



## 景象匹配无人机视觉定位

袁媛 孙柏 刘赶超

### Drone-based Scene Matching Visual Geo-localization

YUAN Yuan, SUN Bo, LIU Gan-Chao

在线阅读 View online: <https://doi.org/10.16383/j.aas.c230778>

---

## 您可能感兴趣的其他文章

### 基于旋翼无人机近地面空间应急物联网节点动态协同部署

Dynamic Cooperative Deployment of Emergency Internet of Things Near Ground Space Based on Drone

自动化学报. 2021, 47(8): 2002–2015 <https://doi.org/10.16383/j.aas.c180146>

### 面向无人机自主着陆的视觉感知与位姿估计方法综述

A Survey on Vision-based Sensing and Pose Estimation Methods for UAV Autonomous Landing

自动化学报. 2024, 50(7): 1284–1304 <https://doi.org/10.16383/j.aas.c230557>

### 无人机反应式扰动流体路径规划

UAV Reactive Interfered Fluid Path Planning

自动化学报. 2023, 49(2): 272–287 <https://doi.org/10.16383/j.aas.c210231>

### 基于显著图融合的无人机载热红外图像目标检测方法

Object Detection Method Based on Saliency Map Fusion for UAV-borne Thermal Images

自动化学报. 2021, 47(9): 2120–2131 <https://doi.org/10.16383/j.aas.c200021>

### 舰载无人机自主着舰回收制导与控制研究进展

Research Development in Autonomous Carrier-Landing/Ship-Recovery Guidance and Control of Unmanned Aerial Vehicles

自动化学报. 2019, 45(4): 669–681 <https://doi.org/10.16383/j.aas.2018.c170261>

### 面向轻轨的高精度实时视觉定位方法

Real-time Visual Localization Method for Light-rail With High Accuracy

自动化学报. 2021, 47(9): 2194–2204 <https://doi.org/10.16383/j.aas.c200009>

# 景象匹配无人机视觉定位

袁媛<sup>1,2</sup> 孙柏<sup>1,2</sup> 刘赶超<sup>2</sup>

**摘要** 无人机因其高度的灵活性,在临地安防、灾后救援、地质勘测、农业植保等领域发挥着重要作用,因此受到越来越多的关注.定位导航作为无人机中的关键技术,对于无人机能否顺利执行任务至关重要.当前主要的定位导航算法包括全球导航卫星系统、惯性定位以及景象匹配定位导航等.首先,景象匹配定位导航方法利用计算机视觉技术,对无人机飞行时采集的航空影像进行数字化特征编码;随后,通过构建相似性度量与检索模型,将航空影像特征与预先获取的遥感地图库特征进行相似性度量,从而完成景象匹配;最后,根据无人机航空影像与遥感卫星地图的匹配结果,获取相应的地理位置信息,并将其更新为无人机的定位结果.景象匹配定位导航方法摆脱了定位系统对定位信号的依赖,实现了无人机飞行定位的自主化.鉴于此,以景象匹配算法中的特征提取方式为线索,分别针对基于模板匹配、基于手工特征以及基于度量学习的景象匹配,梳理其发展过程,并总结景象匹配定位导航方法中的关键问题.最后,针对景象匹配算法的发展现状,总结无人机景象匹配定位方法中亟待解决的问题.

**关键词** 临地安防, 无人机, 视觉定位, 景象匹配, 度量学习, 多视角变化

**引用格式** 袁媛, 孙柏, 刘赶超. 景象匹配无人机视觉定位. 自动化学报, 2025, 51(2): 287–311

**DOI** 10.16383/j.aas.c230778 **CSTR** 32138.14.j.aas.c230778

## Drone-based Scene Matching Visual Geo-localization

YUAN Yuan<sup>1,2</sup> SUN Bo<sup>1,2</sup> LIU Gan-Chao<sup>2</sup>

**Abstract** Drones play an important role in vicinagearth security, post-disaster rescue, geological survey, agricultural plant protection, and other fields due to their high flexibility, and they receive increasing attention. As a key technology in drones, positioning and navigation are crucial for whether the drone can successfully perform tasks. Currently, the main positioning and navigation algorithms include the global navigation satellite system, inertial positioning, and scene matching positioning and navigation. Among them, the scene matching positioning and navigation method uses computer vision technology to encode the digital features of aerial images collected during the flight of drones. Then, by constructing a similarity measurement and retrieval model, it measures the similarity between the aerial image features and the pre-obtained remote sensing map library features to complete the scene matching. Finally, based on the matching results of drone aerial images and remote sensing satellite maps, it obtains the corresponding geographic position information and updates it as the positioning result of the drone. The scene matching positioning and navigation method eliminates the dependence of the positioning system on positioning signals and realizes the autonomy of drone flight positioning. This paper follows the feature extraction methods in the scene matching algorithm and outlines the development process of scene matching based on template matching, manual feature-based, and metric learning-based approaches while summarizing the key problems in the positioning and navigation methods of scene matching. Finally, this paper summarizes the urgent problems that need to be solved in drone scene matching localization methods based on the current development status of scene matching algorithms.

**Key words** Vicinagearth security, drone, visual geo-localization, scene matching, metric learning, multi-view changes

**Citation** Yuan Yuan, Sun Bo, Liu Gan-Chao. Drone-based scene matching visual geo-localization. *Acta Automatica Sinica*, 2025, 51(2): 287–311

收稿日期 2023-12-19 录用日期 2024-06-20  
Manuscript received December 19, 2023; accepted June 20, 2024

国家自然科学基金 (62273282, 62201471) 资助  
Supported by National Natural Science Foundation of China (62273282, 62201471)

本文责任编辑 赖剑煌  
Recommended by Associate Editor LAI Jian-Huang  
1. 西北工业大学计算机学院 西安 710072 2. 西北工业大学光电与人工智能研究院 西安 710072

无人驾驶飞机简称“无人机”,是利用无线电遥控设备或自备的程序控制装置操纵的飞行器.因其成本较低、操作灵活等优势,在临地安防<sup>[1]</sup>、灾害救援、地质勘测、农业植保以及交通执法等领域发挥

1. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072 2. School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072

越来越重要的作用<sup>[2]</sup>。“十四五”民用航空发展规划第3节指出,需要创新无人机产业生态,持续推动无人驾驶航空试验区建设和运行,面向运行场景,基于运行风险,开展运行理论、风险评估和技术验证等研究,探索符合无人驾驶特点的监管和服务模式<sup>[3]</sup>。

作为智能航空器,无人机执行复杂任务的前提是精确的定位导航<sup>[4]</sup>,即判断当前所处位置的经纬度、高度、姿态等信息。然后,规划出无人机飞至目标的飞行轨迹,实现后续任务。针对无人机定位算法展开研究,将提升无人机定位导航的韧性,扩展无人机实际应用的场景,增强无人机任务执行的效率。经过梳理,无人机定位算法主要包括以下三类:

1) 以惯性测量单元 (Inertial measurement unit, IMU) 数据解算和视觉里程计 (Visual odometry, VO) 为代表的惯性导航系统 (Inertial navigation system, INS)<sup>[5]</sup>。惯性导航系统通过陀螺仪或视频连续帧获取运动数据,并推理解算当前位置信息,确定无人机在惯性参考坐标系中的相对运动,然后计算出无人机在惯性参考坐标系中的地理位置。该方法通过连续相对运动的解算,推断当前无人机位置与起始位置的相对位置变化,在小范围运行时,能够提供较为准确的位置估计。但是,由于位置估计依赖连续的计算,长时间运行会出现严重的累积误差,因此该方法常与卫星或景象匹配视觉定位系统相结合,形成组合定位算法。

2) 以全球定位系统 (Global positioning system, GPS) 和北斗卫星导航系统 (Beidou navigation satellite system, BDS) 为代表的全球导航卫星系统 (Global navigation satellite system, GNSS)<sup>[6]</sup>。全球导航卫星系统通过接收卫星定位信号,解算差分信号得到高精度的地理位置信息。GNSS 自问世以来,就以其高精度、全天候、全球覆盖、方便灵活等特点,成功应用于各种领域。但是,该方法的定位依赖于卫星信号,因此在复杂电磁波干扰、遮挡时会因为定位信号传输过程中受到干扰或遮蔽,导致无法精确计算出地理位置信息。

3) 以空-天多源影像匹配为代表的景象匹配定位系统 (Scene matching system, SMS)<sup>[7]</sup>。景象匹配定位系统首先建立无人机所处区域的卫星地图数据库,然后通过对无人机视觉传感器获取的航空影像进行数字化编码,经过匹配度量算法检索到与卫星地图特征编码数据库中匹配的图像,最后获取其对应的地理位置信息,将此地理位置信息作为无人机当前位置进行位置更新。该方法的优势在于定位判断独立性强,不依赖于卫星定位信号。并且,作为绝对定位方法,景象匹配定位系统不产生累计误差。但是,此类算法作为无人机定位的新兴算法代表,

其定位精度严重依赖于匹配算法的准确性,目前相较于全球导航卫星系统,仍有巨大的精度提升空间。

上述三类方法中,惯性导航系统作为相对定位算法,通过相对位置变换,推算当前位置。因此,在长时间的飞行任务中,随着时间的增加会不断累积误差。通常,此类算法作为组合定位算法中辅助定位,在短时飞行过程中提供定位修正。全球定位系统与景象匹配定位系统作为绝对定位算法,能够通过定位导航算法计算直接获得无人机所处的地理位置,定位精度只取决于单次定位误差。因此,此类定位方式在无人机长时间飞行过程中的定位具有较高可靠性,如表1所示。

表1 定位算法对比结果

Table 1 Comparison results of localization algorithms

分类	方法	精度	抗干扰性	实时性	发展现状
相对定位	INS	短时精度高	强	强	较为成熟
绝对定位	GPS	较高	弱	强	成熟
绝对定位	SMS	较低	强	强	亟待研究

与同属绝对定位方法的全球定位系统相比,景象匹配定位系统凭借基于度量学习的方法能够挖掘图像的全局语义信息,从而实现自主定位。这种优势使得景象匹配定位系统能够在无人机应用中发挥重要作用,克服了传统全球定位系统对于卫星定位信号的依赖。在复杂环境下,如城市建筑密集区域、山区或森林等地区,卫星信号可能受到遮挡或干扰,而景象匹配定位系统则能够依靠图像中的语义信息进行自主定位,从而实现了无人机在这些地区的可靠定位。这种自主定位能力不仅拓宽了无人机的应用场景,使得其能够在更广泛的环境下执行任务,同时也极大地提升了无人机的智能化程度。

在无人机视觉定位领域,景象匹配技术具有重要性和研究意义。景象匹配是实现无人机在复杂环境下精确定位的关键技术之一。通过识别和匹配地面或空中景物的特征,无人机可以准确确定自身位置,实现精确的导航和定位,从而完成各种任务。在复杂多变的自然环境中,传统的定位方式可能受到各种干扰因素的影响而失效。而景象匹配技术不依赖定位信号,因此具有较强的场景适用性,为无人机应用提供了更广阔的发展空间。

此外,在无人机定位的算法发展过程中,出现过一些其他类别算法。例如地磁定位、星光定位、地形等高线匹配等。地磁定位技术是一种利用地球磁场进行定位的技术。地球磁场是不均匀的,因为地球内部的液态外核是在不断运动的,所以磁场的强度和方向都会发生变化。利用这个磁场进行定位,

需要测量磁场的强度和方向, 然后根据这些数据计算出位置. 星光定位旨在根据天体在天空的运行规律提供的信息确定飞行器运动参数, 将飞行器引向目标的自主制导, 又称天文制导. 地形等高线匹配技术又称地形匹配, 是一种通过将飞行器飞越的地形剖面与机上存储的数字地形模型相比较, 计算得到位置信息以修正基本导航系统的导航技术. 上述方法由于定位原理的特殊性, 一般作为极端条件下无人机定位的备选方法.

本文旨在对无人机景象匹配定位算法进行综述, 梳理无人机景象匹配定位算法的发展历程, 分析不同阶段代表性算法. 最终, 对复杂条件下景象匹配定位算法领域待攻克问题进行梳理, 为相关研究人员对无人机景象匹配定位算法的研究提供支持.

## 1 问题概述

无人机景象匹配定位算法的实现流程是: 将无人机航拍影像作为查询图像, 通过建立特征提取与相似性度量模型, 比对查询图像与卫星地图库中遥感卫星影像的特征相似程度, 最终完成准确的景象匹配<sup>[6]</sup>. 在此之后, 即可获取匹配的遥感卫星影像的经纬度等位置信息, 从而将匹配结果转化为实现可靠的定位导航结果.

