

分类号 _____

密 级 _____

U D C _____

编 号 10486

武汉大学
硕 士 学 位 论 文

基于深度学习多视密集匹配方法的
卫星立体影像三维重建

研 究 生 姓 名 : 高建

学 号 : 2020202130034

指导教师姓名、职称 : 季顺平 教授

专 业 名 称 : 模式识别与智能系统

研 究 方 向 : 多视遥感影像三维重建

二〇二三年五月

Master Dissertation of Wuhan University

3D Reconstruction from Satellite Stereo
Images Based on Deep Learning Multi-View
Stereo Method

By
Gao Jian

Supervised by
Prof. Ji Shunping

School of Remote Sensing and Information Engineering
Wuhan University
May, 2022

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已发表或撰写的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：

年 月 日

学位论文使用授权书

本论文作者完全了解学校关于保存、使用学位论文的管理办法及规定，即学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，接受社会监督。本人授权武汉大学可以将本学位论文的部分或全部内容编入学校有关数据库进行信息服务，也可以采用影印、缩印或扫描等复制手段保存或汇编本学位论文。

本论文提交□当年/□一年/□两年/□三年以后，同意发布。

若不选填则视为一年以后同意发布。

注：保密学位论文，在解密后适用于本授权书。

作者签名：

导师签名：

年 月 日

武汉大学研究生学位论文作者信息

论文题目				
姓 名		学号		答辩日期
论文级别	博士 <input type="checkbox"/> 硕士 <input checked="" type="checkbox"/>			
院/系/所	遥感信息工程学院	专 业		
联系电话		E_mail		
通信地址(邮编)：湖北省武汉市珞喻路 129 号武汉大学遥感信息工程学院				
备注：				

本论文创新点

本文研究的关注点主要是深度学习多视密集匹配网络在光学卫星影像上的实际应用，以期为光学卫星影像三维重建领域带来新的突破。本文主要创新点有以下几个方面：

(1) 将卫星影像RPC模型表达成张量运算形式，实现基于RPC模型的高维特征变形方法，基于该方法搭建了基于深度学习的多视光学卫星影像密集匹配网络Sat-MVSNet，制作并公布了一套三线阵多视角密集匹配数据集WHU-TLC，成功将深度学习多视密集匹配方法引入到光学卫星影像处理中。

(2) 构建了基于深度学习多视密集匹配方法的光学立体卫星影像三维重建框架Sat-MVSF，涵括成像参数优化、大幅宽影像多视密集匹配，以及将匹配结果转化为最终产品的全面处理流程，并初步验证了其应用潜力和实用价值。

(3) 引入一种基于自我训练的深度学习模型自优化策略，在不需要任何迁移目标场景真值标签的情况下，帮助网络模型完成迁移。在预训练数据与真实数据相差很大的情况下，仍达到略优于商业软件和开源解决方案的重建效果，进一步提升了Sat-MVSF框架在实际场景下的应用能力。

摘 要

随着我国高分辨率对地观测系统重大专项的启动实施, 我国对地观测体系不断完善, 自主获取光学立体卫星影像数据的能力不断增强, 使得现有的处理系统面临着巨大的挑战。目前, 传统的双目立体匹配算法仍然是生产之中的主流算法。传统算法基于设计者对密集匹配问题的经验性建模, 在设计之初设定了一些与真实情况不相符的假设前提以简化问题, 而这种基于经验性的假设前提往往成为算法的性能瓶颈, 限制了算法表现。而深度学习方法可以从大量样本中自动学习低维和高维特征, 从而避免经验性假设的局限性, 在密集匹配领域显现出巨大的优势, 也将成为未来发展的新趋势。

然而, 将深度学习多视密集匹配方法运用到光学立体卫星三维重建实际生产任务中依然面临着一些关键问题: 首先, 卫星影像常用的 RPC (Rational Polynomial Camera) 模型与现有的深度学习多视密集匹配网络中采用的中心投影成像模型存在差异, 导致现有的深度学习模型无法直接应用于卫星影像。其次, 为了从光学立体卫星影像重建三维地形, 还需要完善除了密集匹配网络外的其他必要模块, 建立一个完整的基于深度学习方法的卫星影像三维重建框架, 以及考虑如何对大幅宽卫星影像进行处理等问题。此外, 基于有限数据训练得到的模型很可能会出现性能退化的问题, 且由于地表三维真值获取困难, 实际生产中难以获取真值以进行微调, 从而影响网络模型在实际应用中的表现。

本文针对以上问题开展研究, 主要的研究内容和创新点如下:

(1) 设计了基于深度学习的多视光学卫星影像密集匹配网络 Sat-MVSNet。

首先, 将 RPC 模型进行张量化的表达, 实现了基于 RPC 模型的高维特征变形 (feature Warping based on RPC Model, RPC-Warping); 基于此, 搭建了基于深度学习的多视光学卫星影像密集匹配网络 Sat-MVSNet, 成功将深度学习多视密集匹配方法用于多视光学卫星影像上。此外, 制作并公布了一套三线阵多视卫星影像密集匹配数据集 WHU-TLC, 以进行网络的训练和算法的评估。

(2) 构建基于深度学习多视密集匹配方法的光学立体卫星影像三维重建框架 Sat-MVSF。 以 Sat-MVSNet 网络为核心完善了一套完整的光学立体卫星影像三维重建框架, 以期用于实际生产任务。整个框架分为地表稀疏场景重建、地表稠密场景重建和摄影测量产品生产三大部分, 分别完成 RPC 模型精化、稠密场

景点云重建和 DSM 生产任务。

(3) 引入一种应对模型迁移时性能退化问题的自优化策略，在无需目标数据任何真值标签的情况下，实现网络模型在不同场景间的迁移，提升了网络模型在实际生产数据上的性能。所提出的自优化策略先在公开数据集训练，得到预训练模型，再利用预训练模型对实际生产数据进行推理，并通过几何一致性验证剔除错误匹配以生成伪标签，最后用伪标签进行再训练，完成迁移。

在实验上，本文首先验证了相比于将 RPC 模型拟合成中心投影成像模型的处理方式，基于 RPC-Warping 的网络模型更具优势；之后，在 WHU-TLC 测试集上进行评估，证明相比于现有的商业软件和开源解决方案，所提出的 Sat-MVSF 在精度和效率上都展现出了一定的优势，具有较强的实用潜力；最后，在 MVS3D 数据集上，以航空影像上的预训练模型为基础进行自优化，使得 Sat-MVSF 达到略优于商业软件和开源解决方案的重建效果，证明了自优化策略的有效性，进一步增强了 Sat-MVSF 框架的实用性。

关键词：光学立体卫星影像，多视图立体几何，密集匹配，三维重建，深度学习

ABSTRACT

With the implementation of China's High-Resolution Earth Observation System major project, the country's Earth observation system has been continuously improved. And the increasing abundance of optical stereo satellite image data acquired independently poses significant challenges to existing processing systems. Currently, traditional stereo matching algorithms are still widely used in production as the mainstream approach. These traditional algorithms are based on empirical modeling and are manually designed by developers, who often make assumptions that may not align with the real-world scenarios in order to simplify the problem. However, these empirical assumptions often become the bottlenecks and limit the algorithm's performance. Deep learning approaches, with their ability to learn from large-scale data without relying on empirical assumptions, have demonstrated significant advantages in the field of dense matching.

However, there are some critical challenges when it comes to integrating deep learning-based multi-view dense matching methods into practical production tasks for optical stereo satellite images based 3D reconstruction. Firstly, there are differences between the commonly used RPC (Rational Polynomial Camera) model for satellite images and the projection model used in deep learning-based MVS (Multi-View Stereo) methods, which hinder the direct application of existing deep learning models. Secondly, in order to reconstruct 3D terrain from optical stereo satellite images, it is necessary to develop other essential modules apart from dense matching network, and consider how to process large-scale high-resolution satellite images. Thirdly, models trained on limited data are likely to suffer from performance degradation, and the difficulty in obtaining accurate 3D ground truth makes it challenging to fine-tune the models, which can impact the performance of the models in practical applications.

In this paper, we conducted research to address the aforementioned issues. The main research content and innovative points are outlined as follows:

(1) This paper designs a deep learning-based MVS network called Sat-MVSNet for optical satellite images. Firstly, the RPC model is expressed as a series of tensor operations, enabling high-dimensional features warping between views efficiently, called as RPC warping. Then, a deep learning-based MVS network, Sat-MVSNet, is constructed, successfully integrating deep learning-based MVS

methods into optical stereo satellite images. Additionally, a three-line array multi-view satellite image dense matching dataset, WHU-TLC, is created and published for network training and algorithm evaluation.

(2) This paper presents a framework, Sat-MVSF, for optical stereo satellite image based 3D reconstruction using deep learning-based MVS methods. Considering the challenge of large size of satellite images, the framework is built around the Sat-MVSNet network as the core, and further enhanced for practical production tasks. The framework is divided into three main components: sparse reconstruction, dense reconstruction, and photogrammetric production. These components are responsible for refining the RPC model, generating dense point clouds, and producing DSMs (Digital Surface Models), respectively.

(3) This paper introduces a self-optimization strategy to address the issue of performance degradation during model transfer, allowing network models to be transferred between different scenarios without requiring any ground truth labels of target data. The proposed self-refinement strategy includes training a pre-trained model using publicly available datasets, using the pre-trained model for inference on actual production data, generating pseudo-labels by removing erroneous matches through the geometric consistency check, and retraining the model using these pseudo-labels.

In the experiments, this paper first validates that the network model based on RPC warping outperforms the approach of fitting the RPC model to a pinhole model. Then, the proposed Sat-MVSF framework is evaluated on the WHU-TLC test dataset, demonstrating its advantages in terms of accuracy and efficiency compared to existing commercial software and open-source solutions. Furthermore, on the MVS3D dataset, the self-refinement strategy based on pre-trained models from aerial images is applied to Sat-MVSF, resulting in slightly superior reconstruction results compared to commercial software and open-source solutions. This validates the effectiveness of the self-refinement strategy and further enhances the practicality of the proposed Sat-MVSF framework.

Key words: optical stereo satellite images, multi-view stereo, dense matching, 3D reconstruction, deep learning

目 录

摘 要.....	I
ABSTRACT	III
第二章 绪论	1
2.1 研究背景与意义.....	1
2.2 国内外研究进展.....	2
2.2.1 基于深度学习的影像密集匹配.....	2
2.2.2 基于手工设计特征的光学立体卫星影像三维重建.....	4
2.2.3 基于自我训练的半监督学习方法.....	6
2.3 论文的研究内容与章节安排.....	7
2.3.1 主要研究内容.....	7
2.3.2 章节安排.....	7
第三章 坐标系统与坐标转换关系	9
3.1 坐标系统.....	9
3.1.1 影像坐标系.....	9
3.1.2 地理坐标系.....	10
3.1.3 空间直角坐标系.....	10
3.1.4 站心坐标系.....	11
3.1.5 投影坐标系.....	11
3.2 有理多项式相机模型.....	12
3.2.1 有理多项式相机模型.....	12
3.2.2 有理多项式相机模型参数的解算.....	13
3.2.3 基于正向 RPC 模型的迭代逆向坐标转换	13
3.2.4 基于正向 RPC 模型的多视前方交会	14
3.3 其他坐标系间的转换.....	14
3.3.1 地理坐标系与空间直角坐标系的转换.....	14
3.3.2 空间直角坐标系与站心坐标系的转换.....	14
3.3.3 地理坐标系与投影坐标系的转换.....	15
3.4 实验分析与讨论.....	16

3.5 本章小结.....	18
第四章 基于深度学习的多视光学卫星影像密集匹配方法	19
4.1 基于有理多项式相机模型的高维特征可微变形.....	19
4.2 多视光学卫星影像密集匹配网络 Sat-MVSNet.....	21
4.2.1 网络结构.....	21
4.2.2 多尺度特征提取.....	22
4.2.3 多尺度代价体构建.....	23
4.2.4 代价聚合.....	24
4.2.5 高度回归.....	26
4.2.6 训练损失.....	26
4.3 三线阵多视卫星影像密集匹配数据集 WHU-TLC	26
4.3.1 数据源简介.....	26
4.3.2 WHU-TLC 数据集及制作过程	27
4.4 实验分析与讨论.....	28
4.4.1 评价指标.....	28
4.4.2 RPC 模型拟合为中心投影模型的误差分析	29
4.4.3 多视光学卫星影像密集匹配网络的有效性验证.....	30
4.5 本章小结.....	33
第五章 多视光学卫星影像三维重建框架	34
5.1 地表稀疏场景重建.....	35
5.1.1 基于均匀分块的大幅宽卫星影像关键点提取.....	35
5.1.2 几何约束下的卫星影像连接点匹配.....	36
5.1.3 基于并查集的连接点自动转刺.....	37
5.1.4 基于像方仿射补偿模型的光束法平差.....	38
5.2 地表稠密场景重建.....	40
5.2.1 大幅宽多视卫星影像的 MVS 推理.....	40
5.2.2 基于多视几何一致性的误匹配剔除与匹配结果融合	41
5.3 摄影测量产品生产	42
5.4 实验分析与讨论.....	43
5.4.1 Sat-MVSF 框架在 WHU-TLC 上的性能表现.....	43

5.4.2 产品生产展示.....	45
5.5 本章小结.....	47
第六章 基于自我训练的深度学习模型自优化策略	48
6.1 自优化策略.....	48
6.2 实验分析与讨论.....	49
6.2.1 Sat-MVSF 与自优化策略在 MVS3D 数据上的表现.....	49
6.2.2 自优化精度与自优化轮次之间的关系.....	54
6.2.3 成对处理和真多视处理方式下的性能对比.....	55
6.3 本章小结.....	56
第七章 总结与展望	57
7.1 全文总结.....	57
7.2 研究展望.....	58
参考文献	59
攻硕期间发表的科研成果目录	66
攻硕期间参与的课题情况	67
致 谢.....	68

图索引

图 2-1 瞬时像平面坐标系与影像行列号坐标系	9
图 2-2 地理坐标系与空间直角坐标系	10
图 2-3 空间直角坐标系与站心坐标系	11
图 2-4 横轴墨卡托投影与横轴墨卡托投影坐标系	11
图 3-1 基于有理多项式相机模型的高维特征可微变形示意图	20
图 3-2 系数张量 T	21
图 3-3 多视角光学卫星影像密集匹配网络 Sat-MVS 结构图	22
图 3-4 多尺度特征提取器	23
图 3-5 基于循环编码-解码 RED 结构的代价聚合	25
图 3-6 光学卫星影像构成立体像对的两种方式	27
图 3-7 三线阵全色影像与数字高程模型	28
图 3-8 多视卫星影像瓦片与高度图	28
图 3-9 用中心投影模型拟合 RPC 参数的误差分布图	30
图 3-10 不同的网络模型在 WHU-TLC 测试集上的 DSM 可视化结果	32
图 4-1 基于深度学习多视密集匹配的光学立体卫星三维重建通用框架流程图 .	34
图 4-2 卫星影像对交会角估计	37
图 4-3 追踪路线的不同几何型态	37
图 4-4 基于几何约束的大幅宽多视卫星影像同名瓦片裁剪	40
图 4-5 多视几何一致性与一致点融合	42
图 4-6 保留格网中的最大值以生成数字表面模型 DSM.....	43
图 4-7 各解决方案在 WHU-TLC 测试集上的结果可视化	45
图 4-8 ZY3-01 星三线阵影像 DSM 产品（法国圣马克西姆）	46
图 4-9 ZY3-01 星三线阵影像 DSM 产品（中国香港）	46
图 4-10 TH3 号三线阵影像 DSM 产品（中国山东）	47
图 5-1 自优化策略	48
图 5-2 各方案在 MVS3D 数据集上的 DSM 产品可视化	52
图 5-3WHU-MVS 数据集样本与 MVS3D 数据样本对比图	53
图 5-4Sat-MVSF 自优化前后的 DSM 产品可视化	53

图 5-5 不同预训练模型生产的 DSM 局部剖面图.....	54
图 5-6 自优化策略执行不同轮次时发生的变化	55
图 5-7 成对处理方式与真多视处理方式下的 DSM 产品可视化.....	56

表索引

表 2-1 拟合 RPC 模型的正解精度.....	17
表 2-2 拟合 RPC 模型的迭代反解精度.....	17
表 2-3 拟合 RPC 模型的直接反解精度.....	17
表 2-4 拟合 RPC 模型的迭代反解—正解精度.....	18
表 2-5 拟合 RPC 模型的直接反解—正解精度.....	18
表 3-1 用中心投影模型拟合 RPC 参数的拟合误差（单位：像素）.....	29
表 3-2 不同的网络模型在 WHU-TLC 测试集上的定量比较.....	31
表 4-2 在 WHU-TLC 测试集上不同解决方案的评估.....	44
表 5-1 各方案在 MVS3D 数据集上的精度比较.....	51
表 5-2 成对处理方式与真多视处理方式下的精度对比	55

第一章 绪论

1.1 研究背景与意义

随着航天航空技术的不断发展，高分辨率对地观测卫星已然成为了人类观测地表环境的第三只眼，在国防军事领域、人类活动监测、国土资源调查、城市设计与管理等方面发挥着愈发重要的作用^[1-3]。而高分辨率光学立体卫星，作为“对地观测脑”^[4]中视觉系统的重要组成部分，更是担任着感知地球表面三维几何结构的重要使命。随着高分辨率对地观测系统重大专项的启动实施，我国陆地观测体系和能力得到显著提升^[5]，自主获取了海量的光学立体影像数据，对现有的处理系统提出了巨大的挑战。此外，国内外的各大遥感商业软件服务商也在陆续开发或已经在产品中增加了光学立体卫星三维重建模块，也足可证明市场对光学立体卫星三维重建的广泛需求。

目前，基于光学立体卫星影像进行数字表面模型（Digital Surface Model, DSM）生产的主流算法仍是基于手工设计特征的双目立体匹配算法。在这些手工设计算法设计之初，往往会设定一些与事实不相符的假设条件以简化问题。这些假设往往是手工设计算法的局限所在，也是影响其性能的关键。而深度学习方法将多个网络层组合为计算模型，通过反向传播来从海量数据中自动进行多层次的抽象数据表征学习，从而摆脱了手工设计特征的局限^[6]。凭借这种能力，深度学习方法在目标检测^[7-10]、图像分类^[11-15]、语义分割^[16-18]、实例分割^[19]等计算机视觉领域中大展拳脚，掀起了一场全新的技术变革。特别地，在三维重建领域，随着 Middlebury^[20]、KITTI stereo^[21]、DTU^[22]和 Tanks and Temples^[23]等优质数据集的出现，基于深度学习的密集匹配算法^[24-34]也正如火如荼地展开，并打败了半全局匹配法（Semi-Global Matching, SGM）^[35]、COLMAP^[36]等经典手工设计算法，强势占领了各大密集匹配榜单的前列。将视线转回摄影测量与遥感领域，已经有相关研究^[37-39]表明：在地表几何结构重建任务上，基于深度学习的密集匹配方法在精度和效率上较传统算法有较大的提升。聚焦于多视光学卫星影像三维重建这一子领域，深度学习多视密集匹配方法的应用仍存在着两大障碍：其一，现有网络模型多采用中心投影成像模型，而卫星影像中常用 RPC 模型，二者的差异使得现有方法无法直接用于卫星影像；其二，高精地形图的高保密性导致卫星影像三维重建数据集获取困难。

1.2 国内外研究进展

1.2.1 基于深度学习的影像密集匹配

影像密集匹配是计算机视觉领域的一个重要研究方向，其主要目的是从一组影像中估计出每张影像中逐像素的深度值，从而恢复三维场景的结构信息。随着深度学习的发展，基于深度学习的影像密集匹配算法逐渐成为该领域的主流研究方向。本节将回顾近年来深度学习密集匹配方法的发展历程，包括相关数据集、双目方法、多视方法以及在遥感影像上的深度学习密集匹配方法四个方面。

数据是深度学习发展的基石，深度学习密集匹配的发展壮大离不开优质数据集的支撑。Middlebury Stereo 双目数据集^[20]由米德尔伯里学院 (Middlebury College) 于 2001 年首次公布，包含若干个场景的左右视差图、彩色图和相机参数，常常被用于双目视觉算法的测试和评估。卡尔斯鲁厄理工学院 (Karlsruhe Institute of Technology, KIT) 提供了 KITTI 数据集^[21]，其中的双目立体数据集包含了从车载摄像头拍摄的城市街景影像对和对应的真实视差图，是深度学习立体匹配方法评估的重要基准之一。弗莱堡大学 (University of Freiburg) 使用开源 3D 创作软件 Blender 创建了三个具有足够现实感、足够变化性和足够数量的大规模合成立体视频数据集 FlyingThings3D、Monkaa 和 Driving^[40]，以便于训练和评估用于视差估计和光流估计的卷积神经元网络。苏黎世联邦理工学院 (Eidgenössische Technische Hochschule Zürich, ETH Zürich) 公布了 ETH3D 数据集^[41]，提供高分辨率影像对、相机参数、深度图以及视频，其数据覆盖多个室内和室外场景、不同的视角、不同的相机类型和不同的视场角。DTU 数据集^[22]由丹麦技术大学 (Danmarks Tekniske Universitet, DTU) 使用工业 ABB 机器人上搭载的相机和结构光扫描仪采集获取并向大众开放，以用于 MVS 方法的评估；该数据集包含了不同的场景模型，每个场景都包含从 49 个或者 64 个位置采集的、在 7 种不同光照条件下的影像数据，已经成为密集匹配和三维重建算法的标准榜单。Tanks and Temples^[23]也是一个基于影像的三维重建基准测试数据集，覆盖了室内和室外真实场景，包含影像序列、相机参数、视频数据和工业激光扫描仪获取的真值。由武汉大学 (Wuhan University, WHU) 开放的 WHU-MVS 数据集^[38]是一个大规模的多视航拍密集匹配数据集，可用于评估深度学习 MVS 方法在城市场景三维重建任务中的性能；该数据集以上千张真实航空影像生成得到的高精度 3D 模型为数据源，通过合成技术得到严格对应的影像和深度图。

在双目立体匹配领域中，MC-CNN^[42]训练了一个卷积神经网络来预测两个影像块之间的相似性并用于计算立体匹配代价，开启了深度学习立体匹配的篇章。Disp-Net^[40]使用 FlowNet^[43]的基本架构，将立体像对作为输入，以端到端的方式

预测出视差图。GC-Net^[24]利用二维卷积层对左右影像进行特征提取，再将左右特征图在视差方向上错位堆叠以进行代价体的构建，并使用3D卷积层进行代价体正则化，最后回归得到视差值，为后来大多数端到端双目立体匹配网络奠定了基本架构。PSM-Net^[25]使用空间金字塔池化模块，在不同的尺度和位置上聚合上下文信息以构建代价体，从而提升网络感知全局信息的能力。GA-Net^[26]从半全局匹配和代价滤波等传统方法中吸收思想，提出了半全局聚合层和局部引导聚合层以替换广泛使用的3D卷积，达到了更高的匹配精度。Gwc-Net^[27]使用分组相关这种新的方式来构建代价体，为衡量特征相似性提供了更为高效的表达，在减少参数的同时达到了更好的性能。DeepPrunner^[28]中引入了一种可微分的PatchMatch模块，有效地为每一个像素缩小搜索空间，完成了高效的代价体构建，大大节省了内存和时间花费。Liu等^[29]在现有立体匹配网络中引入局部相似性模式（Local Similarity Pattern, LSP）以辅助网络学习到更具判别性的特征，并设计了一种动态的自组合细化策略以减轻过度平滑的问题。CREStereo^[44]使用循环优化的方式从粗到细地更新视差值，并采用自适应分组相关层以减轻影像纠正错误对匹配的影响，在真实影像中展现出了高质量的细节。RAFT-Stereo^[45]将光流估计网络RAFT^[46]作为基本架构，通过引入多层卷积门控循环单元实现了更高效的信息传播，在实时推断中实现了高精度的匹配。

在多视立体匹配领域中，MVS-Net^[47]将多视角几何嵌入到深度学习网络模型中，通过特征提取、代价体构建、代价体正则化、深度回归、深度优化等模块，实现了端到端的多视密集匹配，将深度学习引入基于深度图的多视图几何重建任务中。R-MVSNet^[48]通过卷积门控循环单元对沿深度方向的2D代价图依次进行正则化，抛弃了对整个3D代价体直接进行正则化的方式，大大减少了内存消耗。CasMVSNet^[32]利用特征金字塔构建了多级代价体，通过前一阶段的深度图来缩小每个阶段的搜索范围，实现从粗到细的匹配，从而减少内存和时间开销。UCS-Net^[49]使用了基于方差的不确定性估计来自适应地构建多级金字塔代价体，通过逐渐增加深度分辨率来逐步细分庞大的搜索空间，以粗到细的方式实现高完整性和高精度的重建。PatchMatchNet^[50]引入了迭代的多尺度PatchMatch，并通过一种可学习的自适应传播和评价方案改进了PatchMatch核心算法，表现出优秀的性能和不俗的效率。EPP-MVSNet^[31]引入了一种基于核线的采样间隔划分方式，在核线方向上实现自适应的采样，在高分辨率上将特征精确地聚合到代价体中，从而实现了高效的三维重建。AA-RMVSNet^[33]引入了视图内聚合模块，改善了网络在具有细微结构和低纹理表面等挑战性区域上的性能；提出了视角间代价体聚合模块，自适应保留视角中利于匹配的像素以应对遮挡问题。TransMVSNet^[30]提出了特征匹配Transformer，利用自我注意力和交叉注意力机

制来实现图像内部和图像之间的长程上下文信息聚合，达到了先进的性能。Vis-MVSNet^[34]考虑到了像素级可见性信息，通过匹配不确定性估计推理出像素级的遮挡信息，并将遮挡信息加入多视代价体融合过程中，从而在代价融合中抑制了遮挡像素的不良影响，在具有严重遮挡的重建场景中显著提高了匹配精度。

在遥感影像三维重建领域，刘瑾等^[39]率先将深度学习立体匹配方法应用于遥感数据中，在三组航空影像数据集上将 MC-CNN、GC-Net、DispNet 这三种典型的卷积神经网络模型、传统方法 SGM 与商业软件 SURE 进行比较评估，在航空影像上证明了深度学习立体匹配方法的性能优势和泛化能力。Liu 等^[38]基于循环编码器-解码器 (Recurrent Encoder-Decoder, RED) 结构设计了 RED-Net，并在多个航空影像数据集上验证了其高效性和高精度性，为大规模城市三维重建提供了新的解决方案。Li 等^[37]开放了一套基于高分辨率卫星图像的立体匹配数据集 WHU-Stereo，提供了一个评估高分辨率卫星图像立体匹配模型性能的基准。然而，有关基于多视角光学立体卫星的深度学习密集匹配相关研究却几乎是一片空白。其背后存在两个主要原因：其一，数据的复杂性。卫星影像广泛使用 RPC 模型来进行摄影测量处理，与中心投影成像几何模型差异较大，不利于现有网络的直接迁移应用；其二，数据集的缺乏。高精地形数据的高保密性导致卫星影像三维重建的相关数据难以获取，导致缺乏足够多且具有代表性的数据来训练深度学习模型，使得模型在实际生产中会出现性能退化问题。

1.2.2 基于手工设计特征的光学立体卫星影像三维重建

早期对基于高分辨率光学立体卫星影像三维重建的研究集中于几何成像模型上。使用严格成像几何模型需要考虑到传感器物理构造、成像方式和各种成像参数，不便于处理，且不满足技术秘密保护的需求^[51]。多项研究^[52-54]证明，RPC 模型这种通用传感器模型可以取代严格几何成像模型来进行下游任务，并不会造成精度损失。由于形式简单，独立于传感器，独立于坐标系统等优势，RPC 模型的应用使得基于光学立体卫星影像三维重建的研究得到了进一步的发展^[52,55,56]。

