引文格式:王密,樊仲藜,叶沅鑫.多模态遥感影像匹配进展与展望:从传统模式到深度学习[J].武汉大学学报(信息科学版),2026,51(10):2105-2123.DOI:10.13203/j.whugis20260369

Citation:WANG Mi,FAN Zhongli,YE Yuanxin.Advances in Multimodal Remote Sensing Image Matching: A Survey from Conventional Approaches to Deep Learning[J].Geomatics and Information Science of Wuhan University,2026,51(10):2105-2123.DOI:10.13203/j.whugis20260369

多模态遥感影像匹配进展与展望:从传统模式到深度学习

王 密[1,2] 樊仲藜[1] 叶沅鑫[3]

1 武汉大学测绘遥感信息工程全国重点实验室,湖北 武汉,430079

2 湖北珞珈实验室,湖北 武汉,430079

3 西南交通大学地球科学与工程学院,四川 成都,611756

摘 要:多模态遥感影像匹配是多源遥感信息融合应用的关键基础,也是摄影测量与遥感领域的国际研究热点。多源传感器能够从不同物理维度获取地表互补特征,在现代遥感应用中,多模态遥感影像正逐渐取代单源数据,成为学术界和工业界的主流数据源。然而,由于成像机理截然不同,多模态遥感影像间存在显著的非线性辐射差异和几何畸变,令传统匹配方法面临严峻挑战,确保影像间的几何一致性与严密空间对齐已成为协同处理的关键瓶颈。回顾并剖析过去十年来多模态遥感影像匹配技术的研究进展,构建了涵盖基于区域的方法、基于特征的方法、深度学习驱动的方法三大核心类别的精细化分类架构,系统归纳并总结了当前算法的内在机制与研究现状,探讨了现存的技术瓶颈及未来演进趋势,并详尽阐述了主流开源数据集与定量评价指标体系。该研究力求为相关研究人员根据应用场景选择合适的匹配方法提供可靠的技术参考和研究指导。

关键词:多模态遥感影像;同名点匹配;区域匹配方法;特征匹配方法;深度学习

中图分类号:P237 **文献标识码**:A **收稿日期**:2026-09-22

DOI:10.13203/j.whugis20260369 **文章编号**:1671-8860(2026)10-2105-19

Advances in Multimodal Remote Sensing Image Matching: A Survey from Conventional Approaches to Deep Learning

WANG Mi[1,2] *FAN Zhongli*[1] *YE Yuanxin*[3]

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 Hubei Luojia Laboratory, Wuhan 430079, China

3 Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: Multimodal remote sensing image matching serves as a vital foundation for multi-source remote sensing information fusion applications and represents an international research hotspot in photogrammetry and remote sensing. Multi-source sensors can capture complementary surface characteristics from diverse physical dimensions. Consequently, multimodal remote sensing images are gradually superseding single-source data in modern remote sensing applications, emerging as the dominant data source in both academia and industry. However, due to fundamentally disparate imaging mechanisms, pronounced non-linear radiometric differences and geometric distortions exist across multimodal images, posing severe challenges to conventional matching methods and making the preservation of geometric consistency and rigorous spatial alignment a critical bottleneck in collaborative processing. We aim to systematically review and analyze the research progress of multimodal remote sensing image matching technologies over the past decade. A fine-grained taxonomic architecture comprising three core categories: area-based methods, feature-based methods,

基金项目:国家自然科学基金(62425102);湖北省科技创新人才计划(2024DJA035)。

第一作者:王密,博士,教授,研究方向为高精度智能卫星遥感技术。wangmi@whu.edu.cn

通信作者:樊仲藜,博士,博士后。fanzhongli@whu.edu.cn

and deep learning-driven methods, is established, systematically summarizing the underlying mechanisms and current research status of existing algorithms. Furthermore, current technical bottlenecks and future evolutionary trends are thoroughly discussed, alongside a comprehensive elaboration of mainstream open-source datasets and quantitative evaluation metrics. We strive to provide researchers with reliable technical references and research guidance for selecting appropriate matching methods tailored to specific application scenarios.

Key words: multimodal remote sensing imagery; tie point matching; area-based matching methods; feature-based matching methods; deep learning

高分辨率遥感对地观测作为地球空间信息科学的前沿领域,是实现地表特征精确解译、地球系统多维立体感知与定量化建模的关键科学手段[1-2]。近年来,中国天基对地观测体系实现了跨越式发展,据中国地理信息产业协会最新统计,中国民用在轨对地观测遥感卫星已突破640颗,形成了涵盖高分辨率可见光、多/高光谱、红外、合成孔径雷达(synthetic aperture radar,SAR)以及星载激光测高仪等多种传感器的立体化天基对地观测网[3]。这些多源传感器充分利用各自独特的物理成像机制与谱段特性,实现了对全球陆表、海洋与大气多角度、多频段、全天候、全天时的立体化协同观测。由此产生的多模态遥感影像及测绘地理信息数字衍生产品,在信息域和物理表征上具备极强的时空互补性[4]。推动多模态遥感数据的协同处理与深层次融合,是打破单一数据源信息盲区、赋能自然资源监管、全天候防灾减灾、智慧农业监测及重大工程建设等国家关键领域的必由之路。这不仅是推动中国空天信息产业由数据体量优势向新质生产力效能跨越的根本动力,更是构筑自主可控天基信息底座、服务国家重大战略纵深与全球动态感知治理的必然抉择。因此,突破多模态天基大数据的高效融合与精准集成技术,已上升为抢占未来空天科技战略制高点、保障国家安全与高质量发展的重大战略需求。

在对地观测的科学探索与行业应用中,多模态遥感数据的协同处理已展现出不可替代的显著优势,正加速取代传统单源模式,成为支撑新一代前沿遥感应用的核心主力[5-6]。在智慧农业与粮食安全监测领域,研究人员联合光学多光谱的时间序列反演与SAR影像的结构穿透特性,结合顾及上下文特征的深度学习模型,不仅有效克服了多云多雨天气导致的物候信息缺失,更精准捕捉到农作物全生命周期的关键物候表征,使复杂地块的动态分类与长势预测精度跃升至95%以上[7]。在自然资源监管与灾害应急变化检测领域,研究人员通过在异构跨模态数据之间引入特征深度泛化表征机制,融合微波雷达的全天候穿透能力与光学影像的丰富光谱特性,打破了恶劣大气条件下光学观测质量严重退化的瓶颈,显著增强了复杂地理环境下地表监测的灵敏度与鲁棒性[8]。这些典型领域的实践充分证明,跨平台、多模态遥感数据的协同感知与联合反演,是突破单一传感器观测局限、释放天基遥感应用效能的关键路径[9]。

然而,要充分发挥百星在轨的协同观测优势,实现多源数据的高效联合应用,其首要瓶颈在于消除多平台、多时相、多模态影像之间的几何偏差,建立严密的空间几何一致性[10-11]。若缺乏像素乃至亚像素级别的严格对齐,多模态遥感数据的融合与分析将无从谈起。在这项复杂的系统任务中,多模态遥感影像匹配是构建影像空间几何一致性、实现高精度几何对齐的核心支撑与关键前沿技术[12]。多模态遥感影像匹配是指通过挖掘影像之间的共有信息并构建相似性度量准则,建立像素级同名对应关系的解算过程。在影像空间基准与先验地理参考信息的支持下,通用的影像匹配流程通常涵盖3个递进阶段:

1)初始匹配区域预测。基于粗略地理参考建立特征网格,以确定候选匹配搜索范围。

2)同名对应关系解算。在搜索范围内,利用特定的匹配算子获取初始同名匹配点对,该阶段是决定匹配成败与精度的核心,亦是不同匹配技术路线的分水岭。

3)几何检验与误匹配剔除。基于几何变换约束构建误匹配探测模型,对候选对应关系集合进行几何一致性检验与粗差剔除。

尽管影像匹配在摄影测量与计算机视觉领域已有长期积累,但传统影像匹配技术难以直接迁移至当下的多模态卫星观测场景[13-15]。在实际观测中,多模态影像匹配面临着两大挑战:(1)影像间存在显著的非线性辐射差异(nonlinear radiometric differences,NRD)[16],其源于传感器记录

光学反射、微波后向散射、红外热辐射等信息的成像物理机理的不同,导致地表同名地物在不同影像中的灰度分布迥异乃至出现梯度反向;(2)噪声与时空畸变干扰[17],如SAR影像固有的强相干斑噪声与光学影像的云雾遮挡、多角度立体成像带来的几何畸变、时相差异导致的地物变化等。传统的灰度相关匹配高度依赖于严格的初始定位,且对辐射失真极度敏感,而经典的通用特征提取算子主要面向同构几何不变性设计,在强烈的非线性辐射畸变场景中往往面临着特征描述性能下降甚至失效的困境[18-19]。

随着中国对地观测系统迈入跨模态协同感知的新阶段,亟须突破复杂多模态遥感影像在高鲁棒性、高精度空间配准方面的技术瓶颈。十余年来,国内外科研人员在多模态几何结构提取、跨模态辐射不变表征以及深度学习赋能的特征匹配等方向推陈出新,逐步探索出一套跨越物理成像机制壁垒的通用多模态算法体系。立足于国家重大空天信息技术发展的迫切需求,系统梳理当前主流多模态遥感影像匹配方法的演进路径,明晰不同方法在典型场景下的优缺点与适用边界,对于支撑中国未来多模态航天遥感大数据的一体化处理与协同应用,具有重要的学术价值与实践意义。

早期的影像匹配综述文献多聚焦于单一模态数据。其中,文献[20]的综述最具代表性,系统梳理了1978—2003年间影像配准的理论体系,并明确提出了划分为基于区域与基于特征的两大分类范式,为后续研究奠定了经典分类框架。然而,该文对多模态/多传感器影像匹配的探讨相对有限;受限于当时的技术发展背景,亦未涉及深度学习等智能匹配算法。总体而言,该文主要反映了早期经典匹配算法的发展脉络与基础架构。近年来,针对多模态遥感影像匹配的综述性研究陆续涌现,但在技术分类体系、分析侧重点及前沿归纳上各具视角[21-22]。现有综述多立足于广义图像处理领域,针对遥感多模态影像匹配的专门性总结仍显不足。该领域的突破性进展主要集中于近十年,但目前学界尚缺乏对此阶段研究成果的系统梳理与评述。以往工作多停留在宏观层面的粗略分类,缺乏对先进算法机理与核心特性的细致剖析,分类体系粒度较粗且存在内容同质化倾向,导致实用参考价值相对受限。