无人机景象匹配定位算法流程如图 1 所示. 当无人机拍摄得到当前位置的航空影像  $x_t$  后, 基于视觉特征提取与匹配算法, 在卫星地图影像数据库  $\{y_i | y_1, y_2, \dots, y_n\}$  中寻找与之匹配的卫星影像. 随后, 根据卫星地图匹配结果, 获取其对应的地理位置信息, 将该位置信息中的经纬度信息更新为无人机的位置坐标. 其中, 卫星地图影像数据预先由固定步长的滑动窗口在区域遥感卫星地图中获得.

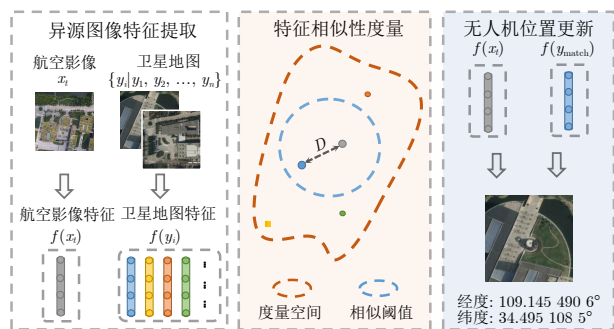


图 1 无人机景象匹配定位算法流程图

Fig.1 Flow chart of drone scene matching geo-localization algorithm

无人机景象匹配定位算法的核心问题可以描述为: 求解与无人机航空影像特征距离  $D$  最小的卫星

图像. 其表达式为

$$\operatorname{argmin}_i D(x, y_i) = \|f(x) - f(y_i)\| \quad (1)$$

其中,  $i$  为求解目标, 对应遥感卫星影像匹配结果;  $f(\cdot)$  表示特征数字化编码, 根据不同的场景和任务选择不同的特征编码方式;  $D(\cdot)$  表示在特征空间中无人机航空影像与卫星图像的特征描述的抽象距离, 常见的抽象距离包括汉明距离、范数距离等. 景象匹配算法根据抽象距离计算结果的数值大小判断图像的相似性.

根据特征提取方式  $f(\cdot)$  的不同, 景象匹配算法主要分为以下三类:

1) 模板匹配算法. 模板匹配是一种简单直观的景象匹配方法, 它通过将待匹配图像与模板图像进行逐像素比较, 寻找最相似的区域来实现匹配. 这种方法的关键在于选择合适的相似度量方法, 常见的包括均方差、相关系数等. 模板匹配适用于目标物体具有明显特征且场景变化较小的情况.

2) 手工描述算法. 手工描述是一种基于特征点或特征描述子的景象匹配方法. 在该方法中, 首先从图像中提取出具有代表性的特征点或特征描述子, 然后通过计算特征点或描述子之间的相似度来实现匹配. 相较于模板匹配, 手工描述方法具有较强的鲁棒性和适应性, 适用于各种不同场景和环境条件下的景象匹配任务.

3) 度量学习算法. 度量学习是一种利用深度学习技术学习图像特征并计算相似度的景象匹配方法. 通过大量已知匹配样本的学习, 系统可以自动学习出适合当前匹配任务的相似度量函数, 从而实现更准确的景象匹配. 度量学习方法在处理大规模数据和复杂场景时具有较好的性能表现.

本文调研无人机景象匹配定位算法的发展, 分析不同类别无人机景象匹配定位算法的特点, 指出无人机景象匹配定位待解决的问题, 为相关研究人员提供解决问题的思路.

根据视觉特征提取方式  $f(\cdot)$  的不同, 将景象匹配算法分为: 基于模板匹配的方法、基于局部不变特征的方法和基于场景语义学习的方法. 本文第 2 ~ 4 节分别综述三种方法, 针对各类代表性算法中存在的优缺点进行分析. 第 5 节整理基于景象匹配的组合适定位导航算法相关的最新研究进展. 第 6 节整理相关算法的数据集与性能指标等. 最后, 根据本文所整理视觉定位算法中存在的问题, 讨论无人机景象匹配定位算法的可行性方法.

## 2 基于模板匹配的景象匹配

基于模板匹配的无人机景象匹配定位算法利用

目标图像 (卫星地图) 与待匹配图像 (无人机影像) 之间的差异建立图像对应关系, 通过相似性度量和数值优化方法来进行几何变换估计和公共区域对齐. 基于模板匹配的无人机景象匹配定位方法由于匹配算法的鲁棒性相对较差, 因此定位效果容易受到噪声、旋转、照度变化等因素的干扰. 并且由于模板匹配较大的计算强度, 在机载计算平台, 算法的实时性难以保证.

基于模板匹配的无人机景象匹配定位算法作为最早出现的无人机视觉定位方法的代表性算法之一, 其核心思路是: 利用像素强度等图像信息, 构建出匹配模型, 完成景象匹配, 算法示意图如图 2 所示. 常用的模板匹配的方法有互相关法 (Cross correlation, CC) 和互信息法 (Mutual information, MI) 等, 本节将分别针对上述两种算法进行总结与分析.

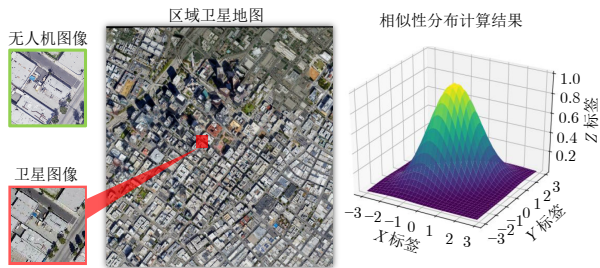


图 2 基于模板匹配的定位算法示意图

Fig. 2 Schematic diagram of location algorithm based on template matching

## 2.1 基于互相关法的景象匹配算法

互相关法是区域匹配方法的经典代表之一, 即通过计算遥感卫星地图的滑动窗口区域与匹配模板的相似性来建立两幅图像之间的对应关系<sup>[9]</sup>. 例如, Le Moigne 等<sup>[10]</sup>利用小波系数的最大值作为配准算法的基本特征, 构建出基于小波的匹配算法. 该方法选择了高频小波特征, 以获得更高的计算速度和相当的准确性. 然而, 这种方法无法克服严重的图像变形, 并且只能应用在轻微的旋转和缩放的时候, 一定程度上限制了算法的应用场景. 为解决匹配过程中的空-天异源匹配问题, 可将原图像转换为另一域, 并在该域上对两幅图像进行对齐. 对此, 基于转换的方法<sup>[11]</sup>和基于傅里叶位移定理的相位相关法<sup>[12]</sup>是解决这一问题的有效方案.

互相关法对频率相关噪声和非均匀、时变照度变化的干扰具有鲁棒性的优点. 然而, 由于该类方法采用频率特征进行匹配, 当面对光谱内容差异较大且重叠面积较小的图像时, 存在一定的局限性.

基于互相关法的无人机视觉定位核心问题是求解基于最大化连续帧之间的互相关函数. 例如, Cain 等<sup>[13]</sup>将帧序列中的每个图像转换为沿图像行和列累加像素值而形成的两个向量投影, 之后通过对投影向量进行一维相关性计算获得多源影像的相似度. 该方法通过将传统二维互相关转换为一维互相关的累加, 降低了匹配算法的复杂度, 并且显著提升了算法对固定模式噪声的鲁棒性. 但是, 对于更加复杂的噪声, 例如区域照度变化, 鲁棒性依然需要进一步提升. van Dalen 等<sup>[14]</sup>提出一种基于归一化互相关的无人机绝对位置估计方法, 将归一化互相关作为无人机影像与全局参考地图密集相似性度量的概率密度函数, 利用单目相机辅助底层视觉的即时定位与地图构建 (Simultaneous localization and mapping, SLAM) 系统确定无人机绝对位置的方法. 该方法在给定该区域的摄影地图后, 在线计算无人机的绝对位置并用于补充基于 SLAM 的导航系统, 融合两种定位方法, 获得更精确的定位结果.

互相关法在受到天气、光照变化、噪声等外部影响时, 匹配结果的置信度难以保证, 因此景象匹配的研究逐渐转向基于更加鲁棒的互信息的方法.

## 2.2 基于互信息法的景象匹配算法

与互相关法类似, 互信息法同样能够实现空-天异源匹配, 实现无人机景象匹配定位. 互信息法利用两幅图像之间的统计相关性进行相似性度量.

例如, Klein 等<sup>[15]</sup>使用互信息和 B 样条 (B-Spline) 实现非刚性图像的匹配, 证明了互信息法对于图像匹配问题的灵活性. Loeckx 等<sup>[16]</sup>通过改进的条件互信息法得到图像匹配结果. 由于互信息方法的判断特征为图像的信息量, 因此该方法能够克服图像匹配时由于无人机视角航空图像影像随着飞行平台旋转而导致的图像旋转方向非对齐问题, 适合于复杂的多源匹配问题. 但是, 由于互信息法依赖于计算不同分布的信息熵, 不同地物对于信息熵的变化影响不一. 某些关键地物对于相似性度量起到重要作用, 但对信息熵仅产生微小影响, 此时, 互信息法难以完成正确的相似性度量.

在互信息方法的基础上, Viola 等<sup>[17]</sup>提出一种基于模板匹配的信息理论方法. 该方法基于模型和图像之间的互信息的公式, 通过统计图像的灰度值来计算基准图和模板之间的相似性. 基于视觉的无人机定位算法的一个主要缺点是缺乏鲁棒性, 大多数方法对场景变化 (如季节或环境变化) 都很敏感. 为了避免这种情况, Yol 等<sup>[18]</sup>提出一种基于场景变化鲁棒的视觉定位算法, 选择使用对局部和全局场

景变化具有很强鲁棒性的互信息. 实验证明了在无人机视觉定位任务中, MI 作为模板匹配的方法代表, 具有最优的算法性能. 但是由于互信息方法缺乏图像语义信息的感知, 在遇到局部变化或多视角问题时, 难以正常运行. 在形如山区这样的特征匮乏场景, 特征的提取、匹配或跟踪比在城市地形中更加困难. Fan 等<sup>[19]</sup>提出了一种新型高效的无人机俯视航拍图像与高分辨率卫星图像配准算法. 该算法通过复合可变模板匹配实现. 为了克服环境变化和不同传感器的限制, 在保留图像信息的同时, 融合图像边缘和熵特征作为图像表示. 根据海拔高度信息, 无人机可以得到俯视航拍图像相对于卫星图像的比例. 接着, 在卫星图像中执行一种有效的搜索策略, 以找到最佳匹配位置. 实验结果表明, 该算法具有良好的鲁棒性和有效性. Levin 等<sup>[20]</sup>提出了基于数字高程图数据的自主地形地貌导航方法. 由于基于对地形特征的认识, 它可以与 GPS 或其他最先进的导航系统集成, 也可独立运行. 该方法从垂直起飞无人机的惯性和视觉传感器组件中获取数据. 该方法作为不依赖 GPS 的相对导航的基础, 仅仅基于不同地形的特征, 即可完成视觉定位.

### 2.3 小结

基于模板匹配的无人机视觉定位方法作为无人机视觉导航方法代表之一, 开辟了不依赖于通信定位信号实现无人机视觉定位的先河, 提升了无人机在飞行过程中的定位导航自主化、智能化的程度. 基于模板匹配的无人机视觉定位算法在应用时, 能够一定程度克服图像因旋转导致的角度不对齐问题. 另外, 该类定位方法能够将无人机与卫星影像从可见光图像转换为频率或信息量中的相关模型, 通过量化分析待匹配图像对的相似性概率, 最终完成无人机景象匹配定位.

然而, 基于模板匹配的视觉定位算法仍存在一些局限性. 首先, 在特征提取的过程中, 这类算法主要关注灰度值或频率特征的相似程度, 而忽略了目标物体的形状和结构信息. 因此, 对于遮挡、非刚性变换、光照、背景变化、背景杂波和尺度变化等因素较为敏感. 当场景中存在以上因素时, 模板匹配可能无法准确识别目标, 导致定位结果不准确或失败. 其次, 由于模板匹配需要建立大量的像素与频率的映射模型, 算法的计算复杂度较高, 导致运算实时性方面出现较为明显的时滞. 特别是在需要实时定位的应用场景中, 时滞会导致定位信息的延迟, 影响系统的实时性和响应性.