目前，基于光学立体卫星的三维重建方法大多以双目立体卫星影像为处理单元，遵循着核线立体纠正、双目立体密集匹配的流程，获得高密度的同名点，完成三维稠密场景重建。学术界首先对卫星立体像对之间的核线几何关系展开研究，并发现卫星影像的核线在立体影像范围内具有“近似直线性”和“局部共轭性”^[57,58]。基于此，一些学者提出了基于 RPC 模型的卫星核线纠正方法：Wang 等^[59]提出了一种基于投影参考平面的卫星核线模型，用于高分辨率推扫式卫星影像的核线重采样，对 SPOT5、IKONOS、IRS-P5 和 QuickBird 立体影像进行处理

并生成了高精度的核线影像；Tatar 等^[60]不借助有理多项式系数、物理传感器模型以及地面控制点，采用自动特征匹配算法进行影像配准，并将影像切分为小块以提升算法的速度和降低内存的消耗，在 GeoEye-1 立体像对上达到了亚像素的精度；Liao 等^[61]提出一种基于正射纠正的卫星核线影像重采样方法，并采用 DSM 从粗到细的生成流程，在不需要额外时间成本和精度损失的前提下克服了卫星核线像对的大视差问题；在密集匹配技术的研究上，SGM 算法^[35]创新性地用多方向的一维代价聚合去近似全局的二维代价聚合，从而高效地获得近似于全局匹配算法的匹配结果，引发了一波对双目立体视觉的研究热潮。众多学者继而将 SGM 算法应用到基于高分辨率卫星影像的稠密三维重建任务中：Eisank 等^[62]使用 SGM 算法对 Pléades 三线阵立体影像进行处理，在复杂的高山地形中生成了精确的 3D 点云和 DSM；Ghuffar^[63]不仅在影像空间中使用 SGM 算法得到视差图，而且将 SGM 匹配直接用于物方地理参考体素空间，省略了核线纠正和三角化步骤，简化了 DSM 生产流程，并在 Worldview-3 立体像对和 Pléades 三线阵立体影像进行了评估；Gong 等^[64]则使用 SGM 的改进版本 t-SGM^[65]进行卫星影像的密集匹配，获得密集点云，并最终生成 DSM。

随着卫星立体核线几何关系、密集匹配相关研究趋于成熟，研究界和工业界都研发出了一些光学卫星影像立体处理系统：德国航空航天中心 (Deutsches Zentrum für Luft- und Raumfahrt, DLR) 开发了全自动处理系统 CATENA^[66]，在分布系统上使用 SGM 匹配来处理大量的光学卫星影像数据并生成全覆盖的 DSM；俄亥俄州立大学 (Ohio State University, OSU) 的 Byrd 极地和气候研究中心 (Byrd Polar and Climate Research Centre, BPCRC) 开发了 SETSM 系统^[67]，使用基于不规则三角格网的搜索空间最小化算法，从高分辨率卫星遥感影像中提取出了高纬度地区和极地地区的 DSM；法国国家地理与森林信息研究所 (The National Institute of Geographic and Forest Information, IGN) 开发了一款免费开源的摄影测量软件 MicMac^[68]，其中包含一个从 RPC 影像生成 DSM 的微型模块，但是缺乏对 RPC 影像位姿进行调整优化的能力；Qin 开发了 RSP (RPC Stereo Processor) 软件包^[69]，实现了基于 RPC 模型卫星影像的 DSM 和正射影像生成，包括 L2 级别纠正、地理参考、点云生成、全色融合、DSM 重采样和正射校正等步骤。在商业界，大多数也遵循相似的技术路线，如：CATALYST^[70]是加拿大 PCI Geomatics 公司所开发的软件，采用多视角 SGM 技术从卫星和航空影像中提取三维信息；美国 Esri 公司所开发的 ArcGIS 软件包在 10.5 版本后的 Ortho Mapping 工具包^[71]中提供了一系列工具，可以从航空或卫星立体影像对中生成高分辨率的 DSM；俄罗斯 Agisoft 公司发布的 Metashape 软件^[72]自 1.7 版本后也具备了从全色和多光谱卫星影像提取地表三维结构信息的能力。

此外，另有一些学者发现：在局部区域，用中心投影成像模型对 RPC 模型进行拟合的拟合误差在可容忍范围内^[73]。基于这一事实，一些学者开始探索将卫星影像切分为瓦片，在瓦片范围内将 RPC 模型拟合为中央投影成像模型，并利用成熟的计算机视觉工具从卫星影像进行三维重建。其中的典型代表是 S2P^[74]和 Adapted-COLMAP^[75]。其中，Adapted-COLMAP 将 RPC 模型近似为透视相机模型，采用 COLMAP^[36]中基于平面扫描的 MVS 方法来处理多视卫星图像，完成三维重建，从而避免核线重采样。

综上而言，目前生产作业环境中的主流方法还停留于 SGM 等传统手工建模算法，限制了生产作业环境中自动化和智能化水平的提升。

1.2.3 基于自我训练的半监督学习方法

基于自我训练的半监督学习方法是一种利用少量标注样本和大量未标注样本来进行模型训练的方法。模型首先使用标注样本进行监督训练，然后使用已经训练好的模型对未标注样本进行推理，并将这些预测结果作为伪标签 (pseudo-labels) 来扩充训练集，再次进行模型训练。这个过程不断迭代，直到模型收敛。

Yarowsky^[76]在 1995 年将这种方法用于词义歧义消除任务中，通过在未标注的英文文本上训练，达到依靠手工标注的监督方法相媲美的性能。Blum 等^[77]在应对网页分类任务时，考虑到虽然有标签的网页文本很少，但是没标签的网页文本却可以轻易地使用网络爬虫批量地获取，于是设置了一正一负两个分类器，并进行协同训练 (Co-Training)，即用一个分类器在未标记数据上的预测来扩充另一个分类器的训练集，通过实验证明了使用未标注样本可以显著提高性能。Lee^[78]将标记和未标记数据同时用于深度学习网络模型的训练中；对于未标记的数据，选择具有网络输出的最大可能类别作为伪标签参与训练，展现了不错的性能。Doersch 等^[79]在只有大量未标记影像的情况下，从每个图像中提取随机的图像块对，并训练卷积神经网络来预测第二个图像块相对于第一个图像块的位置，充分挖掘影像之间的空间邻域知识，为丰富的视觉特征表达提供了免费而大量的监督信号。Kingma 等^[80]利用深度生成模型和近似贝叶斯推断，并结合变分方法，显著改善生成方法的性能，使得从小规模标记数据集到大规模未标记数据集的有效泛化成为可能。Tarvainen 等^[81]提出了“平均教师”的方法，对模型权重进行平均而不是对标签预测进行平均，提高了测试准确性，并且对标签数量的需求进一步降低。在半监督图像识别任务上，Qiao 等^[82]受到协同训练框架的启发，训练多个深度神经网络，并利用对抗性示例来鼓励不同网络之间的差异，在 SVHN、CIFAR-10/100 和 ImageNet 数据集上的性能上大幅超过了先前的最先进方法。

由于高精度地形数据的高保密性，在实际生产中使用的卫星影像很可能与训练数据集差异巨大，这是将本文所提出的多视光学卫星影像深度学习密集匹配方法推向实际生产应用过程中必须考虑的问题。而基于自我训练的半监督学习方法有可能能够缓解这一问题。

1.3 论文的研究内容与章节安排

1.3.1 主要研究内容

本文针对光学立体卫星影像，构建了一个基于 RPC 模型的深度学习多视密集匹配网络，并以此为核心构建了一套完整的多视光学立体卫星影像三维重建框架。同时，考虑到在迁移应用时，深度学习预训练模型可能出现性能退化、且难以获得相近真值进行模型微调的问题，本文提出一种基于自我训练的模型自优化策略，以保证网络模型在迁移过程中的性能，从而提升所提出框架的实用性。

本研究需要解决的关键问题有三个：**其一，将 RPC 模型进行张量运算式表达。**RPC 模型与中心投影成像模型之间的差异是将现有的多视深度学习密集匹配方法应用于光学立体卫星影像的直接障碍，将 RPC 模型表达成张量运算的形式是实现端到端多视卫星影像密集匹配方法所面临的首要问题。**其二，构建一个完整的多视光学立体卫星影像三维重建框架。**除了核心的多视密集匹配网络外，还需要完善其他必要模块，以及考虑如何对大幅宽卫星影像进行处理的问题，实现从多视卫星影像到三维地形产品生产的全过程。**其三，缓解模型迁移时可能存在的性能退化问题。**由于高精地形数据的保密性，地形真值往往很难获取，这意味着能够所能获取的训练或微调数据极其有限。这也导致了训练数据集与真实场景数据之间的分布可能存在着巨大的差异，并最终使得网络模型在应用场景中出现性能退化的问题。

与关键问题对应，本文的研究内容涉及以下三个关键点：**其一，搭建基于深度学习的多视光学卫星影像密集匹配网络**，并制作一套三线阵多视角密集匹配数据集以进行网络的训练和评估；**其二，构建基于深度学习多视密集匹配方法的光学立体卫星影像三维重建框架**，验证其生产能力；**其三，在真值标签难以获取的情况下，使用基于自我训练的自优化方式进一步精化网络模型**，提升网络在实际生产数据上的性能。

1.3.2 章节安排

本文将分为六章进行介绍：

第一章 绪论。本章主要介绍高分辨率光学立体卫星三维重建研究的背景和意义；国内外研究进展，包括三个部分：基于手工设计特征的光学立体卫星影像三维重建，基于深度学习的影像密集匹配，基于自我训练的半监督学习方法；以及本文的主要研究内容与章节安排。

第二章 坐标系统与坐标转换关系。本章介绍高分辨率卫星影像三维重建任务中涉及到的一些坐标系统、有理多项式相机模型以及各个坐标系之间的坐标转换关系，为后续章节的展开奠定基础。

第三章 多视光学卫星影像密集匹配网络。本章主要介绍所提出的基于有理多项式模型的高维特征可微变形 RPC-Warping、所搭建的多视光学卫星影像密集匹配网络 Sat-MVSNet 和本研究中制作并开源的三线阵多视卫星影像密集匹配数据集 WHU-TLC，并通过实验证明 Sat-MVSNet 网络的有效性。

第四章 基于深度学习多视密集匹配方法的光学立体卫星三维重建框架。本章以多视光学卫星影像密集匹配网络 Sat-MVSNet 为核心，扩展出完整的光学立体卫星三维重建框架 Sat-MVSF。本章主要介绍该框架的处理流程，即：地表稀疏场景重建、地表稠密场景重建和摄影测量产品生产，并通过实验证明其实际生产能力。

第五章 基于自我训练的深度学习模型自优化策略。本章介绍一种无需任何生产目标场景中真值的深度学习网络模型自优化策略，以应对模型迁移过程中由于训练数据与真实生产数据之间差异而造成的模型性能退化问题。

第六章 总结与展望。本章中首先对研究工作进行总结，再对未来进一步的研究方向和可能的改进进行展望。

第二章 坐标系统与坐标转换关系

摄影测量学是一门与坐标紧密连接的学科，坐标系统与不同坐标系之间的坐标转换是其基石。本章主要介绍在基于有理多项式相机 (Rational Polynomial Camera, RPC) 模型的光学立体卫星影像三维重建任务中常涉及到的坐标系和它们之间的转换关系。其中，坐标系统包含影像坐标系、地理坐标系、空间直角坐标系、站心坐标系和投影坐标系；在转换关系中，本章对 RPC 模型单独成节进行介绍，其他坐标系间的转换关系另作一节。

2.1 坐标系统

2.1.1 影像坐标系

与常见的数码相机不同，卫星上搭载的光学成像设备采用线阵电荷耦合器件 (Charge Coupled Device, CCD) 进行推扫式成像，在每一时刻得到一条影像线。而随着卫星的在轨运行，卫星影像由线构面，从而组成二维数字矩阵形式。为了便于成像模型的构建，影像坐标系往往有多种形式。

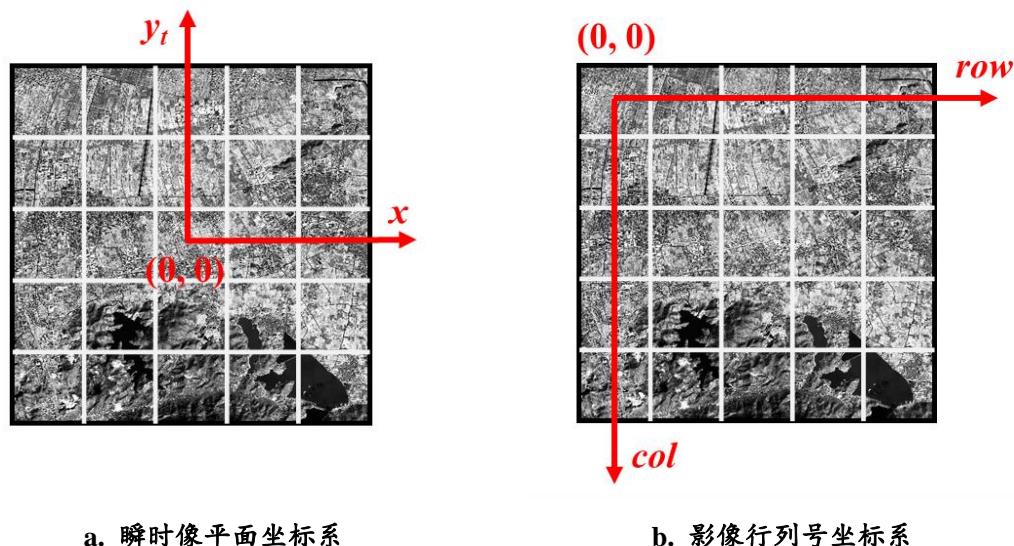


图 2-1 瞬时像平面坐标系与影像行列号坐标系

当基于共线条件方程建立像点坐标和物方坐标关系时，由于每一条影像线完全满足严格的共线条件方程，往往将 CCD 扫描行方向视为 x 轴，将推扫方向视为 y 轴，为每一行建立起瞬时像平面坐标系，如图 2-1 a 所示。像素点的坐标在瞬时像平面坐标系中可表示为 $(x, 0)$ ，而在整个二维影像中可以表示为 (x, y_t) ，其中 y_t 为 t 时刻成像的影像扫描线号。当基于 RPC 模型建立像点坐标和物方坐标

关系时，往往直接使用二维数字矩阵中像素的行列号 (r, c) 作为影像坐标，如图 2-1 b 所示。

2.1.2 地理坐标系

为了便于对地球上的位置进行数学描述，地球球体被近似为一个椭球体，称为参考椭球体。地理坐标系，又称大地坐标系(Geodetic Coordinates System)，则是以参考椭球体为基准所建立起的坐标系统，使用大地经度(Geodetic Longitude)和大地纬度(Geodetic Latitude)来对地球椭球面上的一点进行定位，并附加大地高(Geodetic Height)以准确描述物体在三维空间中的位置，如图 2-2 a。其中，大地经度定义为待定位点所在的子午面与起始子午面所构成的二面角，由起始子午面起算，向东为东经 0 至 180 度，向西为西经 0 至 -180 度；大地纬度定义为待定位点的法线与赤道面的夹角，向北为北纬 0 至 90 度，向南为南纬 0 至 -90 度；大地高定义为待定位点沿椭球法线方向到椭球面的距离，向内为负，向外为正。

2.1.3 空间直角坐标系

与地理坐标系相同，空间直角坐标系(Geocentric Coordinate System)是以参考椭球体为基准所建立，但采用三维坐标 (X, Y, Z) 来描述待定位点的三维位置，如图 2-2 b 所示。其中， Z 轴指向地球平均自转轴， X 轴指向本初子午面与赤道面的交点， Y 轴与 XOZ 平面垂直且指向东。实际应用中，地理坐标和空间直角坐标都有着自身的优势。使用地理坐标能够直观描述出物体在三维空间中的位置，而使用空间直角坐标却能为两个物体之间空间距离的计算带来莫大的方便。

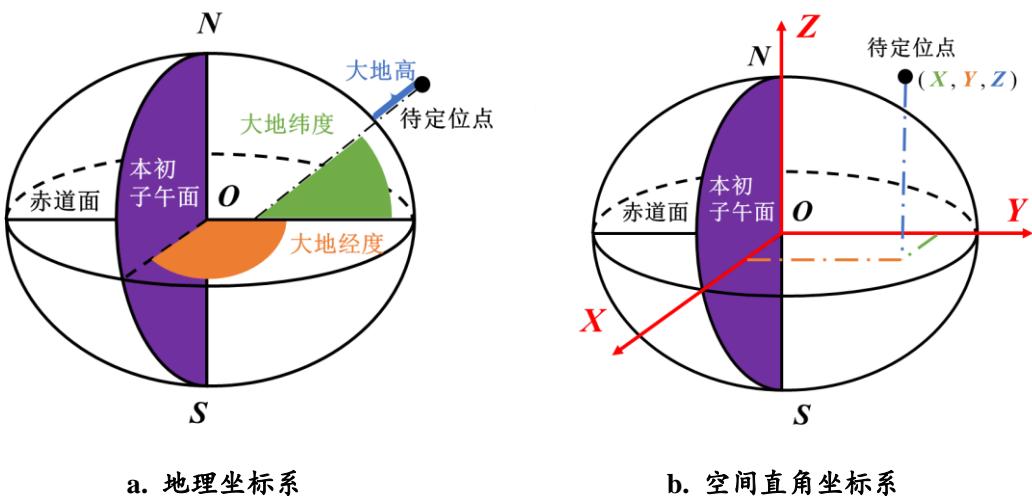


图 2-2 地理坐标系与空间直角坐标系

2.1.4 站心坐标系

站心坐标系(Topocentric Coordinate System)是以点 $(\varphi_o', \lambda_o', h_o')$ 为原点，在地球表面或附近构建起的三维直角坐标系，其X轴指向东(East)，Y轴指向北(North)，Z轴指向上(Up)，因而又被称为东北天坐标系或者ENU坐标系。

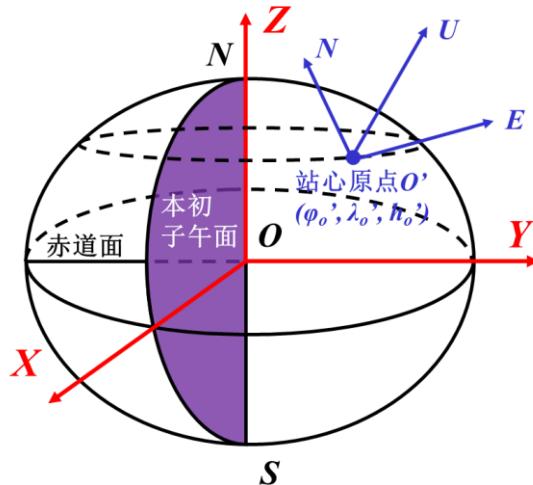


图 2-3 空间直角坐标系与站心坐标系

2.1.5 投影坐标系

为了日常生产生活的便利，特别是为了满足地图制作和使用的需求，需要使用局部平面坐标系对位置进行描述。投影坐标系是一种建立在某种地理坐标系的基础之上，使用某种地图投影算法将地球表面映射到一个平面上而形成的平面直角坐标系。

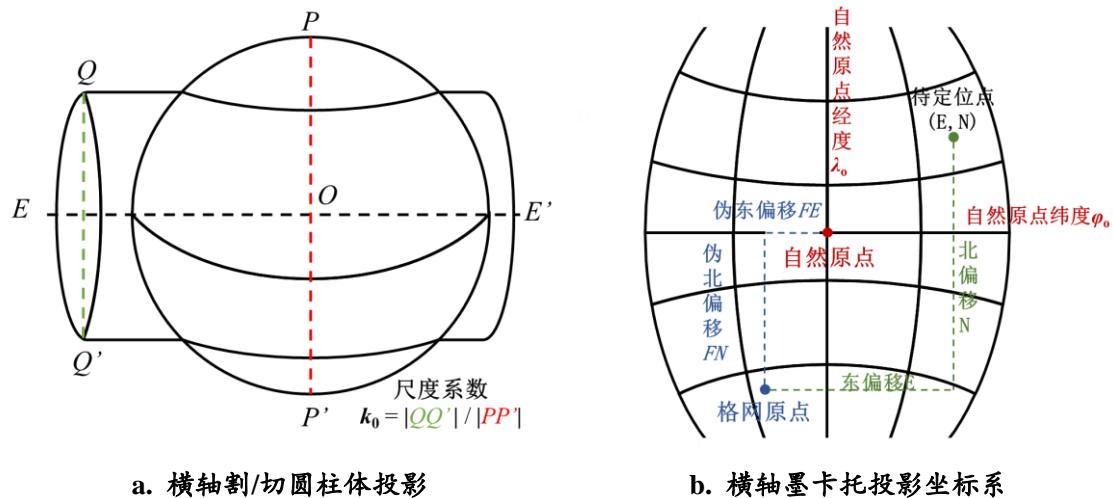


图 2-4 横轴墨卡托投影与横轴墨卡托投影坐标系

横轴墨卡托(Transverse Mercator)投影是使用最为广泛的一类地图投影，其中的典型代表有通用横轴墨卡托(Universal Transverse Mercator, UTM)投影，高斯一

一克吕格(Gauss-Krüger)投影。在定义上，横轴墨卡托投影是一个横轴割/切圆柱体投影，如图 2-4 a 所示。而完全且明确地定义一个横轴墨卡托投影需要以下五个参数：自然原点纬度 φ_0 ，自然原点经度 λ_0 ，自然原点上的尺度系数 k_0 ，伪东偏移 FE 以及伪北偏移 FN 。

2.2 有理多项式相机模型

2.2.1 有理多项式相机模型

有理多项式相机 (Rational Polynomial Camera, RPC) 模型是一种通用成像几何模型，广泛应用于高分辨率卫星影像处理任务中。RPC 模型通过数学函数直接建立起像点与对应物方点（一般为地理坐标系）之间的坐标关系，其中的参数只具有数学含义而不具备任何的几何或物理含义。因而，RPC 模型有着与传感器无关、与坐标系无关、保密性好等诸多优点。

在形式上，RPC 模型可分为正向 RPC 模型和逆向 RPC 模型，其中正向 RPC 模型 $RPC: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 建立了物方点 (φ, λ, h) 到像方点 (r, c) 的转换关系；而逆向 RPC 模型 $RPC^{-1}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 则建立了像方点 (r, c, h) 到物方点 (φ, λ) 的转换关系，如式 (2-1) 所示。值得注意的是，虽然使用正向 RPC 模型和逆向 RPC 模型进行坐标相互转换的精度很高，但是正向 RPC 模型和逆向 RPC 模型之间并不是严格的函数与反函数关系。

$$RPC : \mathbf{R}^3 \rightarrow \mathbf{R}^2 \begin{cases} r_n = \frac{P_1(\varphi_n, \lambda_n, h_n)}{P_2(\varphi_n, \lambda_n, h_n)} \\ c_n = \frac{P_3(\varphi_n, \lambda_n, h_n)}{P_4(\varphi_n, \lambda_n, h_n)} \end{cases} \quad RPC^{-1} : \mathbf{R}^3 \rightarrow \mathbf{R}^2 \begin{cases} \varphi_n = \frac{P_1^{-1}(r_n, c_n, h_n)}{P_2^{-1}(r_n, c_n, h_n)} \\ \lambda_n = \frac{P_3^{-1}(r_n, c_n, h_n)}{P_4^{-1}(r_n, c_n, h_n)} \end{cases} \quad (2-1)$$

式中，下标 n 代表了使用 RPC 模型归一化参数进行归一化； P 代表了三阶多项式，上下标的不同代表了三阶多项式的不同。三阶多项式如式 (2-2) 所示。

$$P(x, y, z) = \sum_{i=0}^3 \sum_{j=0}^3 \sum_{k=0}^3 a_{ijk} x^i y^j z^k \quad (2-2)$$

然而，除了少数的卫星影像，如法国的 Pléades 卫星，同时提供了正向 RPC 模型参数和逆向 RPC 模型参数以外，国际上的 IKONOS 影像、WorldView 影像，以及国产的 ZY-3 影像、GF-7 影像、GF-14 影像等都只提供正向 RPC 模型参数。在没有提供逆向 RPC 模型参数的时候，可利用正向 RPC 模型构建虚拟控制点格网，并按照 2.2.2 节所述进行逆向 RPC 模型的解算。

2.2.2 有理多项式相机模型参数的解算

RPC 模型的参数解算分为与地形无关和与地形有关的解算方案，本文在此只关注前者。注意，此节中解算出的 RPC 模型同时包含正向和逆向参数。

假设物方点与像方点之间的坐标关系已经由一个模型表达出，这个模型可以是严格几何成像模型、已有的 RPC 模型或者是带有仿射补偿模型的 RPC 模型。整个参数解算过程分为以下三步：

(1) 构建虚拟控制点格网。将影像覆盖范围划分为 $m \times n$ 的均匀格网，将高程范围划分为 k 层，构成 $m \times n \times k$ 的均匀物方格网，并使用已知模型获得每一个格网点的虚拟观测值，由物方格网点和其虚拟观测值构成虚拟控制点；

(2) 模型参数解算。使用新建立的 RPC 模型对虚拟控制点格网进行最小二乘拟合，解算出新的模型参数；

(3) 检查。使用已知模型重新构建虚拟检查点格网，对拟合出的新 RPC 模型进行检查。虚拟检查点格网的划分方式如同虚拟控制点，但是划分的格网数量为 $m' \times n' \times k'$ 。

2.2.3 基于正向 RPC 模型的迭代逆向坐标转换

在像点坐标 (r, c) 和高程 h 已知的情况下，求解物方坐标 (φ, λ) 。通常情况下，若逆向 RPC 模型 $RPC^{-1}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 已知，则可直接按式 (2-1) 进行计算。若仅正向 RPC 模型 $RPC: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 参数已知，可使用两种方式计算：(1) 按 2.2.2 节解算出逆向 RPC 模型参数，再按式 (2-1) 进行计算；(2) 通过迭代求解以进行坐标的转换。本节对迭代求解的方式进行介绍。

为了描述的方便，将式 (2-1) 中 $RPC: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 简写为：

$$\begin{cases} r = f_1(\varphi, \lambda, h) \\ c = f_2(\varphi, \lambda, h) \end{cases} \quad (2-3)$$

可列出如下误差方程式：

$$V = AX - l \quad (2-4)$$

其中， $A = \begin{bmatrix} \frac{\partial f_1}{\partial \varphi} & \frac{\partial f_1}{\partial \lambda} \\ \frac{\partial f_2}{\partial \varphi} & \frac{\partial f_2}{\partial \lambda} \end{bmatrix}$, $X = \begin{bmatrix} \Delta \varphi \\ \Delta \lambda \end{bmatrix}$, $l = \begin{bmatrix} r - f_1(\varphi_0, \lambda_0, h) \\ c - f_2(\varphi_0, \lambda_0, h) \end{bmatrix}$ 。初值 φ_0, λ_0 为 RPC

模型中的纬度和经度归一化平移系数。对于一个点，由于未知量的数量为 2，方程数量为 2，因而存在唯一解，使用最小二乘方法求解即可。

2.2.4 基于正向 RPC 模型的多视前方交会

若已知 $\{(r_v, c_v)\}$ 为同名点，则可使用基于正向 RPC 模型的多视前方交会解算出其物方坐标。以 RPC 模型中的归一化系数作为初值 $(\varphi_0, \lambda_0, h_0)$ ，对视角 v 可列出以下误差方程：

$$V_v = A_v X - l_v \quad (2-5)$$

其中， $A_v = \begin{bmatrix} \frac{\partial f_1}{\varphi} & \frac{\partial f_1}{\lambda} & \frac{\partial f_1}{h} \\ \frac{\partial f_2}{\varphi} & \frac{\partial f_2}{\lambda} & \frac{\partial f_2}{h} \end{bmatrix}$ ， $X = \begin{bmatrix} \Delta\varphi \\ \Delta\lambda \\ \Delta h \end{bmatrix}$ ， $l_v = \begin{bmatrix} r_v - f_1(\varphi_0, \lambda_0, h_0) \\ c_v - f_2(\varphi_0, \lambda_0, h_0) \end{bmatrix}$ 。对于一个点，