本文构建了“宏观大类-特性子类”的层级分类架构,系统梳理了近十年间涌现的多模态遥感影像匹配方法,并对典型代表性算法的关键机制展开剖析;同时,整理了主流的开源基准数据集与定量评估指标,探讨了现阶段的典型工程应用场景,以期为该领域的研究人员提供系统性的技术指引与研究启发。

1　多模态遥感影像匹配方法

结合多模态遥感影像的核心物理特性与演进脉络[23-24],本文将主流的匹配方法划分为3个大类:基于区域的匹配方法、基于特征的匹配方法以及结合深度学习的匹配方法,典型方法分类及其发展过程如图1所示。为了更清晰地呈现技术演进的细节,本文将每个大类精细化拆解为多个具体的子类,重点探讨各算法在应对NRD时的核心贡献与发展脉络。



图1　多模态遥感影像匹配方法分类及其发展过程

Fig. 1　Taxonomy and Evolutionary Process of Multimodal Remote Sensing Image Matching Methods

1.1 基于区域的匹配方法

基于区域的多模态遥感影像匹配方法是传统模板匹配方法在多模态场景下的拓展与延申[25]。传统模板匹配方法在纯光学及其他单模态遥感影像匹配任务中应用广泛,其核心是通过窗口影像匹配策略实现同名点对的稳健获取[26]。具体流程如下:首先,在参考影像上以目标点为中心构建模板窗口;随后,在待匹配影像中划定搜索区域,该步骤通常需基于先验位置约束;接着,在搜索区域内滑动模板窗口,基于选定的相似性测度计算各位置相似度得分;最终,通过得分峰值定位对应匹配点,完成同名点对识别。随着多模态匹配技术的发展,现代基于区域的方法已不再局限于传统的纯灰度相关计算,而是广泛引入了各种特征信息的辅助。因此,有必要对此类方法进行详细的分类和讨论,以实现更好的区分。

1.1.1 基于灰度相关的模板匹配方法

灰度相关是基于区域的匹配技术中最经典且直观的度量范式,其代表性方法包括归一化互相关(normalized cross-correlation,NCC)[27]、均方误差(mean square error,MSE)[28]以及绝对差和(sum of absolute differences,SAD)[29]等。此类方法直接根据像元灰度的统计相关性构建相似性测度,具有数学模型简洁、物理意义明确及计算易于硬件加速等优势,因而在同模态影像匹配中得到了广泛应用。然而,在多模态遥感场景下,传感器成像机理的本质差异会导致同一地物呈现强烈的NRD、灰度反转及结构断裂,直接应用传统灰度相关度量极易导致相关峰退化或误匹配。为此,相关研究主要从抗辐射畸变改进与层次化多策略协同两个维度对其展开了演进与拓展。

在抗辐射畸变与鲁棒性度量改进方面,针对多模态影像间复杂的灰度非线性映射问题,研究人员致力于发展对辐射差异更具鲁棒性的相关度量准则。传统NCC虽然具备抗线性辐射变化的能力,但难以应对例如SAR与光学影像间的局部灰度反转问题。针对此类问题,部分研究引入了局部NCC(local NCC,LNCC)[30]及加权相关测度,通过将大范围非线性映射近似分解为局部分块内部的线性变化,在一定程度上缓解了辐射畸变对匹配峰值锐度的负面影响。此外,针对大范围相干斑噪声与极端光照差异,相关研究进一步将零均值NCC(zero-mean NCC,ZNCC)[31]与局部结构先验相结合,抑制异常像元对相关系数全局统计分布的干扰,显著提升了在弱纹理及异质区域的相关峰辨识度。

然而,基于灰度相关的模板匹配方法在多模态遥感影像匹配任务中依然面临难以逾越的瓶颈。首先,该类方法从根本上建立在同名地物灰度线性相关的先验假设条件之上,而多模态遥感影像之间的辐射差异本质上是强非线性的,甚至伴随多类噪声影响与几何透视畸变,这导致灰度相关曲面往往多峰且信噪比极低;其次,该类方法对影像间的旋转、尺度以及局部仿射变形极度敏感,通常仅适用于严格纠正后仅存微小视差的影像对;此外,滑窗相关计算在处理大尺寸影像时仍受限于复杂度,对搜索范围和初始重叠度要求苛刻,难以应对现代遥感跨传感器、大视场、宽基线的多模态匹配需求。

1.1.2 基于互信息的模板匹配方法

互信息(mutual information,MI)作为匹配准则,最初由文献[32]及文献[33]分别独立提出,并成功应用于多模态医学图像配准。此方法的优势在于它不依赖于物体的表面属性信息,且能够有效抵抗光照变化的影响,使其适用于各种成像条件下的影像处理任务,因而成为多模态匹配的经典模型基础。

1)空间信息与搜索模式优化。为了解决当多模态影像间存在充足的场景重叠区域时,MI算法可能无法获得最佳匹配的问题,文献[34]引入了库尔贝克-莱布勒散度的一种对称形式,即Jeffrey散度,作为相似性度量指标。在遥感影像上的结果表明,该方法改善了潜在的搜索空间,提升了最优变换参数的获取能力,并提高了影像配准的精度。文献[35]认为MI在计算概率分布时忽略了空间信息,仅考虑了影像灰度的统计信息,因此在MI的基础上引入了区域信息,并考虑了局部灰度变化和旋转变化对概率密度函数计算的影响,提出了基于旋转不变区域互信息(rotationally invariant regional mutual information,RIRMI)的方法。RIRMI在中低分辨率影像上的表现优于MI,但缺乏在高分辨率影像上的测试结果。

2)特征辅助与多尺度协同。文献[36]提出了一种结合局部自相似描述符和互信息的匹配方法。该方法是在基于区域的匹配过程中引入特征信息的典型代表,先通过局部自相似性获取影像间的初始对应关系,随后利用粒子群优化算

法优化点与区域之间的对应关系,并以此为基础实现了MI效率和精度的提升。为了实现高分辨率光学与SAR卫星影像的配准,文献[37]提出了一种基于MI的由粗到精的匹配方法。首先对待处理影像进行下采样,并基于MI和直接同步扰动随机逼近方法估计影像的全局几何变换,以进行粗匹配;随后,在原始分辨率下,基于MI和连续同步扰动随机逼近方法处理局部畸变,以实现精匹配。利用6对高分辨率光学-SAR卫星影像对该方法的有效性进行了验证。

然而,基于MI的匹配方法在实际应用中仍存在显著的局限性。首先,由于需要频繁计算联合概率分布,其计算复杂度较高;其次,在寻优过程中,MI的目标函数表现出高度的非线性和非凸性,导致多数场景下极易陷入局部极值。此外,该方法对初始定位依赖性较强,通常仅适用于存在微小偏移量的局部最优求解,在缺乏强验算约束的情况下,无法有效保证全局搜索的最优性[38]。

1.1.3　基于结构属性的模板匹配方法

为应对NRD对多模态影像匹配的干扰,实现稳健匹配,近年相关研究聚焦于影像模板窗口的创新设计[39]。该策略较相似性度量设计在多模态场景中展现出更优的适配性,即通过构建结构化影像模板来增强对影像的跨模态共性表征,由此构建了一套以影像模板设计为核心的多模态影像匹配方法体系,为突破传统模板匹配的NRD敏感瓶颈提供了新途径。与传统的基于区域的匹配方法不同,该类方法先根据特征描述子的计算模式,将二维窗口影像转换为特征描述向量;随后,通过计算模板窗口向量与搜索窗口中不同像素向量之间的距离,完成同名点对应关系的搜索。

1)将局部自相似性机制引入模板设计。文献[40]基于局部自相似性(local self-similarity,LSS)描述子[41]实现了具有显著NRD的多光谱遥感影像匹配。首先,基于尺度限制的尺度不变特征变换算法[42]实现了影像间的粗对齐,以消除较大的几何差异;随后,基于LSS描述子计算模板窗口的特征向量;最后,利用NCC度量实现模板匹配。尽管该方法仅是对已有算法的组合嵌套,但其取得了良好的匹配精度,并为基于一维模板特征的方法奠定了基础。进一步地,文献[43]提出了一种用于光学和SAR卫星影像匹配的密集LSS(dense LSS,DLSS)算法。DLSS将模板窗口划分为多个具有重叠区域的单元。每个单元覆盖一个较小的影像区域,其描述向量由多个LSS描述子的组合构成。这种描述方式显著提高了光学与SAR影像特征向量之间的相似性。文献[44]则指出原始LSS易受噪声和灰度剧烈波动干扰,遂引入斯皮尔曼秩相关系数替代原始灰度输入,提出了基于秩的DLSS(dense rank-based LSS,DRLSS)方法,进一步增强了算法对SAR斑点噪声的抵抗力。

2)基于相位一致性的几何结构重构。基于相位一致性(phase congruency,PC)模型[45],2017年,文献[46]提出了著名的方向PC直方图(histogram of oriented PC,HOPC)算法,是此类算法中最经典的方法之一。HOPC的提出深受定向梯度直方图(histogram of oriented gradients,HOG)算法[47]的启发。HOG是一种基于梯度的行人检测算法,而HOPC首次将其框架引入多模态遥感影像匹配任务。由于PC模型具有光照和对比度不变性,HOPC使用PC模型代替梯度计算。“PC模型更适用于多模态遥感影像匹配”这一观点对后续许多算法产生了深远影响。在算法层面,HOPC实现了PC模型的扩展,给出了一种计算PC方向的方法,并以此取代了方向梯度。在此基础之上,文献[48]提出了一种用于光学和SAR影像匹配的相位一致性结构描述子(phase congruency structural descriptor,PCSD)算法。为了抵御相干斑噪声对特征检测的影响,PCSD引入了基于非线性扩散的Harris角点检测器[49],该检测器能够检测到许多分布均匀的显著特征。在描述向量生成过程中,PCSD结合非线性扩散和PC建模以获取多个PC结构影像,并按照类似于LSS的模式为每个影像窗口构建模板信息。

然而,上述影像模板均基于稀疏网格构建,对跨模态细微几何结构的表征能力仍显不足。为此,相关研究逐步转向逐像素密集采样的模板设计,以进一步提升结构表征的精度。

3)密集采样表征的初步引入。文献[50]提出了经典的密集结构化影像模板,即方向梯度通道特征(channel feature of oriented gradients,CFOG)。CFOG通过高斯核卷积优化梯度通道,以增强边缘响应,并首次采用逐像素密集采样模式取代传统的稀疏网格,从而精细化捕获跨模态影像结构;同时,该方法引入快速傅里叶变换,在频域内加速相似性测度的计算,实现了高维结构特征的高效捕获与快速定位。CFOG的出现标志着模板设计由稀疏采样转向稠密采样,将多模