综上所述, 基于模板匹配的视觉定位算法在面对遮挡、光照变化、背景杂波等复杂场景时表现出

局限性, 并且在计算复杂度和实时性方面仍然存在一定的提升空间. 因此, 基于模板匹配的视觉定位算法需要结合其他更为鲁棒和高效的定位算法, 以提高定位系统在各种复杂环境下的性能和稳定性.

## 3 基于局部不变特征的景象匹配

相较于基于模板匹配的景象匹配算法, 基于局部不变特征 (也称作手工特征) 的景象匹配算法展现出更强的环境适应能力. 局部不变特征即手工特征描述符, 在算法设计时通常依赖于专家的先验知识, 这些知识的补充在诸多应用中呈现优于模板匹配方法的效果, 因此得到广泛使用.

基于局部不变特征的景象匹配的算法流程为: 首先, 提取无人机和卫星地图影像的局部信息, 主要为图像的强度或梯度; 随后, 利用规范化策略 (如池化、统计和比较) 生成描述符, 对图像数据进行区别性描述; 最后, 给出相似性估计.

本节对基于局部不变特征的景象匹配算法进行综述, 然后总结该类方法.

### 3.1 基于梯度统计方法景象匹配算法

梯度统计方法作为基于局部不变特征的匹配算法中的代表, 在多种场景中得到应用. 梯度统计方法旨在形成浮点型描述符, 例如, 尺度不变特征转换方法 (Scale-invariant feature transform, SIFT)<sup>[21-22]</sup> 中引入了梯度方向直方图 (Histogram of oriented gradient, HOG), 它们广泛应用于现代视觉任务中, 算法示意图如图 3 所示. 在 SIFT 中, 特征的尺度和方向分别由高斯差分 (Difference of Gaussian, DoG) 计算和梯度方向直方图中的极大值确定, 从而实现尺度和旋转不变性.

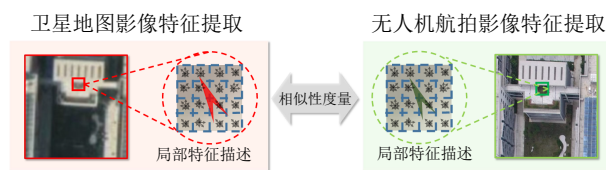


图 3 SIFT 特征点匹配算法示意图

Fig. 3 Schematic diagram of SIFT feature point matching algorithm

在景象匹配任务中, 局部不变特征能够很好地应对旋转等变化. 相较于模板匹配的景象匹配算法, 算法鲁棒性得到了显著提升. 因此作为真实场景下较为可靠的算法, 在实际无人机飞行任务中, 具有广泛的应用. 例如, Shan 等<sup>[23]</sup>设计了一个结合方向梯度直方图 HOG<sup>[24]</sup> 和光流法 (Optical flow, OF)<sup>[25]</sup> 的多源特征匹配框架, 通过度量遥感卫星影像和无

人机影像的 HOG 特征, 实现特征匹配. 其中, 光流估计用于估计无人机相对位置, 从而缩小滑窗搜索范围. 除此之外, 基于人工设计的经典特征点匹配方法还有 ORB<sup>[26-27]</sup>, AKAZE<sup>[28]</sup>, BRISK<sup>[29]</sup> 等. 近年来, 多种手工设计的局部特征用于视觉定位的特征描述. 在此基础上, 为了满足真实场景的运行需要, 一些新的手工特征匹配方法被提出, 使得手工描述子能够在复杂的真实场景下运行. 例如, Chiu 等<sup>[30]</sup>提出了一种新的视觉辅助航空导航方法, 融合了惯性测量与 2D-3D 匹配算法. 该方法将相机观测与地理参考信息集成到一个框架中. 不同于以往的视觉辅助定位方法, 该方法将每一次观测视为一个单独的测量. 通过连续建立观测模型, 得到了每次观测的飞行位置信息, 利用飞行位置信息之间的相关性, 实现了无人机飞行过程的精确建模. 这种方法将绝对的地理注册信息紧密地整合到测量中, 在没有卫星定位信号的情况下进行精确的无人机姿态估计. 由于采用统一的相机观测与地理参考信息更新框架, 使得定位算法的封装性与可用性得到了巨大提升, 有力推动了视觉定位算法的落地.

### 3.2 基于改进特征描述子的景象匹配算法

由于传统手工特征存在局限性, 因此, 改进特征描述子对于视觉定位任务具有积极作用.

例如, Mantelli 等<sup>[31]</sup>设计了一种使用卫星图像进行绝对定位的系统. 该系统使用了一个向下的单目相机, 因此横滚角度和俯仰角度被认为接近于零. 无人机图像基于改进的二值鲁棒特征 (Binary robust independent elementary features, BRIEF) 与卫星地图进行匹配. 作者使用 BRIEF 描述符, 在整个图像中随机选择固定数量的像素对, 而不是在特征点周围的区域中选择. 为图像生成一个全局描述符, 并允许算法跳过角点检测, 以减少计算时间. 然后, 基于蒙特卡罗定位 (Monte Carlo localization, MCL) 估计无人机在状态空间中的位置, 图像的相似性由描述符与当前无人机图像描述符之间的汉明距离组成. 该方法在基于传统特征描述的基础上, 针对性地改进描述符, 并且在采样时选择合适数量的特征, 提升图像匹配的实时性, 在保证精度的同时, 提升了视觉定位算法的实时运算性能.

同时, 优化匹配算法的结构, 设计和改进传统的滤波模型, 同样能够使得视觉定位中的位置更新过程更加准确.

例如, Masselli 等<sup>[27]</sup>提出了一种基于地形分类和粒子滤波的方法. 将无人机视角获取的航空影像分为四类: 草地、灌木丛、道路和建筑物. 为此, 从图像上定义的网格的每个单元中提取 ORB 特征描

述符集. 然后使用一个随机森林模型对每个描述符进行分类. 与特征描述符关联最多的类被认为是单元中存在的地形类型. 然后将该分类过程应用于其他谷歌地图数据, 构建目标环境的参考地图. 将相同的地形分类方法应用于无人机图像, 对无人机进行定位估计. 为此目的, 将参考地图和经过处理的无人机航空影像用于 MCL 系统内部<sup>[32]</sup>. 该方法借用图像中的类别信息, 对场景信息进行捕获后, 筛选出四分之一的匹配场景地图影像, 有效解决了匹配算法在检索时的复杂度问题. 但是受限于场景分类算法, 难以再细分场景归属, 进一步提升匹配算法的效率. 因此, 场景语义信息的充分挖掘, 将是无人机视觉定位算法在提升精度、实时性方面的有力帮助.

### 3.3 基于多视觉特征融合的景象匹配算法

单一的手工特征难以适应真实场景中复杂的环境变化. 对此, 多种视觉特征融合的视觉定位方法作为一种新的思路, 将融合前的多种方法的短板进行弥补, 组成更加可靠的视觉定位方法.

例如, Shan 等<sup>[33]</sup>开发了一种同时使用特征检测和提取的方法, 将无人机图像与参考地图匹配. 采用最大自异相 (Maximal self-dissimilarities, MSD) 检测特征点, 采用局部自相似 (Local self-similarities, LSS) 提取描述子. 利用 MSD 检测特征点, 然后利用 LSS 计算特征点周围的相关曲面, 并将相关曲面转换为描述符. 两种算法内部都使用差异平方和 (Sum of squared differences, SSD) 作为相似性度量, 且能够同时执行特征点检测和特征描述符的提取. 用光流法将参考地图的搜索空间缩小到一个有限的区域. 然后在光流预测位置周围的滑动窗口内匹配特征点. 对于每个可能的窗口, 计算匹配特征点的描述符的欧氏距离之和, 最小距离之和所对应的窗口被认为是无人机的位置. 此外, 估计的位置与光流预测的位置超过一定距离会作为异常值被拒绝. 作者利用无人机的实际飞行数据验证了该方法的有效性. 虽然没有平均位置估计误差的报告, 但飞行路径的对比图显示, 结果相对接近 GPS 定位, 远远优于单独的光流法. 该方法作为两种视觉定位方式的结合, 利用光流计算得到有限的待匹配区域, 在此基础上, 采用粒子滤波的方式, 分别计算所有待卫星影像与当前无人机航空影像的相似性计算结果, 最终比较得到位置更新. 由于采用了“由粗到细”的匹配策略, 其精度与实时性得到了保证. 后续的视觉定位算法在结构上大多沿用此方式, 完成视觉定位算法的设计.

大多数基于视觉的无人机导航算法利用电荷耦合器件 (Charge coupled device, CCD) 相机数据提

取城市地形中结构良好的人工特征, 如建筑物或道路。但在山区, 特征的提取、匹配或跟踪比在城市地形中更困难。同时, 无人机在夜间或黑暗环境下, CCD 相机无法实现有效的信息采集。对此, Woo 等<sup>[34]</sup>提出了一种基于视觉的无人机定位新方法, 通过数字高程地图 (Digital elevation map, DEM) 完成无人机定位, 该方法获取高程地图中有关山区天际线信息, 借此天际线作为视觉定位中的匹配定位线索, 通过辨别山区天际线和山峰线以提升定位的精度。通过借助辅助信息, 完成非可视环境下的定位, 有效地扩展了无人机工作场景, 使得无人机能够在更具有挑战性的场景下运行。

### 3.4 小结

根据专家经验设计的局部结构匹配方法将特征显式表达, 相比模板匹配的方法, 在克服旋转、照度变化、噪声等环境变化时更加鲁棒。

但是, 基于局部不变匹配的视觉定位方法在实际应用中仍存在一些局限性。尽管这种方法在简单场景下表现良好, 但在复杂场景下, 其局限性变得更加显著。首先, 复杂场景中的物体往往具有丰富多样的外观和形状, 在这种情况下, 能够被显式表达的特征可能存在不足, 大多为局部特征, 缺乏对全局语义级特征的描述能力。因此, 在复杂场景中容易发生特征波动, 导致定位结果不稳定。

尽管局部不变特征在某些情况下具有一定的鲁棒性, 但它们缺乏对图像语义信息的全面理解, 因此在复杂场景中的适用性受到限制。在这种情况下, 即使特征随环境变化, 代表图像中物体分布模式和语义属性的特征也能够提供更强的抗干扰能力, 有助于克服多种干扰并提高匹配的准确性。

此外, 为了提升手工特征的精度性能, 往往需要增加特征描述的复杂度, 这会导致手工特征算法的计算复杂度成倍上升, 从而降低了算法的实时性。这对于需要快速响应和准确定位的实时应用, 如自动驾驶和无人机导航等, 是一个严峻的挑战。

综上所述, 基于局部不变匹配的视觉定位方法在复杂场景下存在局限性, 主要表现在对全局语义信息的描述能力不足以及计算复杂度较高导致实时性降低。因此, 未来的研究应该致力于克服这些局限性, 提出更加鲁棒和高效的视觉定位方法, 以满足各种复杂环境下的定位需求。

## 4 基于场景语义学习的景象匹配

随着计算机视觉的发展, 神经网络凭借其能够捕捉图像中的高维语义特征的特点, 提升了匹

配算法的鲁棒性, 使得无人机景象匹配定位算法能够在更加复杂的环境中运行。

早期度量学习主要关注的是来自同一数据源的图像匹配, 例如, Han 等<sup>[35]</sup>基于类似于 AlexNet 的特征提取网络结构, 建立了一个双分支的孪生网络模型, 以完成同源图像相似性计算。由于网络结构的优越, 随后的基于度量学习的视觉定位算法<sup>[36-38]</sup>大多同样采用 MatchNet 框架。此外, 许多度量学习模型在特征学习和相似性度量方面都有了改进。例如, 当使用不同的特征提取网络构建不同的孪生网络模型时, 最终的结果会有显著的差异。究其原因, 不仅是骨干网络的特征感知能力不同, 而且在建立连体网络时, 不同骨干网络对复杂环境下的图像建立度量关系的能力也不同。

与传统的度量学习不同的是, 基于度量学习的无人机景象匹配定位方法旨在通过度量学习模型完成空-天异源图像匹配。相较于模板匹配和局部不变特征方法, 基于度量学习的景象匹配方法的优势在于深度神经网络具有语义信息提取能力, 因此可以更好地建立图像匹配关系。同时, 借助于语义信息的提取, 景象匹配算法能够应对更加复杂的视角变化, 克服多种可能面临的干扰, 例如照度变化、噪声等。

本节根据无人机所处环境的不同, 分别针对“多源无人机视觉定位”与“多视角无人机视觉定位”问题展开综述, 然后分析汇总无人机视觉定位算法中存在的共性问题与难点。

### 4.1 多源语义学习的景象匹配算法

对于多源无人机景象匹配定位, 主要问题在于异源图像特征不对齐, 难以建立图像对应关系的匹配模型。如图 4 所示, 多源景象匹配问题中的时相差异、照度差异等差异问题是指在不同时间或不同源头获取的景象之间存在的差异。时相差异指的是同一区域在不同时间拍摄的图像之间的差异, 可能

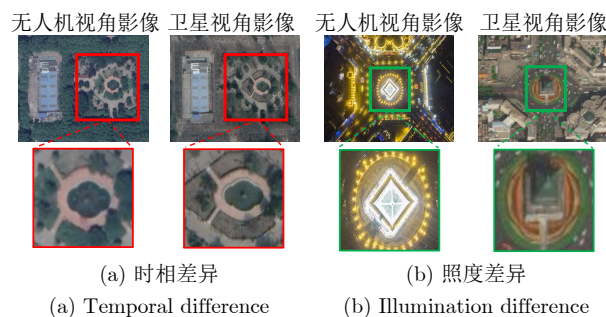


图 4 多源影像差异

Fig. 4 Multi-source image differences



由于季节、天气、光照等因素造成;照度差异则是指不同图像之间的光照条件不同所导致的差异,这可能由于阳光角度、云量、阴影等因素引起.这些差异会影响到图像的视觉特征,进而影响到景象匹配的准确性和稳定性.因此,解决多源景象匹配问题中的时相差异、照度差异等差异问题是提高景象匹配算法性能的关键挑战之一.在算法设计中,需要考虑如何对不同时相、不同光照条件下的图像进行特征提取和匹配,以克服这些差异带来的挑战.

