的作用，对于火星表面的目标探索有着重要的意义。所以本课题使用分割+双目立体匹配的方案进行火星车周围高风险目标的定位，整体框架图为：

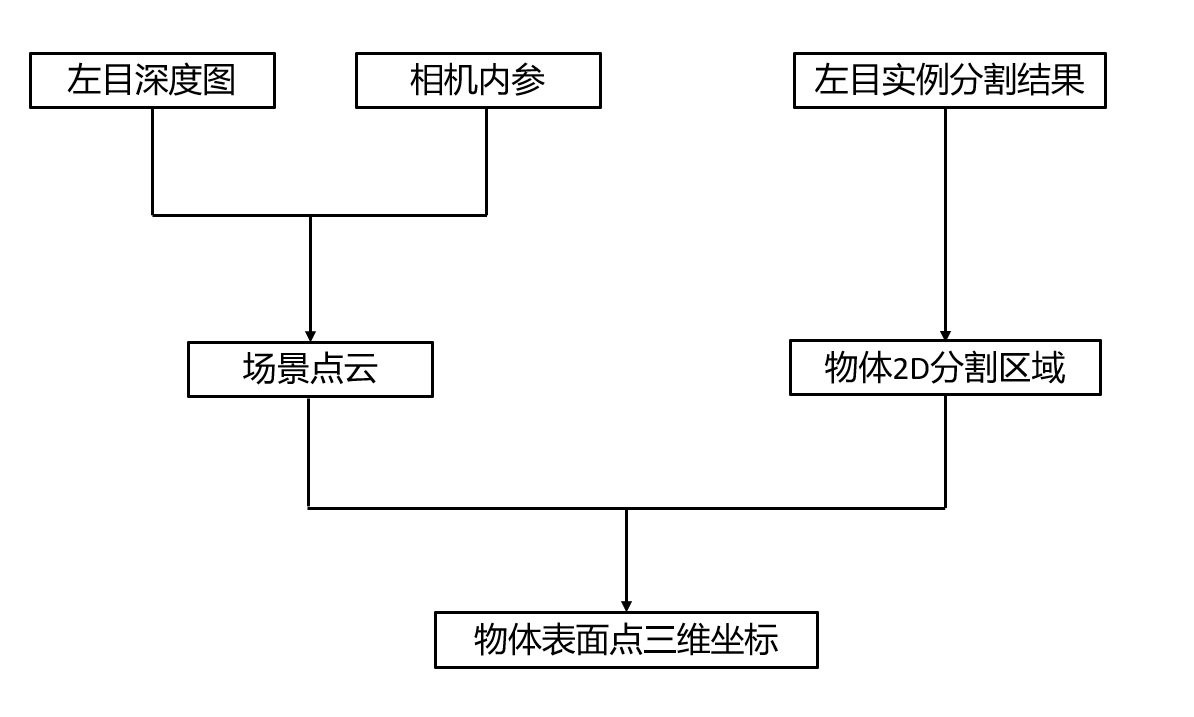


图6.12 双目视觉空间定位算法流程框图

通过获取物体的空间坐标，计算物体的方向包围盒，获取物体的质心以及高度进行高风险目标定位。为检测算法可行性，分割出高风险物体进行空间定位效果如下图：

|  |  |  |
| --- | --- | --- |
| C:/Users/绿油油如你/AppData/Local/Temp/picturecompress_20211123193124/output_42.pngoutput_42 | C:/Users/绿油油如你/AppData/Local/Temp/picturecompress_20211123193124/output_105.jpgoutput_105 |  |
| (a) 原图 | (b) 分割结果 | (c) 高风险目标 |

图6.13 高风险目标定位结果

本课题基于航天院定义的高风险指标进行的实验，对应的指标在图6.13中对应的是岩石高度超过15cm则被认为是高风险。上述图6.13中图(c)的两块被框起来的岩石被定义为高风险。由于距离越远，误差带来的影响越大，所以本课题中只考虑5m以内的物体。对于图(c)两个石头通过算法获得的距离与高度分别为rock1:距离4.18m、高度18.82cm；rock2：距离3.62m、高度25.59cm。

为了验证双目空间定位的精度，本课题还以TOF相机的原始点云下的对应的高风险目标的深度作为真值，算法生成的深度图作为预测值，进行对应的误差分析。定位误差计算示意如下图所示：