态影像模板匹配的精度和效率推至新的高度,并进一步激发了许多相关方法。

4)加权与多尺度、多方向拓展。文献[51]提出了角度加权方向梯度(angle-weighted oriented gradients,AWOG)影像模板。该方法基于梯度方向将幅值加权分配至两个最相关预设方向,进一步提升密集采样对复杂结构的表征能力;同时,采用去归一化三维相位相关作为相似性度量,实现同名点快速识别。AWOG的匹配性能优于CFOG,尤其是在纹理丰富区域,其效果更优。文献[52]提出了多尺度、多方向奇对称Gabor模板(multi-scale and multi-directional features of odd Gabor,SDFG)。SDFG通过奇对称Gabor滤波器提取梯度信息,同时结合高斯卷积抑制噪声,相比传统梯度计算更能保留结构细节,并采用三维相位相关匹配提升速度与鲁棒性。然而,其局限性在于依赖显著结构特征,当影像缺乏轮廓信息时,性能明显下降。

5)高阶信息融合与效率优化。为同时兼顾宏观轮廓与微观细节,文献[53]进一步拓展出一二阶可调通道模板(steerable filters of first- and second-order channels,SFOC)。SFOC通过融合一、二阶梯度实现了地物边缘捕获与微观细节增强的协同优化,结合多尺度策略构筑多通道特征表达,以全面抵抗剧烈的NRD与噪声干扰。在匹配阶段,SFOC将传统NCC度量与快速傅里叶变换及积分影像技术无缝结合,在保持极高匹配精度的同时,降低了算法的空间与时间复杂度。

总体而言,基于区域的多模态影像匹配方法在匹配精度与计算效率上展现出优势,但其固有局限性显著制约了实际应用。对初始匹配位置的先验准确性依赖极强,几乎仅能适配平移变换场景,且需依赖先验知识划定搜索范围,通常限定在半径100像素内;一旦初始对应位置与真实匹配点偏差较大,匹配性能便会急剧衰减。此外,该类方法对旋转与尺度变化高度敏感,需预先校正显著的几何畸变之后才能生效,如大幅旋转、缩放等,这一前提严重制约了其跨场景通用性,导致应用场景局限于小幅平移的多模态影像配准此类几何变形简单的特定任务,难以适配复杂场景下的多模态协同观测需求[54]。

1.2 基于特征的匹配方法

基于特征的匹配方法通过设定筛选条件提取影像中具有特殊性质的显著特征,包含点特征[55]、线特征[56]、面特征[57]等,为这些特征生成可区分的描述信息,再通过计算并比较描述信息间的距离确定匹配点对。一般地,线特征与面特征需以点特征为基元构建,且需影像场景中存在明确的线元或面元,其适用范围相较于点特征而言受限明显。因此,此类方法以点特征为核心逐步发展,尤其在多模态影像匹配领域,绝大多数方法均以点特征为核心设计,线、面特征因场景依赖性强而应用较少[58]。

影像特征匹配是摄影测量与机器视觉领域的核心任务之一,其本质是通过提取两幅或多幅影像中的局部特征(如角点、斑点、边缘点等),为这些特征生成可区分的描述信息,并基于选定的距离测度估计这些特征之间的相似性,从而建立跨影像的几何对应关系(即同名点对)的过程[59]。

1.2.1 基于SIFT架构的特征匹配方法

文献[60]提出的尺度不变特征变换(scale-invariant feature transform,SIFT)算法,作为经典的基于特征的影像匹配算法,在早期被作为多模态特征匹配方法的核心基础。其核心设计包括:通过构建高斯差分金字塔检测尺度空间极值点,结合梯度方向直方图分配关键点主方向;采用归一化梯度方向直方图描述子增强对光照变化的鲁棒性。然而,SIFT应用于多模态场景的局限性在于对剧烈的NRD十分敏感。

针对SIFT的局限性,相关研究人员提出了一系列改进算法:

1)方向限制与导数重定义。文献[61]针对光学与红外影像间的辐射逆转现象,提出了梯度方向限制SIFT(gradient orientation restriction SIFT,GOR-SIFT)。GOR-SIFT在特征描述时将梯度方向严格限制在$[0,\pi]$范围内,从而在根源上避免了因辐射逆转引起的特征点主方向反转,显著提升了跨模态匹配的正确率。文献[62]则从梯度建模的抗噪性切入,提出了位置-尺度-方向SIFT(position-scale-orientation SIFT,PSO-SIFT)。该算法改用Sobel滤波器构建尺度空间,通过基于导数重定义关键点的方向与幅值,并在最终描述子中定义了尺度比、方向差与平移差的一致性直方图模式,在保持几何不变性的同时大幅增强了对辐射不均匀性的抵抗力。

2)跨模态的异构梯度构建。针对最具挑战性的光学与SAR卫星影像匹配,文献[63]提出了光学至SAR SIFT(optical-to-SAR SIFT,OS-SIFT)算法。该方法摒弃了统一的梯度计算模式,采取异构计算策略。针对光学影像采用多尺

度Sobel算子,而针对伴随相干斑噪声的SAR影像则量身定制了指数加权平均值比率算子,以精准提取边缘。在特征检测阶段,OS-SIFT基于异构梯度构建了Harris尺度空间,以精确捕获并定位稳定角点;在特征描述阶段,通过融合方向限制(0°～180°防反转)、多区域融合(多尺度三邻域)以及圆形分区(对数极坐标采样)3项核心技术,在一定程度上解决了光学与SAR影像之间成像机理迥异的特征点难对齐问题,在多模态卫星数据匹配中达到了优于3像素的均方根残差。

可以看出,以SIFT为基础的改进算法大多数使用高斯差分金字塔极值检测作为特征点提取的核心,带有方向限制的影像梯度作为特征描述基础,虽在一定程度上缓解了跨模态辐射畸变,但其本质上仍依赖对影像灰度或一阶梯度的直接建模。在剧烈的NRD影响及弱纹理干扰下,这类方法极易因梯度方向估计错误或伪结构而导致特征点大量漏检或误匹配。

1.2.2　基于定向滤波的特征匹配方法

随着相关研究的深入推进,研究者逐渐摒弃传统的灰度或梯度框架,转向探索一类利用多尺度、多方向滤波提取结构几何特征的定量描述新范式。

1) 多方向滤波初期探索。文献[64]率先引入多方向频域滤波,提出了对数-Gabor直方图描述子(log-Gabor histogram descriptor,LGHD),通过捕捉可见光与长波红外影像中的共性频域特征,显著提升了跨模态匹配的稳定性。然而,LGHD在特征检测阶段仍沿用加速段测试特征(features from accelerated segment test, FAST)算子[65],导致其在弱光照、低对比度区域极易发生特征点锐减或几何响应中心偏移;LGHD不具备主方向与多尺度设计,因此不具有旋转与尺度差异的鲁棒性;此外,由于缺乏特征降维机制,生成的高维描述子极大牺牲了匹配的计算效率。

2) 辐射与旋转抗性进阶。为克服辐射突变对特征点检测的干扰,文献[66]提出了辐射不变特征变换(radiation variation-insensitive feature transform,RIFT),将PC模型与FAST结合构成PC-FAST检测子,有效实现了跨模态的稳定几何定位,并设计了端到端环路结构的特征描述模式,以模拟影像旋转。然而,该模式的计算代价巨大,旋转不变性的获取严重依赖于对环形链路节点的逐级遍历,计算负荷过重;此外,RIFT算法缺乏多尺度金字塔架构,在面对大尺度差异的遥感影像时性能显著下降。

自相似性空间表征。文献[67]提出了相邻自相似(adjacent self-similarity,ASS)方法。该方法通过挖掘像素邻域的局部几何拓扑稳定性生成最小自相似图,并结合Harris算子实现特征点提取,同时构建方向信息图指导对数极坐标网格采样,为旋转不变特征描述提供了新的空间约束机制。

模态变换与规整。文献[68]则提出了局部归一化影像特征变换(locally normalized image feature transform,LNIFT)。该算法尝试通过局部归一化将异构影像映射至近似的同构模态,进而无缝衔接改进的ORB(oriented FAST and rotated BRIEF)检测子[69]与简化方向梯度直方图。然而,其依然沿用了SIFT体系的逐点角度归一化策略,在面临剧烈NRD时,其旋转不变性仍表现出明显的不稳定性。

3) 滤波机理与描述子结构优化。随着对多向滤波机理理解的加深,相关研究开始向尺度适应性与高性能几何描述子演进。

尺度空间与方向索引融合。文献[70]提出了基于方向滤波器的匹配方法(oriented filter-based matching method,OFM)。OFM利用多尺度PC-FAST构建起严格的影像几何金字塔,以解决尺度缩放问题,并融合方向索引图与自适应分箱SIFT描述子,在多模态匹配性能上全面超越了RIFT、ASS及LNIFT。

高阶几何结构聚合。文献[71]设计了加权相位方向直方图(histogram of the orientation of weighted phase,HOWP)。该算法构建了高阶的聚合相位一致性特征实现精确定位,并完全以log-Gabor小波替代传统的一阶梯度,以保留骨架结构,将经典的梯度位置方向直方图改进为48子区域划分,大幅强化了算法对尺度与旋转畸变的鲁棒性。

采样优化与分步寻优。文献[72]提出了位置-方向-尺度引导的几何与强度不变特征变换(position-orientation-scale guided geometric and intensity-invariant feature transformation, POS-GIFT)。该算法引入了标准化强度与梯度密集聚合描述子[73]结构和点采样模式,但在特征向量构建中直接涉及了对辐射敏感的幅值绝对值,导致其对抗剧烈NRD的鲁棒性未达预期,且因忽略了输入影像的异构特性,其复杂的几何自适应机制反而在实际卫星数据中受到了一定限制。

4)高阶各向异性滤波与流形嵌入。最新研究进一步融合了统计学习中的流形结构与数字信号中的高阶边界保持滤波。

流形约束与均匀分布优化。文献[74]提出了预训练嵌入流形(pre-trained embedded manifolds,PTEM)算法。在特征点检测步骤中,该方法采用对非线性扩散更具保持力的KAZE算子[75],并结合均匀竞争力算法[76]消除集群效应;在特征描述环节,通过预训练卷积层提取高维特征图,借助局部线性嵌入实现非线性流形降维,结合自适应分箱空间网格描述子结构与LSS模型分配主方向,确立了稳健的旋转不变拓扑描述。