#### 4.1.1 基于度量学习的多源景象匹配算法

作为数据驱动的算法,利用大规模样本学习,能够使得度量学习模型获得空-天多源影像一致性度量关系的判别能力.

在度量学习算法的早期研究阶段,为了实现准确的图像匹配, Noh 等<sup>[39]</sup>面向大规模图像匹配检索任务,提出了基于注意力机制的局部关键特征学习模型.该方法通过挖掘图像中关键区域的语义信息,建立图像匹配联系,借助语义信息作为图像匹配检索的线索,克服了图像数据来源差异而导致的匹配难题.该方法充分利用了图像中的关键区域信息,相较于传统的特征点方法,进一步提升了深度度量特征中语义信息占比,从而帮助度量模型提升度量判定精度.与之类似, Teichmann 等<sup>[40]</sup>通过引入区域聚合选择匹配核,提出了一种聚合局部特征的方法,有效提升了图像匹配算法的性能.

Hinzmann 等<sup>[41]</sup>将无人机视觉定位任务分解成特征描述和定位追踪两步.首先通过构建基于深度网络的无人机视觉匹配模型,将不同时间、不同数据源的遥感影像进行一致性的特征描述,消除位姿、光照变化导致的差异,之后结合视觉里程计实现位置定位.该方法在图像预处理阶段,对可能影响后续匹配精度的干扰因素进行约束,预处理得到的特征在相似性匹配时排除了复杂环境干扰,为算法精度提升提供了有力保障. Goforth 等<sup>[42]</sup>结合卷积神经网络(Convolutional neural network, CNN)和视觉里程计,提出了一种基于影像深度特征的对齐影像匹配映射模型,实现大规模无人机视觉定位.该方法通过融合视觉惯性定位方法,使得最终的定位结果的置信度得到了提升.除此之外,视觉里程计的加入,在卫星地图检索时,给予了待搜索范围约束,提升了算法的运算效率.

Kinnari 等<sup>[43]</sup>开始重点关注季节变化对无人机影像视觉定位的影响,并提出一种基于卷积神经网络的季节变化一致性映射模型,该模型可有效克服航空影像与遥感卫星影像之间显著的季节性外观差异,拓展了视觉定位算法的应用场景.

为了进一步提升视觉定位效果, Amer 等<sup>[44]</sup>提出了深度城市特征的概念,其中卷积神经网络用于根据不同城市区域的视觉外观计算其独特特征.使用谷歌地图<sup>[45]</sup>的数据对 CNN 进行训练和验证,然后使用 Bing 地图<sup>[46]</sup>的数据对模型进行测试,以模拟无人机图像.这项工作利用不同城市地区或地区的结构和组织之间的差异,通过分类的方法高效辨别无人机位置.事实上,城市发展规划、发展速度、建筑材料等在不同地区存在显著差异.该方法包括两个步骤:地区级定位和邻域级定位.该本地化框架使用在 ImageNet 数据集上预训练的单个 VGG16<sup>[47]</sup>网络.使用迁移学习,训练全连接层.对于区域级定位步骤,训练网络对七个区域之一的查询图像进行分类,以提升搜索效率;对于邻域级定位步骤,使用来自同一 CNN 的特征地图进行最近邻搜索,提取深层卷积层的特征映射,将查询图像和参考图像进行对比,以最接近查询图像的卫星地图的位置作为无人机的位置估计.

度量学习使得景象匹配算法能够通过深度神经网络捕捉影像中不同尺度的物体甚至全局级别的语义特征.相较于基于模板匹配,手工特征描述的景象匹配方法,其影像相似性度量的可靠性得到了巨大提升,因此可得到更加精确的视觉定位结果.

#### 4.1.2 融合传统特征的多源景象匹配算法

基于度量学习的景象匹配算法本质在于构建空-天多源匹配模型,因此存在计算机视觉中的基于学习算法的共性问题,即对大规模数据样本的依赖.当场景中的环境变化并未出现在训练样本中时,模型的适应能力成为制约算法泛化性能提升的主要原因.传统算法与深度学习方法的结合能够弥补不同算法的短板,对于模型在复杂场景的适应能力提升具有较好的效果.

在框架构建方面, Nassar 等<sup>[48-49]</sup>将传统的计算机视觉技术与 CNN 结合在统一框架中.用已知的初始位置在参考卫星地图  $M$  上定义一个窗口  $W$ .每三帧执行一次校准,以保持无人机的当前视图在参考地图上的准确配准.然后从无人机图像和  $W$  中检测 SIFT 特征点,提取其描述子,使用随机抽样一致性(Random sample consensus, RANSAC)算法<sup>[50]</sup>估计描述  $W$  到无人机图像变换的单应矩阵,计算得到的单应性被应用到  $M$  上.同时,使用 ORB<sup>[26]</sup>将当前无人机影像与校准的卫星地图进行配准.所得到的单应性用于更新  $W$ .每三帧使用 SIFT 和每帧使用 ORB,使配准既计算高效又准确. SIFT 在处理尺度变化方面具有优势,但是 ORB 处理速度更快.因此,作者的方法可视为两种特征提取方

法的结合. 随后, 使用 U-Net<sup>[51]</sup> 进行语义分割. 将无人机图像和已配准的  $M$  送入 U-Net, 得到精确的建筑物和道路掩模, 以填补空白的斑点和消除噪声. 将较小的形状块和斑点过滤掉, 对剩下的斑点计算 Hu 矩. 然后, 对两个图像之间的斑点进行匹配, 最终得到计算精确单应性的最佳匹配.

作为模板匹配与场景信息学习结合的典型工作, Schleich<sup>[52]</sup> 利用条件生成对抗网络 (Conditional generative adversarial network, cGAN)<sup>[53]</sup> 和模板匹配开发了一种新方法. 与其他一些定位方法不同的是, 为了获得纯粹的定位结果, 作者有意避免在解决方案中嵌入任何视觉里程计的概念. 该算法的第一步是将无人机当前图像转换为类似地图的表示, 类似于没有卫星图像覆盖的地理地图. 通过使用 cGAN 三种类型 (建筑、道路和背景) 实现分割图像. 训练过程中, 从 OpenStreetMap<sup>[54]</sup> 中提取卫星数据, 并使用相应的元数据生成匹配的类地图表示. 训练数据集由 125 平方公里的地图数据组成, 测试数据集由 9 平方公里的地图数据组成. 生成后, 使用模板匹配技术将类地图图像与该区域的类似的类地图参考地图进行匹配. 模板匹配技术使用归一化平方和差分 (Sum of squared difference, SSD). 假设模板的比例和方向是已知的, 它们可以通过 IMU 获得. 因此, 模板首先被缩放和旋转以匹配参考地图. 然后计算参考地图中每个可能位置的模板之间的 SSD 值, 最终以最小值作为无人机位置.

待匹配图像的拼接能够有效提升数据的信息量, 使得度量模型在位置估计时拥有更多有效信息, 从而提升算法的可靠性. Lin 等<sup>[55]</sup> 强调, 利用机载传感器获取的连续图像进行拼接是图像分析人员必不可少的手段. 严格的局部拼接方法常常会积累误差, 导致图像失真. 为了克服这一限制, 作者采用参考图像 (例如高分辨率的地图图像) 映射方法. 在作者提出的方法中, 使用了帧到帧的迭代配准, 将图像序列中的每一帧都映射到一个统一的参考图上. 在帧对帧的配准过程中, 由于当前帧的前一帧已经映射到参考图中, 因此可以从当前帧到前一帧的关系中计算一个转换矩阵. 在配准过程中, 作者通过这种转换将当前帧投射到参考地图上, 以补偿比例和旋转差异. 随后, 使用互信息执行基于区域的匹配, 以找出当前帧和参考地图之间的对应关系. 这些对应关系可以视为局部拼接与地图的对应关系. 通过这种两步配准, 每个连续帧之间的误差将不会累积. Huang 等<sup>[56]</sup> 考虑在无人机图像序列与谷歌卫星图像质量不匹配问题的情况下, 开发一种鲁棒的图像配准系统. 首先, 作者尝试开发一个图像配准系统,

将无人机图像配准到卫星图像上. 此外, 为了克服质量不匹配问题, 在将无人机图像配准到卫星图像时, 使用互信息归一化变体来度量无人机与卫星图像之间的相似性.

为了挖掘照度变化过程中对基于学习的景象匹配具有提升作用的描述, Wan 等<sup>[57]</sup> 提出了一种基于机载无人机图像序列与参考卫星图像匹配的无人机自主导航定位算法. 由于无人机图像和参考图像不一定是在相同的光照条件下拍摄的, 因此光照不变图像匹配是必不可少的. 通过数学推导和实验研究了相位相关的照度不变性, 提出了一种基于相位相关的快速鲁棒照度不变性无人机导航定位算法. 该算法即使在无人机机载图像的光照条件与参考卫星图像不同的情况下, 也能准确地确定当前和下一个无人机位置. 引入基于狄拉克函数的配准质量评估和风险报警准则, 使无人机在偏离计划航路时能够进行自我修正. 利用模拟地形阴影图像和遥感图像的无人机导航实验已经证明, 在太阳运动引起的非常不同的光照条件下, 所提出的基于 PC 的定位算法具有鲁棒的高性能. 与其他两种广泛应用的图像匹配算法互信息和归一化相关系数相比, 该算法具有较高的匹配精度和较快的处理速度. Patel<sup>[58]</sup> 提出了一种基于视觉的自主路线跟踪系统, 用于多旋翼无人机. 由于多旋翼是欠驱动的, 不能保证相机的视点在相同位置, 这导致视觉定位具有挑战性. 该论文的第一部分论证了采用适当的主动指向策略的三轴框架摄像机可以提高视觉定位性能和鲁棒性; 在第二部分, 作者提出了一种利用信息论方法将真实图像与三维谷歌地球上渲染的地理参考图像配准的方法来估计无人机的全局姿态. 结果表明, 该方法能够准确估计无人机的姿态, 其定位精度能够达到 GPS 同等水平.

遥感图像配准技术在地理信息系统领域占有重要地位, Pluckter 等<sup>[59]</sup> 针对不同视点的遥感图像, 提出了一种配准方法. 利用提取的 SIFT 点对两幅不同视点的遥感图像进行特征表征, 通过提出的基于混合特征的非刚性对应估计对其进行配准, 以解决地面起伏度变化和成像视点变化的问题. 通过仿射变换定义两幅图像最终映射的图像转换模型. 作者通过从谷歌地球和无人机获得的两组遥感卫星图像来评估所提出方法的性能, 并与三种最先进的方法进行比较, 作者的方法在大多数情况下显示出最佳的对齐效果.