由于未知量的数量为 3，方程数量为 2，因而需要至少 2 个视角上的同名点才能进行最小二乘求解。方程的求解使用普通最小二乘求解即可。

2.3 其他坐标系间的转换

2.3.1 地理坐标系与空间直角坐标系的转换

假设椭球的长半轴为 a ，短半轴为 b ，则在同一地理基准下，从地理坐标系转换到空间直角坐标系可通过式 (2-6) 完成：

$$\begin{aligned} X &= (v + h) \cos \varphi \cos \lambda \\ Y &= (v + h) \cos \varphi \sin \lambda \\ Z &= [(1 - e^2) v + h] \sin \varphi \end{aligned} \quad (2-6)$$

式中， λ 为经度， φ 为纬度， h 为大地高， $v = a / (1 - e^2 \sin^2 \varphi)^{0.5}$ ， $e^2 = (a^2 - b^2)/a^2$ 。

而其逆向运算为可通过式 (2-7) 完成：

$$\begin{aligned} \varphi &= \arctan[(Z + e^2 v \sin \varphi) / (X^2 + Y^2)^{0.5}] \\ \lambda &= \arctan(Y / X) \\ h &= X \sec \lambda \sec \varphi - v \end{aligned} \quad (2-7)$$

式中，纬度 φ 的解算是迭代的，初值为 $\arctan[Z / (X^2 + Y^2)]$ 。

2.3.2 空间直角坐标系与站心坐标系的转换

空间直角坐标系与站心坐标系之间仅仅存在一个刚体变换。将空间直角坐标系转换为站心坐标系可通过式 (2-8) 完成：

$$\begin{bmatrix} E \\ N \\ U \end{bmatrix} = \begin{bmatrix} -\sin \lambda_o' & \cos \lambda_o' & 0 \\ -\sin \varphi_o' \cos \lambda_o' & -\sin \varphi_o' \sin \lambda_o' & \cos \varphi_o' \\ \cos \varphi_o' \cos \lambda_o' & \cos \varphi_o' \sin \lambda_o' & \sin \varphi_o' \end{bmatrix} \begin{bmatrix} X - X_o' \\ Y - Y_o' \\ Z - Z_o' \end{bmatrix} \quad (2-8)$$

其中, (X_o', Y_o', Z_o') 为站心原点 $(\varphi_o', \lambda_o', h_o')$ 的空间直角坐标。

反之, 从站心坐标系到空间直角坐标系的转化可通过式 (2-9) 完成:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\sin \lambda_o & -\sin \varphi_o \cos \lambda_o & \cos \varphi_o \cos \lambda_o \\ \cos \lambda_o & -\sin \varphi_o \sin \lambda_o & \cos \varphi_o \sin \lambda_o \\ 0 & \cos \varphi_o & \sin \varphi_o \end{bmatrix} \begin{bmatrix} E \\ N \\ U \end{bmatrix} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} \quad (2-9)$$

2.3.3 地理坐标系与投影坐标系的转换

各种形式的横向墨卡托投影是世界地形和近海测绘中应用最广泛的投影坐标系。由于其重要性, 本节只对地理坐标系与横轴墨卡托投影坐标系之间的转换进行介绍。假设椭球的长半轴为 a , 短半轴为 b , 则可以得到: 扁率为 $f = (a - b) / a$, 第一偏心率为 $e = (a^2 - b^2)^{0.5} / a$, 第二偏心率为 $e' = (a^2 - b^2)^{0.5} / b$ 。

地理坐标系到横轴墨卡托投影坐标系的转换可由式 (2-10) 完成:

$$\begin{aligned} E &= FE + k_o v [A + (1 - T + C)A^3 / 6 + (5 - 18T + T^2 + 72C - 58e'^2)A^5 / 120] \\ N &= FN + k_o \{M - M_o + v \tan \varphi [A^2 / 2 + (5 - T + 9C + 4C^2)A^4 / 24 \\ &\quad + (61 - 58T + T^2 + 600C - 330e'^2)]A^6 / 720\} \end{aligned} \quad (2-10)$$

式中的变量计算方式如下:

$$\begin{aligned} T &= \tan^2 \varphi \\ C &= e^2 \cos^2 \varphi / (1 - e^2) \\ A &= (\lambda - \lambda_o) \cos \varphi \\ v &= a / (1 - e^2 \sin^2 \varphi)^{0.5} \\ M &= a [(1 - e^2 / 4 - 3e^4 / 64 - 5e^6 / 256 - ...) \varphi \\ &\quad - (3e^2 / 8 + 3e^4 / 32 + 45e^6 / 1024 + ...) \sin 2\varphi \\ &\quad + (15e^4 / 256 + 45e^6 / 1024 + ...) \sin 4\varphi - (35e^6 / 3072 + ...) \sin 6\varphi + ...] \\ M_o &= a [(1 - e^2 / 4 - 3e^4 / 64 - 5e^6 / 256 - ...) \varphi_o \\ &\quad - (3e^2 / 8 + 3e^4 / 32 + 45e^6 / 1024 + ...) \sin 2\varphi_o \\ &\quad + (15e^4 / 256 + 45e^6 / 1024 + ...) \sin 4\varphi_o - (35e^6 / 3072 + ...) \sin 6\varphi_o + ...] \end{aligned}$$

反之, 从横轴墨卡托投影坐标系到地理坐标系的转换可由式 (2-11) 完成:

$$\begin{aligned}
\varphi &= \varphi_1 - (\nu_1 \tan \varphi_1 / \rho_1) [D^2 / 2 - (5 + 3T_1 + 10C_1 - 4C_1^2 - 9e^{12}) D^4 / 24 \\
&\quad + (61 + 90T_1 + 298C_1 + 45T_1^2 - 252e^{12} - 3C_1^2) D^6 / 720] \\
\lambda &= \lambda_o + [D - (1 + 2T_1 + C_1) D^3 / 6 + (5 - 2C_1 + 28T_1 - 3C_1^2 + 8e^{12} \\
&\quad + 24T_1^2) D^5 / 120] / \cos \varphi_1
\end{aligned} \tag{2-11}$$

式中的变量计算方式如下：

$$\begin{aligned}
\nu_1 &= a / (1 - e^2 \sin^2 \varphi_1)^{0.5} \\
\rho_1 &= a(1 - e^2) / (1 - e^2 \sin^2 \varphi_1)^{1.5} \\
\varphi_1 &= \mu_1 + (3e_1 / 2 - 27e_1^3 / 32 + ...) \sin 2\mu_1 + (21e_1^2 / 16 - 55e_1^4 / 32 + ...) \sin 4\mu_1 \\
&\quad + (151e_1^3 / 96 + ...) \sin 6\mu_1 + (1097e_1^4 / 512 - ...) \sin 8\mu_1 + ...
\end{aligned}$$

而其中，

$$\begin{aligned}
e_1 &= [1 - (1 - e^2)^{0.5}] / [1 + (1 - e^2)^{0.5}] \\
u_1 &= M_1 / [a(1 - e^2 / 4 - 3e^4 / 64 - 5e^6 / 256 - ...)] \\
M_1 &= M_o + (N - FN) / k_o \\
T_1 &= \tan^2 \varphi_1 \\
C_1 &= e^{12} \cos^2 \varphi_1 \\
D &= (E - FE) / (\nu_1 k_o)
\end{aligned}$$

2.4 实验分析与讨论

在本节中，按照 2.2.2 节所述，对多种传感器获取的光学卫星影像所附带的 RPC 参数进行拟合，以验证拟合 RPC 模型的精度与可靠性。所使用的 RPC 参数来自 WorldView3、BJ3、SuperView1、Pléades1A、GF7、GF14、ZY3 和 TH1 卫星影像。

首先，以影像附带的原始正向 RPC 模型为已知模型，按照 2.2.2 节进行正向 RPC 和逆向 RPC 模型的解算。在获取虚拟控制点格网时， $m = n = 10$, $k = 5$ ，而在获取虚拟检查点格网时， $m' = n' = 11$, $k' = 7$ 。之后，统计以下各项精度：

(1) 正向 RPC 模型的正解精度：将检查点的物方坐标通过拟合的正向 RPC 模型转换到像方空间，统计其与虚拟检查点观测值的像方误差，包括列方向中误差 m_x 、行方向中误差 m_y 和中误差 m_{xy} 。

(2) 正向 RPC 模型的迭代反解精度：将检查点的像方坐标和高度通过拟合后的正向 RPC 模型以迭代的方式转换到物方空间，统计其与虚拟检查点在 ENU 坐标系中的物方误差，包括 E 方向上的中误差 m_E , N 方向上的中误差 m_N , U 方向上的中误差 m_U 和中误差 m_{ENU} 。

(3) 逆向 RPC 模型的直接反解精度: 将检查点的像方坐标和高度通过拟合的逆向 RPC 模型直接转换到物方空间, 统计其与虚拟检查点在 ENU 坐标系中的物方误差, 包括 E 方向上的中误差 m_E , N 方向上的中误差 m_N , U 方向上的中误差 m_U 和中误差 m_{ENU} 。

(4) RPC 模型的迭代反解—正解精度: 将检查点的像方坐标和高度通过拟合的逆向 RPC 模型迭代转换到物方空间, 再通过拟合的正向 RPC 模型转换到像方空间, 统计重投影误差, 包括列方向中误差 m_x 、行方向中误差 m_y 和中误差 m_{xy} 。

(5) RPC 模型的直接反解—正解精度: 将检查点的像方坐标和高度通过拟合的逆向 RPC 模型转换到物方空间, 再通过拟合的正向 RPC 模型转换到像方空间, 统计重投影误差, 包括列方向中误差 m_x 、行方向中误差 m_y 和中误差 m_{xy} 。

表 2-1 拟合 RPC 模型的正解精度

	分辨率 (m)	m_x (pixel)	m_y (pixel)	m_{xy} (pixel)
WorldView3	0.3	1.630e-004	1.711e-004	2.364e-004
BJ3	0.5	2.741e-005	9.446e-005	9.740e-005
SuperView1	0.5	6.418e-005	1.293e-004	1.444e-004
Pléades1A	0.5	6.229e-005	4.291e-005	7.564e-005
GF7	0.6	1.202e-004	2.409e-004	2.693e-004
GF14	0.6	2.998e-004	1.745e-004	3.469e-004
ZY3	2.5	7.371e-005	3.189e-005	5.929e-005
TH1	5.0	4.998e-005	7.053e-005	1.020e-004

表 2-2 拟合 RPC 模型的迭代反解精度

	分辨率 (m)	m_E (m)	m_N (m)	m_U (m)	m_{ENU} (m)
WorldView3	0.3	1.515e-003	1.515e-003	1.869e-006	1.541e-003
BJ3	0.5	8.389e-004	8.389e-004	8.221e-007	8.423e-004
SuperView1	0.5	6.152e-005	8.149e-005	8.149e-005	1.021e-004
Pléades1A	0.5	8.649e-003	9.389e-003	9.389e-003	1.277e-002
GF7	0.6	1.683e-004	1.683e-004	2.460e-007	2.491e-004
GF14	0.6	4.123e-004	4.123e-004	7.344e-007	7.053e-004
ZY3	2.5	1.053e-003	5.266e-004	5.266e-004	1.178e-003
TH1	5.0	1.021e-004	1.073e-003	1.073e-003	1.292e-003

表 2-3 拟合 RPC 模型的直接反解精度

	分辨率 (m)	m_E (m)	m_N (m)	m_U (m)	m_{ENU} (m)
WorldView3	0.3	1.510e-004	1.510e-004	1.510e-004	1.636e-004
BJ3	0.5	2.118e-003	8.423e-004	1.593e-006	2.124e-003
SuperView1	0.5	6.014e-004	3.036e-004	4.359e-007	6.737e-004
Pléades1A	0.5	3.168e-005	2.122e-005	2.122e-005	3.813e-005
GF7	0.6	1.888e-004	1.908e-004	2.940e-007	2.685e-004
GF14	0.6	7.625e-003	7.217e-005	1.695e-005	7.626e-003
ZY3	2.5	2.176e-004	1.224e-004	5.652e-007	2.497e-004
TH1	5.0	3.836e-002	6.028e-004	7.949e-005	3.837e-002

表 2-4 拟合 RPC 模型的迭代反解—正解精度

	分辨率 (m)	m_x (pixel)	m_y (pixel)	m_{xy} (pixel)
WorldView3	0.3	3.894e-003	6.422e-004	3.946e-003
BJ3	0.5	1.361e-003	1.829e-004	1.373e-003
SuperView1	0.5	1.144e-004	7.577e-005	1.373e-004
Pléades1A	0.5	1.685e-002	1.831e-002	2.488e-002
GF7	0.6	2.387e-004	1.498e-004	2.818e-004
GF14	0.6	6.429e-004	8.867e-004	1.095e-003
ZY3	2.5	1.191e-004	4.258e-005	1.265e-004
TH1	5.0	1.324e-004	2.170e-004	2.542e-004

表 2-5 拟合 RPC 模型的直接反解—正解精度

	分辨率 (m)	m_x (pixel)	m_y (pixel)	m_{xy} (pixel)
WorldView3	0.3	4.361e-004	1.947e-004	4.776e-004
BJ3	0.5	3.461e-003	3.954e-004	3.484e-003
SuperView1	0.5	1.267e-003	6.919e-004	1.444e-003
Pléades1A	0.5	1.317e-002	1.317e-002	2.742e-006
GF7	0.6	3.681e-004	2.095e-004	4.235e-004
GF14	0.6	1.293e-002	2.483e-003	1.317e-002
ZY3	2.5	1.191e-004	4.258e-005	1.265e-004
TH1	5.0	7.592e-003	1.596e-003	7.758e-003

从表 2-1 可知，拟合 RPC 模型的正解精度基本都在 0.01 像素内；从表 2-2 和表 2-3 可知，拟合 RPC 模型的迭代反解精度和直接反解精度能够达到影像分辨率的 1%；而从表 2-4 和表 2-5 可知，拟合 RPC 模型的迭代反解—正解和直接反解—正解重投影误差基本在 0.01 像素内。虽然迭代反解和直接反解在精度上相差不大，但是相比于迭代反解，直接反解却有着不需要迭代计算以及与正解形式共轭的特点，在进行视角间大量的坐标变换时具有优势。

2.5 本章小结

本章主要介绍了两部分内容：高分辨率卫星影像三维重建任务中涉及到的坐标系及其转换关系。其中所介绍的坐标系包括：影像坐标系、地理坐标系、空间直角坐标系、站心坐标系和投影坐标系；所介绍的转换关系包括：影像坐标系与地理坐标系之间的转换（即 RPC 模型）、地理坐标系与空间直角坐标系之间的转换、空间直角坐标系与站心坐标系之间的转换以及地理坐标系与横轴墨卡托投影坐标系之间的转换。在实验部分，本文使用了来自多种传感器的光学影像 RPC 模型进行 RPC 模型拟合，并对拟合的 RPC 模型进行了精度评价，证明了拟合模型的精度，为后续工作的开展奠定基础。

第三章 基于深度学习的多视光学卫星影像 密集匹配方法

在计算机视觉领域, 基于深度学习的多视密集匹配的快速发展得益于对多视几何关系之间的深入研究, 而大量优势的数据集也同样提供了强劲的动力。然而, 这两个方面同样是阻碍深度学习多视密集匹配方法与光学卫星结合的两大障碍。针对这种情况, 一方面, 本研究自主提出基于有理多项式相机模型的高维特征可微变形 (feature Warping based on RPC model, RPC-Warping), 并将其嵌入网络中, 发展出多视光学卫星影像密集匹配网络 Sat-MVSNet; 另一方面, 本研究还制作并开放了一个大规模的三线阵卫星密集匹配数据集 WHU-TLC, 以期填补国际上的空白。在实验部分, 本研究提供了将 RPC 模型局部拟合成中心投影策略的误差分析, 并在数据集上进行了验证, 证明了基于 RPC-Warping 的多视光学卫星影像密集匹配网络的有效性。

3.1 基于有理多项式相机模型的高维特征可微变形

密集匹配的核心问题是如何根据不同视角下观测到的二维影像, 重建出最接近真实情况的三维几何结构。作为现代深度学习多视密集匹配方法的基础, 平面扫描算法^[83]的核心是, 给定物方搜索空间内一系列的假设平面, 检测在这些平面假设下, 视角间的对应影像块之间是否足够相似, 取使得对应影像块最为相似时的假设平面作为真实三维结构的近似解。本文提出的基于有理多项式相机模型的高维特征可微变形就是完成在平面已知的假设下, 将邻近视角的卫星影像特征与参考视角下的特征进行对齐这一过程, 以便于后期代价的计算, 如图 3-1。

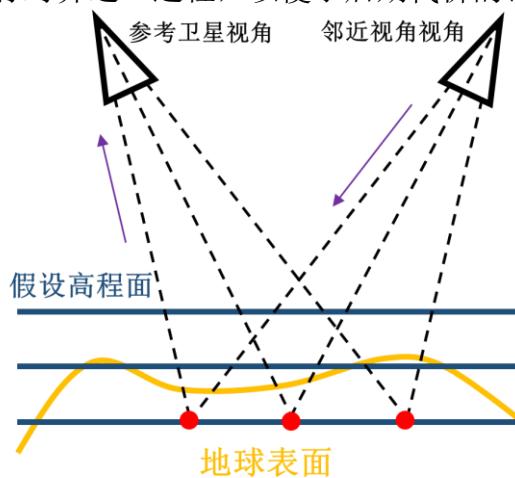


图 3-1 基于有理多项式相机模型的高维特征可微变形示意图

在中心投影成像几何模型中，当假设平面已知时，视角间影像上同名点之间的坐标对应关系可由单应矩阵(Homography Matrix)来描述。在现有的深度学习多视密集匹配方法中，往往使用基于单应矩阵的高维特征可微分变形(Differential Homography Warping, DHW)将邻近影像的高维特征变形到参考视角下，以便进行进一步的处理。然而，卫星影像采用 RPC 模型，本质上仅是一个数学模型，无物理几何含义。相机的真实几何姿态无法从模型中得知，单应矩阵无法直接适用。有研究表明在局部将 RPC 模型拟合成中心投影模型是在误差允许范围内的^[73]，但是拟合方案必定带来误差，且限制了处理时影像瓦片的尺寸，在处理上不够自然。为此，本研究实现了基于 RPC 模型的高维特征可微变形(RPC warping)：基于 RPC 模型，在给定假设平面时，将邻近卫星影像特征变形到参考卫星影像视角下。整个过程可分为三个部分：（1）通过逆向 RPC 模型将邻近影像特征投射到三维空间中；（2）通过正向 RPC 模型将三维空间中的影像特征投射到参考视角；（3）高维特征的采样，一般采用双线性重采样。其中，需要回答两个关键问题：其一，如何选取假设平面；其二，如何对 RPC 模型进行张量化的表达。

由于从 RPC 模型中无法得知相机的真实姿态，更无法推导像素点深度值到物方三维坐标的转换关系，因而在 RPC-Warping 中无法对深度面进行假设。而在正向 RPC 模型和逆向 RPC 模型中，物方高度都直接参与了计算并显式存在于 RPC 模型中，因而本文将物方高程面设定为 RPC-Warping 中的假设平面。

由于正向 RPC 模型 $RPC: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 和逆向 RPC 模型 $RPC^{-1}: \mathbf{R}^2 \rightarrow \mathbf{R}^3$ 具有共轭的形式，在不丧失一般性的前提下，下文中仅以正向 RPC 模型为例进行张量化。RPC 正向模型的分子和分母是数学形式相同的三元三次多项式，可统一表达为：

$$f(x_1, x_2, x_3, x_4) = \sum a_{ijk} x_i x_j x_k \quad (i, j, k \in \{1, 2, 3, 4\}) \quad (3-1)$$

其中， (x_1, x_2, x_3) 为式 (2-1) 中的归一化坐标 $(\varphi_n, \lambda_n, h_n)$ ， $x_4 \equiv 1$ ， a_{ijk} 是多项式系数。将归一化坐标表示为齐次坐标 $\mathbf{X} = (x_1, x_2, x_3, x_4)$ ，并用一个系数张量 $\mathbf{T} \in \mathbf{R}^{4 \times 4 \times 4}$ 来记录多项式系数。系数张量中位置 (i, j, k) 处的元素 $\mathbf{T}(i, j, k)$ 与多项式系数 a_{ijk} 之间的关系如图 3-2 和公式 (3-2) 所示。

$$T_{ijk} = \begin{cases} a_{ijk}, & i = j = k \\ a_{ijk} / 3, & i = j \text{ or } j = k \text{ or } i = k \\ a_{ijk} / 6, & i \neq j \text{ and } j \neq k \text{ and } i \neq k \end{cases} \quad (3-2)$$

则 f 可以表示为四元三次型(Quaternary Cubic Form, QCF)：

$$f(\mathbf{X}) = \sum_{i,j,k=1}^4 \mathbf{T}^{(i,j,k)} \mathbf{X}^{(i)} \mathbf{X}^{(j)} \mathbf{X}^{(k)} \quad (3-3)$$

其中, $\mathbf{X}^{(i)}$ 代表矢量 \mathbf{X} 的第 i 个分量, $\mathbf{T}^{(i,j,k)}$ 代表系数张量在索引 (i, j, k) 处的元素。分子和分母之间再进行逐元素的除法, 即可张量化实现 $RPC^{-1}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 。此外, 为了实现高维特征图的批量高效 RPC-Warping, 本文将四元三次型进行扩展, 如式 (3-4) 所示。

$$f^{(b,d,c,u,v)} = \sum_{i,j,k=1}^4 \mathbf{T}^{(b,i,j,k)} \mathbf{X}^{(b,d,c,u,v,i)} \mathbf{X}^{(b,d,c,u,v,j)} \mathbf{X}^{(b,d,c,u,v,k)} \quad (3-4)$$

其中, b, d, c, u 和 v 分别代表了高维特征的批次、假设面、通道、行和列索引号。借助扩展的四元三次型和逐元素相除, 即可并行批量地将多个批次中具有多个通道的高维特征从三维空间中的多个假设高度平面上投射到参考视角下。

在此需要强调一下, RPC-Warping 需要正向 RPC 模型 $RPC: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 和逆向 RPC 模型 $RPC^{-1}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 的同时参与。在数据本身不提供逆向 RPC 模型时, 需按 2.2.2 节所述, 使用基于正向 RPC 模型解算出逆向 RPC 参数, 再参与到 RPC-Warping 中。如 2.4 节所述, 相比基于正向 RPC 模型的迭代反解, 基于逆向 RPC 模型的直接反解不仅能够达到相似的精度, 而且拥有着无需迭代以及与正解形式共轭的优点, 在视角间进行密集型坐标转换任务更具优势。

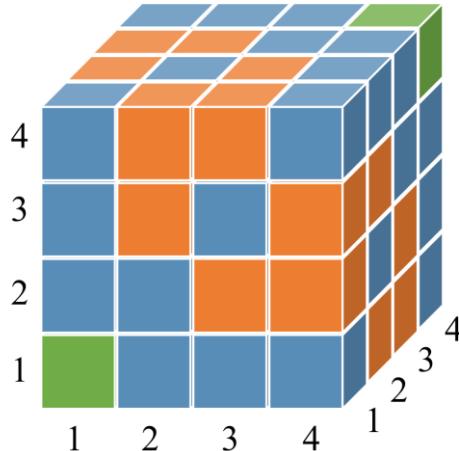


图 3-2 系数张量 T

说明: 系数张量 T 记录了多项式系数。在图 3-2 中, 绿色网格处, i, j, k 三者都相等, 故 $T_{ijk} = a_{ijk}$; 蓝色格网处, i, j, k 中只有其中两个是相等的, 故 $T_{ijk} = a_{ijk}/3$; 而橙色格网处, i, j, k 全部不相等, 故 $T_{ijk} = a_{ijk}/6$ 。

3.2 多视光学卫星影像密集匹配网络 Sat-MVSNet

3.2.1 网络结构

基于上节提出的 RPC-Warping, 本文设计了一个用于多视光学卫星影像的密集匹配网络 Sat-MVSNet, 旨在利用深度学习方法实现多视角卫星影像的匹配, 从而服务于三维地表重建任务。Sat-MVSNet 网络整体采用 3 阶段的金字塔结构,

主要组成部分包括多尺度特征提取器，多尺度代价体构建，代价聚合和高度回归，如图 3-3 所示。其中，在代价体构建过程中，利用 RPC-Warping 实现多视卫星影像特征的对齐。整个网络模型的输入是多视角卫星影像 $\{I_i\}$ 和对应的 RPC 参数 $\{RPC_i\}$ ，以一个视角的影像为参考影像，从其余的邻近视角影像中同时寻求支持，从而实现针对参考影像的高度推理，输出是最后一个阶段参考视角下的高度图 HM ：

$$HM = \mathbf{Z}(\omega, \{I_i, RPC_i\}) \quad (3-5)$$

其中， \mathbf{Z} 为网络模型； ω 为网络权重； HM 为网络输出，即参考视角下的高度图。

相比于依靠手工设计和经验参数设置的传统卫星影像三维重建方案，Sat-MVSNet 网络的优势在于两点：其一，无需核线重采样，直接使用 RPC-Warping 完成影像特征之间的对齐以进行后续代价的计算；其二，能够在匹配的时候利用更多的视角信息，从而提升匹配的鲁棒性。

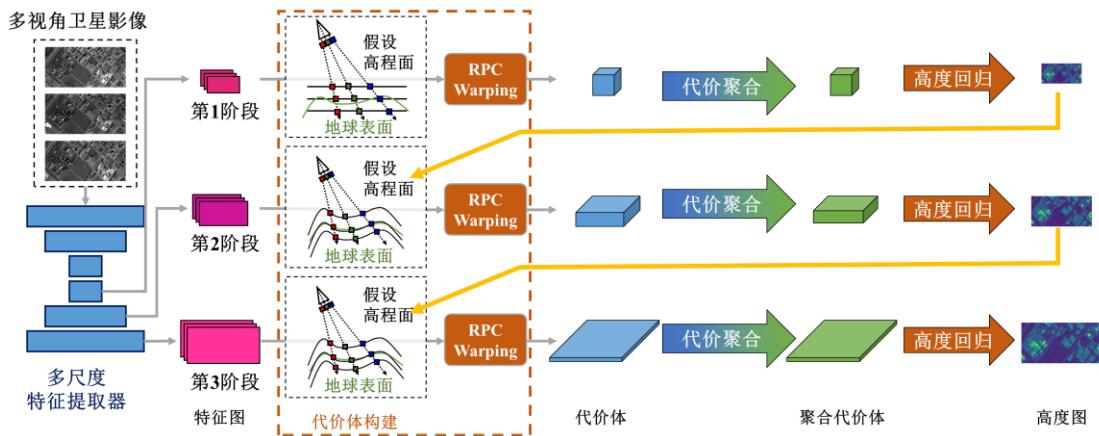


图 3-3 多视角光学卫星影像密集匹配网络 Sat-MVS 结构图

3.2.2 多尺度特征提取

在传统密集匹配方法中，窗口灰度^[84]、census^[85]变换等手工设计特征被广泛应用。然而，受限于研究者对数据集中像素灰度的人工观察和理解，这些手工特征通常无法很好地涵盖数据中的所有变化，具有局限性。而卷积神经元网络 (Convolutional Neural Network, CNN) 具有在大量数据中自动学习生成更具鲁棒性特征的能力，克服了手工设计特征的局限性，在密集匹配中逐渐受到青睐。

为了从影像中提取利于不同尺度上匹配的适当特征，本文采取的多尺度特征提取器为一个简化版的 3 阶段 U-Net^[18]，如图 3-4 所示。其中，编码-解码结构将高分辨率的特征与低分辨率的特征进行融合，以达到同时保留丰富细节和全局信息的效果；跳跃连接将浅层特征进行拷贝并与同一尺度的深层特征串联，以保留空间信息和上下文信息；在每一个阶段，都附加一个卷积层将解码器的特征进一步融合。对于每张输入影像，多尺度特征提取器均输出三个阶段的特征图 $\{F_1, F_2, F_3\}$ ，