各向异性边缘保持与对数极坐标分区。文献[77]则提出了侧窗高斯匹配(matching using side-window Gaussian,MSG)算法。该方法在检测阶段将边缘保持性能极佳的侧窗高斯Sobel算子与Harris检测子融合,从根源上规避了弱纹理下的边界模糊;在描述阶段,通过一阶高斯方向可调滤波器与Sobel二阶梯度联合计算幅值与相位,并基于对数极坐标分区设计了提升的梯度位置方向直方图,显著增强了复杂遥感成像条件下算法对剧烈NRD与旋转仿射畸变的抵抗能力。

尽管当前多模态影像匹配研究取得了显著进展,但如何应对复杂NRD与显著几何畸变仍是匹配方法性能提升的主要关注点。首先,基于特征点的匹配范式对显著NRD与较大几何畸变展现出更强鲁棒性,相比区域匹配方法而言具有更好的灵活性;其次,此类方法对软硬件条件依赖性低,在各种计算平台均能高效运行。方法设计遵循真实世界的物理及视觉原理,具有良好的可解释性,且方法完全不依赖任何标注数据[78-79],节省大量人力成本,可在各模态类型的数据上实现即插即用。这些层面上的优势使特征点匹配的范式在当前深度学习技术的冲击下仍具有巨大的研究潜力。然而,当前现有方法在处理显著几何差异时仍存在不足,如影像旋转角度变化时性能波动显著,且因抗NRD能力有限,处理不同类型多模态影像数据的泛化性仍存局限。基于特征的多模态影像匹配方法虽在精度上存在局限,但其应对复杂多模态匹配的潜力仍具有重要研究价值。

1.3 结合深度学习的多模态影像匹配方法

随着人工智能领域的跨越式发展,将深度学习强大的非线性特征表达能力融入传统多模态影像匹配框架,已成为克服剧烈辐射畸变与几何形变的核心发展方向。通过端到端网络架构、模态不变特征协同学习以及跨模态映射等策略,基于深度学习的方法有效突破了传统人工设计度量准则及特征表征的瓶颈。

1.3.1 基于深度学习的模板匹配方法

基于深度学习的影像模板构建与匹配方法,旨在利用深度神经网络提取高阶异构共性特征,从而重构对NRD鲁棒的模板设计与相似性测度模式。

1)稠密表征与阶段式搜索空间。早期研究主要聚焦于利用全卷积网络重构特征度量空间。文献[80]提出深度稠密特征网络(deep dense features network,DDFN),针对光学-SAR影像匹配设计7层卷积Siamese网络,每层采用3×3卷积核,后接批归一化与线性整流激活函数,输出9层像素级稠密影像特征,以保持输入影像空间分辨率,基于误差平方和度量计算相似性得分,并通过快速傅里叶变换技术在频域加速匹配过程。

为了解决无初始定位时全图窗口搜索导致的计算负荷大、误匹配率高的问题,文献[81]开发了一种三阶段递进式深度学习匹配架构。首先,利用评价网络自适应识别影像中富含显著跨模态共性结构的区域,大幅压缩候选搜索范围;其次,在感兴趣区域内利用对应网络提取影像多尺度特征信息,深度融合浅层高分辨率几何细节与深层强语义模板,并借由特征空间互相关计算实现高精度匹配得分获取;最后,通过误匹配剔除网络净化几何拓扑关联,输出高可靠性的匹配结果。

2)结构注意力增强与跨模态对抗生成。近年来,高级注意力机制与跨模态协同生成技术的融入,将深度模板表征推向了空间结构自适应对齐的新高度。

3)多维注意力协同。文献[82]提出了注意力增强结构特征(attention-enhanced structural features,AESF)方法。该方法首先利用CFOG构建基础几何模板,随后引入多分支全局注意力模块,将多尺度特征提取与通道注意力(用于筛选表征模态不变性的关键通道)、空间注意力(用于精确聚焦跨模态共享骨架区域)深度耦合,显著提升了算法在复杂城市、多特征平原及多变水域等全地形场景下的泛化性能。

4)监督对抗流形对齐。通过在标准循环一致生成对抗网络(cycle-consistent generative ad-

versarial network,CycleGAN)框架[83]中引入监督学习机制,文献[84]提出了光学影像-SAR的CycleGAN (optical-SAR CycleGAN, OS-CycleGAN)网络,使用预配准的SAR-光学影像块训练,设计SAR到光学与光学到SAR的双向映射机制,实现跨模态结构对齐,同时引入多层感知机辅助影像模板生成,其匹配精度与鲁棒性显著优于AESF,为多模态遥感数据融合提供可靠基础。这些方法通过深度学习技术显著提升了影像模板的跨模态表征能力与匹配效率,推动多模态影像匹配向智能化方向发展。

1.3.2　基于深度学习的特征匹配方法

与全局模板构建不同,深度特征提取与匹配方法的目的是从网络底层或高层语义空间中直接提取具有多尺度、旋转及仿射不变性的特征点与鲁棒描述子。

1) 移位窗口注意力与递进匹配。针对几何非线性变形与辐射反转的联合干扰,近年来基于Transformer架构的多尺度交互模型取得了突破性进展。文献[85]提出了基于位移窗口变换器(shifted window transformer, Swin Transformer)[86]的通用匹配器(Swin Transformer-based universal matcher, SwinMatcher)。SwinMatcher在特征提取阶段采用Swin Transformer骨干网络,通过层级式窗口自注意力与移位窗口交叉注意力机制实现多尺度特征交互;在匹配阶段设计3级分层策略:粗匹配通过内积相似度计算与双向最近邻筛选候选点,精匹配采用局部窗口特征与双向Softmax优化位置,亚像素优化通过回归卷积块预测位置残差,并结合多级监督损失,包括粗匹配聚焦损失、细匹配对称焦点损失和亚像素重投影误差等,增强模型对NRD与复杂几何畸变的鲁棒性。

2) 分层交叉映射与多级语义嵌入。基于分层式交互表征的思路,文献[87]提出了一种基于分层提取实现匹配的变换器模型MatchFormer。通过利用自注意力进行特征提取、交叉注意力进行特征匹配,MatchFormer实现了一种符合人类直觉的提取与匹配方案。该模型在分层架构中整合多尺度特征的自注意力与交叉注意力,进一步提升了匹配鲁棒性。

全流程语义引导。文献[88]提出了一种以语义引导全局(semantics to lead all, SemLA)的网络,用于红外与可见光影像的级联匹配、对齐及融合。SemLA的核心思想是将语义信息嵌入神经网络的每个阶段,生成的语义感知图与空间结构信息的描述相结合,以实现精准匹配。

3) 交互式金字塔与跨模态知识蒸馏。以动态窗口调制为基础,文献[89]提出了基于模态自适应特征调制与稀疏变换器融合的跨模态图像匹配方法(cross-modal image matching with modal-adaptive feature modulation and sparse transformer fusion, CM2-STNet)。CM2-STNet在特征提取阶段采用多尺度特征金字塔结合交互式Swin Transformer模块,通过动态窗口划分策略实现跨模态特征交互;在匹配阶段设计3级分层策略:粗匹配通过双向Softmax置信度矩阵筛选候选点,精匹配采用局部窗口特征与双向注意力优化位置,亚像素优化通过回归卷积块预测残差,并结合多级监督损失增强模型对NRD与复杂几何畸变的鲁棒性。

异构教师集群知识蒸馏。文献[90]提出跨模态影像匹配的高效模态不变网络(efficient modality-invariant network, EMINet),EMINet采用知识蒸馏框架,通过教师网络(包括无标签自蒸馏语义编码器[91]、SuperPoint关键点检测器[92]和SuperGlue图匹配器[93]),与轻量级学生网络——18层残差骨干网络[94],进行协同训练实现模态不变特征学习,其核心创新包括语义窗口关系蒸馏,通过教师网络提取多模态影像块间语义关联特征指导学生网络建立跨模态窗口对应关系以增强判别力,跨三元组描述子蒸馏强制学生学习模态不变描述符,以解决NRD导致的特征偏移以及多级监督训练策略。

4) 跨模态训练数据集生成。文献[95]提出跨模态图像匹配的数据生成迭代模型(data-generating iterative model, DGIM)。DGIM采用级联动态数据生成框架,通过动态数据引擎,集成随机单应性变换模块与基于自适应实例归一化的轻量级生成模型,利用感知损失(内容损失+风格损失)优化多风格跨模态图像对生成质量,解决模态差异导致的特征分布偏移,采用混合架构匹配模型,结合残差可变形卷积网络提取多尺度局部特征,变换器编码器捕获全局上下文依赖并通过步进聚合策略平衡计算效率,特征金字塔网络解码器生成细粒度特征图,协同优化实现鲁棒特征匹配。

当前结合深度学习的多模态影像匹配方法展现出一定潜力,但其局限性同样显著,主要体现在3个方面:

首先,训练数据瓶颈突出[96]。现有方法训练高度依赖大规模、高精度标注的多模态遥感数据集,但当前公开数据集在数量与规模上均显不足,且多模态影像的获取与标注成本高昂;更关键的是,现有数据集的几何配准多基于非学习方法,其配准精度存在固有上限,而模型训练需以预配准影像对为基础,这种缺陷将直接制约最终匹配结果的可靠性。

其次,跨模态泛化性不足[97-98]。受传感器制造工艺、搭载平台设计差异等因素影响,多模态影像间存在显著异质性(如光学-SAR的辐射差异、光学-深度图的结构差异),而深度学习模型对数据分布变化高度敏感。例如,基于光学-SAR影像训练的模型迁移至光学-深度图匹配任务时性能可能骤降,即便更换光学-SAR数据源,性能亦可能因数据分布偏移而显著波动。这种泛化性缺陷导致方法更适于特定场景,难以应对数据类型多样的跨模态匹配需求。

最后,对软硬件环境的适配性挑战[99-100]。随着模型参数量激增,部分研究转向大模型架构,却忽视了无人机在线处理、卫星星上处理等资源受限场景对轻量化的迫切需求,这些场景计算与存储资源有限,而大模型的高复杂度难以部署。因此,实际应用中需向自监督学习与轻量化网络结构转型,以平衡精度与效率。这些局限性表明,深度学习在多模态影像匹配中的应用仍需突破数据、泛化与部署的三重瓶颈,方能实现从实验室潜力到实际应用的跨越。

2 多模态遥感影像匹配数据集

高质量、标准化的基准数据集是进行方法定量性能评估与深度学习模型训练的基础。相较于同模态遥感任务,多模态遥感影像因传感器机制的本质差异,普遍存在剧烈的NRD、几何视角偏差以及局部形变,构建亚像素级精细几何真值的技术难度极高。近年来,随着高分辨率多传感器对地观测系统的发展,学术界与工业界相继构建并开源了一系列针对特定传感器组合的基准数据集,覆盖光学-SAR、光学-红外、光学-电子地图等典型多模态场景。本文系统整理了当前主流的12个代表性多模态遥感影像数据集,详细阐述其模态构成、几何分辨率、真值构建等特性,具体信息汇总于表1。