#### 4.1.3 多任务知识共享的多源景象匹配算法

网络的知识能够在多任务模型中迁移和共享, 若采取多任务的融合训练策略, 能够使得单个任务

额外学习到其余任务中的信息. 另外, 由于多任务的并行训练, 使得所获得的模型具有很好的多任务适应性. 这样的训练方式对于多任务的智能载体 (如无人机) 具有积极意义.

例如, Marcu 等<sup>[60]</sup> 提出了一种多阶段多任务神经网络体系结构, 可以同时进行航空图像影像分割和地理定位. 该网络使用一个编码器和两个解码器来完成. 第一阶段构建一个语义分割网络, 用于获得道路分割. 网络的其余部分被划分为两个分支. 第一个分支是学习识别输入图像的位置, 并输出相应的经纬度信息; 在第二个分支中, 生成最终的分割结果. 考虑用分割掩模中最大连通分量的质心来估计无人机的位置, 如果分割失败, 则使用经纬度回归的输出. 该方法将编码器和解码器一起训练, 直到收敛后冻结其权值. 此后, 对两个并行分支的解码器进行独立训练. 经纬度回归器使用均方根 (Root-mean-square error, RMSE) 损失进行训练, 此外, 用于分割路径的损失函数是二元交叉熵和 Dice 损失函数的线性组合. 在网络之外, 加入了一个额外的对齐步骤, 其目的是将道路分割与估计无人机位置附近的 OpenStreetMap<sup>[53]</sup> 中的道路精确对齐.

与定位导航一样, 无人机自主降落是自主飞行的必要组成部分, 确保无人机能够回到无人机起飞的同一地点. Pan 等<sup>[61]</sup> 针对由于测量的不准确性, 引导无人机无法到达初始起飞位置展开研究. 提出了一种使用向下鱼镜头相机的方法来精确地降落无人机至起飞地点. 这种方法使用相对于无人机起飞路径的位置估计来引导无人机返回. 利用鱼镜头提供的较大视场, 从着陆开始前的较大位置误差处开始提供视觉反馈, 直到着陆结束. 实验结果表明, 该算法能较好地修正状态估计中的漂移误差, 并以 40 cm 的精度着陆.

#### 4.1.4 基于图像时序关系的多源景象匹配算法

无人机定位时, 所面对的场景在一定时间内是固定不变的. 因此, 挖掘航空影像相邻帧时序关系成为提升匹配度量可靠性的一个可行方法. 例如, 当飞行器的飞行高度极低时, 对于场景的观察视角近似于地面视角. 地面视角的图像与遥感卫星影像的度量同样属于多源匹配问题.

例如, 来自社交媒体的照片通常包含一些可用于建立多源匹配模型的线索, 如地标、天气、植被、道路标记或建筑细节. 有效地将这些信息结合起来可以推断照片的拍摄地点. 常见的做法是用图像检索的方法来解决这个问题, 即通过将地球表面细分为数千个地理单元, 将问题视为分类问题. 由此产

生的模型 PlaNet<sup>[62]</sup> 在某些情况下甚至达到了超越人眼的识别精度. 通过与长短期记忆 (Long short term memory, LSTM) 架构相结合, 学习利用时间一致性来定位不确定的照片, 所得到的模型比单幅图像定位模型取得了 50% 的性能提升.

基于图像匹配的方法大多要求匹配图像之间有较好的相似度, 但由于参考图像与无人机图像存在异源差异, 因此很难满足要求. 现有的方法很难在配备了廉价定位装置的无人机上实现精确的图像地理定位. 为了解决这一问题, Wu 等<sup>[63]</sup> 提出了一种由粗到细的无人机图像地理定位算法. 首先, 利用无人机的运动传感器实现无人机图像的粗定位. 然后, 提出了一种多阶段 Lucas-Kanade 深度学习算法, 对无人机图像的地理定位结果进行迭代优化. 该方法可以充分利用无人机图像的全局纹理. 作者在两个真实的无人机数据集上验证了所提方法的有效性. 结果表明, 本文提出的无人机图像地理定位方法可以提高传统基于摄影测量的方法的定位效果.

#### 4.2 多视角语义学习的景象匹配算法

无人机由于飞行过程中姿态的调整, 航拍视角发生变化, 景象匹配算法应对的问题从多源匹配变成了多源多视角匹配问题. 对于多源多视角无人机景象匹配定位, 由于加入额外的视角变化, 正确提取图像特征并建立正确匹配模型变得更为困难. 无人机飞行过程中因姿态调整导致的视角差异, 使得无人机视角图像在同一区域呈现巨大的视觉差异, 如图 5 所示. 在多视角条件下, 除了考虑时相差异外, 还需要处理不同视角下的图像之间的差异. 这意味着除了考虑光照、季节等因素带来的影响外, 还需要考虑拍摄角度、视角、遮挡等因素对图像特征的影响. 因此, 多视角条件下的景象匹配问题更加复杂, 需要更精密的算法和更灵活的处理方法来应对各种不同情况, 以提高匹配准确性和稳定性.

作为基于度量学习的多视角景象匹配算法的代

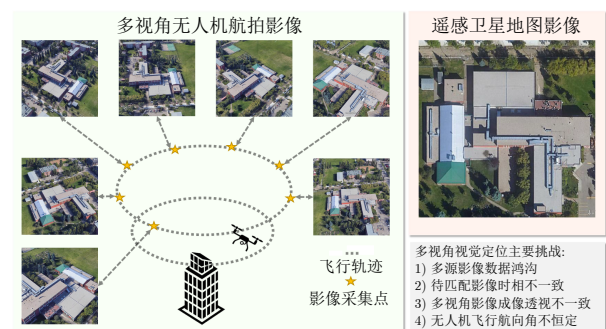


图 5 多源多视角无人机影像差异

Fig. 5 Multi-source multi-view drone image differences

表, Zheng 等<sup>[64]</sup> 公开了首个多源多视角景象匹配无人机视觉定位数据集 University-1652, 为多视角景象匹配研究提供了支持. 在此基础上, 作者构建了一个三分支的孪生神经网络模型, 通过建立多视角无人机影像、街景影像、卫星影像对应关系, 完成景象匹配. 由于额外的街景影像分支的加入, 使得网络模型在特征关系建立时, 补充了更多的信息. 并且由于孪生网络的共享特征提取网络权重的策略, 使得无人机影像与卫星影像的匹配精度得到了提升. 由于特征提取模型采用了经典的 ResNet 结构, 并未增加额外的改进. 此时, 在度量学习模型中, 单纯依靠 ResNet 特征提取网络结构难以有效完成视角干扰的去除, 使得相似性度量结果置信度下降. 同样, Workman 等<sup>[65]</sup> 提出了一种利用深度卷积神经网络来解决跨视点图像地理定位问题的方法, 即通过与地理参考航拍图像的匹配来估计地面查询图像的地理位置. 通过对地面图像使用最先进的特征表示, 引入一种交叉视图训练方法来学习航空图像的联合语义特征表示. 为了支持对这些网络的训练, 引入了一个大型数据库, 其中包含来自美国各地的空中和地面图像对. 该方法在两个基准数据集上的性能明显优于当前的技术水平. Hays 等<sup>[66]</sup> 的研究旨在解决一个重要且具有挑战性的计算机视觉问题: 如何从图像中准确估计地理信息. 大量校正过的地理图像数据的出现使得视觉定位成为全球范围内学者的热门研究对象之一. Hays 等<sup>[66]</sup> 提出了一种简单而有效的算法, 利用基于数据驱动的场景匹配方法, 能够从单幅图像中估计地理位置的分布情况. 为了完成这项任务, 作者利用了一个包含超过 600 万幅带有 GPS 标记的图像的互联网数据集. 通过将估计的图像位置表示为地球表面上的概率分布, 对其方法进行了定量评估. 相较于随机选择的方法, 性能提升了 30 倍. 这项研究表明地理位置估计在许多其他图像理解任务中具有重要作用, 例如人口密度估计、土地覆盖估计以及城市与农村的分类等.

#### 4.2.1 多视角信息挖掘的景象匹配算法

在变化的视角影响下, 度量辅助信息的挖掘与补充提供给相似性度量判断更多的依据. 主要的方法有增加学习过程中的细粒度标签和补充视角不变的元素作为辅助信息.

例如, Yang 等<sup>[67]</sup> 为了解决无人机视觉定位算法在匹配图像时的视角差异, 充分挖掘了特征的位置编码. 利用位置编码信息, 指导双分支孪生网络建立多视角图像对之间的区域对应关系. 借助位置编码, 使得图像透视关系得到了标记, 网络能够学

习到透视变化后的区域对应关联, 使得算法的多视角鲁棒性得到提升. 但是, 由于选择了 Transformer<sup>[68]</sup> 作为特征提取骨干网络, 并且空间位置对应模块使得网络具有较大的复杂度, 在训练和移动平台部署方面都存在一定的限制.

为了更充分地挖掘隐藏在多源多视角图像特征中的关联模式, Shi 等<sup>[69]</sup> 开发了一种新的网络, 通过将不同视角的图像映射到同一个空间, 有效降低了地面图和鸟瞰图之间的内在差异. 通过引入特征聚合策略, 建立特征的空间对应关系, 有效克服了视角差异在特征度量中的影响. 但是, 在特征映射过程中, 信息的损失问题将会影响模型的精度, 仍需要进一步细化分析.

Liu 等<sup>[70]</sup> 通过在图像中挖掘颜色编码的方向, 使网络学习到旋转不变性特征, 增强了跨视角地理定位算法的方向鲁棒性. 但是, 依靠特征挖掘的方法具有明显的应用约束, 即需要挖掘的特征必须具有视角变化鲁棒性. 当无人机视角图像仅存在轻微视角变化时, 增加的方向特征在模型中成为冗余计算量, 这将导致算法在轻微视角变化场景下依然需要较多的计算运行时间, 使得算法在应用时不能自适应调整策略, 浪费飞行平台中有限的计算资源.

Hu 等<sup>[71]</sup> 使用一种基于方向的方法来对齐组件特征. 作者提出一种圆形裁剪方法来提高旋转鲁棒性, 利用映射的方式将不同视角图像投影到同一个圆形坐标系中. 作为一种新颖的投影方式, 圆形坐标系有效地解决了旋转条件下的矩形无人机影像变化问题. 但是, 坐标的转换过程中, 不可避免地产生畸变, 使得网络在学习空间对应关系时产生困难, 更加依赖于语义信息补充来弥补空间上的不对齐问题. 多源多视角图像之间的对应关系的建立, 使得即使是人眼, 也很难完全正确地匹配一些复杂场景. 因此, 借助多视角图像中保持一致的特征成为提升算法多视角鲁棒性的可靠方法.

在变化的视角影响下充分挖掘图像中保持一致的特征信息, 主要的方法有语义信息增强和多视角特征分析. 例如, Jin 等<sup>[72]</sup> 通过建立特征增强模型, 充分利用语义特征, 使匹配结果具有更强的多视角鲁棒性, 不受旋转和环境光照变化的影响. 为了解决变化视角下语义信息一致性挖掘不充分的问题, Sun 等<sup>[73]</sup> 提出了一个端到端的地理定位框架 F3-Net, 用于计算多源和多视图图像的相似度. 为了改进多视角的语义特征一致性, 基于目标语义连贯的不变性原则, 引入了目标特征增强模块. 在多视角特征学习后, 利用 EM 距离计算非对齐特征的相似度. 与传统的孪生网络不同, F3-Net 将多视角图像

的特征视为一个概率分布,因此 F3-Net 可以在学习过程中量化并消除多视角图像的特征差异。

#### 4.2.2 多视角信息补充的景象匹配算法

除了增加额外的辅助信息,通过多视角信息补充同样能够补足因透视问题产生的信息缺失。

例如, Ding 等<sup>[74]</sup>采用了多幅不同视角无人机影像作为查询输入,补充了不同的视角信息,弥补了单幅多视角无人机影像对于地物信息的缺失,有效提升了算法的精度。该方法要求无人机在飞行过程中,对同一区域进行多角度拍摄,获取多角度航空影像。显然,该方法在实际应用时存在一定限制。但是,多输入查询的方式如果转换为通过连续视频帧进行相邻视角信息补充,则实际应用时能够提供帮助。无人机飞行过程中,因姿态的变化调整传感器拍摄视角,连续变化的视角中的差异可以被相邻视角影像弥补。但是,相邻视角的透视差异因为变化的连续性,往往只提供有限的视角差异信息。在无人机平稳飞行时,由于没有出现视角的变化,多幅输入反而增加了移动平台的计算负担。