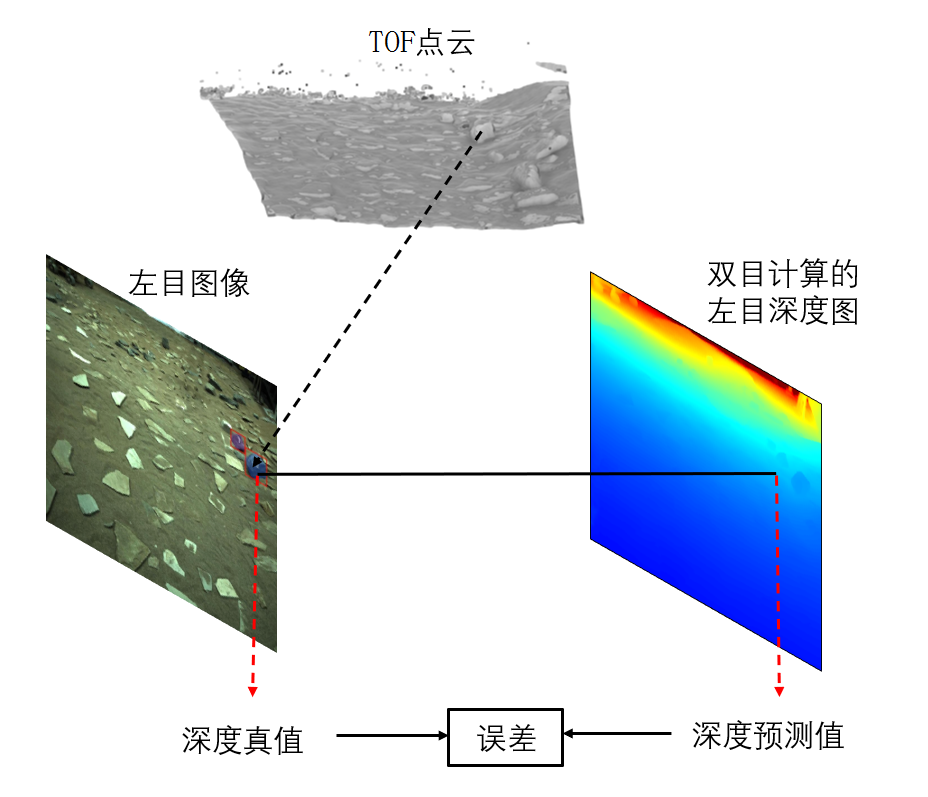


图6.14 定位误差计算示意图

在高风险目标区的分割区域内采样部分像素点，它们对应的真实距离为 预测距离, 然后计算这些像素点的平均真实距离和平均预测距离。最后采用以下公式计算定位误差：

利用上述定位误差计算方法，对图6.13时刻的图像以及高风险目标进行误差分析结果如下表所示：

表6.3 高风险目标定位误差分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 物体编号 | 高度(cm) | 距离预测值(m) | 距离真实值(m) | 定位误差 |
| 1 | 18.82 | 4.181 | 4.183 | 0.048% |
| 2 | 25.59 | 3.618 | 3.620 | 0.055% |

* 1. 本章小结

本章主要基于航天院火星场的场景以及双目视觉系统进行三维空间环境的感知。与常见的端到端的激光点云检测算法不同的是，本章节通过2D分割+3D映射的方式，实现基于两步走的方式完成对三维环境中特定目标的定位。具体方式是首先在左目图像上进行二维的图像分割，然后利用双目立体匹配网络生成的深度图，将左目图像分割得到的二维目标映射到三维空间中，并进行一定条件的高风险目标的判断，最终以两步走的方式实现三维空间环境的感知。此种方式相较于端到端的点云检测、分割网络的优势在于，训练时不需要大规模的进行3D标注，并且可以实现小物体的三维感知。本章节使用的方案识别精度较高，并且对于航天院火星场数据进行目标定位误差为毫米级。

# 总结与展望

《正文》×××××

《正文》×××××

# 

# 参考文献

[1] Zhang H Y, Zheng L, Cai L. Design and Analysis of Hierarchical Physical Layer Network Coding[J]. IEEE Transactions on Wireless Communications, 2017, 16(12): 7966-7981.

[2] 张贤达, 保铮. 通信信号处理[M]. 北京:国防工业出版社, 2000: 30-50.

[3] Larimore M, Treichler J. Convergence behavior of the constant modulus algorithm[C]. IEEE International Conference on ICASSP,2017:13-16.

**按照“GB/T 7714”格式列参考文献，请删除此行。**

# 附录1 攻读硕士学位期间撰写的论文

[1] Li C, Sun X,Zhou X, et al. Optimal biased association scheme with Networks[J]. IEEE Transactions on Wireless Communications, 2017, 6(8):360-371. （SCI二区：000407686300001）；

[2]Li C, Ji Y. Modeling and analysis of epidemic spreading on community network [J]. Journal of China Universities of Posts & Telecommunications, 2016, 23(5):82-87.（EI：20170303250062）

[3]李春，网络路由技术，电子学报，已录用。

**百度学术，搜索论文，点击“引用”，按照“GB/T 7714”格式 复制出论文引用，后面加上（检索 信息），SCI论文需要写上几区。请删除此行。**

# 附录2 攻读硕士学位期间申请的专利

[1] 作者一，作者二，作者三. 专利名称，专利申请号，专利申请日期，专利授权日期；

[2] 张三，李四. 一种面向代理的安全传输方法，2007062.5，2016.1,2017.10；

[3] 张三. 一种实用的网络路由方法，200845610.5，2017,1,2018.1。

# 附录3 攻读硕士学位期间参加的科研项目

（1）项目类型，项目名称（项目编号）；

（2）国家自然科学基金，基于安全移动代理新一代网络研究(702710456)；

（3）国家自然科学基金，移动代理机制关键技术研究(601732578)。

# 致谢

主要感谢导师和对本论文学术研究有特别贡献的组织或个人：

对提供资助或者支持的基金、合同单位、企业、组织或者个人（基金项目应该包括基金名称、项目名称、项目编号、项目负责人、研究起止年月）；

协助完成研究工作或提供便利的组织或个人。