SEN1-2数据集[101]:数据模态为SAR+光学,数据源自哨兵1号(Sentinel-1)双极化SAR卫星与哨兵2号(Sentinel-2)多光谱光学卫星,分辨率为10 m,该数据集由慕尼黑工业大学基于谷歌地球引擎平台提取构建,按春夏秋冬划分为4个子集,共包含282 384对严格对齐的正射切片,建立了影像间的像元级空间对应关系。

SARptical数据集[102]:高空机载超级摄像头航拍超高分辨率光学影像与TerraSAR-X聚束模式SAR影像,影像分辨率分别达0.2 m和1 m。数据集关注于德国柏林、法兰克福等城市复杂都市核心区的密集高层建筑物,包含逾10 000对标准分块影像,提供了严密的亚像素级空间对应真值。

OSDataset 2.0[103]数据集:国产高分三号(Gaofen 3,GF-3)SAR卫星影像与谷歌地球影像,覆盖14个国家的6 476对0.43 m分辨率的512×512影像切片,提供了高精度、空间分布均匀的同名特征点对及通用评估模型。

3MOS数据集[104]:光学卫星影像与多源SAR卫星数据集,光学影像来自谷歌地球,SAR卫星涵盖GF-3、ALOS、Sentinel-1、RADARSAT、RCM等5类雷达影像,分辨率跨越3.5~12.5 m。数据集包含113 000对光学-SAR影像切片,通过人工精选控制点使整体配准残差小于1像素,为跨分辨率、跨模态的影像匹配算法提供了评测基准。

SpaceNet 6数据集[105]:Maxar商业星座高分辨率光学影像(WorldView系列,分辨率约0.3~0.5 m)与UMBRA高分辨率X波段商业SAR影像(分辨率约0.3~0.5 m)。此数据集专为跨模态全自动高精配准竞赛构建,包含大尺寸场景影像及密集人工精标同名特征点真值表。

WHU-SEN-City数据集[106]:Sentinel-1 SAR与Sentinel-2多光谱光学卫星,空间分辨率为10 m。聚焦全球数十座典型特大城市的核心建成区,利用精细局部残差矫正提取成对样本,高密度建筑物几何纹理极其丰富。

SOPatch数据集[107]:整合重构自SEN1-2、OSDataset及WHU-SEN-City的混合源光学-SAR影像对,分辨率覆盖1~10 m。数据集包含超过650 000对精选匹配图像块,并系统性地引入了人工位置平移扰动与可控仿射/旋转变换噪声。

MapData数据集[108]:栅格化高分辨率矢量电子地图与高分辨率可见光遥感正射影像。原始影像尺寸为7 000×5 000像素至20 000×15 000像素,覆盖全球233个具有代表性的地理采样区,

表 1　典型多模态遥感影像匹配基准数据集汇总

Table 1　Summary of Typical Benchmark Datasets for Multimodal Remote Sensing Image Matching

数据集名称	模态组合	数据集细节特性	获取途径
SEN1-2	SAR＋光学	数据覆盖全球各有人居住大洲的四季场景，影像分辨率为 10 m，尺寸为 256×256 像素，共 282 384 对影像	https://mediatum.ub.tum.de/1436631
SARptical	SAR＋光学	德国柏林密集城区的高层建筑与高复杂性城市场景，SAR 影像分辨率为 1 m，光学影像分辨率为 0.2 m，尺寸为 112×112 像素，超过 10 000 对影像	http://www.sipeo.bgu.tum.de/downloads/SARptical_data.zip
OSDataset 2.0	SAR＋光学	包含平原农田、复杂高层城市建筑群、山地水系、丘陵与水陆交界区等典型困难匹配场景，影像分辨率为 0.43 m，尺寸为 512×512 像素，共 6 476 对影像	https://radars.ac.cn/web/data/getDatanewsColumnId=992fa882-30f1-4bda-9553-36c56a8b457d
3MOS	5 源 SAR＋光学	场景主要覆盖密集城市建筑区、工业厂区、复杂立交枢纽及起伏地貌，影像分辨率为 3.5～12.5 m，尺寸为 256×256 像素，共 113 000 对影像	https://github.com/3M-OS/3MOS
SpaceNet 6	SAR＋光学	最初面向多源建筑物提取竞赛，后被用作跨模态匹配，影像分辨率为 0.3～0.5 m，尺寸为 900×900 像素，共 3 401 对影像	https://spacenet.ai/sn6-challenge/
WHU-SEN-City	SAR＋光学	以中国现代化城市环境为主，涵盖密集路网、工业厂房等，影像分辨率为 10 m，尺寸为 512×512 像素，共 13 674 对影像	https://github.com/whu-csl/WHU-SEN-City
SOPatch	SAR＋光学	涵盖城市建筑群、山地、农田裸地等典型地物，影像分辨率为 1～10 m，尺寸为 64×64 像素，超过 650 000 对影像	与发布者联系获取
MapData	电子地图＋光学	聚焦城市与城郊环境，包含高密度建筑足迹、规则与不规则道路网网格、绿地公园、交叉路口及水系，影像分辨率为 0.5～2 m，尺寸为 512×512 像素，共 121 781 对影像	https://github.com/PeihaoWu/MapGlue
RAMMR	光学、红外、SAR 等 8 种模态	覆盖平原农田、山区丘陵、城市密集建筑群、沿海与水陆交界面等多样化典型地貌，影像分辨率为 1～10 m，尺寸为 500×500 像素，共 24 对影像	https://github.com/RSmfmr/multimodal-dataset
NIR-VIS-RS	红外＋光学	重点选取对近红外波段极其敏感的场景，包括林地与密集植被、水体与滩涂、农田与温室等区域，影像分辨率为 10 m，尺寸为 512×512 像素，共 90 对影像	https://github.com/JasonLi-UCAS/NIR-VIS-RS
CityMMReg-data	光学、SAR、红外	重点覆盖城市路网、工业厂区、大型公共基础设施等，影像分辨率为 0.5～2 m，尺寸为 512×512 像素，超过 10 000 对影像	https://github.com/QAlana/CityMMReg-data
University-1652	跨平台（卫星、无人机、街景）	覆盖全球 1 652 所知名大学校园，影像分辨率为厘米级至亚米级，尺寸为 512×512 像素，超过 119 715 对影像	https://github.com/layumi/University1652-Baseline
WHU-OPT-SAR	SAR＋光学	地理范围覆盖中国湖北省及周边典型区域，包含平原水网、丘陵山地及密集城镇等多样地貌，影像分辨率为 5 m，尺寸为 5 556×3 704 像素，共 100 对影像	https://github.com/AmberHen/WHU-OPT-SAR-dataset

处理后共包含 121 781 对标准切片，采用自动化拓扑校正与人工复核相结合的混合真值标注。

RAMMR 数据集[79]：系统涵盖 8 种地理影像模态组合：多时相可见光、可见光-近红外、可见光-短波红外、近红外-短波红外、可见光-SAR、可见光-数字高程模型、可见光-电子地图、可见光-深度图。图像切片统一为 500×500 像素，给出经精密几何配准校正的标准参考矩阵与特征点集。

NIR-VIS-RS 数据集[109]：机载及微纳卫星平台采集的可见光与近红外波段影像切片。数据集聚焦极具挑战性的“弱纹理区”，包含数百对经严格筛选的成对样本，提供亚像素级特征对应点与单应变换矩阵标签。

CityMMReg-data 数据集[110]：高分辨率光学遥感影像、X/C 波段高分辨率 SAR 影像以及航空长波红外影像。包含两组子数据集：(1)跨模态

风格迁移训练集,支持图像重构网络训练;(2)复杂城市场景配准测试集,提供了高精度控制点对真值与严格几何映射变换矩阵,用于跨模态图像匹配与几何配准算法验证。

University-1652数据集[111]:涵盖全球1 652所大学校区的卫星级正射影像、无人机多高度/多视角倾斜影像以及地面车载街景影像,空间分辨率跨度从厘米级至亚米级,总计119 715对影像。数据集提供多层级地理空间拓扑对应标签,主要用于多模态跨视角无人机地理重定位与宽基线多视匹配。

WHU-OPT-SAR数据集[112]:国产高分一号光学多光谱影像与GF-3 C波段多极化SAR影像;空间重采样统一为5 m分辨率,由武汉大学团队构建并开源,覆盖中国湖北省范围。全集包含100对幅面达5 556×3 704像素的大尺寸成对影像,基于精密轨道参数及严格正射校正对齐,近年来被广泛作为大幅面遥感影像跨模态特征匹配的基准评估数据。

3 多模态遥感影像匹配定量评价指标

科学、客观、多维度的定量评价指标体系是检验多模态影像匹配算法的关键[38]。现有评价指标通常围绕匹配精度、匹配数量与有效性、空间几何分布质量以及计算效率与资源消耗4个核心维度构建,以全面反映算法性能。

1)匹配精度指标。其用于衡量预测匹配位置与真实对应点之间的几何偏离程度,通常依赖于参考同名点对或已验证的真实几何变换模型(如单应矩阵H、仿射变换矩阵或严密有理多项式系数模型)。

均方根误差(root mean square error,RMSE):衡量匹配点几何误差的最基础指标。设通过粗差点剔除后保留的匹配点对集合为$\left\{\left(p_i,q_i\right)\right\}_{i=1}^{N}$,其中$p_i$与$q_i$分别为参考影像与待匹配影像上的像元坐标,真实几何投影模型为T,则其计算公式为:

$$P_{\mathrm{RMSE}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\| T\left(p_i\right)-q_i\right\|_2} \tag{1}$$

RMSE直接反映了算法的配准精度能力,其值越小,表明配准后影像的几何对齐残差越低。

2)匹配数量与有效性指标。多模态影像协同处理任务中,算法输出的正确匹配点对的数量与比例直接决定了后续几何模型平差求解的可靠性。

正确匹配数(number of correct matches,NCM):匹配点总数(number of total matches,NTM)表示完成算法输出的点对总数,NCM则指通过预设误差容忍阈值ϵ验证为真实同名点的集合规模,ϵ通常在遥感任务中设定为1.5~3像素。其计算公式为:

$$N_{\mathrm{NCM}}=\sum_{i=1}^{N_{\mathrm{NTM}}}I\left(\left\| T\left(p_i\right)-q_i\right\|_2\leqslant\epsilon\right) \tag{2}$$