Arjovsky 等<sup>[75]</sup>根据分布对深度特征进行建模,旨在更准确地计算特征之间的相似性。他们分析了特征分布之间的差异,并采用推土距离 (Earth mover, EM) 来生成与源图像分布相似的图像。通过隐式学习相似图像的分布模式,并利用其生成新图像,EM 距离在最优路径规划下具有最小消耗,相较于传统的散度,其计算更为平滑。相反,JS (Jensen-Shannon) 散度具有突变特性,在两个分布无重叠部分时,即使在优化过程中逐渐靠近,也难以在训练中体现。与之相比,EM 距离能够实现更为平滑的收敛过程。

除了前述方法外,充分挖掘多视角特征中的隐藏信息同样对提升度量模型的鲁棒性具有积极作用。Chan 等<sup>[76]</sup>利用神经辐射场隐式表示对象,并学习了多视图图像特征的潜在分布。目前,多视角图像特征处理存在两个主要不足: 1) 缺乏底层的 3D 表示或依赖于视图不一致的渲染,导致合成的图像缺乏多视图的一致性; 2) 通常依赖于特征网络架构,而这些架构的感知能力有限,导致结果在图像质量上不尽如人意。作者提出的 SIREN 周期激活函数充分利用了视角与特征的对应关系,从而量化整个多视角影像与特征之间的关系。然而,对于无人机移动平台而言,复杂的训练过程显然成本过高,不利于在移动端部署。

Shi 等<sup>[77]</sup>提出了一种跨视图特征传输方法,旨在明确建立跨视图域传输,以便更好地对齐多源图像之间的特征。该方法通过构建一个网络层,负责

将特征从一个域传输到另一个域,以完成特征相似性度量。通过传输矩阵重新排列从一个分支中提取的特征映射,使得从一个域转换出来的特征与另一个域中对应的特征在位置上更加接近,便于进行相似性比较。此外,空间布局信息也被嵌入到特征中,从而使得特征更具辨别力。该方法证明了从特征层面进行视角迁移在跨视角匹配任务中的可行性。类似于景象匹配任务,跨视角匹配任务也是通过匹配街景视角影像与卫星地图来实现定位。在训练过程中,引入更多的监督信息有助于建立正确的特征迁移与对齐模型。

Wang 等<sup>[78]</sup>提出了局部模式网络,采用方环特征划分策略,以获取描述中心目标和周围环境信息的特征。通过多次测量对应的特征对,得到基于中心和边缘方环的相似度结果,最终通过融合得到相似度量结果。然而,在对图像进行区域划分的过程中,相似度量的环形分区模式是完全固定的,这使得算法难以自我调整,在不同情况下难以作出最优决策,比如在特征分布密集的城市和特征分布稀疏的村庄中。

#### 4.3 小结

基于度量学习的景象匹配能够充分挖掘图像中的语义信息,有效提升景象匹配在复杂环境下的可靠性。主流的景象方法充分挖掘并利用深度特征进行特征提取,从而更全面地捕获图像中的语义和结构信息。在极端复杂场景下,这种方法通过识别地物特征,判断图像是否匹配。相较于模板匹配与手工特征方法,具有更强的适用性,能够应对多源多视角复杂场景的无人机视觉定位问题。此外,基于度量学习的方法还可以利用视角一致信息感知和多视角信息补充等技术,实现对多视角图像的一致性挖掘,提高定位的准确性和稳定性。并且,随着计算平台运算能力的提升,这类方法的边缘计算部署也得到了有力的保障,使得无人机在复杂环境中的自主定位能够更加高效可靠地实现。

然而,基于度量学习的景象视觉定位方法在实际应用中仍存在一些局限。由于深度学习属于数据驱动的方法,模型的性能取决于所使用的训练样本。然而,目前在景象匹配领域可用的数据集存在着明显的局限性。这主要体现在样本量有限以及场景种类有限等方面。受限于数据的质量和数量,现有的模型无法涵盖所有可能的场景和变化,从而导致在实际应用中的泛化能力不足,特别是在新领域或未知场景下的性能表现可能会受到影响。针对这一问题,解决方案之一是收集更多、更丰富的景象匹配

数据样本,以扩展模型的训练集,从而提高模型的泛化能力和适应性.另外,开展模型迁移和泛化性能研究也变得至关重要.通过在不同场景和环境下进行模型迁移和验证,可以评估模型在不同条件下的性能表现,并进一步优化模型设计和训练策略,从而更好地适应各种复杂环境的目标.

## 5 基于视觉的组合定位算法

景象匹配定位方法在连续位置估计中存在局限性,即连续的飞行状态信息难以得到有效利用,这使得短时连续定位结果之间不具备相关性,从而使连续定位精度与效率不理想.

基于视觉的组合定位算法<sup>[79]</sup>主要以景象匹配定位结果为主,通过与相对定位算法的运算结果融合,能够提升定位精度与鲁棒性,进一步提升定位的稳定性.由于逐渐提升的景象匹配绝对位置估计的准确性,使得基于景象匹配的组合定位方法将逐渐达到甚至超越传统组合定位方法.

### 5.1 传统组合定位算法

传统组合定位算法以 GNSS 作为绝对定位,同时借助惯性导航系统估计相对位置变化,最终形成绝对/相对组合定位.其中,惯性导航系统 (INS) 利用运动传感器 (例如加速度计)、旋转传感器 (例如陀螺仪) 和计算机通过航迹推算来连续计算移动物体的位置、方向和速度 (运动方向和速度).惯性测量单元 (IMU) 通常包含三个正交速率陀螺仪和三个正交加速度计,分别针对角速度和线加速度进行测量.

传统的惯性导航方法中存在的问题包括位置更新误差较大、漂移误差的累积、精确的初始化要求等.对此, Castañeda 等<sup>[80]</sup>提出一种改进的惯性导航系统.该方法基于陀螺仪测量的模糊逻辑步进检测器和卡尔曼滤波来校正漂移误差的角度.与传统的步进检测器不同,所提出的步进检测器对额外的传感器检测初始化阶段并无要求.同时,该算法利用 IMU 的刚性连接特点,以解算初始状态.参考姿态被用于测量,输入到卡尔曼滤波模型中,每次可以检测到一个步长,因此该滤波器可以计算出准确的漂移误差.

无人机导航通常采用 GNSS/INS 融合进行定位. GNSS 作为一种紧凑且经济的方法用于限制由 INS 传感器引起的定位上的无界误差<sup>[81]</sup>.但是, GNSS 已被证明在一些特殊情况下 (信号遮挡、屏蔽) 可靠性不足.这种方法的缺点在于严重依赖获取定位数据所需的通信信号.组合定位导航方法有

效解决了在接收不到通信定位信号时,无人机无法定位的问题.例如, Couturier 等<sup>[82]</sup>提出了一种利用下向二维单目摄像机和惯性测量单元在没有 GPS 的环境下进行相对视觉定位的方法.该解决方案嵌入一个自适应粒子滤波器,利用特征点匹配图像,以实现无人机的定位.

除此之外,惯性导航算法的运算效率同样至关重要. Carlone 等<sup>[83]</sup>对传感器数据进行优先级排序,以提升视觉导航中的计算效率.所提算法特别注重于运动估计,考虑到由于机载计算资源的限制,机器人只能利用环境中的少量视觉特征来支持运动估计,对此他们设计了一种视觉注意机制,通过选择合适的视觉特征集来最大化提升定位的精度与可靠性.与传统的视觉特征选择方法不同,作者认为特征的作用不仅仅是特征本身的内在属性,还取决于环境和观察者状态的交互.他们提出的方法能够有效捕捉特征的视觉显著性和任务依赖性.所提出的注意机制本质上是可预测的.在确定哪个特征更有用时,该方法会快速模拟机器人的状态,以便做出能感知机器人动态和“意图”的特征选择.

Tan 等<sup>[84]</sup>设计了一种基于加速度计的惯性导航系统,以解决传统惯性导航系统受到传感器选择影响的问题.该系统仅利用加速度计进行测量,用于计算刚体的线性和角运动.作者通过推导出刚体相对于固定惯性系的线性运动和角运动的加速度计输出方程,提出了确定加速度计结构是否可行的充分条件.在满足这一条件的情况下,利用输入与输出动力系统的两个解耦方程,即角速度的状态方程和线性加速度的输出方程,可以分别计算出角运动和线性运动.

### 5.2 视觉组合定位算法

视觉组合定位算法指以景象匹配为主、视觉里程计为辅组成的定位方法.该类方法能够充分发挥作为绝对定位的景象匹配在大范围定位时的可靠性,并且能利用相对定位的相对位置估计信息,最终实现可靠的定位.其中,视觉里程计 (VO) 是指通过使用视觉信息来估计相机在连续图像帧之间的运动,并从中推断出相机的轨迹和位置.它是机器人视觉和自主导航领域的重要技术之一.视觉组合定位算法相较于传统组合定位方法,摆脱了对于通信定位信号的依赖,其自主性得到了有效提升.然而,视觉组合定位方法存在一些挑战,例如,快速运动、光照变化、纹理缺乏和重复结构等因素可能导致特征点提取和匹配的困难.

作为景象匹配组合定位方法的代表, Zhao 等<sup>[85]</sup>

提出一种将视觉里程计与景象匹配相结合的集成导航系统,以提升无人机的导航精度和可靠性.通过摄像头捕捉图像并利用特征提取和匹配算法,视觉里程计计算出无人机的运动轨迹,而场景匹配则通过比对当前视图与预存参考场景来校正位置误差,从而克服单一技术在导航中的局限性.实验结果显示,该系统在各种复杂环境中表现出良好的实时性和精度,尤其在GPS不可用时尤为有效.该集成系统有效解决了单独使用视觉里程计或场景匹配技术时面临的一些问题,如误差累积和匹配失败.

将传统的惯性导航算法与视觉里程相结合,在计算相对位置时可获得更精确的结果.例如, Li 等<sup>[86]</sup>提出了一种基于场景匹配的扩展卡尔曼滤波 SLAM 视觉导航系统,旨在提高小型无人飞行器的自主导航能力.具体而言,该研究结合了场景匹配技术和扩展卡尔曼滤波器用于同步定位与地图构建任务. SLAM 技术使无人机能够在未知环境中实时构建环境地图并同时确定自身位置,而场景匹配技术通过将实时捕获的图像与预先存储的参考图像进行匹配,进一步提升了定位精度和鲁棒性.作者详细介绍了系统架构和实现方法.首先,无人机通过摄像头获取环境图像,并提取图像中的特征点.接着,这些特征点与数据库中的参考图像进行匹配,得到匹配结果后,利用扩展卡尔曼滤波器融合场景匹配的结果和其他传感器数据(如惯性测量单元的数据),实现精确定位和地图构建.扩展卡尔曼滤波通过预测和更新步骤,不断修正无人机的位置信息,减小累积误差,提高导航精度.

选择超广视角的传感器对于提升视觉定位方法的感知范围具有积极作用. Liu 等<sup>[87]</sup>提出了一种基于鱼眼相机的视觉里程计算法.该算法实现了同步的运动估计和半密集重建.该方法由两个线程组成:跟踪线程和映射线程.在跟踪线程中,作者通过半密集的直接图像对齐来估计相机的姿态.为了获得更宽的视场,使用广角图像,而不将它们转换为传统的针孔图像,从而提升了图像的视场.另外,为捕捉立体匹配的不确定性特征,对选定的像素点跟踪多个不同的深度假设.然后利用时间运动立体视觉对深度进行细化,去除假阳性深度假设.实验结果表明,该方法能够以较低的漂移估计 6D 位姿,同时进行精度较高的半致密三维重建.

精确的建立噪声模型能够提升组合定位方法的精度.例如, Babu 等<sup>[88]</sup>提出了一种基于概率传感器噪声模型的高密度视觉里程测量新方法.与传统基于匹配的视觉特征估计相机姿态的稀疏视觉里程计方法不同,该模型是一种充分利用 RGB-D 相机所

有像素信息的密集视觉里程计.在此之前,为了减少异常值对优化的影响,使用  $t$  分布对光度和几何误差进行建模.但是,这种方法有其局限性,即只使用误差值来确定异常值,而没有考虑物理过程.因此,作者应用一个概率传感器噪声模型,通过线性化的传播不确定性来权衡每个像素.此外,作者发现几何误差可以很好地借助传感器噪声模型表示,而光度误差则不能.据此,提出了一种混合方法,以结合  $t$  分布的光度误差和概率传感器噪声模型的几何误差.