$F_2, F_3\}$, 特征图的空间尺寸分别为输入影像空间尺寸的 $\{1/16, 1/4, 1\}$, 特征图的通道数分别为 $\{64, 32, 8\}$, 构成一个 3 阶段的特征金字塔。此外, 为了提升泛化性并减少参数量, 不同视角影像的特征提取器之间权重共享。

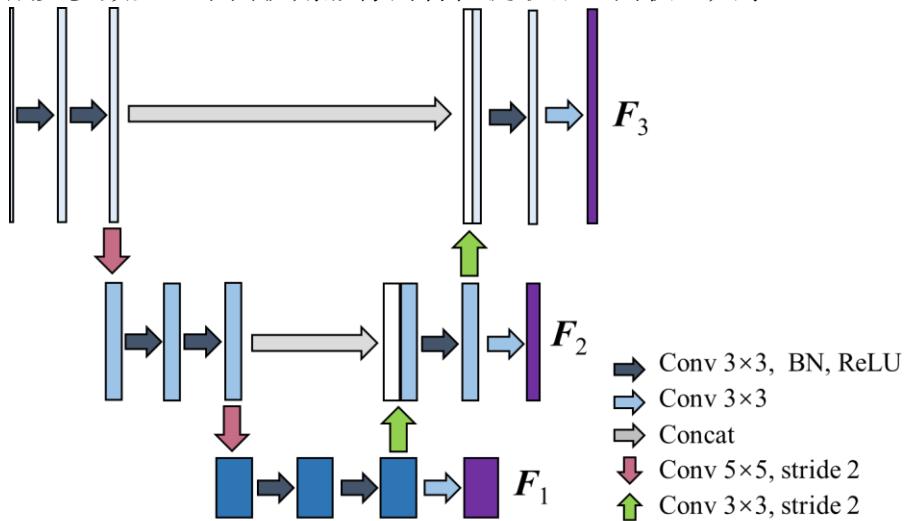


图 3-4 多尺度特征提取器

3.2.3 多尺度代价体构建

本文为 3 个阶段的特征图构建 3 个尺度的代价体, 以实现多层次由粗到细的匹配。代价体的构建需要解决三个关键问题: 其一, 如何确定并划分搜索空间范围; 其二, 如何在物方平面假设已知的情况下, 寻找到对应的多视特征; 其三, 如何从多视角的特征信息中计算代价。

搜索空间范围的确定与匹配所处的阶段相关。在第 1 阶段, 搜索空间范围 SR_1 可由一些场景的先验信息确定。这种先验信息可来源于稀疏场景重建结果、公开 DEM 或 RPC 模型中的高程归一化系数, 其中从稀疏场景重建结果获取搜索空间范围最为准确。稀疏场景重建结果源于多视角特征关键点的交会, 能够覆盖整个场景高程范围的大部分区间, 再将稀疏场景点的高程范围进行一定外扩, 即可确定第 1 阶段的搜索范围, 如式 (3-6) 所示:

$$SR_1 = [H_{s\min} - \Delta h_{s1}, H_{s\max} + \Delta h_{s2}] \quad (3-6)$$

其中, $H_{s\min}$ 和 $H_{s\max}$ 为稀疏场景点的最小高程与最大高程, Δh_{s1} 和 Δh_{s2} 分别向下和向上扩展的高程范围; 从公开 DEM 获取搜索空间范围则需要考虑到公开 DEM 自身可能存在的系统误差等问题, 在这里可假设场景高程在公开 DEM 附近, 将公开 DEM 高程分别向上和向下进行一定扩展, 构建缓冲区以作为搜索空间范围, 如式 (3-7) 所示:

$$SR_1 = [\text{DEM} - \Delta h_{dem1}, \text{DEM} + \Delta h_{dem2}] \quad (3-7)$$

其中, Δh_{dem1} 和 Δh_{dem2} 分别 DEM 向下和向上扩展的高程范围; 而从 RPC 模型中的高程归一化系数获取的搜索范围最为宽松, 假设 RPC 解算时虚拟控制点的高程范围能够覆盖整个场景, 已知 RPC 模型通过高程归一化系数将虚拟控制点高程归一化到-1 至 1, 因而可从归一化系数反向推测出场景的高程范围, 如式 (3-8) 所示:

$$SR_1 = [H_{offset} - H_{scale}, H_{offset} + H_{scale}] \quad (3-8)$$

其中, H_{offset} 和 H_{scale} 为 RPC 模型中的高程归一化系数。在第 2 阶段和第 3 阶段, 搜索空间范围均是以上一阶段的匹配结果高度图为中心的高程缓冲区。在每一个阶段的高程搜索范围内, 按一定步长进行均匀采样得到一系列假想高程面。第 1 阶段、第 2 阶段和第 3 阶段的采样次数固定为 64, 32, 8。由于每个场景的高程变化范围不一, 第 1 阶段中的采样步长被设定为可浮动, 由第 1 阶段的搜索范围和采样数量确定, 而第 2 阶段和第 3 阶段的采样步长固定为地面采样间隔 (Ground Sampling Distance, GSD) 的 2 倍和 1 倍。

在假想物方高程面采样完成后, 使用 3.1 节提出的 RPC-Warping 批量完成多视影像特征之间的对齐, 得到多视角的特征体 $\{Vf_i\}$, $Vf_i \in \mathbf{R}^{d \times b \times c \times h \times w}$ 。

当假设高程面接近真实场景三维结构时, 多视角特征之间差异应当最小, 即离散程度最低。因此, 跟随以往研究^[32,38,47], 本文采用方差来计算多视角的特征之间的代价, 将多视角特征体融合为一个代价体 C , 如式 (3-9):

$$C = \frac{\sum_{i=1}^n (Vf_i - Vf_*)}{n} \quad (3-9)$$

其中, Vf_* 为所有视角特征体的平均特征体, n 为视角数。

3.2.4 代价聚合

在 3.2.3 节中, 每一个阶段所进行的代价计算仅考虑了多视角间局部特征的相似性, 却忽略了真实场景表面具有连续性的假设, 缺乏全局一致性, 易受到噪声的干扰。而这一问题可使用代价聚合来解决。所谓的代价聚合即是考虑相邻像素之间匹配结果应具有相似性, 在邻域或者全局范围内寻求支撑。在深度学习密集匹配方法中, 这一过程也被称为代价体正则化。在传统手工设计算法时代, 最为著名的就是半全局匹配(Semi-Global Matching, SGM)^[35]中使用的方案, 使用多方向的一维代价聚合来近似模拟二维全局的代价聚合, 并用动态规划算法进行求解。而在深度学习时代, 常利用 3D 卷积强大的学习能力从代价体中自动寻求最优的代价聚合方式, 如 GC-Net^[24]、MVS-Net^[47]等。然而, 3D 卷积对内存的需求

随着输入代价体分辨率的增长而呈立方次增长，在应对高分辨率大幅宽的遥感影像时受到阻碍。

本文采用循环编码-解码(Recurrent Encoder and Decoder, RED)^[38]结构来取代3D卷积进行代价聚合。RED结构将行、列、高程上的代价聚合分为两部分：行列空间上的聚合和高程方向上的聚合。其中，行列空间上的聚合主要由2D卷积层组成编码-解码结构来完成；而高程空间上的聚合由卷积门控循环单元(Convolutional Gated Recurrent Unit, ConvGRU)来实现。沿着高程方向，代价体被切分成一张张代价图，这些代价图按顺序依次通过RED结构进行聚合，最后重新叠合成为聚合后的代价体，如图3-5所示。当一张代价图被输入到RED结构时，RED中的编码部分首先会生成四个尺度的特征图；之后，这四张特征图将分别被四个ConvGRU结构处理，并与解码部分生成的对应特征图相加以进行融合；最后，通过一个卷积层输出聚合后的代价图。而当下一个代价图被输入到RED结构后，四个ConvGRU结构通过其隐藏参量 $State^{\{1, 2, 3, 4\}}$ 将上一特征图中的聚合信息进行注入，从而实现高程方向上的代价聚合。

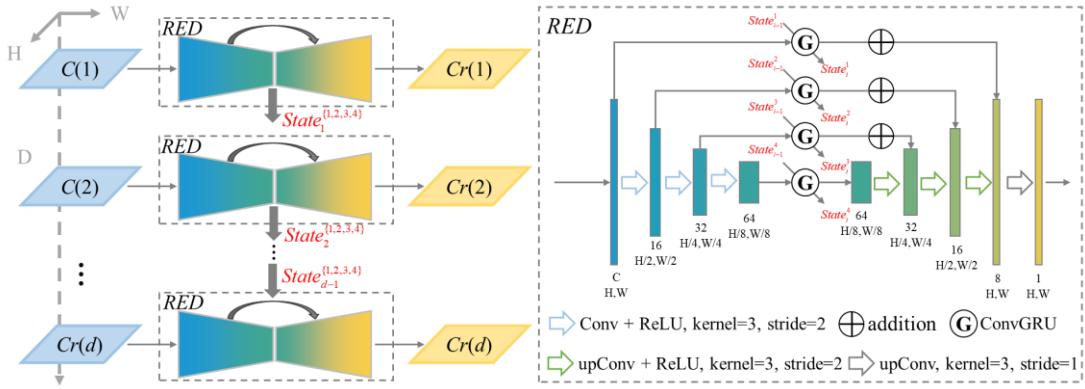


图 3-5 基于循环编码-解码 RED 结构的代价聚合

在应对高分辨率的遥感影像时，RED结构比3D卷积更具优势。本文借用大O记法来从理论上分析两者各自的内存占用情况。具体而言，在使用3D卷积进行代价聚合时，无论是训练过程，还是推理过程中，其空间复杂度均为 $O(HWD)$ ，其中 H 和 W 分别为特征的行列数， D 为假想高程面的数量。然而，在使用RED结构时，代价体是沿着高程方向分割的并依次进行处理的。RED一次仅处理一个切片（即一张代价图），并将沿D维的聚合信息交予ConvGRUs的隐藏参量进行汇总。在训练过程中，所有的变量和梯度都需要被记录下来用于反向传播，因此RED结构进行代价聚合的空间复杂度仍为 $O(HWD)$ 。但是，在推理过程中，由于只进行正向计算，所以只需要存储前一张代价图的隐藏状态和当前处理的代价图，空间复杂度骤降为 $O(HW)$ 。从以上分析可以看出，RED结构在推理时，空间复杂度与代价体的假想高程面的数量无关，因而可以在硬件条件相同的情况下使用更精细的高程假设平面采样，以处理更高分辨率的影像。

3.2.5 高度回归

从代价体中获得匹配结果，最直观的做法就是为每一个像素选取最小代价的高程面作为匹配结果，这便是传统方法中常用的“赢家通吃”(Winner Takes All, WTA)策略。然而，WTA 策略是典型的“硬决策”，无法进行微分，为端到端的深度学习网络设计带来了障碍；此外，WTA 只是进行选取，需要额外的插值手段以达到子像素级别的匹配精度。为了解决这两个问题，本文跟随^[47]采用 soft argmin 来进行高度值的回归，如式 (3-10) 所示。

$$y = \sum_{d=0}^{D-1} d \times \sigma(C_d) \quad (3-10)$$

其中， y 为最终选取的假想高程面索引号， D 为假想高程面的数量， σ 为 softmax 函数。最终，回归得到的高度图 HM 可由式 (3-11) 计算得到：

$$HM = SR_{\min} + S \times y \quad (3-11)$$

其中， SR_{\min} 是搜索范围的下界， S 为假想高程面的采样步长。

3.2.6 训练损失

当真值已知时，网络训练的损失函数定义为最小化估计高度值和真值之间的差异。特别地，本文在 3 个阶段都使用平滑 L1 损失，如式 (3-12) 所示：

$$L = \begin{cases} \sum 0.5(HM - HM_{gt}), & \text{if } |HM - HM_{gt}| < 1 \\ \sum(|HM - HM_{gt}| - 0.5), & \text{otherwise} \end{cases} \quad (3-12)$$

而整个网络的损失定义为 3 阶段损失的加权和，如式 (3-13) 所示：

$$Loss = \sum_{i=1}^3 w_i \cdot L_i \quad (3-13)$$

其中，本文为 3 阶段损失赋予的权重分别为 0.5, 1 和 2。

3.3 三线阵多视卫星影像密集匹配数据集 WHU-TLC

3.3.1 数据源简介

国际上已有一些卫星影像三维重建相关的数据集，如 MVS3D^[86]和 US3D^[73]，其影像数据源为 WorldView-3 卫星。然而，WorldView-3 卫星自身只搭载了单线阵推扫式 CCD 相机，通过对同一场景进行重访拍摄，获取不同时相的影像数据构成立体模型，如图 3-6 a 所示。由于立体模型中的影像获取存在时间窗口，影像之间存在巨大的季节和场景差异，并不适合密集匹配任务自身。而本文所提出

的数据集，其影像数据源为资源三号 (ZY3) 02 星。ZY-3 号卫星上搭载了三线阵推扫 CCD 式相机，几乎同时对场景进行拍摄获取多视角卫星影像，如图 3-6 b 所示。由于三线阵直接获取影像的时间几乎是同时的，不存在场景变化和季节差异的干扰，从而有利于聚焦密集匹配任务自身。

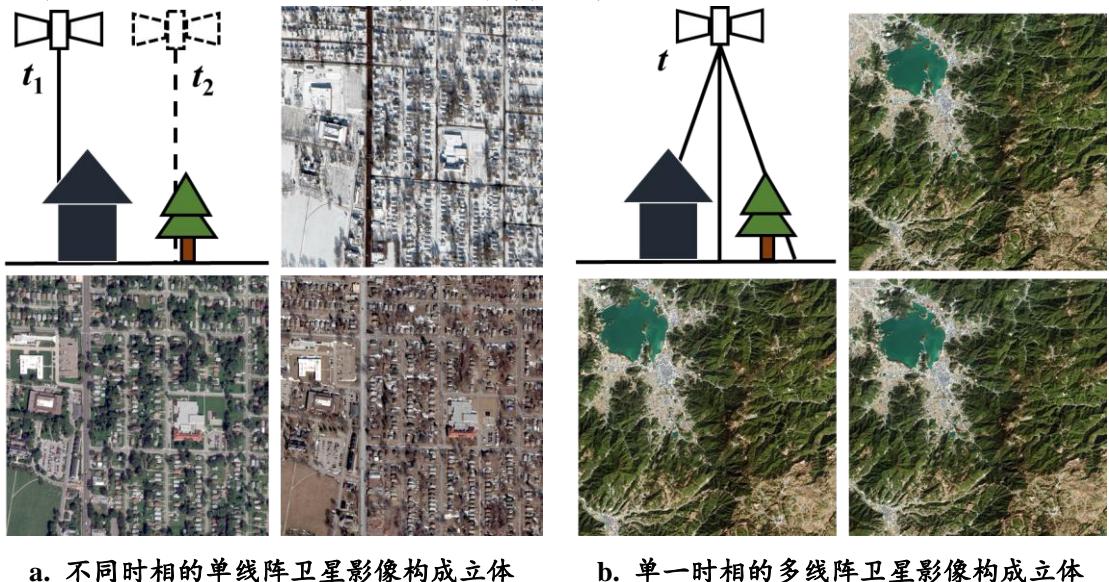


图 3-6 光学卫星影像构成立体像对的两种方式

本文所提出的 WHU-TLC 数据集，其影像数据由安装在 ZY-3 卫星上的三线阵相机(Three-Line array Camera, TLC)采集。一套三线阵影像由前视、下视、后视影像各一张组成，拍摄的高度约为 506 km。其中，前视影像和后视影像的地面分辨率为 2.5 米，下视影像的分辨率为 2.1 米，前视与下视、后视与下视之间的夹角约为 22 度。作为专业测绘卫星，ZY-3 卫星自身搭载的三线阵卫星从硬件设计上保证了其几乎在同一时间对同一场景进行立体拍摄，不受光照和季节变化的影响，为密集匹配与三维重建任务的开展提供了极大的便利。在本文使用的数据中，影像的 RPC 参数已经进行了校准，达到了亚像素级的重投影误差，并与 DSM 真值对齐。而作为真值的 DSM 是由高精度 LiDAR 观测数据和摄影测量软件生产得到，使用 WGS-84 椭球和 UTM 投影，存储格网的分辨率为 5 米。

3.3.2 WHU-TLC 数据集及制作过程

本文构建了三套 WHU-TLC 数据集：第一套数据被称为原始数据集，是为了便于用户对完整的重建流程进行测试和评估而创建，共包括 127 组训练样本和 46 组测试样本。一个样本中包括有 5120×5120 像素的 16 位全色三视角卫星影像、RPC 参数和 DSM 产品，如图 3-7 所示。数据所覆盖的地表总面积约为 125 平方公里，三视角卫星影像之间的重叠度超过 95%；第二套数据被称为 **RPC 瓦片数据集**，是为了方便在主流的硬件上对 Sat-MVSNet 进行训练，将第一套数据进

行处理得到。具体而言，将每张 5120×5120 像素的影像裁剪成 768×384 像素的瓦片，瓦片之间的水平和垂直方向的重叠率均为 5%。总计 5011 组样本用于训练，1791 组用于测试。每一组样本包含多视卫星影像瓦片、RPC 参数和高度图，如图 3-8 所示；第三套数据被称为拟合瓦片数据集，将第二套数据中的 RPC 模型根据^[75]中的描述，在 UTM 坐标系下拟合为中心投影成像模型，并将 DSM 按照拟合后的模型投影到影像瓦片上获得深度图。第三套数据的样本数量与第二套一致，每一组样本包含多视卫星影像瓦片、中心投影模型参数和深度图。

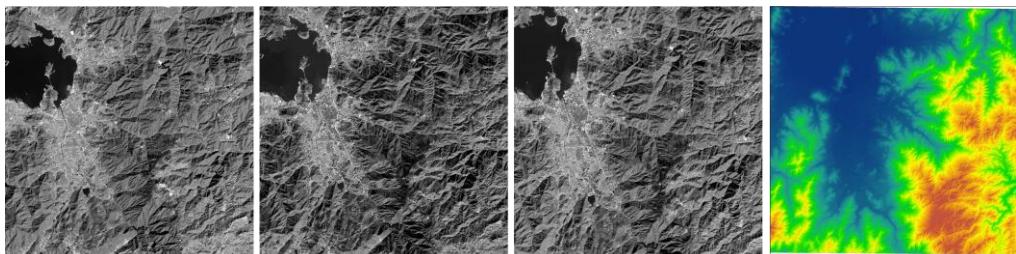


图 3-7 三线阵全色影像与数字高程模型

对于第二套数据的制作过程，以下进行进一步的介绍。第一步，检查逆向 RPC 参数是否存在，如果不存在，则按照 2.2.2 节进行逆向 RPC 参数的解算；第二步，按照指定的重叠度和格网尺寸，将多视影像中的一张划分为均匀格网，将每一个格网范围内的影像裁切为瓦片；第三步，将每一个瓦片按照平均高程投影到物方空间，裁切得到对应的 DSM 瓦片；第四步，将 DSM 瓦片按照 DSM 自身的高程投影到其他视角，以投影区域的中心为中心，裁切指定的格网尺寸，得到邻近视角的影像瓦片；第五步，对于每一张影像瓦片，将影像的 RPC 参数中行列归一化系数中的偏移系数分别减去瓦片在原始影像中的行列坐标，获得每一个影像瓦片的 RPC 参数；第六步，对于每一张影像瓦片，将 DSM 瓦片按照影像瓦片的 RPC 参数投影到瓦片上，保留最大高程值即可获得高度图。

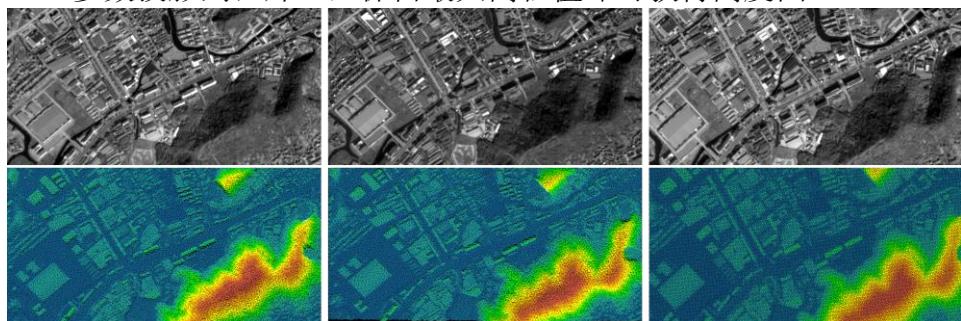


图 3-8 多视卫星影像瓦片与高度图

3.4 实验分析与讨论

3.4.1 评价指标

本文使用三种精度指标来评价 DSM 的精度。

(1) 平均绝对误差(Mean Average Error, MAE): 指的是真值 DSM 和估计 DSM 在所有网格单元上的 L_1 距离平均值, 如式 (3-14) 所示。

$$MAE = \frac{\sum_{(i,j) \in G \cap G_{gt}} |x_{ij} - y_{ij}|}{\sum_{(i,j) \in G \cap G_{gt}} I((i,j) \in G \cap G_{gt})} \quad (3-14)$$

其中, G 和 G_{gt} 分别估计 DSM 和真值 DSM 中的有效格网单元, x_{ij} 和 y_{ij} 分别代表了估计 DSM 和真值 DSM 位于第 i 行第 j 列的格网单元上的高程值, $I(A)$ 表示艾弗森括号, 在 A 为真时取值为 1, 否则取值为 0。

(2) 均方根误差(Root Mean Square Error, RMSE): 指的是真值 DSM 和估计 DSM 之间残差的标准差, 如式 (3-15) 所示。

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in G \cap G_{gt}} (x_{ij} - y_{ij})^2}{\sum_{(i,j) \in G \cap G_{gt}} I((i,j) \in G \cap G_{gt})}} \quad (3-15)$$

(3) 准确网格占总比(Percentage of Accurate Grids in total, PAG): 将真值 DSM 和估计 DSM 之间残差小于阈值 α 的格网定义为准确格网, 统计准确格网数量与真值 DSM 中有效格网数量的比值, 如式 (3-16) 所示。

$$PAG_\alpha = \frac{\sum_{(i,j) \in G \cap G_{gt}} I(|x_{ij} - y_{ij}| < \alpha)}{\sum_{(i,j) \in G_{gt}} I((i,j) \in G_{gt})} \quad (3-16)$$

3.4.2 RPC 模型拟合为中心投影模型的误差分析

为了强调 RPC-Warping 提出的必要性, 本文首先进行了一项数值模拟实验。本文选取一景 TH-1 号卫星影像的 RPC 模型为代表, 将 RPC 模型按照^[75]中的方案拟合成中心投影模型, 并使用原始 RPC 模型构建虚拟检查点格网以用于拟合误差分析。

本文选取一系列不同尺寸的方形影像块, 在影像块的范围内拟合成中心投影模型, 并将检查点的物方坐标按照拟合的中心投影模型投影到影像上, 统计其最大投影误差, 如表 3-1 所示。拟合误差在 XY 平面上的误差分布如图 3-9 所示。

表 3-1 用中心投影模型拟合 RPC 参数的拟合误差 (单位: 像素)

影像大小	768^2	4608^2	9216^2	13824^2	18432^2	23040^2
最大误差	0.16	1.09	2.43	3.90	5.36	6.62

如表 3-1 所示，拟合误差是不可避免的，且拟合误差随着影像块尺寸的增大而逐渐增加。在尺寸为 768×768 像素的瓦片分片中，拟合误差较小，几乎可以忽略不记；而在分块尺寸接近 5000×5000 像素，最大拟合的误差就已经大于 1 个像素；在分块尺寸为 23040×23040 像素时，最大拟合误差会达到 6 个像素以上。由此可见，将 RPC 模型拟合成中心投影的策略只能局限于小尺寸瓦片，而一旦分块尺寸接近 5000×5000 像素，就会为重建带来不可忽视的精度损失。

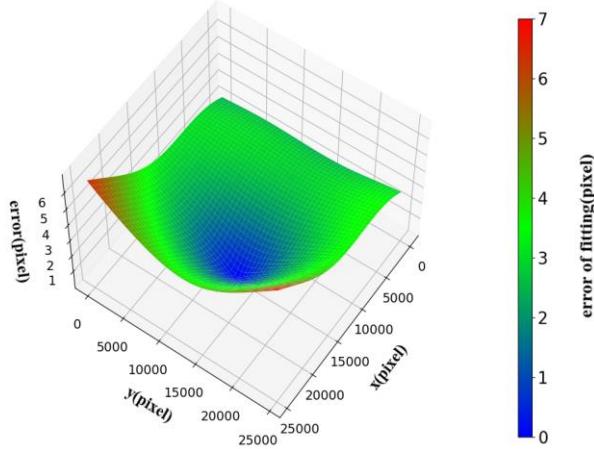


图 3-9 用中心投影模型拟合 RPC 参数的误差分布图

3.4.3 多视光学卫星影像密集匹配网络的有效性验证

1. 实验设置

使用 WHU-TLC 数据集的测试集来验证本文提出的多视光学卫星影像密集匹配网络 Sat-MVSNet 的有效性。对比的方法包括：基于 DHW 和 3D 卷积代价聚合的 CasMVSNet^[32]，将 DHW 改为 RPC-Warping 后的 CasMVSNet (被记作 CasMVSNet / RW)，基于 DHW 和 RED 代价聚合的 MS-RED-Net^[87]以及本文提出基于 RPC-Warping 和 RED 代价聚合的 Sat-MVSNet。

所有的网络模型都使用 PyTorch 框架进行实现，并在单张 NVIDIA TITAN RTX GPU(24GB)上完成训练。在各个模型的训练过程中，超参数遵循相同的设计：在训练阶段，批次大小设置为 1，优化器选择为 RMSprop。所有的网络都被训练了 35 个轮次；初始学习率为 0.001，并在第 10 个轮次后降低为原来的一半。其中，CasMVSNet 和 MS-RED-Net 在 WHU-TLC 的拟合瓦片训练集上进行训练，CasMVSNet / RW 和 Sat-MVSNet 在 WHU-TLC 的 RPC 瓦片训练集上进行训练。

在预测时，使用 WHU-TLC 原始测试集，并进行 DSM 的生产，在 DSM 上进行精度评价。对于 CasMVSNet 和 MS-RED-Net，先将影像切分成瓦片，将瓦片 RPC 参数拟合为中心投影成像模型；再将影像和对应的成像参数输入到网络模型中，推理得到深度图；之后，经过几何一致性检验，将深度图按照中心投影成像模型投影到物方空间的均匀格网中，取每个格网中的最高点的高程值，生成

DSM，并与真值 DSM 进行评价。对于 CasMVSNet / RW 和 Sat-MVSNet，先将影像切分成瓦片，保留瓦片的 RPC 参数；再将影像和瓦片的 RPC 参数输入到网络模型中，推断得到高度图；之后，经过几何一致性检验，将高度图按照 RPC 模型投影到物方空间的均匀格网中，取每个格网中的最高点的高程值，生成 DSM，并与真值 DSM 进行评价。在训练和预测过程中，网络输入的视角数固定为 3。

2. 实验结果

将影像裁剪成不同尺寸的瓦片，输入不同网络模型，在 WHU-TLC 测试集上的定量比较结果如表 3-2 所示，此处采用的评价指标有 MAE、RMSE、PAG_{2.5m}、PAG_{7.5m} 以及从输入一组影像到 DSM 生产结束为止的运行时间。

表 3-2 不同的网络模型在 WHU-TLC 测试集上的定量比较

瓦片尺寸 (像素)	网络模型	MAE (m)	RMSE (m)	PAG _{2.5m} (%)	PAG _{7.5m} (%)	运行时间 (min)
2048×1472	CasMVSNet	2.031	4.351	63.72	79.47	4.00
	CasMVSNet / RW	2.020	3.841	62.61	78.87	12.33
	MS-RED-Net	2.171	4.514	60.65	78.47	9.25
	Sat-MVSNet	1.945	4.070	64.13	79.48	13.87
5120×5120	CasMVSNet			OOM		
	CasMVSNet / RW			OOM		
	MS-RED-Net	2.517 (0.346)	4.873 (0.358)	54.09 (6.56)	77.80 (0.67)	4.28 <u>(4.97)</u>
	Sat-MVSNet	1.946 (0.001)	4.224 (0.153)	64.13 (0.00)	79.50 (0.02)	5.87 <u>(8.00)</u>