其中,$I(\ )$为计数函数。充足且可靠的NCM是稳定解算影像间几何变换关系的必要前提。

正确匹配率(correct matching ratio,CMR):正确匹配点数占匹配总点数的百分比,计算公式为:

$$P_{\mathrm{CMR}}=\frac{N_{\mathrm{NCM}}}{N_{\mathrm{NTM}}}\times 100\% \tag{3}$$

配准成功率(success rate,SR):在测试数据集当中,匹配成功的影像对占测试集总量的百分比。通常定义当某对影像解算出的$P_{\mathrm{RMS\,E}}\leqslant\epsilon$(例如$\epsilon$=3或5)时,判定该对影像配准成功。该指标直观体现了算法在大批量测试条件下的泛化鲁棒性。

3)空间几何分布质量指标。仅依赖点数与均方根误差往往会产生“局部过度聚簇而全图欠约束”的假象。若所有同名点均集中在局部高对比度区域(如规则道路交叉口或孤立强建筑),解算出的几何矩阵将在全局边缘区域产生严重发散。

空间均匀度分布系数:常用的量化手段是将影像空间划分为$K\times M$个均匀网格,统计落入各网格内的匹配点数量分布。利用信息熵或变异系数衡量网格点数密度的均匀程度,信息熵越高或变异系数越低,表明匹配点在几何空间上的支撑越均匀。

网格覆盖率(grid coverage ratio,GCR):统计存在至少一个正确匹配点($N_{\mathrm{NCM}}\geqslant 1$)的有效网格占全图划分总网格数的比例:

$$P_{\mathrm{GCR}}=\frac{\sum_{k=1}^{K\times M}I\left(N_{\mathrm{grid}_k}\geqslant 1\right)}{K\times M}\times 100\% \tag{4}$$

较高的GCR能够有效保障全局多项式或局部三角网仿射变换在全图范围内的几何刚性,防止无点约束区域产生局部扭曲。

4)计算效率与资源消耗指标。运行时间:分解记录特征检测、特征描述提取、特征匹配搜索

以及后处理平差等各阶段的绝对时间开销,对于高通量卫星地面站批量生产系统及星载计算平台而言是核心约束。

计算复杂度与硬件占用:针对深度学习端到端模型,通常以浮点运算量、网络可学习参数量以及图形处理器显存峰值消耗作为度量准则,用以评价模型在轻量化部署与实时处理任务中的工程可行性。

4　多模态遥感影像匹配的未来研究趋势

尽管多模态遥感影像匹配技术在特征表达建模、深度网络架构以及跨域泛化策略等方面取得了显著进展,但在面对海量多模态遥感数据的广域、全天候自主几何对准需求时,现有理论与方法仍面临诸多瓶颈。结合遥感对地观测系统的演变趋势与人工智能前沿技术,未来该领域的研究将主要聚焦于以下4个核心方向:

1)物理成像先验与深度特征表征的深度融合。当前主流的深度学习匹配方法多依赖于纯数据驱动的数据集拟合,在遇到未标注的极端地理场景或未知传感器对时,网络的几何感知能力容易出现退化。未来研究需将遥感传感器严格的成像物理机制显式或隐式地编码嵌入神经网络模型。通过在特征提取与注意力分配机制中引入传感器入射角、视线几何、光照方位角等物理先验,构建“物理机理约束-数据特征自适应”的双驱动匹配范式,从机理上抑制NRD对特征一致性的破坏。

2)遥感多模态大模型引导的通用匹配。随着具有广域时空感知能力的遥感视觉大模型及多模态基础模型的突破,利用大模型的强泛化语义抽象能力为低层几何匹配提供高层语义引导将成为新趋势。传统特征在面对跨度极大的异源影像时极易出现纹理失配。未来趋势应倾向于探索“大模型语义掩码+拓扑先验定位+细粒度亚像素几何微调”的级联架构,借助大模型的零样本与小样本迁移能力,实现跨任意传感器、分辨率的即插即用型通用匹配模型。

3)弱监督与仿真驱动的数据生成范式。针对多模态影像匹配中高精度亚像素同名点真值极度匮乏这一核心问题,未来研究一方面应深入探索自监督、弱监督及对比学习机制,摆脱网络训练对密集人工标注或有理多项式平差标签的强依赖;另一方面,依托物理渲染引擎与高保真生成式扩散模型,构建基于真实地表数字孪生模型的跨模态影像合成范式。通过在仿真环境中自动生成涵盖任意传感器视角、光照及气象条件的配准数据与精确坐标映射,实现“以虚代实、以虚强实”的模型预训练与实测域自适应迁移。

4)轻量化实时匹配与自主协同定位。随着低轨智能卫星星座与无人机自主集群协同观测任务的激增,针对海量遥感数据传输受限与突发灾害应急响应的时效性要求,边缘端实时处理已成为必然需求。现有多阶段端到端网络计算负载繁重、显存占用高,难以直接部署于低功耗星载或机载嵌入式计算平台。未来发展将重点聚焦于匹配模型的结构重参数化、稀疏注意力机制、网络量化剪枝以及硬件协同加速技术,开发低延迟的高精度轻量化跨模态匹配算法,推动多源载荷从地面后处理几何配准向在线准实时几何定位与校正的跨越。

5　结　语

随着传感器制造技术与对地观测系统的飞速发展,多源遥感数据融合已成为摄影测量与遥感领域的主流趋势。针对多模态遥感影像协同处理中几何一致性这一底层支撑技术,本文从基于区域的匹配方法、基于特征的匹配方法以及结合深度学习的匹配方法3个阵营及其精细化子类,对过去十年来多模态影像匹配方法的研究进展进行了系统、全面的回顾与分析。

多模态影像的几何一致性解算作为现代遥感应用的关键前提,由于其数据源间存在显著的非线性辐射差异与复杂的几何畸变,其共性问题受到了学术界的长期关注。总体而言,基于区域和特征的传统几何方法遵循真实世界的物理及视觉原理,具有低软硬件依赖性与强解释性,在当前仍展现出极大的研究潜力和即插即用的工程实用价值。此外,深度学习技术强大的非线性特征表达能力为多模态匹配带来了全新思路。虽然深度网络在模拟人类视知觉由粗及精、跨模态流形映射等层面表现优异,但其在工程领域落地时仍受限于大规模高精度标注多模态数据集的匮乏,且对数据分布变化高度敏感导致跨模态泛化性不足,参数量激增亦带来软硬件环境的适配性挑战,因此,目前多模态匹配依然以传统特征/区域几何方法为主。但不可否认,深度学习在跨模态协同生成与自监督轻量化网络结构转型中依然具有巨大的潜力。

回望十余年的发展,多模态遥感影像匹配正朝着处理方式高效化、应对复杂辐射、几何畸变鲁棒化以及多源信息融合一体化的方向快速前进。从早期的同源匹配方法到如今逐步成型的多源通用算法体系,多模态影像匹配正成为现代遥感应用中的新主力,并在数据融合、目标识别与变化检测等诸多国家发展和科学探索应用中发挥更加显著的作用。在这种大背景下,越来越多不同学科的学者会参与到多模态影像匹配与处理的发展浪潮中,不断拓宽应用边界,注入新的技术和思想,多模态遥感协同观测与处理技术将迎来全方位快速发展期。

参 考 文 献

[1] 李德仁,张洪云,金文杰. 新基建时代地球空间信息学的使命[J]. 武汉大学学报(信息科学版),2022,47(10):1515-1522.
LI Deren, ZHANG Hongyun, JIN Wenjie. Mission of Geospatial Information Science in the Era of New Infrastructure[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(10): 1515-1522.

[2] 李德仁. 从珞珈系列卫星到东方慧眼星座[J]. 武汉大学学报(信息科学版),2023,48(10):1557-1565.
LI Deren. From the Luojia Series Satellites to the Oriental Smart Eye Constellation[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(10): 1557-1565.

[3] 李德仁,王密,申文斌,等. 论人工智能时代的测绘学发展[J]. 测绘学报,2025,54(12):2107-2115.
LI Deren, WANG Mi, SHEN Wenbin, et al. On the Development of Surveying and Mapping in the Era of Artificia Intelligence[J]. *Acta Geodaetica et Cartographica Sinica*, 2025, 54(12): 2107-2115.

[4] 冷成财,洪雅萌. 加强局部自相似性描述符在多模态遥感图像匹配中的应用[J]. 西北大学学报(自然科学版),2025,55(6):1253-1266.
LENG Chengcai, HONG Yameng. Application of an Enhanced Local Self-Similarity Descriptor for Multimodal Remote Sensing Images Matching[J]. *Journal of Northwest University* (*Natural Science Edition*), 2025, 55(6): 1253-1266.

[5] 杨永民,邱建秀,苏红波,等. 基于热红外的4种土壤含水量估算方法对比[J]. 红外与毫米波学报,2018,37(4):459.
YANG Yongmin, QIU Jianxiu, SU Hongbo, et al. Estimation of Surface Soil Moisture Based on Thermal Remote Sensing: Intercomparison of Four Methods[J]. *Journal of Infrared and Millimeter Waves*, 2018, 37(4): 459.

[6] 曹思珩,陈旭,杨仁敏,等. 基于可见光-近红外-短波红外的互花米草湿地土壤含水量预测[J]. 湿地科学,2025,23(5):905-914.
CAO Siheng, CHEN Xu, YANG Renmin, et al. Predicting Soil Water Content in Spartina Alterniflora Wetland with VIS-NIR-SWIR Spectroscopy and Random Forest[J]. *Wetland Science*, 2025, 23(5): 905-914.

[7] JIANG Q, TANG Z G, ZHOU L H, et al. Mapping Paddy Rice Planting Area in Dongting Lake Area Combining Time Series Sentinel-1 and Sentinel-2 Images[J]. *Remote Sensing*, 2023, 15(11): 2794.

[8] LIU Q, REN K, MENG X C, et al. Domain Adaptive Cross Reconstruction for Change Detection of Heterogeneous Remote Sensing Images via a Feedback Guidance Mechanism[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4507216.

[9] 刘慧,何玉青,胡秀清,等. 联合微光辅助红外遥感数据的夜间微小火点识别算法[J]. 光学学报,2024,44(8):0828001.
LIU Hui, HE Yuqing, HU Xiuqing, et al. Nighttime Tiny Fire Recognition Algorithm by Joint Low-Light-Assisted Infrared Remote Sensing Data[J]. *Acta Optica Sinica*, 2024, 44(8): 0828001.

[10] 王密,樊仲藜,皮英冬,等. 一阶高斯方向可调滤波器引导的多源卫星遥感影像配准方法[J]. 武汉大学学报(信息科学版),2024,49(5):681-690.
WANG Mi, FAN Zhongli, PI Yingdong, et al. First-Order Gaussian Steerable Filter-Guided Multisource Satellite Remote Sensing Image Registration Method[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(5): 681-690.