构建基于最优估计的定位模型,相较于传统的滤波模型具有更强的可靠性.例如, Buczko 等<sup>[89]</sup>提出了一种基于立体视觉里程计的离群值去除方法,适用于深度值较大环境下的高速位姿变化估计.首先,作者研究了在给定位姿变化的固定误差下,重投影误差在三维位置上的方差,从而得出基于重投影误差固定阈值的异常值检测是不合适的结论.然后,提出了一种光流相关的特征自适应缩放重投影误差,以保证每个特征的 3D 位置几乎不变.这种特征自适应缩放是从显示相机纵向姿态变化、光流绝对值和特征距离之间关系的近似中推导出来的.利用这种缩放,作者开发了一个迭代交替方案来指导内值与异常值的分离.它优化了在基于给定的位姿变化中找到一个良好的去除异常值的准则和基于当前的内列集改进位姿变化假设之间的权衡. Steinbrücker 等<sup>[90]</sup>提出了一种能量最小约束的方法实现密集的 RGB-D 图像视觉里程计.关键思想是通过最小化反向投影误差来解决潜在的逆问题:其目标是从表示摄像机运动的特殊欧氏群  $SE(3)$  中找到一个刚体变换向量,使配准的第二幅图像完全匹配第一幅图像.通过顺序凸优化逼近这个非凸能量的最小值.然后,将能量线性化,并求解表示所需刚体运动的三维扭转坐标的法方程.由于线性化只适用于小的扭转,作者应用了从粗到细的方法以应付较大的相机运动.将该方法的性能与迭代最近点算法 (Iterative closest point, ICP)<sup>[91]</sup> 进行比较,作者发现,ICP 对较大的摄像机位移更有效,而作者的方法在小相机运动的情况下提供了更好的结果.

Hesch 等<sup>[92]</sup>针对视觉辅助惯性导航系统中估计量不一致问题,从系统可观察性的角度展开研究.作者假设不一致的主要原因来自不可观测方向所获得的虚假信息,这导致了较小的不确定性和较大的估计误差.作者开发了一个可观察性约束的视觉惯性导航系统 (Visual-inertial navigation system, VINS),它显式地强制约束系统的不可观察方向,防止虚假信息增益和减少不一致性.该框架适用于

VINS 问题的几种变种, 如视觉同步定位和测绘, 以及使用多状态约束卡尔曼滤波器的视觉惯性里程计. 作者分析了真实系统和估计系统的结构, 并证明了对于真实系统存在四个不可观测方向, 而用于估计目的的系统只包括三个不可观测的方向. 此外, 作者假设 VINS 中不一致的一个主要来源是当方向信息沿围绕重力矢量旋转的方向被错误地投影时所获得的虚假信息. 对此, 提出了一个估计量修改, 它可以明确地禁止这种不正确的信息增益. 当使用线性化估计器 (如扩展卡尔曼滤波) 时, 所提方法足够通用, 可以应用于多个 VINS 中. 作者同时提供了大量的证据来证明标准 VINS 方法的不一致性, 并通过蒙特卡罗模拟验证了其方法有效性, 以表明与标准 VINS 相比, 该方法提高了一致性并减少了估计误差.

针对巡飞弹弹载微机电系统在长航时过程中航向易发散的问题, 朱启举等<sup>[93]</sup>提出一种基于景象匹配的低精度 MEMS 航向修正与组合算法. 首先, 基于匹配点几何关系与匹配重投影设计了两类航向误差模型, 前者根据实时图与基准图匹配点连线与图像纵轴计算反正切值, 作差后计算出航向修正值. 后者利用实时图与基准图的匹配关系构建重投影误差, 通过优化计算出航向最优值. 然后, 基于上述两种模型, 根据匹配点数量设计一种航向修正与组合算法. 实验结果表明, 与传统算法相比, 该算法提供了一种航向修正方法, 且对传感器精度依赖性较低, 具备一定的通用性.

结合深度学习的方法, 能够有效克服传统方法中存在的一些问题. Yin 等<sup>[94]</sup>提出一种新的尺度恢复方法, 用于解决视觉里程计中单目传感器缺乏深度信息的问题. 该方法利用深度神经网络来估计图像深度, 并据此计算翻译比例尺. 该方法考虑了连续图像和运动约束对深度预测的影响. 为此, 采用卷积神经网络和条件随机场相结合的卷积神经网络进行深度估计. 卷积神经网络从输入图像中回归粗深度图, 随后, 条件随机场通过约束条件对粗深度进行细化. 在细化深度的基础上, 得到了运动的尺度. 通过迭代计算深度和尺度, 这一框架在计算精度上为视觉里程计模型提供了可靠的深度补充信息支持. Ye 等<sup>[95]</sup>提出了一种新的全景视觉里程计框架, 以实现更全面的场景和全景分割信息的建模. 所提方法将视觉里程测量和视频全光分割建模在一个统一的视图中, 使两者相互受益. 在图像全景分割的指导下, 引入了一个全景更新模块. 该全景增强 VO 模块可以通过全景感知动态掩模减轻动态物体对相机姿态估计的影响. 另一方面, 还利用从 VO 模块

获取的相机姿态、深度、光流等几何信息, 将当前帧的实时全光分割结果融合到相邻帧中, 从而提高分割精度.

### 5.3 小结

组合定位算法融合绝对定位算法与相对定位算法, 实现多场景复杂任务的定位. 随着视觉技术的发展, 基于视觉的组合定位方法逐渐实现了精度上的提升, 在一般场景下的定位精度已经能够满足常规任务的需求.

但是对于组合定位算法, 仍然存在一些问题. 首先, 误差的累积是一个重要的问题, 特别是在长时间运行的情况下, 传感器误差可能会积累导致定位精度下降. 其次, 环境变化可能影响传感器数据的准确性, 如天气、地形、建筑物等因素可能会影响定位结果. 此外, 多传感器数据融合需要考虑多种因素, 设计合适的融合算法是一项复杂的工程任务. 在动态环境下, 传感器数据可能会发生突变, 需要及时响应以保证定位的准确性和稳定性. 同时, 一些高精度的组合定位系统可能需要大量的计算资源和能耗, 对移动设备和嵌入式系统来说是一个挑战. 此外, 组合定位涉及到用户位置信息的采集和传输, 可能引发安全和隐私问题. 最后, 地图数据的质量和更新频率也直接影响定位的精度和稳定性. 综上所述, 虽然组合定位技术在提高定位精度和可靠性方面取得了进展, 但仍然面临一些问题, 需要不断地研究和改进.

## 6 代表性方法与评价指标

景象匹配无人机视觉定位算法经历多年发展, 逐渐形成了成熟的理论体系与评价方法. 本节汇总各个类别中的代表性无人机视觉定位算法与算法评价指标.

### 6.1 代表性方法汇总

本节将文中提及的相关算法汇总后, 以特征提取的不同方式将其分为三类, 即基于模板匹配、基于手工特征和基于度量学习的景象匹配算法.

如表 2 所示, 为了获得符合真实情况的无人机视角数据, 学者们大多在真实环境中采集数据, 用于模型的训练与测试. 但是, 受限于复杂的拍摄与数据标注过程, 往往得到的数据集规模存在明显的局限性. 同时, 由于算法提出的时间不同, 且采用了不相同的测试方法、测试指标以及测试数据集, 想要客观地进行比较并非易事.

基于模板匹配和特征点法的景象匹配算法集中



表 2 代表性方法汇总  
Table 2 Summary of representative methods

方法	算法分类	实现方式	地图数据来源	无人机数据来源	航拍影像尺寸 (像素)
Loeckx 等 <sup>[16]</sup>	模板匹配	NCC	谷歌地图	真实拍摄	—
Yol 等 <sup>[18]</sup>	模板匹配	MI	谷歌地图	真实拍摄	—
Fan 等 <sup>[19]</sup>	模板匹配	NCC	谷歌地图	谷歌地球	—
Levin 等 <sup>[20]</sup>	模板匹配	CC	DEM 数据	DEM 数据	—
Lin 等 <sup>[55]</sup>	模板匹配	MI	谷歌地图	谷歌地球	720 × 480
Huang 等 <sup>[56]</sup>	模板匹配	MI	谷歌地图	真实拍摄	640 × 480
Wan 等 <sup>[57]</sup>	模板匹配	PC	卫星数据	真实拍摄	3 648 × 2 736
Patel <sup>[58]</sup>	模板匹配	NID	谷歌地图	真实拍摄	560 × 315
Shan 等 <sup>[23]</sup>	特征点法	HOG	谷歌地图	真实拍摄	850 × 500
Masselli 等 <sup>[27]</sup>	特征点法	ORB	谷歌地图	真实拍摄	640 × 480
Chiu 等 <sup>[30]</sup>	特征点法	2D-3D 点	DARPA	真实拍摄	—
Mantelli 等 <sup>[31]</sup>	特征点法	abBREF	谷歌地图	真实拍摄	—
Shan 等 <sup>[33]</sup>	特征点法	MSD+ LSS	谷歌地图	真实拍摄	—
Woo 等 <sup>[34]</sup>	特征点法	角点	谷歌地图	真实拍摄	—
Pluckter 等 <sup>[59]</sup>	特征点法	ORB	谷歌地图	真实拍摄	—
Pan 等 <sup>[61]</sup>	特征点法	SIFT	谷歌地图	真实拍摄	586 × 452
Couturier 等 <sup>[81]</sup>	特征点法	ORB	—	真实拍摄	—
Couturier 等 <sup>[82]</sup>	特征点法	SURF	—	真实拍摄	1 920 × 1 080
Goforth 等 <sup>[42]</sup>	深度学习	VGG16	谷歌地图	真实拍摄	4 608 × 2 592
Amer 等 <sup>[44]</sup>	深度学习	VGG16	谷歌地图	Bing 地图	500 × 500
Nassar 等 <sup>[48]</sup>	深度学习	U-Net	谷歌地图+Bing 地图	谷歌地球	—
Marcu 等 <sup>[60]</sup>	深度学习	MSMT	OpenStreetMap	—	1 500 × 1 500
Schleiss <sup>[52]</sup>	深度学习	cGAN+SSD	—	真实拍摄	—
Zheng 等 <sup>[64]</sup>	深度学习	ResNet	谷歌地图	谷歌地球	512 × 512
Workman 等 <sup>[65]</sup>	深度学习	—	谷歌地图	谷歌街景/Flickr	—
Hays 等 <sup>[66]</sup>	深度学习	—	网络爬取	Flickr	—
Weyand 等 <sup>[62]</sup>	深度学习	LSTM	谷歌地图	谷歌地球	—
Wu 等 <sup>[63]</sup>	深度学习	Lucas-Kanade	真实拍摄	仿真数据	5 632 × 5 376
Li 等 <sup>[96]</sup>	深度学习	channel attention	真实拍摄	真实拍摄	—
Kinnari 等 <sup>[97]</sup>	深度学习	正交投影	谷歌地图	真实拍摄	4 800 × 2 987
Wen 等 <sup>[98]</sup>	深度学习	SiamRPN	谷歌地图	真实拍摄	—
Wang 等 <sup>[78]</sup>	深度学习	LPN	谷歌地图	谷歌地球	512 × 512
Dai 等 <sup>[99]</sup>	深度学习	FSRA	谷歌地图	谷歌地球	512 × 512
Tian 等 <sup>[100]</sup>	深度学习	PCL	谷歌地图	谷歌地球	512 × 512
Zhu 等 <sup>[101]</sup>	深度学习	SUES-200	谷歌地图	真实拍摄	512 × 512

于解决遥感卫星地图与航空影像保持一致的视角与缩放比例情景下的问题, 这样就将景象匹配简化为配准问题. 随着算法研究的深入, 出现了基于手工特征的匹配方法, 由于局部特征在旋转、尺度变化等条件下具有较好的鲁棒性, 使得景象匹配算法的应用场景从完全理想逐渐转化为近似真实复杂环境. 后来, 学者们利用深度学习方法捕捉空-天异源图像的语义一致性, 有效地提升了景象匹配对于多源影像分布差异的理解能力. 直至近些年来, 基于

度量学习的景象匹配算法已经延伸至解决低空飞行时产生的多视角问题.

基于度量学习的景象匹配算法在适应复杂场景、提高判断精度等方面取得了进步. 然而, 当前缺乏统一的数据采集标准, 导致不同算法采集到的数据质量不一致, 进而难以实现对视觉定位算法效果的统一比较. 这种数据质量的不一致性可能来自于不同飞行平台、不同天气条件、不同光照情况等因素的影响, 使得数据具有较大的差异性.