说明：表格中加粗的数字代表在该瓦片尺寸设置下该指标中表现最优；括号中斜体、下划线分别代表了分块尺寸设置为 5120×5120 像素时候，相比 2048×1472 像素，在该指标上性能下降和上升的数值；OOM 代表模型运行对内存的需求超过了硬件内存。

从表 3-2 中，可以得到一些结论：

(1) 相比于拟合后再使用 DHW，使用 RPC-Warping 作为高维特征变形模块更具优势。与 3.4.2 节中的模拟实验结论相同，将 RPC 模型拟合成中心投影成像模型不可避免会引入误差。在处理小瓦片影像的时候，由于引入的误差可以忽略，对 RPC 模型进行拟合再使用 DHW 的网络模型与直接使用 RPC-Warping 的网络模型在性能上表现几乎相当，甚至前者能达到更高的运行效率。然而，当瓦片尺寸变大时，由于拟合误差变大，基于拟合 DHW 模型的网络模型性能急剧下降，在 MAE、RMSE 和 PAG_{2.5m} 三个指标上分别下降约 15.94%、7.93% 和 10.82%。而使用 RPC-Warping 的模型，其性能几乎不受瓦片尺寸变化的影响。

(2) 使用大瓦片尺寸能够带来效率上的提升。相比使用 2048×1472 像素的瓦片大小，使用 5120×5120 像素的瓦片大小分别为 MS-RED-Net 和 Sat-MVSNet 带来了 53.72% 和 57.68% 的效率提升。

(3) RED 结构更适合大幅宽的高分辨率遥感影像处理。CasMVSNet 采用 3D 卷积进行代价聚合，其内存消耗对输入瓦片尺寸极其敏感，在处理大瓦片尺寸瓦片时不占优；而 RED 结构由于其分片逐层聚合的特点，对输入瓦片尺寸不敏感，对大幅宽的高分辨率遥感影像处理更为友好。

图 3-10 中展示通过不同的网络模型所生产得到的 DSM 产品。从图中，可以得到结论：多视光学卫星影像密集匹配网络是有效的，所有的网络模型都恢复出了三维地形的几何结构。此外，从局部等高线图来看，不同的网络模型在细节上有些许差异。

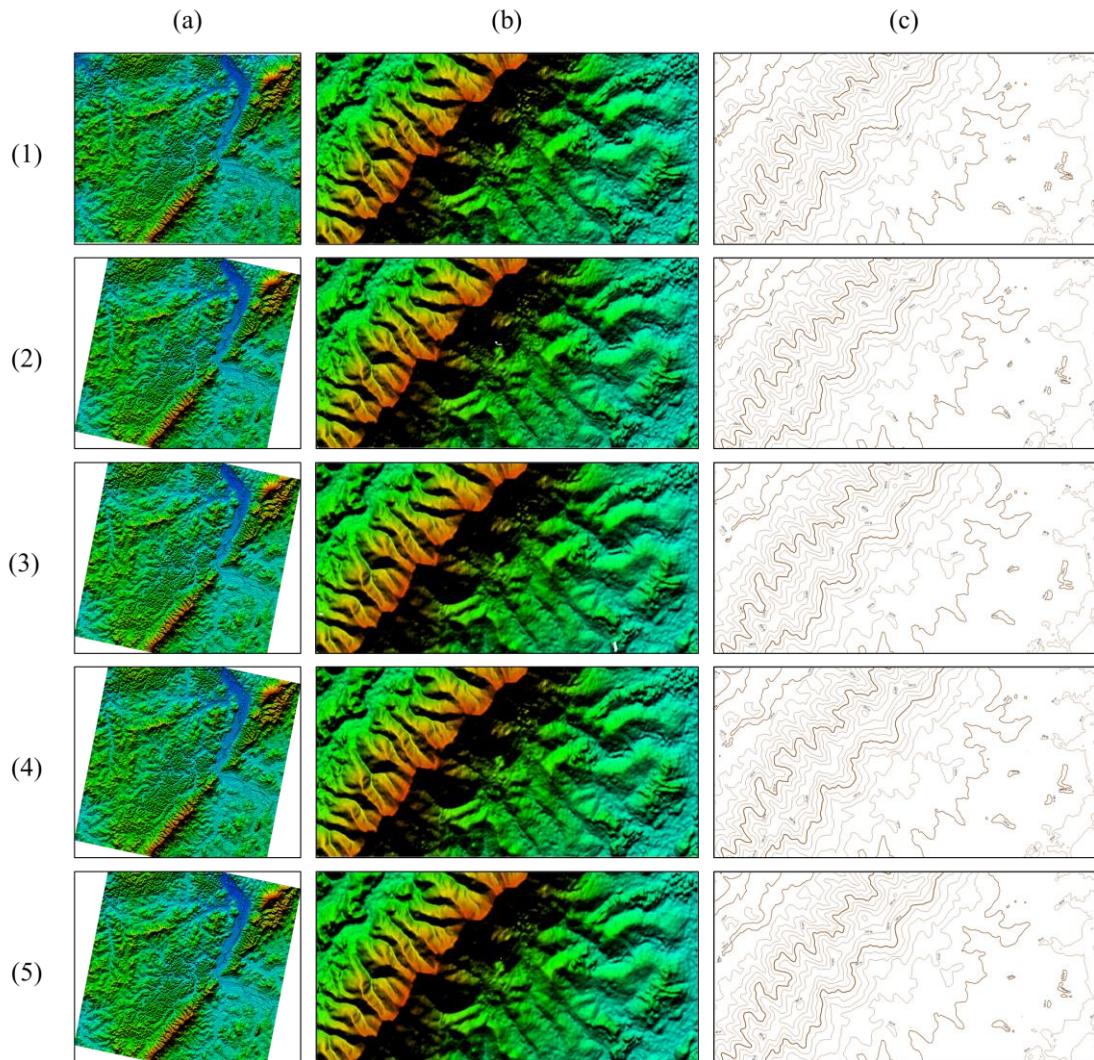


图 3-10 不同的网络模型在 WHU-TLC 测试集上的 DSM 可视化结果

說明：行(1), (2), (3), (4)和(5)分別代表真值、通过 CasMVSNet 模型生产的 DSM、通过 CasMVSNet / RW 模型生产的 DSM、通过 MS-RED-Net 模型生产的 DSM 和通过 Sat-MVSNet 模型生产的 DSM；列(a), (b), (c)分別代表 DSM 渲染结果图、DSM 渲染结果图局部和与局部渲染图对应的等高线图局部。其中，等高线图由 DSM 格网自动跟踪生成，等高距为 20m。

3.5 本章小结

本章首先将 RPC 模型改写成了张量运算的形式，提出了基于 RPC 模型的高维特征可微分变形 RPC-Warping；紧接着，对所提出的多视光学卫星影像密集匹配网络 Sat-MVSNet 进行了详细的介绍；此外，还介绍了三线阵多视卫星影像密集匹配数据集 WHU-TLC 及制作过程。在实验部分，首先介绍了对 DSM 的评价指标；之后，采用模拟实验对 RPC 模型拟合为中心投影模型的误差进行了分析；然后，使用 WHU-TLC 数据集进行了多视光学影像密集匹配网络的有效性验证，并得到以下结论：（1）在瓦片足够小的时候，使用中心投影模型拟合 RPC 模型是有效的，但这种有效性无法在大尺寸瓦片上得到保证；（2）处理大尺寸的瓦片时，RPC-Warping 比拟合后使用 DHW 更具性能上的优势。（3）将影像裁切为大尺寸瓦片能带来效率上的提升。（4）在高分辨率遥感影像处理任务上，基于 RED 结构的代价聚合方式比 3D 卷积更具优势。

第四章 多视光学卫星影像三维重建框架

为了完成大幅宽多视卫星影像的三维地形重建任务，除了基于深度学习的多视卫星影像密集匹配方法的设计外，还需要应对多个技术问题：一方面，原始光学影像的定位精度有限，需要进行成像参数的优化，以实现精确的相对位姿定位；另一方面，目前的硬件水平不足以支撑大幅宽影像的直接处理，需要完善针对大幅宽卫星影像的处理策略；此外，多视密集匹配结果的表现形式为高度图，需要进行进一步处理以生成最终的摄影测量产品。在本章中，以基于深度学习的多视卫星影像密集匹配方法为核心，搭建光学立体卫星三维重建通用框架，并将其应用到光学立体卫星三维重建任务中。

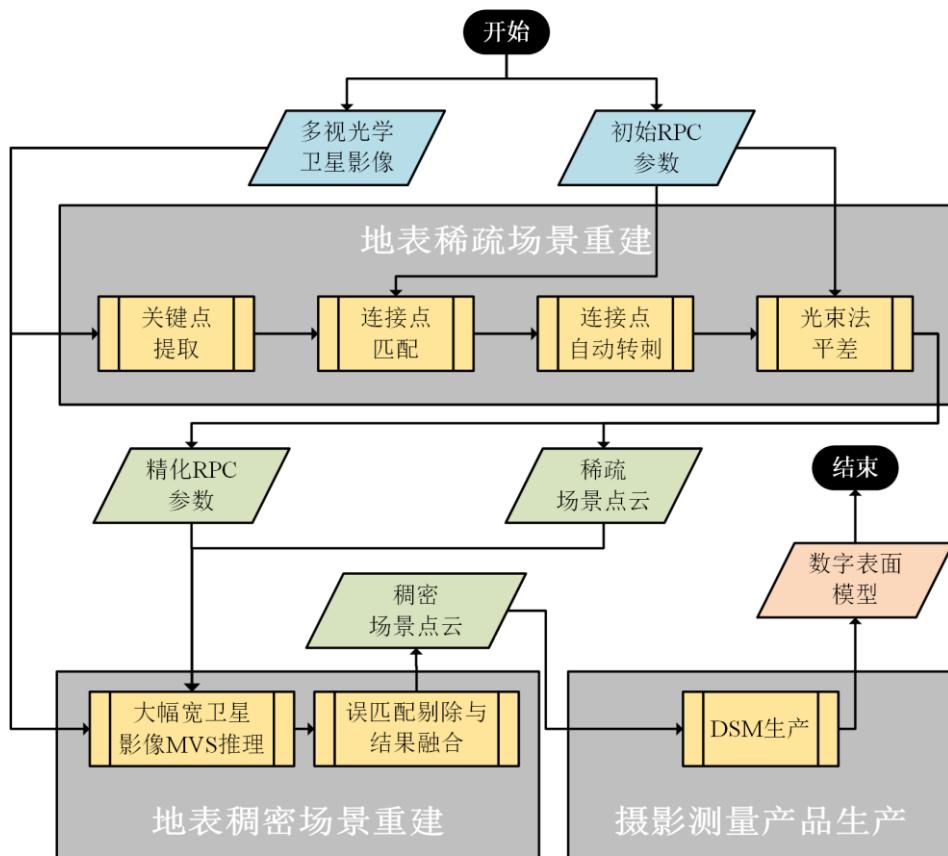


图 4-1 基于深度学习多视密集匹配的光学立体卫星三维重建通用框架流程图

基于深度学习多视密集匹配方法的光学立体卫星三维重建通用框架 Sat-MVS-F 整体流程如图 4-1 所示，分为三大部分：地表稀疏场景重建、地表稠密场景重建和摄影测量产品生产。其中，在地表稀疏场景重建部分中，进行成像参数的优化与稀疏场景的重建，为稠密重建提供准确的成像参数和搜索范围约束；在地表稠密场景重建中，利用多视卫星密集匹配网络完成大幅宽卫星影像密

集匹配，并进行粗差剔除和冗余融合，生成稠密场景点云；而在摄影测量产品生产部分中，利用稠密场景点云生成数字高程模型(Digital Surface Model, DSM)。

4.1 地表稀疏场景重建

4.1.1 基于均匀分块的大幅宽卫星影像关键点提取

为了应对大幅宽卫星影像的处理场景，本文采用一种基于均匀分块策略的大幅宽卫星影像关键点提取方案，以获得数量足够的、分布均匀的、精确的关键点。

首先，对参与特征提取的影像波段，提取均值 μ 、方差 σ^2 、最大值 Mx 、最小值 Mn 以及直方图 $Hist$ 等灰度统计特征(Intensity Statistical Features, ISF)。金字塔低分辨率波段的灰度信息与原始分辨率波段仅在高频区域差异明显，其灰度分布相似。因而，为了提高效率，并不直接对原始的参与波段进行统计，而是采用预先构建的影像金字塔中参与波段的较低分辨率采样进行统计，以获取波段灰度统计特征的近似值 $ISF_0=\{\mu_0, \sigma_0^2, Mx_0, Mn_0, Hist_0\}$ 。

其次，使用简单均匀分块策略对大幅宽卫星影像进行瓦片切分。在每一块瓦片上，利用式 (4-1) 所示的截断百分比线性拉伸等方法对每块瓦片进行对比度增强，以提高瓦片影像的对比度，从而提取出更多的特征点。其中，对比度增强方法中采用全局的波段灰度统计特征 ISF_0 ，而非瓦片影像的统计特征。

$$Ix = \frac{(Ix - Mn_\tau)}{(Mx_\tau - Mn_\tau)} \cdot 255 \quad (4-1)$$

其中， τ 为截断百分比点位， Mn_τ 和 Mx_τ 分别为直方图中累计频率达到 τ 和 $1-\tau$ 所对应的灰度值，默认 τ 值取 0.02。

随后，对每一个瓦片影像提取 SIFT^[88] 特征点。为了获得分布尽可能均匀的特征点，规定每一个瓦片影像上需提取的 SIFT 特征点数量与其像素数成正比，如式 (4-2) 所示。

$$Nf_{tile} = \frac{Np_{tile}}{Np_{total}} \cdot Nf_{total} \quad (4-2)$$

其中， Nf_{tile} 和 Nf_{total} 分别为瓦片影像和影像全图需要提取的特征点数量； Np_{tile} 和 Np_{total} 分别为瓦片影像和影像全图的像素数。

最后，对所提取的 SIFT 描述子 $desc$ 按 (4-3) 式进行处理，得到 RootSIFT 描述子^[89]，以获得更好的匹配效果。

$$desc' = \sqrt{desc / \sum_{i=0}^{128} desc_i} \quad (4-3)$$

其中， $\sqrt{}$ 代表向量中每一个元素进行开方运算。

4.1.2 几何约束下的卫星影像连接点匹配

当进行特征点匹配时，由于卫星影像中存在噪声、遮挡和变形等因素，容易导致匹配错误。虽然现代卫星影像的初始 RPC 参数定位精度有限，但仍不失为一种可靠的先验信息。因此，本文采用初始 RPC 模型作为几何约束，以引导和检验卫星影像之间的特征点匹配，提高匹配的精度和可靠性。

首先，进行像对的筛选。交会角过大的影像对之间变形大、遮挡严重，匹配困难；交会角过小的影像对之间虽容易匹配，但交会精度却很低，不利于重建；重叠度小的影像对匹配点很少，失去匹配的价值。在像对筛选中，仅有交会角 α 在 $[\tau_{\alpha 1}, \tau_{\alpha 2}]$ 区间内且重叠度 ω 大于 τ_ω 的像对被保留，从而减小匹配的搜索空间，以提高匹配的准确性和效率。影像的重叠度定义如式 (4-4) 所示：

$$\omega = \frac{R_1 \cap R_2}{R_1 \cup R_2 - R_1 \cap R_2} \quad (4-4)$$

其中， R_1 和 R_2 分别为两张影像各自覆盖地面范围的最小外接矩形。影像对之间的交会角由投影轨迹法确定，如图 4-2 所示。假设影像 1 为参考影像，取影像 1 的中点记作 P_1 。将 P_1 通过影像 1 的 RPC 模型按照最大高程和最小高程投影到物方空间得到点 A 和 B 点，即确定一根假想光线 AB 。将 B 点投影到影像上得到点 P_2 。将 P_2 通过影像 2 的 RPC 模型按照最大高程投影到物方空间得到点 C ，即确定了同名光线 BC 。则影像 1 与 2 之间的交会角定义为站心坐标系下，假想光线 AB 和其同名光线 BC 的夹角，如式 (4-5) 所示。

$$\alpha = \arccos \frac{\mathbf{AB} \cdot \mathbf{BC}}{|\mathbf{AB}| |\mathbf{BC}|} \quad (4-5)$$

其中，假想光线 AB 和其同名光线 BC 中各点的坐标均以转换到站心坐标系之中。

之后，对每一对像对中的匹配点进行相互交叉检验(Cross Check)，即对两幅影像中提取的特征点进行双向匹配，保证匹配点互为最优匹配点，以进一步减少匹配的错误率。同时，检验匹配点的独特性(uniqueness)，即匹配点在特征空间中的距离不仅应是与特征点最近邻，且其距离与次近邻匹配点的距离之比还应当小于阈值 τ_u ，以提高匹配的精度和鲁棒性。

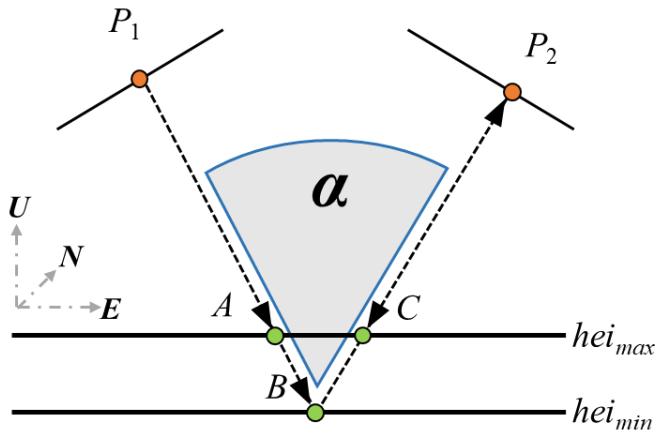


图 4-2 卫星影像对交会角估计

4.1.3 基于并查集的连接点自动转刺

在获得两两像对之间大量的同名点之后，如何从大量无序的两两连接关系得到多视连接关系成为下一个需要解决的问题。在摄影测量学中，这个过程被称为“连接点的转刺”；而在计算机视觉中，被称为“特征追踪”(Feature Tracking)。

为了便于描述，本文将一组匹配描述为一条追踪边(Edge)，边的端点(Node)分别是影像 I_1 上的特征点 f_1 和影像 I_2 中的同名匹配点 f_2 ，记作 $f_1 \rightarrow f_2$ ；而多个追踪边首尾相连可以构成一条追踪路线(Track)，且称追踪路线 $f_1 \rightarrow f_2$ 经过影像 I_1 和 I_2 。每一条追踪线路上的每一个端点都可以通过两两匹配关系最终连接到追踪线路上的某一个端点。本文将此端点记为根端点 f_{root} ，而与之连接的端点都称为子端点 f_{child} ，并且规定父端点有一个子端点为自身。由于正确的追踪路线之间绝不可能相交，因而可知根端点与追踪线路一一对应。

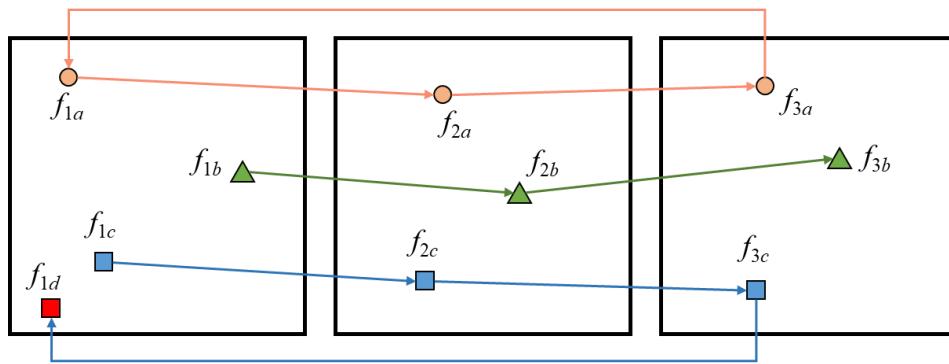


图 4-3 追踪路线的不同几何型态

根据追踪路线构成的几何图形，可以形象地将追踪路线分为树(Tree)和环(Ring)：如果追踪路线的所有子路线都不重复经过同一张影像，则该追踪路线呈树状，如图 4-3 中的绿色追踪路线 $f_1b \rightarrow f_2b \rightarrow f_3b$ ；而如果追踪路线中存在某一子

路线经过了某张影像两次，则称该子路线呈环状，且称该追踪路线中有环。进一步，如果环中端点的数量等于经过的影像数量，则该环被称作闭合环(Closed Ring)，如图 4-3 中的橙色追踪路线 $f_{1a} \rightarrow f_{2a} \rightarrow f_{3a} \rightarrow f_{1a}$ ；否则，称作开口环(Open Ring)，如图 4-3 中的蓝色追踪路线 $f_{1c} \rightarrow f_{2c} \rightarrow f_{3c} \rightarrow f_{1d}$ 。而开口环违背了匹配的唯一性(Uniqueness)原则，不满足多视角之间的一致性，是由于匹配错误造成的追踪失败，需要进行识别与剔除。

本文采用并查集(Union Find)算法^[90]来完成这一任务：对于每一个匹配对 $f_a \rightarrow f_b$ ，如果两个端点 f_a 和 f_b 都不是某个根端点的子端点，则将 f_a 作为新的根端点，并将 f_b 作为 f_a 的子端点；如果 f_a 为某一根端点的子端点，则将端点 f_b 同样作为该根端点的子端点，反之亦然。最终，每一个端点都会成为某个根端点的子端点，而根端点的所有子端点构成一条跟踪线路。然而，本文还考虑到了开口环这一实际情况。根据分析可得，如果跟踪线路中存在开口环，则其中必然存在至少两个端点出现在同一张影像上。据此可对开口环进行检测并剔除，以确保错误跟踪不会对后期处理造成影响。

算法 4-1 基于并查集的连接点自动转刺算法

算法：基于并查集的连接点自动转刺算法

```

输入：无序匹配对  $Matches = \{f_{ai}, f_{bi}\}$ 
输出：跟踪路线  $Tracks$ 
1   初始化父端点列表  $Vecf_{root}$  和子端点列表  $Vecf_{child}$ 
2   for  $f_a, f_b$  in  $Matches$  do
3       for  $froot$  in  $Vecf_{root}$  do
4           if  $f_a$  属于  $Vecf_{child}[froot]$  do
5               | 将  $f_b$  插入  $Vecf_{child}[froot]$ 
6           else if  $f_b$  属于  $Vecf_{child}[froot]$  do
7               | 将  $f_a$  插入  $Vecf_{child}[froot]$ 
8           else
9               | 将  $f_a$  插入  $Vecf_{root}$ 
10              | 将  $f_a$  插入  $Vecf_{root}[f_a]$ 
11              | 将  $f_b$  插入  $Vecf_{child}[f_a]$ 
12    $Tracks = Vecf_{child}$ 
13   for  $track$  in  $Tracks$  do
14       for  $f$  in  $track$  do
15           if  $track$  中存在与  $f$  视角相同的端点 do
16               | 从  $Tracks$  中剔除  $track$ 
17   return  $Tracks$ 

```

4.1.4 基于像方仿射补偿模型的光束法平差

为了提升初始 RPC 模型的定位精度，本文采用像方仿射模型^[52]对 RPC 模型误差补偿，并进行整体光束法平差(Bundle Adjustment)提升定位精度。

首先，采用重投影误差来对每一条追踪路线进行检验，以尽可能降低错误匹配对平差过程带来的误差影响。理想情况下，一条追踪路线的所有端点是对应于

同一物方点的多视同名。对这些端点 $\{f_v\}$ 如 2.2.4 节进行基于 RPC 正向模型的前方交会, 得到物方点坐标 $(\varphi_{t0}, \lambda_{t0}, h_{t0})$ 。再通过每个视角的 RPC 正向模型 $RPC_v: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 投影到影像上, 得到坐标 $\{f_v'\}$ 。若其中存在 $\|f_v - f_v'\| > \tau_{re}$, 则认为跟踪线路的误差超过了观测值测量误差和定位模型的误差所能带来的影响, 存在错误匹配, 需要被剔除。阈值 τ_{re} 默认为 5 个像素。

之后, 在某一个视角对应的 RPC 模型上施以如式 (4-6) 的像方补偿仿射模型, 进行整体平差, 提升定位精度并获得稀疏三维场景点。

$$\begin{cases} r_b = g_1(\mathbf{a}, \varphi, \lambda, h) = a_1 f_1(\varphi, \lambda, h) + a_2 f_2(\varphi, \lambda, h) + a_3 \\ c_b = g_2(\mathbf{a}, \varphi, \lambda, h) = a_4 f_1(\varphi, \lambda, h) + a_5 f_2(\varphi, \lambda, h) + a_6 \end{cases} \quad (4-6)$$

其中, $\mathbf{a} = \{a_i\}$ ($i = 1, \dots, 6$) 为仿射系数, $f_1(\varphi, \lambda, h)$ 和 $f_2(\varphi, \lambda, h)$ 参考式 (2-3), 初值为 $\mathbf{a}_0 = \{1, 0, 0, 0, 1, 0\}$ 。对于每一条追踪路线 $\{f_v = (r_v, c_v)\}$, 其物方初值设定为前方交会结果 $(\varphi_{t0}, \lambda_{t0}, h_{t0})$, 可列出以下误差方程:

$$V_t = A_t X_t - l_t \quad (4-7)$$

$$\text{其中, } A_t = \begin{bmatrix} \frac{\partial g_{1(t)}}{\partial a_1} & \frac{\partial g_{1(t)}}{\partial a_2} & \frac{\partial g_{1(t)}}{\partial a_3} & 0 & 0 & 0 & \frac{\partial g_{1(t)}}{\partial \varphi} & \frac{\partial g_{1(t)}}{\partial \lambda} & \frac{\partial g_{1(t)}}{\partial h} \\ 0 & 0 & 0 & \frac{\partial g_{2(t)}}{\partial a_4} & \frac{\partial g_{2(t)}}{\partial a_5} & \frac{\partial g_{2(t)}}{\partial a_6} & \frac{\partial g_{2(t)}}{\partial \varphi} & \frac{\partial g_{2(t)}}{\partial \lambda} & \frac{\partial g_{2(t)}}{\partial h} \end{bmatrix},$$

$$l_t = \begin{bmatrix} g_{1(t0)} - r_v \\ g_{2(t0)} - c_v \end{bmatrix}, \quad X_t = [\Delta a_1, \Delta a_2, \Delta a_3, \Delta a_4, \Delta a_5, \Delta a_6, \Delta \varphi, \Delta \lambda, \Delta h]^T, \quad g_{1(t)} \text{ 和 } g_{2(t)} \text{ 分别}$$

为 $g_1(\mathbf{a}, \varphi_t, \lambda_t, h_t)$ 和 $g_2(\mathbf{a}, \varphi_t, \lambda_t, h_t)$ 的缩写, $g_{1(t0)}$ 和 $g_{2(t0)}$ 为将初始值 \mathbf{a}_0 和 $(\varphi_{t0}, \lambda_{t0}, h_{t0})$ 分别带入 $g_{1(t)}$ 和 $g_{2(t)}$ 中计算所得值。而对于每个控制点 $(\varphi_c, \lambda_c, h_c)$, 其像方观测值为 $\{o_v = (r_c, c_c)\}$ 。可列出以下误差方程式:

$$V_c = A_c X_c - l_c \quad (4-8)$$

$$\text{其中, } A_c = \begin{bmatrix} \frac{\partial g_{1(c)}}{\partial a_1} & \frac{\partial g_{1(c)}}{\partial a_2} & \frac{\partial g_{1(c)}}{\partial a_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial g_{2(c)}}{\partial a_4} & \frac{\partial g_{2(c)}}{\partial a_5} & \frac{\partial g_{2(c)}}{\partial a_6} \end{bmatrix}, \quad l_c = \begin{bmatrix} g_{1(c0)} - r_c \\ g_{2(c0)} - c_c \end{bmatrix},$$