[11] 刘玉轩,张力,樊仲藜,等. 面向多模态图像的非线性辐射和几何不变匹配[J]. 光学学报,2025,45(6):0628010.
LIU Yuxuan, ZHANG Li, FAN Zhongli, et al. Nonlinear Radiation and Geometric Invariant Matching for Multimodal Imagery[J]. *Acta Optica Sinica*, 2025, 45(6): 0628010.

[12] 南轲,齐华,叶沅鑫. 深度卷积特征表达的多模态遥感影像模板匹配方法[J]. 测绘学报,2019,48(6):727-736.
NAN Ke, QI Hua, YE Yuanxin. A Template Matching Method of Multimodal Remote Sensing Images Based on Deep Convolutional Feature Repre-

sentation[J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(6): 727-736.

[13] FAN Z L, ZHANG L, LIU Y X, et al. Exploiting High Geopositioning Accuracy of SAR Data to Obtain Accurate Geometric Orientation of Optical Satellite Images[J]. *Remote Sensing*, 2021, 13(17): 3535.

[14] 张祖勋, 张剑清, 廖明生, 等. 遥感影像的高精度自动配准[J]. 武汉测绘科技大学学报, 1998,28(4): 320-323.
ZHANG Zuxun, ZHANG Jianqing, LIAO Mingsheng, et al. Automatic Precision Registration of Multi Resolution Remote Sensing Imagery[J]. *Geomatics and Information Science of Wuhan University*, 1998,23(4): 320-323.

[15] 王蒙蒙, 叶沅鑫, 朱柏, 等. 基于空间约束和结构特征的光学与SAR影像配准[J]. 武汉大学学报(信息科学版), 2022, 47(1): 141-148.
WANG Mengmeng, YE Yuanxin, ZHU Bai, et al. An Automatic Registration Method for Optical and SAR Images Based on Spatial Constraint and Structure Features[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(1): 141-148.

[16] 彭韬, 杨培臻, 周亮, 等. 一种旋转和尺度不变的异源景象匹配方法[J]. 弹箭与制导学报, 2024, 44(5): 38-46.
PENG Tao, YANG Peizhen, ZHOU Liang, et al. A Rotation and Scale Invariant Heterogeneous Sence Matching Approach [J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2024, 44(5): 38-46.

[17] 叶沅鑫, 王蒙蒙, 杨超, 等. 基于方向相位稠密特征的多传感器遥感影像配准方法和系统[J]. 遥感学报, 2024, 28(6): 1525-1538.
YE Yuanxin, WANG Mengmeng, YANG Chao, et al. Multisensor Remote Sensing Registration Method and System Based on Dense Feature of Orientated Phase[J]. *National Remote Sensing Bulletin*, 2024, 28(6): 1525-1538.

[18] 朱柏,叶沅鑫. 多模态遥感图像配准方法研究综述[J]. 中国图象图形学报,2024, 29(8): 2137-2161.
ZHU Bai, YE Yuanxin. Multimodal Remote Sensing Image Registration: A Survey[J]. *Journal of Image and Graphics*, 2024, 29(8): 2137-2161.

[19] ZHANG Y J, WU P H, YAO Y X, et al. Multimodal Remote Sensing Image Robust Matching Based on Second-Order Tensor Orientation Feature Transformation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4701314.

[20] ZITOVÁ B, FLUSSER J. Image Registration Methods: a Survey[J]. *Image and Vision Computing*, 2003, 21(11): 977-1000.

[21] JIANG X Y, MA J Y, XIAO G B, et al. A Review of Multimodal Image Matching: Methods and Applications [J]. *Information Fusion*, 2021, 73: 22-71.

[22] SOMMERVOLD O, GAZZEA M, ARGHANDEH R. A Survey on SAR and Optical Satellite Image Registration[J]. *Remote Sensing*, 2023, 15(3): 850.

[23] FAN Z L, LIU Y X, LIU Y X, et al. 3MRS: an Effective Coarse-to-Fine Matching Method for Multimodal Remote Sensing Imagery[J]. *Remote Sensing*, 2022, 14(3): 478.

[24] LIU Y X, FAN Z L, LI Z, et al. Comparative Analysis of Mainstream Multi-source Image Matching Methods on Remote Sensing Imagery [C]//IEEE International Geoscience and Remote Sensing Symposium. Athens, Greece, 2024.

[25] YE Y X, SHEN L. Hopc: a Novel Similarity Metric Based on Geometric Structural Properties for Multi-modal Remote Sensing Image Matching[J]. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, 3: 9-16.

[26] 张力, 张祖勋, 张剑清. Wallis滤波在影像匹配中的应用[J]. 武汉测绘科技大学学报, 1999, 24(1): 24-27.
ZHANG Li, ZHANG Zuxun, ZHANG Jianqing. The Image Matching Based on Wallis Filtering[J]. *Journal of Wuhan Technical University of Surveying and Mapping*, 1999, 24(1): 24-27.

[27] LEWIS J P. Fast Normalized Cross-Correlation[J]. *Vision Interface*, 1995,10(1): 120-123.

[28] SZELISKI R. Image Alignment and Stitching: A Tutorial[J]. *Foundations and Trends in Computer Graphics and Vision*, 2007, 2(1): 1-104.

[29] BARNEA D I, SILVERMAN H F. A Class of Algorithms for Fast Digital Image Registration [J]. *IEEE Transactions on Computers*, 1972, C-21(2): 179-186.

[30] SHEN D G,DAVATZIKOS C. HAMMER:Hierarchical Attribute Matching Mechanism for Elastic Registration[J]. *IEEE Transactions on Medical Imaging*, 2002, 21(11): 1421-1439.

[31] HIRSCHMULLER H, SCHARSTEIN D. Evaluation of Stereo Matching Costs on Images with Radiometric Differences[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(9): 1582-1599.

[32] VIOLA P, WELLS III W M. Alignment by Maxi-

mization of Mutual Information[J]. *International Journal of Computer Vision*,1997,24(2):137-154.

[33] MAES F, COLLIGNON A, VANDERMEULEN D, et al. Multimodality Image Registration by Maximization of Mutual Information[J]. *IEEE Transactions on Medical Imaging*, 1997, 16(2): 187-198.

[34] XU X C,LI X,LIU X P,et al. Multimodal Registration of Remotely Sensed Images Based on Jeffrey's Divergence[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 122: 97-115.

[35] CHEN S H, LI X R, ZHAO L Y, et al. Medium-Low Resolution Multisource Remote Sensing Image Registration Based on SIFT and Robust Regional Mutual Information[J]. *International Journal of Remote Sensing*, 2018, 39(10): 3215-3242.

[36] LIU X P, CHEN S L, ZHUO L, et al. Multi-sensor Image Registration by Combining Local Self-similarity Matching and Mutual Information [J]. *Frontiers of Earth Science*, 2018, 12(4): 779-790.

[37] PAUL S, PATI U C. High-Resolution Optical-to-SAR Image Registration Using Mutual Information and SPSA Optimisation[J]. *IET Image Processing*, 2021, 15(6): 1319-1331.

[38] FAN Z L, PI Y D, WANG M, et al. GLS-MIFT: A Modality Invariant Feature Transform with Global-to-Local Searching [J]. *Information Fusion*, 2024, 105: 102252.

[39] YE Y X, BRUZZONE L, SHAN J, et al. Fast and Robust Structure-Based Multimodal Geospatial Image Matching [C]//IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 2017.

[40] YE Y X,SHAN J. A Local Descriptor Based Registration Method for Multispectral Remote Sensing Images with Non-linear Intensity Differences [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 90: 83-95.

[41] SHECHTMAN E, IRANI M. Matching Local Self-Similarities Across Images and Videos [C]// IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA, 2007.

[42] YI Z, ZHIGUO C, YANG X. Multi-spectral Remote Image Registration Based on SIFT[J]. *Electronics Letters*, 2008, 44(2): 107-108.

[43] YE Y X, SHEN L, HAO M, et al. Robust Optical-to-SAR Image Matching Based on Shape Properties [J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(4): 564-568.

[44] XIONG X, XU Q, JIN G W, et al. Rank-Based Local Self-Similarity Descriptor for Optical-to-SAR Image Matching[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(10): 1742-1746

[45] KOVESI P. Phase Congruency: a Low-Level Image Invariant[J]. *Psychological Research*, 2000, 64(2): 136-148.

[46] YE Y X, SHAN J, BRUZZONE L, et al. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2941-2958.

[47] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005.

[48] FAN J W, WU Y, LI M, et al. SAR and Optical Image Registration Using Nonlinear Diffusion and Phase Congruency Structural Descriptor[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(9): 5368-5379.

[49] HARRIS C, STEPHENS M. A Combined Corner and Edge Detector[C]//The Alvey Vision Conference, Manchester, 1988.

[50] YE Y X, BRUZZONE L, SHAN J, et al. Fast and Robust Matching for Multimodal Remote Sensing Image Registration[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 9059-9070.

[51] 樊仲藜,张力,王庆栋,等. SAR影像和光学影像梯度方向加权的快速匹配方法[J]. 测绘学报,2021,50(10):1390-1403.
FAN Zhongli, ZHANG Li, WANG Qingdong, et al. A Fast Matching Method of SAR and Optical Images Using Angular Weighted Orientated Gradients[J]. *Acta Geodaetica et Cartographica Sinica*, 2021, 50(10): 1390-1403.

[52] ZHU B, YE Y X, ZHOU L, et al. Robust Registration of Aerial Images and LiDAR Data Using Spatial Constraints and Gabor Structural Features[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 181: 129-147.

[53] YE Y X, ZHU B, TANG T F, et al. A Robust Multimodal Remote Sensing Image Registration Method and System Using Steerable Filters with First- and Second-Order Gradients[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 188: 331-350.

[54] 杨超,刘畅,唐腾峰,等. 基于深度特征重构增强的光学和SAR图像鲁棒匹配[J]. 遥感学报,2025,29(8):2616-2626.
YANG Chao, LIU Chang, TANG Tengfeng, et

al. Robust Matching of Optical and SAR Images Based on Deep Feature Reconstruction Enhancement [J]. *National Remote Sensing Bulletin*, 2025, 29(8): 2616-2626.

[55] YE Y X, WANG M M, HAO S Y, et al. A Novel Keypoint Detector Combining Corners and Blobs for Remote Sensing Image Registration[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(3): 451-455.