$X_c = [\Delta a_1, \Delta a_2, \Delta a_3, \Delta a_4, \Delta a_5, \Delta a_6]^T$, $g_{1(c)}$ 和 $g_{2(c)}$ 是 $g_1(\mathbf{a}, \varphi_c, \lambda_c, h_c)$ 和 $g_2(\mathbf{a}, \varphi_c, \lambda_c, h_c)$ 的缩写, $g_{1(c0)}$ 和 $g_{2(c0)}$ 分别代表将初始值 \mathbf{a}_0 和 $(\varphi_c, \lambda_c, h_c)$ 分别带入 $g_{1(c)}$ 和 $g_{2(c)}$ 中计算所得值。在构建完基于 RPC 模型的光束法平差系统之后, 采用列文伯格-马夸尔特(Levenberg-Marquardt, LM)算法进行求解, 得到每一条追踪路线的场景点坐标以及每一张影像的仿射系数。在缺少地面控制点时, 基于 RPC 模型的光束法平差模型自由度过高, 法方程系数矩阵很可能秩亏, 并出现解算不收敛或者误差过

度累积等问题。为缓解这一问题，本文采用^[91]提出的方法，利用影像的初始 RPC 模型生成虚拟控制点，并加入整个平差系统中。

最后，对于每一张影像，使用带有像方补偿仿射的 RPC 模型构建虚拟控制点，按 2.2.2 节解算出精化 RPC 参数。

4.2 地表稠密场景重建

4.2.1 大幅宽多视卫星影像的 MVS 推理

在使用基于深度学习的多视密集匹配网络对大幅宽多视卫星影像进行推理时，如何裁剪得到具有足够重叠度的多视同名瓦片是一个关键问题。本文使用稀疏场景点、精化 RPC 参数作为几何约束，将大幅宽多视卫星影像裁切成一系列的同名瓦片组，以输入深度学习多视光学卫星影像密集匹配网络进行推理，如图 4-4 所示。

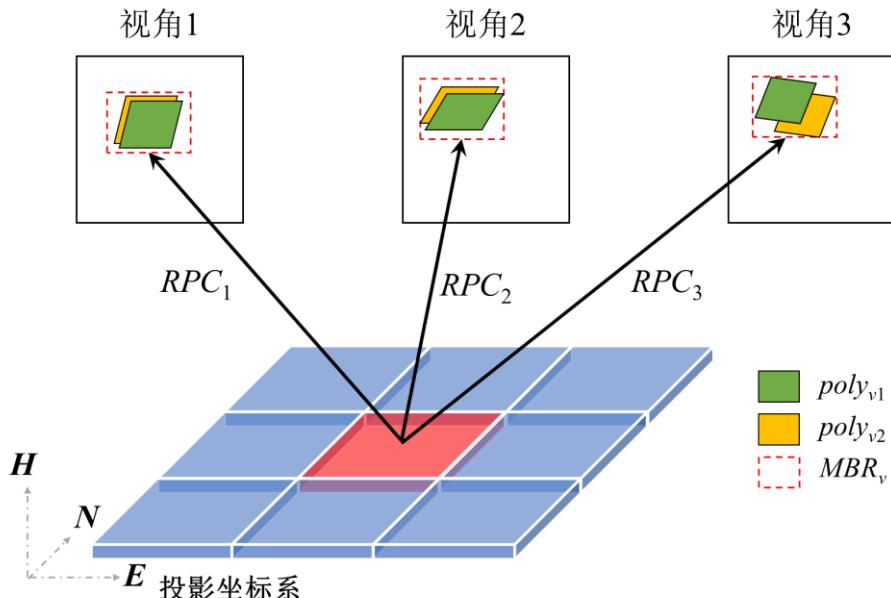


图 4-4 基于几何约束的大幅宽多视卫星影像同名瓦片裁剪

首先，将影像覆盖的地表范围在平面上划分为均匀格网。对于其中一个格网，以稀疏场景点的高程范围作为约束，使用视角 v 的精化 RPC 参数将格网四角点按稀疏场景点的最大高程 H_{Mx} 和最小高程 H_{Mn} 投影到各个视角的影像上，获得两个多边形区域 $Poly_{v1}$ 和 $Poly_{v2}$ ；之后，获取每一个视角上两多边形的最小外接矩形 MBR_v ；而后，保持所有 MBR_v 的中心不变，将所有 MBR_v 的长和宽各自统一扩展为其中的最大值，裁剪获得同名多视瓦片。

将同名多视瓦片送入网络模型进行推理之后，将每一组瓦片的推理结果进行镶嵌，得到大幅宽多视卫星影像的推理结果，即多视角高度图。

4.2.2 基于多视几何一致性的误匹配剔除与匹配结果融合

本文将每一张影像都作为参考影像，推理得到对应的高度图。而由于噪声、遮挡、光照变化等因素，多视匹配结果之间难免会出现一些误匹配点。为了去除这些误匹配点，本文采用几何一致性进行误匹配点的剔除和冗余匹配的融合，如算法 4-2 所示。

算法 4-2 基于多视几何一致性的误匹配剔除与匹配结果融合算法表

算法 2：基于多视几何一致性的误匹配点剔除与匹配结果融合算法

输入： RPC 模型 $\{RPC_i\}_{i=1}^V$, 高度图 $\{HM_i\}_{i=1}^V$, 距离阈值 τ_d , 可视数阈值 τ_v
输出： 点云 *Point*, 可视关系 *Visibility*

```

1  初始化 Point, Visibility 为 空
2  for ref = 1 to V do
3      初始化 pVec, GeoVec 为 空
4      for v = 1 to V do
5          for 每一个高度图像素 p do
6              | pVec[v][p] 和 GeoVec[v][p] 设为无效值
7          for 每一个高度图像素 p do (In parallel)
8              | pVec[ref][p] = p
9              | if HMref[p] 为无效值 do
10                 | | continue
11             使用 RPCref 将 p 点按 HMref[p] 转换为 GeoVec[ref][p] (见 2.2.3 节)
12             将 PGeo 按照转换为 PENU (见 2.3.1 和 2.3.2 节)
13             for src = 1 to V do
14                 if ref 与 src 相等 do
15                     | | continue
16                 使用 RPCsrc 将点 PGeo 算到像方并取整得 pVec[src][p] (见 2.2.1 节)
17                 if pVec[src][p] 超出 HMsrc 的范围 do
18                     | | continue
19                 if HMsrc[p] 为无效值 do
20                     | | continue
21                 使用 RPCsrc 将 pVec[src][p] 点按 HMsrc[p] 转换为 GeoVec[src][p] (见 2.2.3 节)
22                 将 GeoVec[ref][p] 转换为 PENU' (见 2.3.1 和 2.3.2 节)
23                 if ||PXYZ - PXYZ'||2 大于 τd do
24                     | | GeoVec[ref][p] 设置为无效值
25         for 每一个高度图像素 p do
26             Pfuse 初始化为 0
27             Visp 初始化为 空
28             for v = 0 to V do
29                 if GeoVec[ref][p] 为无效值 do
30                     | | continue
31                     Visp 插入 v
32                     使用 Welford 方法更新 Pfuse = Pfuse + (GeoVec[ref][p] - Pfuse) / Visp.size()
33                 if Visp.size() 不小于 τv do
34                     | | Point 插入点 Pfuse
35                     | | Visibility 插入 Visp
36                     | | for v = 1 to V do
37                         | | | if GeoVec[v][p] 非 无效值 do
38                             | | | | HWv[pVec[v][p]] 设为无效值
39     return Point, Visibility

```

对于参考视角 v_1 中一点 p_1 , 其高度为 h_1 , 使用正向 RPC 模型 $RPC_1: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 将其投影到物方空间, 得到点 P_1 ; 再将点 P_1 通过逆向 RPC 模型 $RPC_2^{-1}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 投影到视角 v_2 上, 得到点 p_2 ; 由视角 v_2 的推理高度图可知, 点 p_2 对应的高度值为 h_2 , 使用正向 RPC 模型 $RPC_2: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 将其投影到物方空间, 得到点 P_2 。若 $\|P_1 - P_2\| < \tau_d$, 则称该点 P_1 满足双目几何一致性。在多视场景中, 只有当一个点 P_1 满足至少 τ_v 视几何一致时, 才足够被认为是一个正确匹配点。由于误差的存在, 几何一致的点仅是一群相近的点, 而非理想情况下的一点。为此, 对于满足几何一致的所有点, 本文求其均值进行融合。此外, 对于已经参与融合的高度图像素点将不再参与后续融合, 以防止出现冗余点。遍历所有视角, 每一个视角都轮流作为参考视角进行多视融合, 最终生产得到稠密场景点云。

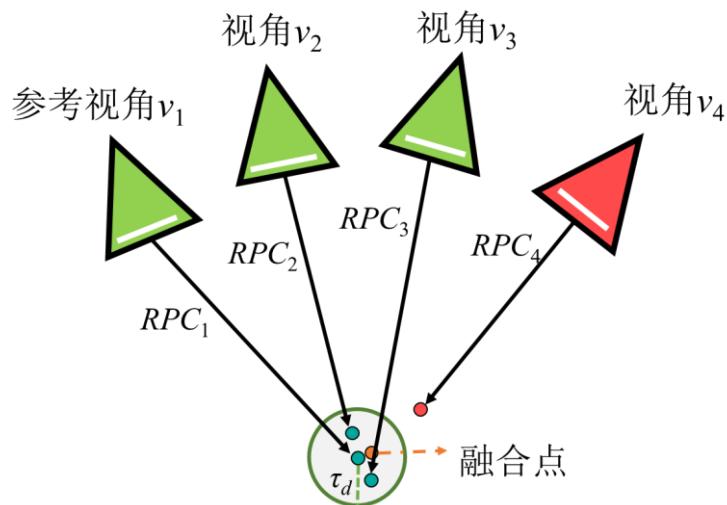


图 4-5 多视几何一致性与一致点融合

说明: 绿色代表参考视角满足双目一致性, 红色反之。

4.3 摄影测量产品生产

本文中的摄影测量产品生产指的是从稠密场景点云得到数字表面模型 (Digital Surface Model, DSM) 产品。考虑到数字表面模型 DSM 存储的是整个地表的高程起伏情况, 包括建筑物等人工建造物以及植被树木等自然地物, 因而本文采用以下方案进行 DSM 生产, 如图 4-6 所示: 首先, 将点云中的点都按其坐标位置归置到对应的 DSM 格网上; 对于每一个格网, 如果有点被置入, 则保留格网中的最大高程值作为该格网的栅格值; 而如果没有点置入, 则格网的栅格值被置为无效值 (一般是-999 或者-9999 等)。

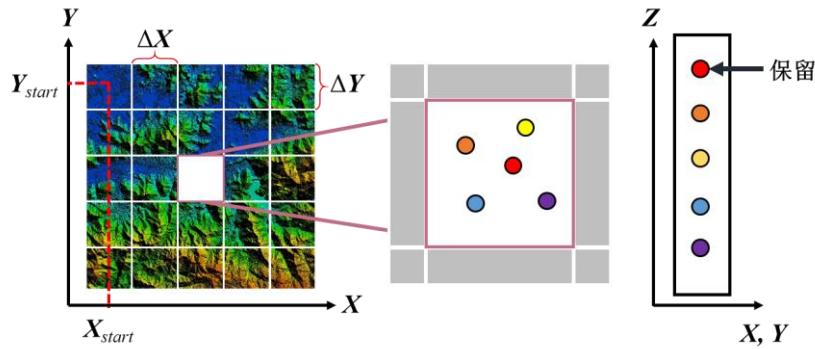


图 4-6 保留格网中的最大值以生成数字表面模型 DSM

4.4 实验分析与讨论

4.4.1 Sat-MVSF 框架在 WHU-TLC 上的性能表现

1. 参数设置

本文在 WHU-TLC 数据集上将所提出的 Sat-MVSF 框架与开源软件、主流的商业软件等进行性能上的对比，以验证基于深度学习多视密集匹配方法的 Sat-MVSF 框架其所具有的优势。Sat-MVSF 中深度学习多视密集匹配方法 Sat-MVSNet 的训练参数同 3.4.3 节。各对比方法的参数设置如下：

S2P^[74] S2P(Satellite Stereo Pipeline)是针对线阵推扫影像立体重建的一个自动化和模块化的处理管道。在处理过程中，影像被切分为小瓦片，之后被分配给多个进程进行并行处理，以提高效率。在本实验中，瓦片宽高被设定为 640 像素；视差搜索策略设定为“根据固定高程进行确定”；融合步骤中的离群点清除阈值选择为“25 米”（注：25 米约为原始影像分辨率的 10 倍，为 S2P 的建议值）；密集匹配方法是默认的 MGM 算法^[92]；将“离群点清理”选项关闭以保证较高的完整度。其他设置仍为默认设置。

SDRDIS^[93] SDRDIS 在 IARPA MVS3D 挑战赛的 Explorer 数据中获得第一名，在 Master 数据中获得第三名。其主要使用 NASA Ames Stereo Pipeline^[94]提供的工具进行卫星影像的几何处理，使用 OpenCV^[95]中提供的立体匹配算法来进行密集匹配。在实验中，由于 WHU-TLC 数据集中的 RPC 参数已提前进行了校准，所以此处禁用了光束法平差模块；密集匹配算法选为“SGBM”；最大视差设置为 288；其他参数保持默认。由于该方案中不包含 DSM 生成算法，因此在点云生成之后，按 4.3 节所述生成 DSM。

Adapted COLMAP^[75] Adapted COLMAP 将局部卫星影像瓦片的 RPC 模型用针孔相机模型进行拟合，而后使用计算机视觉界的立体重建管道进行光学立体卫星影像的三维重建。在实验中，光束法平差被禁用，其他参数保持默认设置。由于 Adapted COLMAP 本身没有能力处理大幅宽影像，本文采用 4.2.1 中所述的

多视影像裁块方案进行影像裁剪，然后送入 Adapted COLMAP 中进行处理，得到一系列 DSM 瓦片块；最终将 DSM 瓦片块进行镶嵌，得到最终的 DSM 结果。

ArcGIS^[71] 实验中使用了 ArcGIS 10.8 的 Ortho Mapping 模块。采用以下流程进行试验：创建镶嵌数据集，向镶嵌数据集中添加栅格，建立立体模型，生成点云，和点云插值。匹配方法设置为 SGM，内插方法设置为不规则三角网插值，并使用核大小为 5×5 的高斯滤波器进行平滑。其他参数被设置为默认值。

CATALYST^[70] 实验中使用了 CATALYST(PCI Geomatica)专业版的试用版。匹配方法设置为 SGM，核线重采样保持原始分辨率。其他参数保持默认。

Metashape^[72] 实验中使用了 Agisoft Metashape Professional 1.75 试用版本。为了实现高质量的重建，在影像对齐和密集匹配中都选择了“超高精度”。此外，在生成 DSM 的过程中选择“不插值”。虽然 Metashape 支持处理多视角卫星图像，但其结果存在巨大的高程偏差。因此，在实验中，只使用了后视和前视两个视角的影像进行重建。

Sat-MVSF 框架的推理设置如下。由于 WHU-TLC 中影像的定位参数已经与 DSM 对齐，本实验中仅需进行地表稠密场景重建和摄影测量产品生产。在基于多视几何一致性的误匹配剔除与匹配结果融合中， τ_d 和 τ_v 分别设置为 5m 和 2。

除了基于 MVS 的 Adapted COLMAP 和 Sat-MVSF 直接能够处理 WHU-TLC 数据集中的三线阵影像外，其他方法都以立体像对为处理单元。

2. 实验结果

在 WHU-TLC 原始数据集中的测试集中，各解决方案的性能表现如表 4-1 所示。

表 4-1 在 WHU-TLC 测试集上不同解决方案的评估

Method	MAE (m) ↓	RMSE (m) ↓	PAG _{2.5m} (%) ↑	PAG _{7.5m} (%) ↑	Time (min)
S2P	3.16	10.09	54.96	73.37	17.04
SDRDIS	4.50	15.01	47.58	73.57	9.41
Adapted COLMAP	<u>2.17</u>	<u>4.71</u>	<u>58.78</u>	76.80	77.45
ArcGIS	4.61	10.69	48.88	77.71	6.82
CATALYST	3.45	7.94	52.31	82.52	3.80
Metashape	2.69	13.05	56.59	75.46	24.51
Sat-MVSF (proposed)	1.90	3.65	64.82	<u>80.05</u>	<u>5.87</u>

说明：其中黑体为该指标中最优，而下划线斜体为该指标中次优。S2P、SDRDIS 和 Adapted COLMAP 为开源解决方案，ArcGIS、CATALYST 和 Metashape 为商业软件。

从表 4-1 中可以推导出一些结论：

(1) 在 MAE 和 RMSE 这两个精度指标上，所提出的 Sat-MVSF 框架明显优于其他解决方案，包括商业软件和开源解决方案。Sat-MVSF 的 MAE 和 RMSE

分别为 1.90 米和 3.65 米，而商业软件方案中最优的 CATAKLYST 其 MAE 和 RMSE 分别为 3.45 米和 7.94 米，最佳的开源方案 Adapted COLMAP 其 MAE 和 RMSE 分别为 2.17 米和 4.71 米；

(2)对于 PAG_{2.5m} 和 PAG_{7.5m} 这两个指标，所提出的 Sat-MVSF 框架在 PAG_{2.5m} 上获得了最好的重建精度，比次优的 Adapted COLMAP 高 6 个百分点；在 PAG_{7.5m} 上，Sat-MVSF 取得了次优的结果，比最优的 CATALYST 低 2.5 个百分点；

(3) 在效率上，所提出的 Sat-MVSF 框架显示出了良好的效率，超过了大多数的解决方案，仅比 CATALYST 软件稍慢。总而言之，在 WHU-TLC 数据集上，基于深度学习的 Sat-MVSF 框架已经展现出了令人满意的性能。

图 4-7 显示了 WHU-TLC 数据集上的可视化结果，其中白色的区域为无效重建。ArcGIS 和 CATALYST 的可视化结果具有很高的完整性。这是由于这两个软件对实际上的无效匹配区域进行了插值处理，以获得更好的可视化效果。但是这种插值操作必定会导致对重建的准确性造成损失，正如表 4-1 中这两种方案在 RMSE 指标上都偏低。

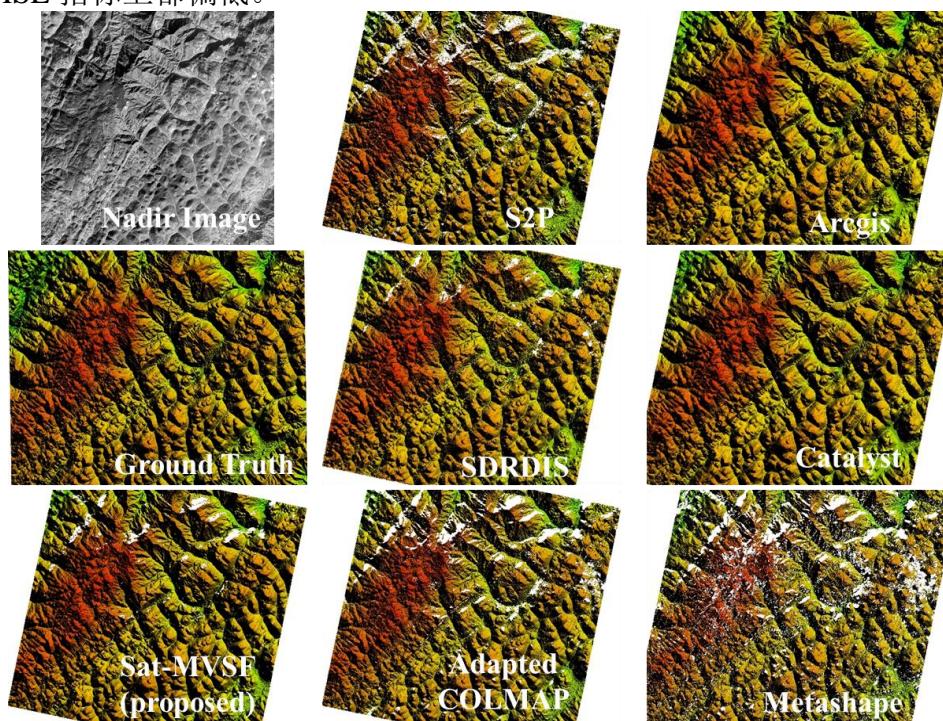


图 4-7 各解决方案在 WHU-TLC 测试集上的结果可视化

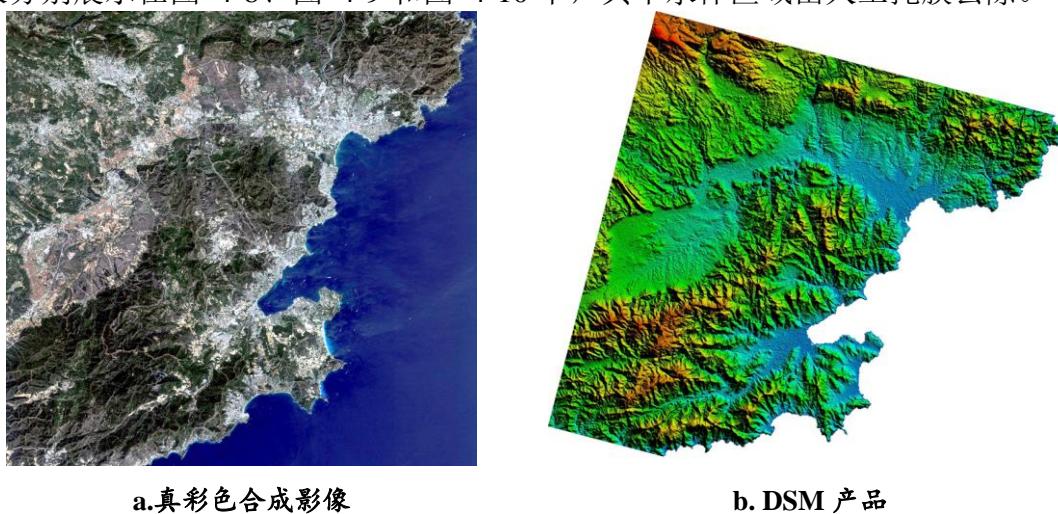
4.4.2 产品生产展示

本文还将 Sat-MVSF 框架应用于覆盖法国圣马克西姆地区、中国香港地区的 ZY3 号 01 星三线阵影像各一景以及覆盖中国山东地区的 TH-1 号卫星三线阵影像两景，以展示 Sat-MVSF 框架的实际生产能力。其中，ZY3 号 01 星三线阵影像中下视、前视和后视影像的分辨率为 2.1 米、3.5 米和 3.5 米，下视影像像素数

约为 16000×16000 ；而 TH1 号三线阵影像的分辨率为 5 米，像素数约为 12000×12000 。

对于法国圣马克西姆地区的影像数据，将其提供的控制点加入光束法平差中对三线阵卫星影像的 RPC 参数进行优化；对于香港地区的影像数据，由于缺乏控制点，引入虚拟控制点格网进行光束法平差；而本研究中所获取 TH1 号三线阵影像数据自身已经经过处理，并达到了相当的定位精度，因此不进行光束法平差；其他参数与第 4.4.1 节相同。而 Sat-MVSF 中的密集匹配网络模型 Sat-MVSNet 在 WHU-TLC 数据集中的 RPC 瓦片训练集上进行训练的。

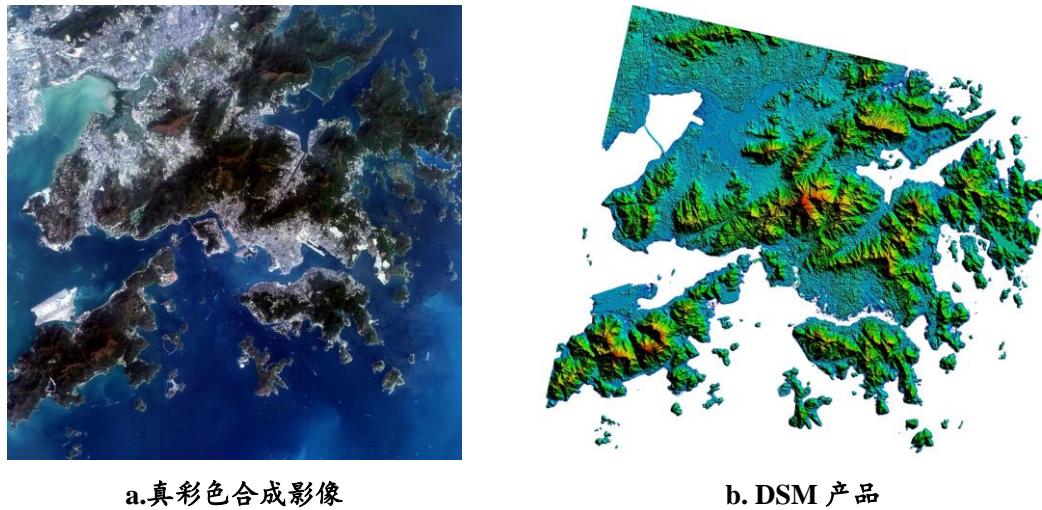
使用所提出的 Sat-MVSF 框架从三套数据中生产得到的 DSM 产品可视化结果分别展示在图 4-8、图 4-9 和图 4-10 中，其中水体区域由人工掩膜去除。



a.真彩色合成影像

b. DSM 产品

图 4-8 ZY3-01 星三线阵影像 DSM 产品（法国圣马克西姆）



a.真彩色合成影像

b. DSM 产品

图 4-9 ZY3-01 星三线阵影像 DSM 产品（中国香港）

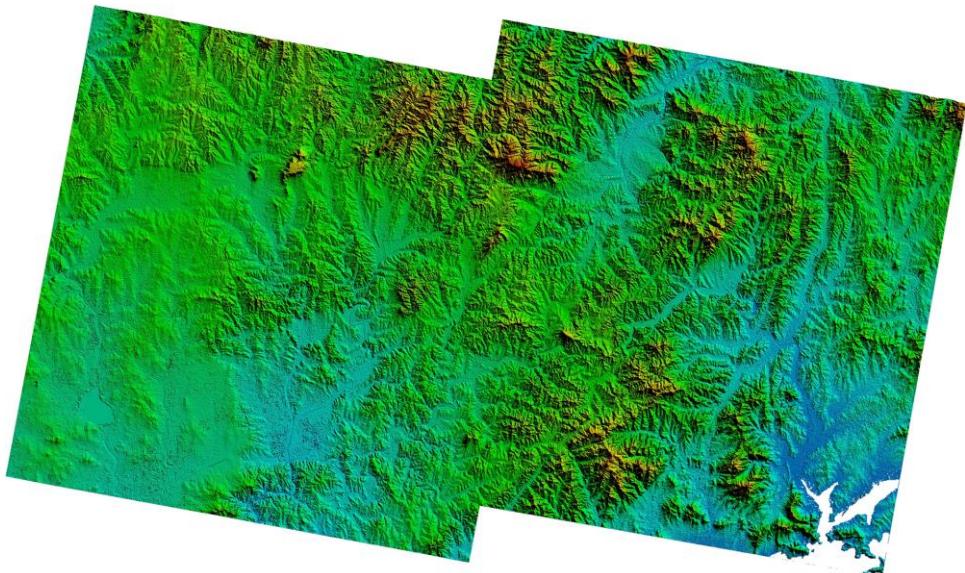


图 4-10 TH3 号三线阵影像 DSM 产品（中国山东）

4.5 本章小结

本章主要介绍如何基于第三章中的多视光学卫星影像密集匹配网络 Sat-MVSNet 搭建一套完整的光学立体卫星三维重建框架 Sat-MVSF，以用于光学立体卫星影像的三维重建实际生产任务。本章介绍了所提出的 Sat-MVSF 框架的基本处理流程，即：地表稀疏场景重建、地表稠密场景重建和摄影测量产品生产三部分，并针对大幅宽卫星影像数据的特点进行了考虑。在实验部分，本章将 Sat-MVSF 与一些商业软件和一些开源软件在 WHU-TLC 数据集上进行了比较，基本验证了基于深度学习多视密集匹配方法的三维重建框架具有实用价值，在精度和效率上展现出了一定的潜力。此外，本章还展示了 Sat-MVSF 在全球其他地区卫星影像上的生产表现，进一步证实了所提出框架的实用性。

第五章 基于自我训练的深度学习模型自优化策略

传感器类型、成像条件、成像时间和地表覆盖等方面差异导致不同的立体卫星影像之间存在着巨大的辐射差异和几何差异，这对深度学习模型的泛化能力提出了巨大的挑战。作为一种数据驱动的方法，如若能够获得具有多样性的训练数据，深度学习方法的泛化能力能得到进一步的保证。然而高精度地形数据的高保密性使得卫星影像三维重建真值标签的获取十分的困难，收集包含有各种场景的大范围卫星影像密集匹配数据集并不是一项简单的任务。卫星影像的复杂性和代表数据集的缺乏性对基于监督学习的深度学习密集匹配方法提出了巨大的挑战，使得在有限数据集上训练的模型在迁移时很有可能出现性能退化问题。那么，是否可以充分挖掘多视影像之间本身蕴含的信息，仅利用影像信息帮助网络模型完成迁移呢？为此，本文引入一种无需迁移数据真值标签参与的自优化策略，以提高所提出的光学立体卫星三维重建框架的实用性。

5.1 自优化策略

所谓的自优化策略，指的是仅利用输入数据自身的信息对网络模型进行微调，以便迁移到输入数据之上。本文所采用的自优化策略共包括预训练、推理、验证和训练四个阶段，如图 5-1 所示。

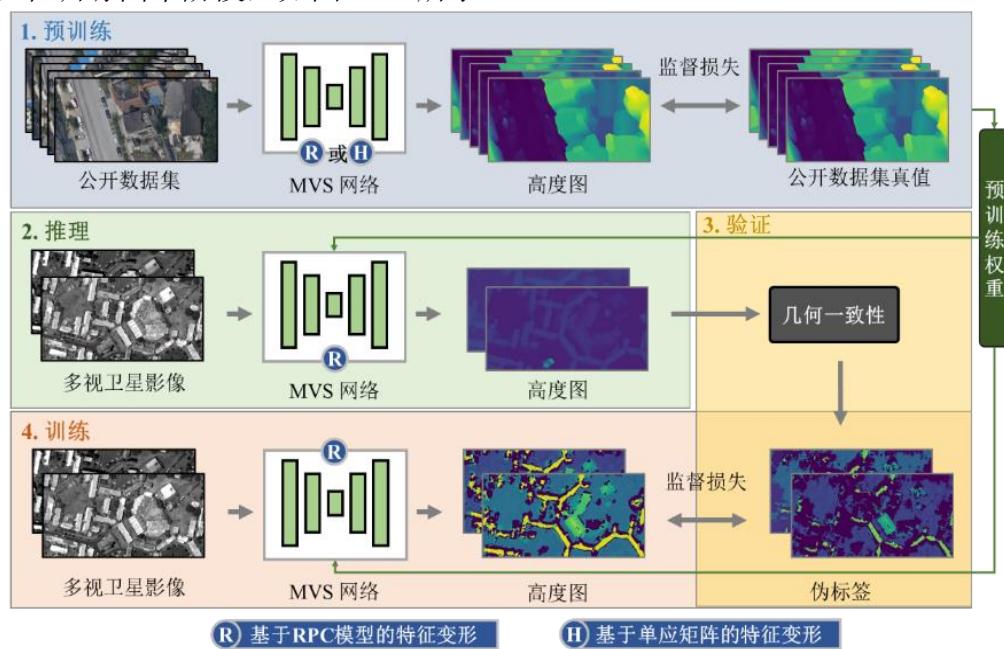


图 5-1 自优化策略

首先，使用公开可获取的数据集进行预训练，获得初始权重。若使用以中心投影为成像模型的数据进行训练，网络训练的目标如式(5-1)所示：

$$\arg \min_{\omega_0} Loss(\mathbf{Z}_H(\omega_0, \{I_{i+}, Cam_{i+}\}), HM_{gt+}) \quad (5-1)$$

其中， $Loss$ 为 3.2.6 节中定义的损失函数； \mathbf{Z}_H 为网络模型，使用 DHW 来完成特征的几何对齐； ω_0 为预训练网络权重； $\{I_{i+}, Cam_{i+}\}$ 为公开数据集中的影像和相机参数； HM_{gt+} 为公开数据集的真值标签。

随后，将预训练权重加载到网络模型中，对目标场景的多视卫星影像进行推理，如式(5-2)所示：

$$HM_0 = \mathbf{Z}_R(\omega_0, \{I_i, RPC_i\}) \quad (5-2)$$

其中， \mathbf{Z}_R 为网络模型，使用 RPC-Warping 来完成特征的几何对齐； $\{I_i, RPC_i\}$ 为目标场景的多视卫星影像和 RPC 参数， HM_0 为预训练网络模型输出的高度图。即便是使用以中心投影为成像模型的数据进行训练，预训练权重也可以直接用于多视卫星影像密集匹配任务中。这是由于 RPC-Warping 和 DHW 均是网络模型中的“硬结构”，不含任何可学习参数。

之后，使用几何一致性进行多视高度图的检验，保留正确匹配结果，生成伪标签。不同于语义分类等任务中基于概率的伪标签生成方式，多视高度图之间存在着严格的几何一致性。因此，按照 4.2.2 节中所述的方法，剔除参考影像中错误的匹配点，保留高度图中的正确匹配点，生成伪标签 $HM_{\#}$ 。

最后，使用伪标签进行监督学习，网络训练的目标如式(5-3)所示：

$$\arg \min_{\omega} Loss(\mathbf{Z}_R(\omega, \{I_i, RPC_i\}), HM_{\#}) \quad (5-3)$$

其中， ω 为网络权重。

推理—验证—训练这三个过程可以持续进行，以不断优化预训练权重，并最终收敛，从而将生产场景中的多视影像信息融入预训练模型之中，完成迁移。

5.2 实验分析与讨论

5.2.1 Sat-MVSF 与自优化策略在 MVS3D 数据上的表现

1. 数据集简介

MVS3D 数据集^[86]原本是 IARPA 多视角立体三维制图挑战赛(IARPA Multi-View Stereo 3D Mapping Challenge, IAPRA MVS3D)的比赛数据，共提供了 50 个视角的 WorldView-3 卫星影像，以及机载 LiDAR 数据生产的 DSM 作为真值。影像分辨率和真值 DSM 格网间隔均为 0.3 米；DSM 使用 UTM 投影坐标系。

然而，该数据中的 RPC 参数没有经过校准，且与 DSM 真值没有实现几何配准；此外，影像的采集时间为 2014 年 11 月至 2016 年 1 月，LiDAR 采集时间为 2016 年 6 月。这使得立体图像之间以及图像与地面实况之间存在较大的季节性差异，使得几何重建任务变得极具挑战性。

2. 自优化设置

本实验采用第 5.1 节中提出的自优化策略：首先，在 WHU-MVS 多视航空密集匹配数据集的训练集上^[38]将 Sat-MVSNet 中的 RPC-Warping 切换为 DHW 进行 30 个轮次的训练，获得一个预训练模型；之后，将 DHW 切换回 RPC-Warping，加载预训练模型，使用 Sat-MVSNet 预测 MVS3D 数据集中影像的高度图；然后，使用阈值 $\tau_d = 0.5$ 与 $\tau_v = 1$ 对推理的高度图进行几何一致性检查，以便从推理得到的高度图中生成伪标签；最后，将多视影像和其对应伪标签作为训练样本，在预训练模型的基础上继续训练一个轮次，以实现自优化。

3. 参数设置

为了进行公平的比较，所有的对比方案都进行了相同的预处理和后处理流程。跟随^[96]的工作，挑选出前五对影像进行实验，具体为：首先，由 50 张影像构成 1225 对影像对，计算影像对的交会角 α 和视角最大入射角 β ，仅保留 α 大于 5° 且小于 45° 和 β 小于 40° 的像对；之后，按照影像对中两影像的获取时间差进行从小到大的排序，取排序后的前五对进行实验。对于所取的每一对影像对，按照 4.1.4 节所述进行光束法平差以提升相对定位精度。由于真值范围有限，采用官方提供的影像裁剪方案^[86]裁剪出能够覆盖真值区域的影像块，以提高处理效率；实验中，影像块的尺寸为 3072×3072 像素。在所有的方法中，点云结果被导出以便与真值进行几何配准。而 CATALYST 软件无点云结果的导出接口，因此，在其生成 DSM 时，将 DSM 格网采样率设置为影像分辨率，再将其转化为点云。然后，使用官方提供的点云—格网配准工具^[86]来实现点云结果和 DSM 真值之间的对齐，并按照 4.3 节所述生成 DSM。最后，将多个像对的 DSM 进行融合，融合的方法参考^[75]。

每种方案的参数设置如下，其中未提及的参数保持默认。在 CATALYST^[70]中，选择 SGM 进行密集匹配；为了提高匹配质量，核线重采样的缩放比例设置为 1。在 Metashape^[72]中，在点云生成模块中选择“超高精度”选项以获得更好的重建效果。在 SDRDDIS^[93]、JHU/APL（比赛官方提供的基准方案）^[86]和 Adapted COLMAP^[75]中，禁用其光束法平差模块。在所提出的 Sat-MVSF 框架中，视角数为 2；通过将 SRTM 的高程值扩展一定范围以确定第一阶段的搜索范围；对于几何一致性检查，阈值 τ_d 与 τ_v 分别被设置为 0.5 和 1。

3. 实验结果

实验结果如表 5-1 所示。需要强调的是，此处的评价指标沿用了 IAPRA MVS3D 比赛的指标 PAG_{1.0m} 和 RMSE，而高程误差的中值(Median)则是点云和 DSM 几何配准的优化目标函数，也一并在表格中列出。

表 5-1 各方案在 MVS3D 数据集上的精度比较

解决方案		CATALYST	Metashape	S2P	SDRDIS	JHU/APL	Adapted COLMAP	Sat-MVSF (SR)	Sat-MVSF (WHU-MVS)
所有 Site 均值	PAG _{1.0m} (%) ↑	58.92	56.73	<u>59.49</u>	56.67	55.19	50.38	60.48	51.09
	Median(m) ↓	0.77	0.50	<u>0.40</u>	0.50	0.88	0.37	<u>0.40</u>	0.37
	RMSE(m) ↓	4.32	<u>3.46</u>	4.78	4.166	4.90	8.40	3.24	2.96
Site1	PAG _{1.0m} (%) ↑	72.31	67.61	74.42	69.82	68.09	63.21	<u>72.51</u>	51.44
	Median(m) ↓	0.35	0.28	0.24	0.30	0.51	0.26	<u>0.26</u>	0.28
	RMSE(m) ↓	2.91	2.50	2.42	2.772	3.16	3.47	<u>2.49</u>	2.08
Site2	PAG _{1.0m} (%) ↑	64.57	65.65	70.46	63.82	61.91	55.64	<u>70.07</u>	69.15
	Median(m) ↓	0.55	0.40	<u>0.35</u>	0.53	0.66	0.26	0.38	0.35
	RMSE(m) ↓	2.04	<u>1.87</u>	1.84	2.04	2.18	5.46	1.89	1.78
Site3	PAG _{1.0m} (%) ↑	58.92	54.22	55.06	57.65	54.05	46.7	<u>58.47</u>	50.34
	Median(m) ↓	0.67	0.51	0.38	0.40	0.83	0.32	<u>0.34</u>	0.34
	RMSE(m) ↓	4.31	3.90	<u>3.87</u>	3.91	4.58	9.60	3.34	3.12
Site4	PAG _{1.0m} (%) ↑	43.86	38.68	40.16	41.37	41.94	28.83	<u>42.03</u>	24.4
	Median(m) ↓	1.47	0.72	<u>0.53</u>	<u>0.53</u>	1.53	0.60	0.44	0.39
	RMSE(m) ↓	10.32	<u>7.21</u>	12.87	7.84	11.75	19.14	7.09	5.96
Site5	PAG _{1.0m} (%) ↑	65.58	66.01	70.48	63.65	61.73	64.22	<u>70.28</u>	70.04
	Median(m) ↓	0.55	0.41	<u>0.38</u>	0.55	0.66	0.35	0.40	0.37
	RMSE(m) ↓	2.13	<u>1.82</u>	1.77	2.06	2.24	2.78	1.89	1.77
Site6	PAG _{1.0m} (%) ↑	63.32	62.85	67.47	61.17	57.44	60.88	<u>66.99</u>	66.23
	Median(m) ↓	0.58	0.45	0.40	0.57	0.74	0.38	0.43	0.40
	RMSE(m) ↓	2.61	2.11	<u>2.10</u>	2.52	2.63	3.29	2.27	2.09
Site7	PAG _{1.0m} (%) ↑	42.26	<u>44.66</u>	44.38	40.91	42.16	37.02	46.59	33.41
	Median(m) ↓	1.36	0.75	<u>0.56</u>	0.77	1.31	0.48	<u>0.56</u>	0.45
	RMSE(m) ↓	6.19	<u>4.65</u>	6.35	<u>4.65</u>	8.16	13.19	3.68	3.27
Site8	PAG _{1.0m} (%) ↑	60.50	54.14	53.50	54.95	54.19	46.51	<u>56.88</u>	43.71
	Median(m) ↓	0.62	0.44	0.38	0.37	0.83	0.32	<u>0.36</u>	0.36
	RMSE(m) ↓	4.08	<u>3.66</u>	7.00	7.54	4.47	10.25	3.30	3.59

说明：Sat-MVSF(WHU-MVS)指的是使用 WHU-MVS 航空多视影像密集匹配训练集中训练的权重；而 Sat-MVSF(RA)代表的是 WHU-MVS 预训练模型进行自优化后的权重。S2P、SDRDIS、JHU/APL 和 Adapted COLMAP 为开源解决方案，CATALYST 和 Metashape 为商业软件。加粗为该指标上的最优，下划线斜体为该指标上的次优。

从表 5-1 中可以得出一些结论：

(1) 针对深度学习多视密集匹配方法而言，在航空数据集上训练得到的预训练模型可以直接用于卫星影像。就总体表现而言，使用 WHU-MVS 航空数据集上预训练的模型进行直接迁移，虽然由于数据差异过大出现了模型性能退化问题，但是还是能够达到超过商业软件和开源解决方案的 RMSE 精度，且在其他指标上达到与其他方案相近的水平。这表明多视密集匹配网络模型从数据中学习的知识是足够通用的，这对实际生产应用具有重要意义。

(2) 基于深度学习多视密集匹配方法的 Sat-MVSF 在自优化策略的加持下，即便是从航空数据集迁移到卫星数据集上，也能够达到不错的效果。就总体表现上，自优化后的 Sat-MVSF 在 PAG_{1.0m} 和 RMSE 指标上都超过了商业软件和开源解决方案。就 PAG_{1.0m} 这一指标，在商业软件和开源解决方案中，S2P 最优，其

结果展现出了最好的完整性，接近于自优化后的 Sat-MVSF 框架；而 Metashape 则取得了最好的 RMSE 精度，稍逊于 Sat-MVSF 框架。

总而言之，在没有与目标数据集相近的真值存在时候，所提出的自优化策略可以仅通过预训练模型的通用知识和目标影像本身信息，来提高性能表现，从而在一定程度上增强了 Sat-MVSF 框架的实用性。

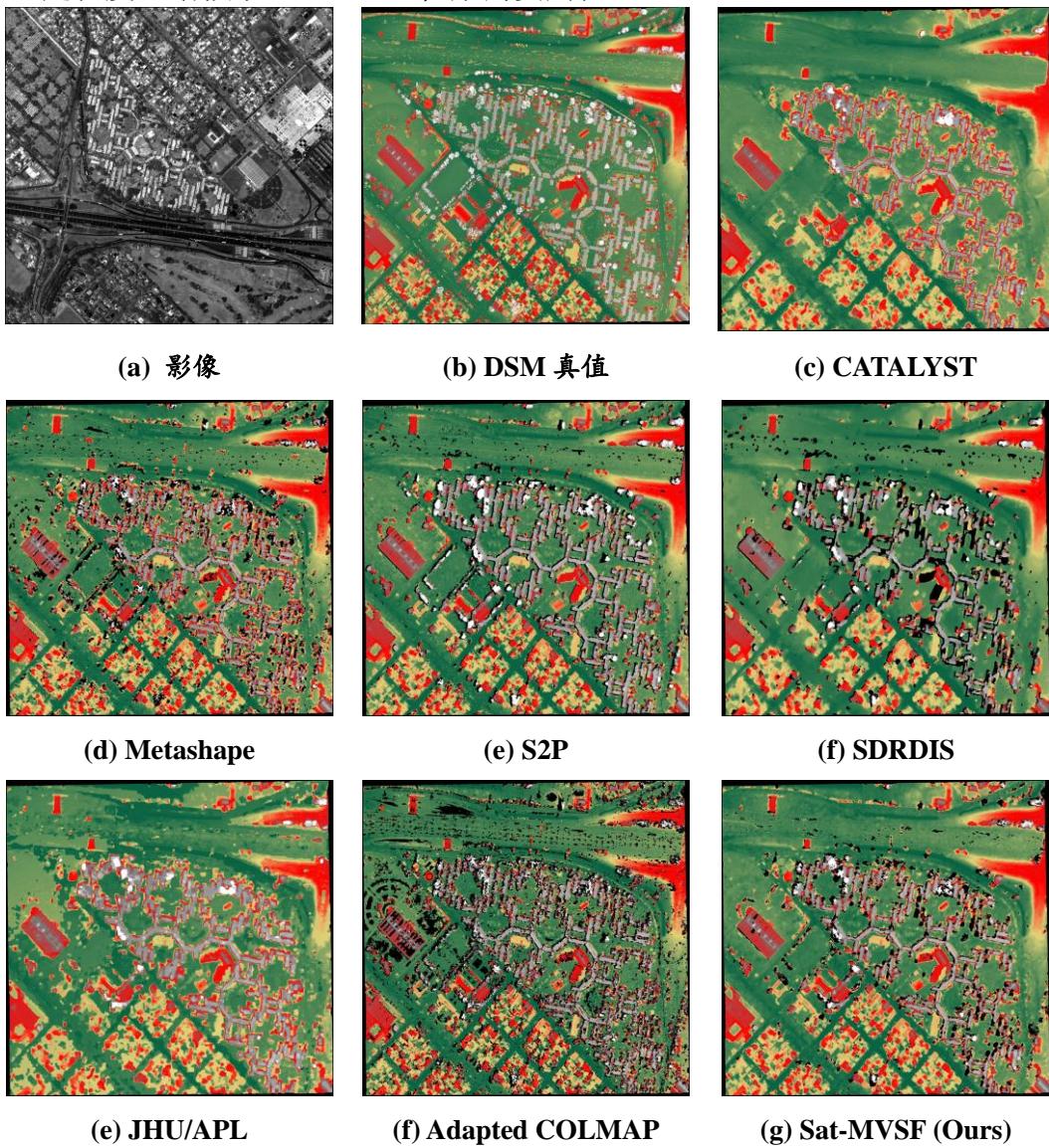


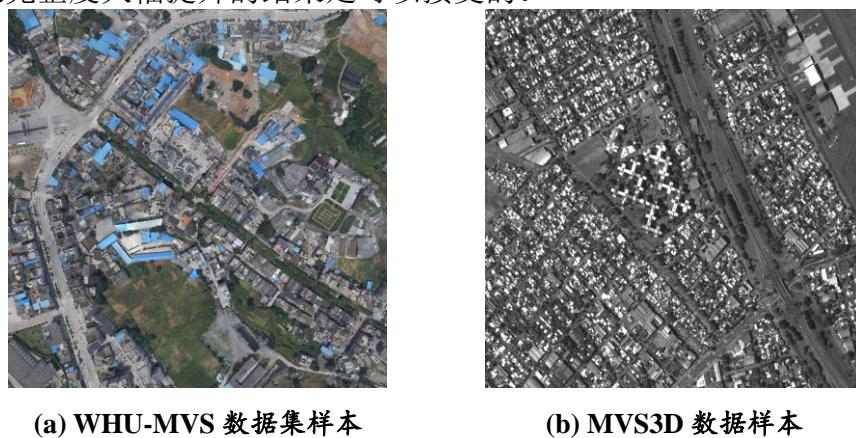
图 5-2 各方案在 MVS3D 数据集上的 DSM 产品可视化

图 5-2 中展示了各方案在 MVS3D 数据集 Site1 上的 DSM 产品。所有的方案都能够恢复该地区的基本地形轮廓。从可视化结果上看，Adapted COLMAP 中含有一些不少的噪声，是表 5-1 中其 RMSE 精度低的主要原因；CATALYST 和 JHU/APL 的结果在建筑物边缘都较为平滑，边缘不够清晰；Metashape 和 SDRDIS 存在较多的空洞；S2P 和 Sat-MVSF 在完整度上最高，但 S2P 可视化效果稍优于所提出的 Sat-MVSF，与表 5-1 中 Site1 上的定量评价结果一致。

4. 讨论

本节中还留有两个问题需要讨论：（1）为什么自优化后 Sat-MVSF 的重建 RMSE 精度下降了？（2）为什么不使用 WHU-TLC 卫星多视密集匹配数据集上的预训练模型进行自优化？以下围绕这两个问题展开讨论。

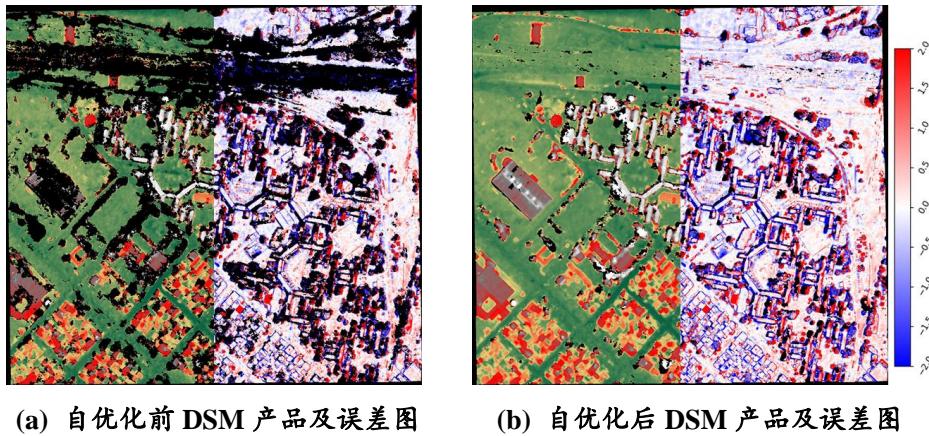
针对第一个问题，虽然从表 5-1 能看出航空预训练模型直接迁移到卫星多视卫星上能达到一定的精度水平，但是 WHU-MVS 航空数据集与目标数据集 MVS3D 之间存在着如视角数、光谱特征、交会角等上的巨大差异，如图 5-3 a 和图 5-3 b 所示。这种差异使得 WHU-MVS 航空数据集上预训练的模型出现不适应和模型性能退化的问题，导致生产的 DSM 结果中含有大量空洞，如图 5-4 a 所示。而自优化通过挖掘多视卫星影像之间的信息，将预训练权重中蕴含的通用知识适应到多视卫星影像密集匹配中，将完整度提升了 9.39 个百分点，得到了更完整的 DSM 产品，如图 5-4 b 所示。然而，这些补充的点中不可避免存在些许错误，导致了 RMSE 精度略微下降了 0.28m。但，总体而言，这种用少量 RMSE 精度换取完整度大幅提升的结果是可以接受的。



(a) WHU-MVS 数据集样本

(b) MVS3D 数据样本

图 5-3 WHU-MVS 数据集样本与 MVS3D 数据样本对比图



(a) 自优化前 DSM 产品及误差图

(b) 自优化后 DSM 产品及误差图

图 5-4 Sat-MVSF 自优化前后的 DSM 产品可视化

针对第二个问题,使用 WHU-TLC 数据集训练得到的预训练模型能够得到正确但是过于平滑的重建结果,如图 5-5 (c)所示。从理论上分析,WHU-TLC 预训练模型生产的 DSM 中的点都经过了几何一致性检验,即匹配正确;然而由于 WHU-TLC 数据集中的影像分辨率为 2.5m,远远大于目标影像的分辨率 0.3m,因而预训练模型对高分辨率精细结构特征的捕捉能力不足,导致边缘和细微结构不够精细。继而,如果使用 WHU-TLC 预训练模型作为初始模型进行自优化,则会导致生成的伪标签将过于平滑,进而将这种过度平滑的不精确信息传递给网络模型,无法达到精细的重建。相比之下,在航空多视密集匹配数据集 WHU-MVS 上训练的预训练模型,其生成的 DSM 虽然完整度偏低,如图 5-4 (a)所示,但是其有效重建区域细节更为丰富,如图 5-5 (d) 所示。

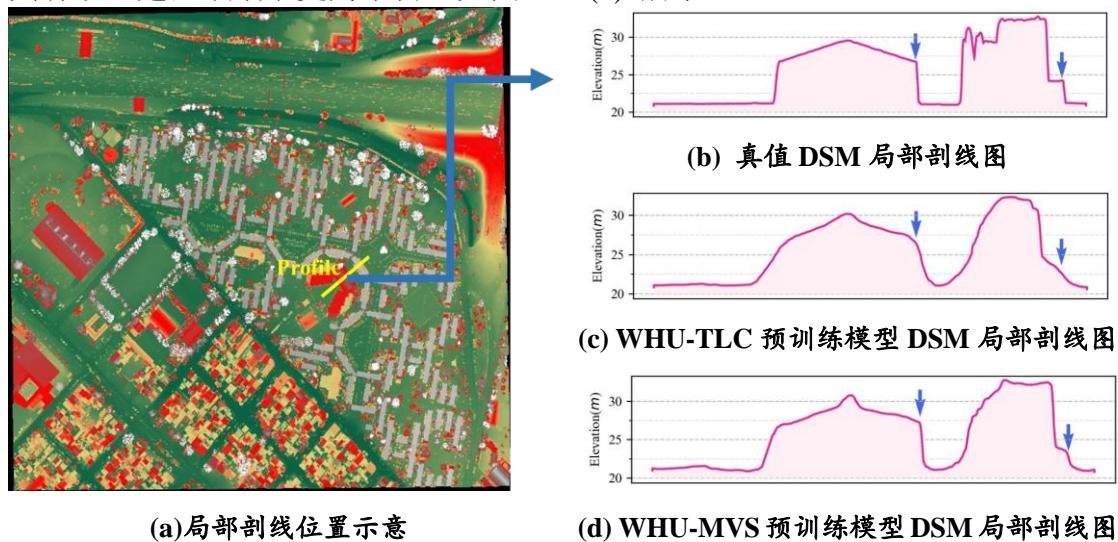


图 5-5 不同预训练模型生产的 DSM 局部剖面图

5.2.2 自优化精度与自优化轮次之间的关系

所采用的自优化策略虽然不需要任何目标数据集的真实标签的参与,但是并非没有成本。自优化策略可以总结为预训练、推理、伪标签生成和再训练四个部分。若假设预训练模型已经存在,可无时间成本获取,那其他三个部分执行一个轮次所需花费的时间就已经接近 2 小时。因而,进一步探究自优化轮次对精度的影响对于实用性验证而言是必要的。

基于在 WHU-MVS 数据集上训练得到的预训练模型,本实验在 MVS3D 数据集上进行了进一步的实验。在实验中,执行多轮次的推理—验证—训练;之后,在每一个轮次结束时,利用该轮次的模型进行 DSM 生产和精度评估。实验结果如图 5-6 (a)所示,其中精度指标与 5.2.1 节保持一致。