[56] LI H, MANJUNATH B S, MITRA S K. A Contour-Based Approach to Multisensor Image Registration[J]. *IEEE Transactions on Image Processing*, 1995, 4(3): 320-334.

[57] GONCALVES H, CORTE-REAL L, GONCALVES J A. Automatic Image Registration Through Image Segmentation and SIFT[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(7): 2589-2600.

[58] 李加元. 鲁棒性遥感影像特征匹配关键问题研究[D]. 武汉: 武汉大学, 2018.
LI Jiayuan. Research on Key Techniques of Robust Remote Sensing Image Feature Matching[D]. Wuhan: Wuhan University, 2018.

[59] 凌霄. 基于多重约束的多源光学卫星影像自动匹配方法研究[D]. 武汉: 武汉大学, 2017.
LING Xiao. Automatic Matching of Multi-sources Optical Satellite Images Based on Multiple Constraints[D]. Wuhan: Wuhan University, 2017.

[60] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.

[61] 叶沅鑫, 单杰, 熊金鑫, 等. 一种结合SIFT和边缘信息的多源遥感影像匹配方法[J]. 武汉大学学报(信息科学版), 2013, 38(10): 1148-1151.
YE Yuanxin, SHAN Jie, XIONG Jinxin, et al. A Matching Method Combining SIFT and Edge Information for Multi-source Remote Sensing Images [J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(10): 1148-1151.

[62] MA W P, WEN Z L, WU Y, et al. Remote Sensing Image Registration with Modified SIFT and Enhanced Feature Matching[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(1): 3-7.

[63] XIANG Y M, WANG F, YOU H J. OS-SIFT: a Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(6): 3078-3090.

[64] AGUILERA C A, SAPPA A D, TOLEDO R. LGHD: A Feature Descriptor for Matching Across Non-Linear Intensity Variations[C]//IEEE International Conference on Image Processing, Quebec City, QC, Canada, 2015.

[65] ROSTEN E, PORTER R, DRUMMOND T. Faster and Better: A Machine Learning Approach to Corner Detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(1): 105-119.

[66] LI J Y, HU Q W, AI M Y. RIFT: Multi-modal Image Matching Based on Radiation-Variation Insensitive Feature Transform[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3296-3310.

[67] XIONG X, JIN G W, XU Q, et al. Robust Registration Algorithm for Optical and SAR Images Based on Adjacent Self-Similarity Feature [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5233117.

[68] LI J Y, XU W Y, SHI P C, et al. LNIFT: Locally Normalized Image for Rotation Invariant Multimodal Feature Matching[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5621314.

[69] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An Efficient Alternative to SIFT or SURF [C]//International Conference on Computer Vision, Barcelona, Spain, 2011.

[70] FAN Z L, WANG M, PI Y D, et al. A Robust Oriented Filter-Based Matching Method for Multisource, Multitemporal Remote Sensing Images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4703316.

[71] ZHANG Y J, YAO Y X, WAN Y, et al. Histogram of the Orientation of the Weighted Phase Descriptor for Multi-modal Remote Sensing Image Matching [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 196: 1-15.

[72] HOU Z L, LIU Y X, ZHANG L. POS-GIFT: a Geometric and Intensity-Invariant Feature Transformation for Multimodal Images[J]. *Information Fusion*, 2024, 102: 102027.

[73] TOLA E, LEPETIT V, FUA P. DAISY: an Efficient Dense Descriptor Applied to Wide-Baseline Stereo[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 815-830.

[74] JOVHARI N, SEDAGHAT A, SHAH-HOSSEINI R, et al. Noise-Robust Multimodal Remote Sensing Image Matching via Geometric Analysis of Embedded Pre-Trained Manifolds [J]. *Neurocomputing*, 2025, 638: 130150.

[75] ALCANTARILLA P F, BARTOLI A, DAVISON A J. KAZE Features[C]//European Conferen-

ce on Computer Vision, Florence, Italy, 2012.

[76] SEDAGHAT A, MOHAMMADI N. Uniform Competency-Based Local Feature Extraction for Remote Sensing Images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 135: 142-157.

[77] ZHENG C Y, LI S S, WANG C Y, et al. MSG: Robust Multimodal Remote Sensing Image Matching Using Side Window Gaussian Space[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4706223.

[78] ZHANG Z Z, WEI P C, SHAO Z F, et al. MS-POFT: Multiscale Phase-Orientation Guided Feature Transform for Multimodal Image Matching[J]. *Geo-spatial Information Science*, 2026, 29(1): 274-296.

[79] REN K, SUN W W, MENG X C, et al. A Robust and Accurate Feature Matching Method for Multimodal Geographic Images Spatial Registration[J]. *Geo-spatial Information Science*, 2025, 28(3): 1014-1033.

[80] ZHANG H, LEI L, NI W P, et al. Optical and SAR Image Matching Using Pixelwise Deep Dense Features[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6000705.

[81] HUGHES L H, MARCOS D, LOBRY S, et al. A Deep Learning Framework for Matching of SAR and Optical Imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 169: 166-179.

[82] YE Y X, YANG C, GONG G Q, et al. Robust Optical and SAR Image Matching Using Attention-Enhanced Structural Features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5610212.

[83] ZHU J Y, PARK T, ISOLA P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks[C]//IEEE International Conference on Computer Vision, Venice, Italy, 2017.

[84] LIAGHAT A, SADEGH HELFROUSH M, NOROUZI J, et al. OS-CycleGAN: Modified CycleGAN-Based Descriptors for Optical and SAR Image Matching[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5214918.

[85] LI W, WENG D S, GAO C Z, et al. SwinMatcher: Universal Cross-Modal Remote Sensing Image Matching with Interactive Swin Transformer[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4705916.

[86] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[C]// IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 2021.

[87] WANG Q, ZHANG J M, YANG K L, et al. MatchFormer: Interleaving Attention in Transformers for Feature Matching[C]//Asian Conference on Computer Vision, Macao, China, 2022.

[88] XIE H S, ZHANG Y K, QIU J H, et al. Semantics Lead All: Towards Unified Image Registration and Fusion from a Semantic Perspective[J]. *Information Fusion*, 2023, 98: 101835.

[89] ZHANG Z Z, WEI P C, WU P L, et al. CM2-STNet: Cross-Modal Image Matching with Modal-Adaptive Feature Modulation and Sparse Transformer Fusion [J]. *Information Fusion*, 2026, 127: 103750.

[90] LIU Y P, LAI W P, GU Y L, et al. Learning Modality-Invariant Feature for Multimodal Image Matching *via* Knowledge Distillation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5623115.

[91] OQUAB M, DARCET T, MOUTAKANNI T, et al. DINOv2: Learning Robust Visual Features Without Supervision[PP/OL]. V2. arXiv (2024-02-02)[2026-05-05]. https://doi.org/10.48550/arXiv.2304.07193.

[92] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: Self-Supervised Interest Point Detection and Description[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 2018.

[93] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: Learning Feature Matching With Graph Neural Networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020.

[94] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016.

[95] WENG D S, LI W, GAO C Z, et al. DGIM: Cascaded Dynamic Data Generation for Robust Cross-Modal Image Matching[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4708616.

[96] XIANG Y M, TAO R S, WANG F, et al. Automatic Registration of Optical and SAR Images via Improved Phase Congruency Model[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 5847-5861.

[97] QUAN D, WANG Z, LV C H, et al. LM-Net: A Lightweight Matching Network for Remote Sensing Image Matching and Registration[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-13.

[98] LI L Z, HAN L, LIU M, et al. SAR-Optical Image Matching with Semantic Position Probability Distribution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.

[99] WANG W, QUAN D, NING H Y, et al. Multi-expert Learning Framework with the State Space Model for Optical and SAR Image Registration[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2026, 64: 5204018.

[100] GAO T, LAN C Z, HUANG W J, et al. SFA-Net: A SAM-Guided Focused Attention Network for Multimodal Remote Sensing Image Matching[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 223: 188-206.

[101] SCHMITT M, HUGHES L H, ZHU X X. The SEN1-2 Dataset for Deep Learning in SAR-Optical Data Fusion[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2018, IV-1: 141-146.

[102] WANG Y Y, ZHU X X. The SARptical Dataset for Joint Analysis of SAR and Optical Image in Dense Urban Area[C]//IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018.

[103] XIANG Y M, CHEN J Y, HONG Z H, et al. OS-Dataset2.0: SAR-Optical Image Matching Dataset and Evaluation Benchmark[J]. *Journal of Radars*, 2025: 15(3): 983-995.

[104] YE Y B, TENG X C, YANG H R, et al. 3MOS: A Multi-source, Multi-resolution, and Multi-scene Optical-SAR Dataset with Insights for Multi-modal Image Matching[J]. *Visual Intelligence*, 2025, 3(1): 19.

[105] HÄNSCH R, ARNDT J, POTNIS A, et al. SpaceNet 9: Cross-Sensor Alignment of Optical and SAR Imagery[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2026, 19: 11491-11502.

[106] WANG L, XU X, YU Y, et al. SAR-to-Optical Image Translation Using Supervised Cycle-Consistent Adversarial Networks[J]. *IEEE Access*, 2019, 7: 129136-129149.

[107] XU W Y, YUAN X H, HU Q W, et al. SAR-Optical Feature Matching: A Large-Scale Patch Dataset and a Deep Local Descriptor[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2023, 122: 103433.

[108] WU P H, YAO Y X, ZHANG W F, et al. MapGlue: Multimodal Remote Sensing Image Matching[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2026, 240: 166-183.

[109] LI J Q, BI G L, WANG X Z, et al. Radiation-Variation Insensitive Coarse-to-Fine Image Registration for Infrared and Visible Remote Sensing Based on Zero-Shot Learning[J]. *Remote Sensing*, 2024, 16(2): 214.

[110] 李晓奎, 杨树文, 李子源, 等. 耦合模态转换和三维空间关系约束的多模高分遥感影像配准方法[J]. 遥感学报, 2026, 30(8): 2301-2316.
LI Xiaokui, YANG Shuwen, LI Ziyuan, et al. Registration Method for Multimodal High-Resolution Remote Sensing Images Coupling Modal Transformation and 3D Spatial Relationship Constraints[J]. *National Remote Sensing Bulletin*, 2026, 30(8): 2301-2316.

[111] ZHENG Z D, WEI Y C, YANG Y. University-1652: A Multi-view Multi-source Benchmark for Drone-Based Geo-Localization [C]//The 28th ACM International Conference on Multimedia, Seattle, WA, USA, 2020.

[112] LI X, ZHANG G, CUI H, et al. MCANet: a Joint Semantic Segmentation Framework of Optical and SAR Images for Land Use Classification[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 106: 102638.