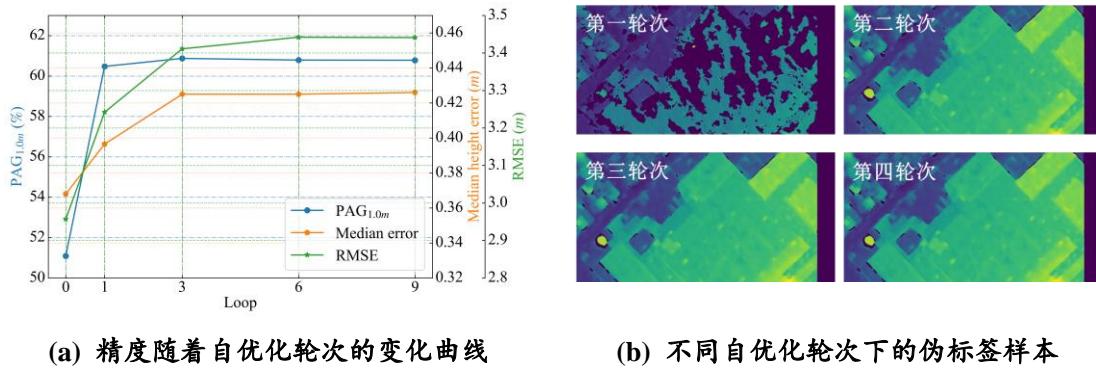


图 5-6 自优化策略执行不同轮次时发生的变化

根据图 5-6 (a)，第一个轮次就带来了 $PAG_{1.0m}$ 的大幅提升；而在第一个轮次之后，从 $PAG_{1.0m}$ 指标的变化上看，完整度得到小幅度的提升，并在第三个轮次之后，几乎不再变化。而根据图 5-6 (b)，伪标签样本也仅仅是在第一个轮次的优化中得到完整度的大幅度提升，之后几乎不变。这说明，自优化策略只需要执行一到两个轮次，就能适应目标影像数据，进一步增强了自优化策略的实用性。

5.2.3 成对处理和真多视处理方式下的性能对比

按照^[83]的划分，MVS 可以分为成对(pair-wise)处理方式以及真多视(true MVS)处理方式。在 5.2.1 节中，只有 Sat-MVSF 和 Adapted-COLMAP 可以在真多视处理模式下工作，即使用多视影像作为输入，而不是一对立体影像。本节进一步比较了这两种方法在真多视处理模式下的性能。本节使用 5.2.1 节中相同的参数设置和自优化后的网络权重。实验结果如表 5-2，DSM 产品可视化于图 5-7。

表 5-2 成对处理方式与真多视处理方式下的精度对比

Method	Mode	$PAG_{1.0m}$ (%)↑	Median(m)↓	RMSE(m)↓
Adapted-COLMAP	Pair-wise	50.38	0.37	8.40
	true MVS	55.35	0.40	4.32
Sat-MVSF	Pair-wise	60.48	0.40	3.24
	true MVS	58.01	0.50	2.98

根据表 5-2，与成对处理相比，使用真多视处理的方式使 Adapted COLMAP 的 PAG 指标提高了 5.0%；使 Sat-MVSF 的 PAG 指标降低了 2.5%；使 Adapted-COLMAP 的 RMSE 指标提高了 48.5%；使 Sat-MVSF 的 RMSE 指标提高了 8.0%。Adapted-COLMAP 在各项指标上的巨大提升源于它采用的逐像素视角选择算法^[36]，能够为每个像素找到适合匹配的最佳多视角影像集，避免因噪声、遮挡等造成的图像间的一些不一致信息，并最大限度地发挥多视角信息注入的优势。而 Sat-MVSF 没有这样的像素级视角选择机制，对多视图信息同等对待，导致了 PAG 指标略有下降。然而，即使在 MVS 模式下，Sat-MVSF 在 PAG 和 RMSE

指标上仍优于 Adapted-COLMAP。此外，从图 5-7 看出，多视角的信息注入进一步提升了 DSM 结果的完整度。

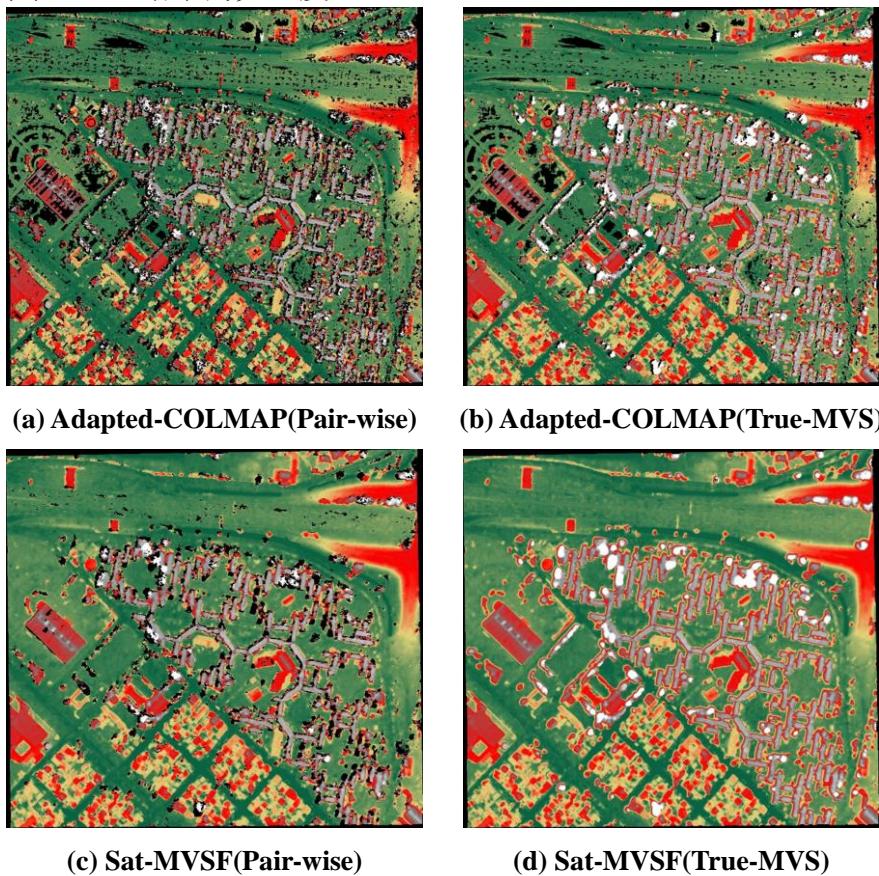


图 5-7 成对处理方式与真多视处理方式下的 DSM 产品可视化

总体而言，所提出的 Sat-MVSF 框架不仅在成对处理方式下表现良好，而且在 MVS 模式下也能展现出优势，表现出了较高的适用性。

5.3 本章小结

为缓解深度学习方法在直接应用预训练模型时性能退化问题，本章提出一种基于自我训练的自优化策略，在不需要目标数据任何真值标签的情况下帮助网络模型完成迁移。所提出的自优化策略包括预训练、推理、验证和训练四个阶段，其中预训练模型在公开数据集上进行训练，推理、验证和训练过程中仅有目标数据中的多视影像参与。在实验中，本章使用航空影像上的预训练模型与提出的自优化策略在 MVS3D 数据集上达到了略优于商业软件和开源解决方案的结果，证明了自优化策略的有效性；之后，进一步探究了自优化精度随自优化轮次的变化趋势，实验表明一个自优化轮次就足以将预训练模型适应到目标数据上，实用性进一步得到展示；最后，对成对处理和真多视处理方式下所提出的 Sat-MVSF 框架其性能的研究进行了分析，证明了多视信息能够提高 Sat-MVSF 框架的表现。

第六章 总结与展望

6.1 全文总结

本文研究的关注点位于将深度学习多视密集匹配网络引入到基于多视光学卫星影像的三维重建任务中。而将深度学习多视密集匹配方法运用到光学立体卫星三维重建实际生产任务中，面临着一些问题：其一，RPC 模型与中心投影成像模型存在差异，缺乏一种基于 RPC 模型的高维特征变形方法，在深度学习网络内部实现卫星影像的特征对齐；其二，为了从光学立体卫星影像重建三维地形，除了密集匹配网络外，还需要完善其他必要模块，以及考虑如何对大幅宽卫星影像进行处理等问题；其三，生产场景中的数据没有相近的真值以进行模型精化，所使用的深度学习方法可能出现性能退化问题。针对以上问题，本文完成了以下工作：

(1) 设计基于深度学习的多视光学卫星影像密集匹配网络 **Sat-MVSNet**，验证了其有效性。将 RPC 模型表达成张量运算，实现基于 RPC 模型的高维特征变形方法(RPC-Warping)，并搭建基于深度学习的多视光学卫星影像密集匹配网络 Sat-MVSNet，成功将深度学习多视密集匹配方法引入到光学卫星影像中。此外，制作并公布了一套三线阵多视角密集匹配数据集 WHU-TLC，以进行网络的训练和评估。实验证明，相比于将 RPC 模型拟合成中心投影成像模型的处理方式，基于 RPC-Warping 的网络模型更具优势。

(2) 构建基于深度学习多视密集匹配方法的光学立体卫星影像三维重建框架 **Sat-MVSF**，验证了其生产能力。围绕多视卫星影像密集匹配网络 Sat-MVSNet，本文完善了稀疏场景重建、摄影测量产品生产等必要组件，并考虑卫星影像大幅宽的特点，构建了一套完整的多视光学立体卫星影像三维重建框架 Sat-MVSF。实验证明相比于现有的商业软件和开源解决方案，Sat-MVSF 在精度和效率上都展现出了一定的潜力，具有一定的实用价值。

(3) 使用自优化的方式在无真值标签监督的情况下提升网络在实际生产数据上的性能，进一步增强 **Sat-MVSF** 框架的实用性。不同的立体卫星影像之间存在着巨大的辐射差异和几何差异，对模型的泛化能力提出了巨大的挑战。而真实应用场景中，与目标场景相近的真值数据往往难以获取。本文引入一种无需迁移数据的真值标签参与的自优化策略来帮助网络模型完成迁移。实验中，本文使用提出的自优化策略在 MVS3D 数据集上能达到略优于商业软件和开源解决方案的结果，证明了自优化策略的有效性。

6.2 研究展望

本文成功地将深度学习多视密集匹配方法与光学立体卫星影像结合,设计了多视卫星密集匹配网络,搭建了基于深度学习多视密集匹配方法的光学立体卫星影像三维重建框架,并提出自优化策略以应对实际生产情况。然而,本文的研究仍存有一些不足,有待进一步深入研究:

(1) 除了现采用的自优化策略外,无监督的网络训练方法也是一种摆脱高精度真值束缚的方案。探究自优化策略和无监督训练两者之间的优劣,也是值得研究的课题。未来的研究还可以探索如何在实际应用中灵活地选择或结合这两种方法,以更好地应对高精度真值数据缺乏的挑战,并提高深度学习网络模型的精度和鲁棒性。

(2) 目前采用的深度学习方法在三维重建中主要专注于密集匹配过程本身,但语义信息的注入却存在不足,线面约束也未能充分考虑。深度学习在语义信息提取方面具有极大的潜力,如何将深度学习中获得的语义信息有效地注入密集匹配过程,使得生成的三维重建结果更加符合真实世界中的几何规律,例如使直线更加直、平面更加平,成为三维重建智能化的下一步重要工作。

(3) 当前,现有三维重建框架的产品仅限于 DSM,但随着卫星影像分辨率不断提高,这种 2.5D 的产品已经无法完全满足应用需求。未来,卫星影像三维重建技术将会向无人机倾斜摄影测量技术靠齐,进入实景三维模型时代。这将会带来更丰富、更真实的三维地理信息,为地图制图、城市规划、遥感监测等领域带来更好的服务。

参考文献

- [1] 陈韬亦, 彭会湘, 王彬. 遥感卫星灾害应急监测效能评估指标体系构建[J]. 计算机测量与控制, 2020, 28(9): 256-261.
- [2] 姜景山. 中国对地观测技术发展现状及未来发展的若干思考[J]. 中国工程科学, 2006(11): 19-24.
- [3] 付伟, 刘晓丽, 宋世杰, 等. 面向应急常态化的卫星任务管控模式研究[J]. 无线电网, 2021, 51(5): 407-410.
- [4] 李德仁, 王密, 沈欣, 等. 从对地观测卫星到对地观测脑[J]. 武汉大学学报(信息科学版), 2017, 42(2): 143-149.
- [5] 江碧涛. 我国空间对地观测技术的发展与展望[J]. 测绘学报, 2022, 51(7): 1153-1159.
- [6] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.
- [8] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [9] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [12] Szegedy C, Wei Liu, Yangqing Jia, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.
- [13] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [14] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[C]//Chaudhuri K, Salakhutdinov R. Proceedings of the 36th International Conference on Machine Learning: Vol. 97. 2019: 6105-6114.
- [15] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]//International Conference on Learning Representations (ICLR). 2015: 1-14.
- [16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic

- segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 3431-3440.
- [17] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention. 2015: 234-241.
- [19] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [20] Scharstein D, Szeliski R, Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[C]//Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001). 2001: 131-140.
- [21] Geiger A, Lenz P, Urtasun R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite[C]//Conference on Computer Vision and Pattern Recognition (CVPR). 2012: 3354-3361.
- [22] Jensen R, Dahl A, Vogiatzis G, et al. Large scale multi-view stereopsis evaluation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 406-413.
- [23] Knapitsch A, Park J, Zhou Q Y, et al. Tanks and temples: Benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics (ToG), 2017, 36(4): 1-13.
- [24] Kendall A, Martirosyan H, Dasgupta S, et al. End-to-End Learning of Geometry and Context for Deep Stereo Regression[C]//2017 IEEE International Conference on Computer Vision (ICCV). 2017: 66-75.
- [25] Chang J R, Chen Y S. Pyramid Stereo Matching Network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 5410-5418.
- [26] Zhang F, Prisacariu V, Yang R, et al. GA-Net: Guided Aggregation Net for End-To-End Stereo Matching[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 185-194.
- [27] Guo X, Yang K, Yang W, et al. Group-Wise Correlation Stereo Network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3268-3277.
- [28] Duggal S, Wang S, Ma W C, et al. DeepPruner: Learning Efficient Stereo Matching via Differentiable PatchMatch[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 4383-4392.
- [29] Liu B, Yu H, Long Y. Local Similarity Pattern and Cost Self-Reassembling for Deep Stereo Matching Networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 36. 2022: 1647-1655.
- [30] Ding Y, Yuan W, Zhu Q, et al. Transmvsnet: Global context-aware multi-view stereo network

- with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 8585-8594.
- [31] Ma X, Gong Y, Wang Q, et al. EPP-MVSNet: Epipolar-assembling based Depth Prediction for Multi-view Stereo[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 5712-5720.
- [32] Gu X, Fan Z, Zhu S, et al. Cascade cost volume for high-resolution multi-view stereo and stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2495-2504.
- [33] Wei Z, Zhu Q, Min C, et al. Aa-rmvsnet: Adaptive aggregation recurrent multi-view stereo network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6187-6196.
- [34] Zhang J, Li S, Luo Z, et al. Vis-MVSNet: Visibility-Aware Multi-view Stereo Network[J]. International Journal of Computer Vision, 2022: 1-16.
- [35] Hirschmüller H. Stereo processing by semiglobal matching and mutual information[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 30(2): 328-341.
- [36] Schönberger J L, Zheng E, Frahm J M, et al. Pixelwise View Selection for Unstructured Multi-View Stereo[C]//European conference on computer vision. 2016: 501-518.
- [37] Li S, He S, Jiang S, et al. WHU-Stereo: A Challenging Benchmark for Stereo Matching of High-Resolution Satellite Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.
- [38] Liu J, Ji S. A novel recurrent encoder-decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 6050-6059.
- [39] 刘瑾, 季顺平. 基于深度学习的航空遥感影像密集匹配[J]. 测绘学报, 2019, 48(9): 1141-1150.
- [40] Mayer N, Ilg E, Hausser P, et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4040-4048.
- [41] Schöps T, Schönberger J L, Galliani S, et al. A Multi-view Stereo Benchmark with High-Resolution Images and Multi-camera Videos[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2538-2547.
- [42] Zbontar J, LeCun Y. Computing the Stereo Matching Cost with a Convolutional Neural Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1592-1599.
- [43] Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: Learning Optical Flow with Convolutional Networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 2758-2766.
- [44] Li J, Wang P, Xiong P, et al. Practical Stereo Matching via Cascaded Recurrent Network with

- Adaptive Correlation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 16242-16251.
- [45] Lipson L, Teed Z, Deng J. RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching[C]//2021 International Conference on 3D Vision (3DV). 2021: 218-227.
- [46] Teed Z, Deng J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow[C]//Vedaldi A, Bischof H, Brox T, et al. Proceedings of the European conference on computer vision (ECCV). 2020: 402-419.
- [47] Yao Y, Luo Z, Li S, et al. Mvsnet: Depth inference for unstructured multi-view stereo[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 767-783.
- [48] Yao Y, Luo Z, Li S, et al. Recurrent MVSNet for High-Resolution Multi-View Stereo Depth Inference[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 5520-5529.
- [49] Cheng S, Xu Z, Zhu S, et al. Deep Stereo Using Adaptive Thin Volume Representation with Uncertainty Awareness[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 2521-2531.
- [50] Wang F, Galliani S, Vogel C, et al. PatchmatchNet: Learned Multi-View Patchmatch Stereo[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14194-14203.
- [51] Tao C, Hu Y. A Comprehensive Study of the Rational Function Model for Photogrammetric Processing[J]. Photogrammetric Engineering and Remote Sensing, 2001, 67: 1347-1357.
- [52] Fraser C S, Hanley H B. Bias compensation in rational functions for IKONOS satellite imagery[J]. Photogrammetric Engineering & Remote Sensing, 2003, 69(1): 53-57.
- [53] Grodecki J. IKONOS stereo feature extraction-RPC approach[C]//ASPRS annual conference St. Louis: Vol. 7. 2001.
- [54] Paderes F. Batch and on-line evaluation of stereo SPOT imagery[C]//1989 ASPRS/ACSM Annual Convention. 1989: 31-40.
- [55] Ahn C H, Cho S I, Jeon J C. Ortho-rectification software applicable for IKONOS high resolution images: GeoPixel-Ortho[C]//International Geoscience and Remote Sensing Symposium: Vol. 1. 2001: 555-557.
- [56] Toutin T. Geometric processing of IKONOS Geo images with DEM[C]//ISPRS Joint Workshop High Resolution from Space. 2001: 19-21.
- [57] Kim T. A Study on the Epipolarity of Linear Pushbroom Images[J]. Photogrammetric Engineering and Remote Sensing, 2000, 66: 961-966.
- [58] Habib A F, Morgan M F, Jeong S, et al. Epipolar Geometry of Line Cameras Moving with Constant Velocity and Attitude[J]. 2005, 27(2): 172-180.
- [59] Wang M, Hu F, Li J. Epipolar Resampling of Linear Pushbroom Satellite Imagery by A New Epipolarity Model[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 347-355.

- [60] Tatar N, Saadatseresht M, Arefi H, et al. Quasi-epipolar Resampling of High Resolution Satellite Stereo Imagery for Semi Global Matching[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015, XL-1/W5: 707-712.
- [61] Liao P, Chen G, Zhang X, et al. A Linear Pushbroom Satellite Image Epipolar Resampling Method for Digital Surface Model Generation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 56-68.
- [62] Eisank C, Rieg L, Klug C, et al. Semi-Global Matching of Pléades Tri-Stereo Imagery to Generate Detailed Digital Topography for High-Alpine Regions[C]//International Journal of Geographical Information Science: Vol. 1. 2015: 168-177.
- [63] Ghuffar S. Satellite Stereo Based Digital Surface Model Generation Using Semi Global Matching in Object and Image Space[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, III-1: 63-68.
- [64] Gong K, Fritsch D. DSM Generation from High Resolution Multi-View Stereo Satellite Imagery[J]. Photogrammetric Engineering & Remote Sensing, 2019, 85(5): 379-387.
- [65] Rothermel M, Wenzel K, Fritsch D, et al. SURE: Photogrammetric surface reconstruction from imagery[C]//Proceedings LC3D Workshop, Berlin: Vol. 8. 2012.
- [66] Krauß T, d'Angelo P, Schneider M, et al. The fully automatic optical processing system CATENA at DLR[C]//ISPRS Hannover workshop: Vol. 1. 2013: 177-181.
- [67] Noh M J, Howat I M. Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions[J]. GIScience & Remote Sensing, 2015, 52(2): 198-217.
- [68] Rupnik E, Daakir M, Pierrot Deseilligny M. MicMac – a free, open-source solution for photogrammetry[J]. Open Geospatial Data, Software and Standards, 2017, 2(1): 1-9.
- [69] Qin R. RPC Stereo Processor (RSP) – A Software Package for Digital Surface Model and Orthophoto Generation from Satellite Stereo Imagery[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, III-1: 77-82.
- [70] Catalyst. Catalyst Professional – CATALYST.Earth[EB/OL]. (2021). <https://catalyst.earth/products/catalyst-pro/>.
- [71] ArcGIS. Ortho mapping with mosaic datasets—Help | ArcGIS Desktop[EB/OL]. (2022). <https://desktop.arcgis.com/en/arcmap/10.5/manage-data/raster-and-images/ortho-mapping-overview.htm>.
- [72] Agisoft. Agisoft Metashape[EB/OL]. (2022). <https://www.agisoft.com/>.
- [73] Bosch M, Foster K, Christie G, et al. Semantic stereo for incidental satellite images[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019: 1524-1532.
- [74] De Franchis C, Meinhardt-Llopis E, Michel J, et al. An automatic and modular stereo pipeline for pushbroom images[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2014: 49-56.
- [75] Zhang K, Snavely N, Sun J. Leveraging Vision Reconstruction Pipelines for Satellite

Imagery[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). 2019: 2139-2148.

[76] Yarowsky D. Unsupervised word sense disambiguation rivaling supervised methods[C]//Proceedings of the 33rd annual meeting on Association for Computational Linguistics. 1995: 189-196.

[77] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]//Proceedings of the eleventh annual conference on Computational learning theory. 1998: 92-100.

[78] Lee D H. Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks[C]//Workshop on challenges in representation learning, ICML: Vol. 3. 2013: 896.

[79] Doersch C, Gupta A, Efros A A. Unsupervised Visual Representation Learning by Context Prediction[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1422-1430.

[80] Kingma D P, Mohamed S, Jimenez Rezende D, et al. Semi-supervised Learning with Deep Generative Models[C]//Advances in Neural Information Processing Systems: Vol. 27. 2014.

[81] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Advances in Neural Information Processing Systems: Vol. 30. 2017.

[82] Qiao S, Shen W, Zhang Z, et al. Deep Co-Training for Semi-Supervised Image Recognition[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 135-152.

[83] Collins R T. A space-sweep approach to true multi-image matching[C]//Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1996: 358-363.

[84] Bleyer M, Rhemann C, Rother C. PatchMatch Stereo - Stereo Matching with Slanted Support Windows[C]//Proceedings of the British Machine Vision Conference 2011: Vol. 11. 2011: 1-11.

[85] Zabih R, Woodfill J. Non-parametric local transforms for computing visual correspondence[C]//European Conference on Computer Vision. 1994: 151-158.

[86] Bosch M, Kurtz Z, Hagstrom S, et al. A multiple view stereo benchmark for satellite imagery[C]//2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). 2016: 1-9.

[87] Yu D, Ji S, Liu J, et al. Automatic 3D building reconstruction from multi-view aerial images with deep learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 155-170.

[88] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[89] Arandjelovic R, Zisserman A. Three things everyone should know to improve object retrieval[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2012: 2911-2918.

[90] Moulon P, Monasse P. Unordered feature tracking made fast and easy[C]//CVMP 2012. 2012:

1.

- [91] 皮英冬. 缺少地面控制点的光学卫星遥感影像几何精处理质量控制方法[D]. 武汉大学, 2021.
- [92] Facciolo G, De Franchis C, Meinhardt E. MGM: A significantly more global matching for stereovision[C]//British Machine Vision Conference. 2015.
- [93] Sdrdis. sdrdis/arpa: My IARPA Contest submission[EB/OL]. (2016). <https://github.com/sdrdis/arpa>.
- [94] Moratto Z M, Broxton M J, Beyer R A, et al. Ames Stereo Pipeline, NASA's open source automated stereogrammetry software[C]//Lunar and Planetary Science Conference. 2010: 2364.
- [95] Bradski G. The openCV library.[J]. Dr. Dobb's Journal: Software Tools for the Professional Programmer, 2000, 25(11): 120-123.
- [96] Facciolo G, De Franchis C, Meinhardt-Llopis E. Automatic 3D reconstruction from multi-date satellite images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 57-66.

攻硕期间发表的科研成果目录

- [1] **GAO J, LIU J, JI S.** Rational polynomial camera model warping for deep learning based satellite multi-view stereo matching[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6148-6157.
- [2] **GAO J, LIU J, JI S.** A general deep learning based framework for 3D reconstruction from multi-view stereo satellite images[J/OL]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 195: 446-461. DOI:10.1016/j.isprsjprs.2022.12.012.

攻硕期间参与的课题情况

1. 基于深度学习和多视遥感影像的建筑物精识别与三维建模（210971417），国家自然科学基金项目，项目参与人员
2. **快速生成与自动精化-2（240371403），部委（纵向）项目，技术骨干
3. **三维框架构建-1（240371404），部委（纵向）项目，技术骨干
4. 基于深度学习的航空多视影像建筑物自动提取（250000762），华为公司横向项目，项目参与人员
5. LuojiaNet 遥感影像解译深度学习框架的典型应用模型研究（250071578），华为公司横向项目，项目参与人员

致 谢

珞珈山下雨初晴，水风清，晚霞明。行文至此，回望江城求学之路，算来樱花已经开了七次，又落了七次。时光匆匆总不愿停下她的脚步，我也即将迎来硕士毕业的这一天。这一路上，有良师相助，有好友相伴，有家人守候，万般庆幸，感激生命的星空之中有你们的陪伴！

首先，我想感谢我的老师季顺平教授！从大四开始，季老师就一直悉心指导我，督促我在学术的道路上前行。随季老师在科研的道路上同行，我被老师源源不绝的创新思想和激情满满的工作精神所折服。感谢季老师带领我在学术的海洋之中领略一番，使我收获颇丰。同时，我还想向陶鹏杰老师和其他帮助过我的诸位老师表示衷心的感谢！

之后，感谢课题组的各位小伙伴！一起漫步东湖，一起奔走出差，一起外出调研，一起开怀畅饮，一起研讨学术，都是我最珍贵的回忆。在紧张的学习路上，感谢你们的欢声笑语，感谢你们的支持鼓励，感谢你们的默契合作！在此，特向师姐、魏大哥、港子哥、韬哥、曾牛、晶晶、小戴老师、牧老师、超超哥、瑄哥等人进行提名感谢。最后，我还想补充一句：“饭，还得跟你们一起吃才香！”

然后，我需要感谢我的家人们！感谢爷爷奶奶对我的牵挂和关爱，感谢父母对我的鼓励与支持，感谢涛、球两兄弟的陪伴及关心。此外，感谢姐姐在我情绪低落之时的陪伴鼓励，在我面对困难时伸出援手，在我欢心喜悦之时共同分享欢笑，有你真好！

最后，感谢各位评审老师，在百忙之中审阅本文，并提出宝贵的意见！

旅途总有一天会迎来终点，但新的篇章总会继续展开。站在这一人生的里程碑处，歇歇脚，看看来时的路，与各位道一声再见，再送上我最真挚的感谢与祝福。学路漫漫，感激有你们相伴！梦想仍在远方，我已准备好背上行囊，继续出发前行。在此，希望未来的我能继续怀揣热忱，向着星辰与大海，勇敢去追寻属于自己的光芒！