例如, 基于度量学习的方法与基于模板匹配的方法相比, 由于多源多视角场景带来的挑战使得问题的困难程度极大提升. 在不同的数据集中测试不同的算法难以客观评价不同类别算法的性能, 如多源多视角景象匹配数据集中衡量景象算法性能时, 由于多源多视角问题的复杂性高于多源匹配, 呈现为表 2 中的结果, 即多源多视角算法测试结果平均定位精度低于模板匹配在多源数据中的测试结果. 对此, 使用虚拟引擎渲染的仿真数据集由于可以通过脚本自动进行数据采集, 不仅能够成倍地扩大数据集的样本数量, 而且由于采集过程中能够按照既定经纬度数值进行影像采集, 极大地简化了数据标注的过程. 同时, 若采用统一的虚拟引擎采集数据, 对于算法的运行效果的对比将更具有说服力. 然而, 虚拟渲染的数据集中, 缺少真实环境中复杂的环境影响, 导致模型在真实环境中运行存在一定的模型鲁棒性缺陷. 与其他视觉任务类似, 引入数据增强算法, 能够扩充已有的训练样本, 提升模型对于人为加入的噪声干扰的抗性. 由于人为加入的噪声分布的不确定性, 模型在应对真实环境中更为复杂的场景时, 依然表现出明显的不足.

充分利用虚拟数据集中丰富的样本学习视觉定位中的模式, 并且在小规模真实数据集中进行环境适应性的提升, 该方式成为目前解决缺少无人机视觉定位数据样本的一个最优的方法.

## 6.2 评价指标汇总

评价指标是衡量模型性能优劣的定量指标. 通常, 一种评价指标只能反映模型一部分性能, 如果选择的评价指标不合理, 那么可能会得出错误的结论, 故而应该针对具体的数据, 模型选取不同的评价指标. 为了更全面地评价视觉定位算法, 本文整理汇总了在无人机视觉定位算法中最常见的评价指标:

1) 定位误差. 定位误差直观反映每次定位结果与真实值之间的距离差距, 反映定位的直观效果. 定位误差可以细分为经纬度方向的距离误差, 与真实距离误差构成矢量三角形.

2) Top- $K$  候选. 将候选的数量  $K$  与来自评估集的查询图像的分值进行统计, Top- $K$  候选返回候选位置的有序列表时, 计算使用固定数量的候选中正确匹配的图像数量.

3) 定位正确率. 定位正确率检索最可能的位置, 并度量至少在给定阈值误差的情况下正确的查询次数. 定位正确率的优点是, 能够直接得到该方法对评估集中给定部分的查询图像的准确性.

4) 地理定位区域/感兴趣的区域. 对于每个查询, 根据地理定位区域/感兴趣的区域判断定位效果. 在候选区域统一的情况下, 此方法与 Top- $K$  候选相同, 能够给出若干次定位中准确匹配定位的概率.

5) 精度/召回率. 该类方法是评价分类和检索方法的标准指标. 在非学习的地理定位算法中, 是一种非常少见的评价方法. 该类评价指标主要出现在近些年开始流行的基于度量学习的方法中.

在评估景象匹配算法的性能时, 单一的评价指标可能无法全面反映算法的效果. 因此, 为了更充分、立体地比较不同算法的性能, 采用多个评价指标进行对比衡量是很重要的.

使用多个评价指标可以从不同的角度对算法进行评估, 更全面地了解其优缺点. 例如, 可以同时考虑匹配精度、计算效率、鲁棒性、泛化能力等指标. 匹配精度可以通过计算正确匹配的数量或误差来衡量; 计算效率可以考虑算法的运行时间或计算复杂度; 鲁棒性可以通过在不同环境下的实验验证; 泛化能力则可以通过在不同数据集或场景下进行验证.

通过综合考虑多个评价指标, 可以更全面地了解不同算法的优劣势, 并选择最适合特定应用场景的算法. 因此, 在进行算法比较和性能评估时, 采用多个评价指标是非常重要和必要的.

当前, 景象匹配领域的挑战之一为缺乏类似于 ImageNet 的标准数据集. 这样的标准数据集不仅需要包含丰富多样的场景, 如城市、乡村、山区等, 还需要考虑各种不同的环境条件, 如晴天、雨天、雪天、雾天等. 构建标准数据集将有助于提高算法模型之间对比的说服力, 使得研究人员能够更准确地评估和比较不同算法的性能. 然而, 建立一个标准的、丰富多样的景象匹配数据集, 需要耗费大量的时间和精力, 并且需要考虑到数据的真实性、多样性和标注准确性等问题. 因此, 这需要学术界和工业界的共同努力来完成.

## 7 问题归纳与解决思路

无人机视觉定位发展经历了模板匹配、手工描述子直至深度学习方法, 方法上从像素级别向场景级别发展, 应用上从特殊标定场景向一般场景甚至极端复杂场景转变. 国内外学者针对无人机视觉定位的精度提升、实时性提升方面开展了大量的研究工作, 取得了一定的进展. 但是, 由于无人机飞行过程中, 复杂的场景、变化的环境、平台的差异等因素, 导致视觉定位任务仍然存在诸多挑战.

纵观无人机定位算法的发展, 从依赖于通信定

位信号的传统方法到完全自主的基于运动数据解算的算法,无人机的自主定位能力、定位算法精度、定位算法实时性得到大幅提升.随着计算机视觉与深度学习的发展,基于景象匹配的定位算法因其能够摆脱定位系统对定位信号的依赖,并且长时运行时不产生累计误差,得到了越来越多的关注.

在整个无人机视觉定位算法发展过程中,设计出更准确、更稳定、更鲁棒的算法模型一直是研究人员的目标.同时,由于缺少丰富场景,海量样本的视觉定位数据集对模型的训练验证与落地都存在一定的约束.研究学者们在不断丰富无人机视觉定位数据集的基础上,开展算法的泛化性能提升相关研究同样重要.

本节归纳景象匹配算法中待攻克问题,结合现今计算机视觉技术的发展现状,进行潜在的解决方法讨论.

## 7.1 空-天多源差异

空-天异源图像指的是来自不同传感器、不同观测平台(如卫星和无人机)以及不同时间的图像.由于这些图像可能具有不同的分辨率、光谱范围、观测角度和大气条件,特征的提取与对齐成为一个复杂且具有挑战性的问题.在空-天异源图像中,由于图像来源的多样性,常规的特征提取方法可能不再适用.传统的颜色、纹理和形状特征在不同传感器和观测条件下可能表现出很大的变化,因此需要针对异源图像开发更具适应性的特征提取方法.这可能涉及到在不同光谱波段中提取特征、使用度量学习方法进行特征学习等.传统的度量学习模型采取共享权重的孪生神经网络实现异源图像特征的统一空间映射.在一些极端场景下,差异的数据源之间存在极大的域鸿沟,使得共享权重的特征提取网络难以实现有效的特征空间统一.本节汇总空-天多源差异中的主要待解决问题,并讨论潜在的解决方法.

### 7.1.1 几何不对齐

几何校正是对遥感图像进行坐标系统一、图像投影、几何变换等处理的过程,旨在使图像能够在地理空间中准确表达.其主要目的是确保原始图像的像素点能够与地球表面上的真实地理位置相对应,以便进行后续的匹配度量.

在实际的景象匹配环境中,不同传感器或观测条件下获取的图像存在几何失真,如投影差异、尺度不一致等.为了在进行图像度量之前获得相对对齐的图像,提升匹配的精度,需要对空-天异源图像进行几何校正<sup>[102]</sup>,以确保在地理坐标上的对应性.

### 7.1.2 辐射不一致

辐射误差是遥感图像处理中常见的误差类型,指的是由于遥感图像中像素值的不一致性引起的误差.这种差异可以是由于不同传感器之间的差异,如不同传感器的光谱响应不同,也可以是由于传感器内部因素,如温度、湿度等环境因素的影响.不同传感器的辐射响应函数可能存在差异,受大气吸收和散射等因素的影响.为了在不同图像之间进行比较,通常需要进行辐射校正<sup>[103]</sup>.首先,通过测量黑体的辐射值来校正图像中的灰度偏移和增益误差;然后,利用白板或均匀灰度面作为参考标准,校正图像中的色彩偏移和对比度误差.

在将数据灌入景象匹配的度量模型之前进行辐射系数校正,有助于确保图像在度量条件上相对一致,减少度量过程中由于辐射差异而导致的度量不一致问题.

### 7.1.3 地物变化

景象匹配问题需要应对不同时间获取的图像中可能存在的地物变化挑战.地物的变化对景象匹配构成重大挑战,度量模型不仅需要原有相似性判断任务中获得相对稳定的结果,还需要有效解决地物迁移后的匹配问题.

当地物发生变化时,原有度量模型中保存的一致性度量知识可能无法实现可靠的相似性判断.根据地物变化的程度,可以分为两类情况:1)变化后仍然保留足够的度量线索;2)变化剧烈难以正确辨别.

对第一种情况,度量模型可以采用多级特征挖掘的方法,利用不同尺度的特征信息建立鲁棒的地物变化适应能力.然而,当面临第二种情况时,度量模型面对的影像数据经历了一场灾难性的改变,其细节纹理信息和高级语义信息由于地物的剧烈变化而难以捕捉.在这种情况下,只能依赖卫星定位信号或者基于飞行过程中的可靠场景进行累积运动位置推断等方式实现对无人机当前位置的计算.这种综合运用多种信息源的方法有助于有效应对地物变化对景象匹配带来的挑战.

### 7.1.4 解决思路

综合来看,空-天异源图像的特征提取与对齐问题涉及多个方面的挑战.对此,可行的方法包括:

1) 设计数据预处理正则化模型,借助于建立神经网络中图像的预处理模型,抑制或消除空-天异源影像的不对齐问题.

2) 构建更可靠的语义一致度量关系模型,通过在双分支孪生神经网络分别建立空间与通道注意力机制,增强异源图像之间的语义对应关系,提升相

似性度量在非对齐条件下的一致性。

3) 构建更可靠的多级特征一致性关系模型, 通过多级特征融合模型与度量决策模型, 充分挖掘神经网络中不同层级特征中包含的局部与全局信息, 提升相似性度量在非对齐条件下的一致性。

总之, 针对空-天多源问题, 解决问题的思路集中于两点。首先, 在数据预处理阶段借助数据矫正方法消除多元差异; 其次, 在特征捕捉时, 挖掘一致性元素作为连接孪生网络不同分支的桥梁。

## 7.2 多视角图像差异

多视角图像匹配在遥感图像处理中扮演着至关重要的角色。在这一过程中, 为了实现准确的匹配, 关键在于提取具有鲁棒性的特征。由于无人机在飞行过程中的姿态变化, 导致获得的图像存在明显的视角差异, 使得度量学习模型难以正确完成相似性度量。同时, 不同视角下的光照条件变化, 导致多视角图像在不同光照条件下呈现出视觉成像结果的显著差异。甚至在极端条件下, 可能会出现遮挡或由于缺少环境光而导致场景部分信息丢失的问题。本节汇总了多视角景象匹配中的主要待解决问题, 并讨论潜在的解决方法。

### 7.2.1 描述子不稳定

传统的描述子在面对多视角图像时存在稳定性不足的问题, 特别是当航空影像出现大视角变化时, 透视效应可能导致关键信息的遮挡和丢失, 甚至可能导致匹配失败。

为了解决这一问题, 目前主流的方法主要采用两种策略来提升度量的可靠性, 即通过充分挖掘影像数据和引入额外的度量辅助信息来增强描述子在多视角复杂环境下的鲁棒性。

针对多视角挑战, 一些方法通过深度学习等技术充分挖掘影像数据, 提升描述子的性能。这包括对模型进行训练, 使其对于大范围的视角变化具有更好的适应性。同时, 通过收集更多的复杂视角变化的影像数据, 可以扩大模型的场景认知范围, 有助于提升算法的性能。另外, 度量学习方法借鉴传统手工特征融合的思想, 通过引入融合策略, 结合多种匹配方法的优势。这种方法可以在处理视角变化时, 充分发挥每种匹配方法的优势, 从而解决景象匹配算法单一手段的局限性。这样的融合策略有助于提高匹配的鲁棒性和准确性, 使得算法在多样的环境中更具适应性。

### 7.2.2 几何信息缺失

几何信息的表征涉及地面物体实际分布情况, 这是一种客观存在的分布特征。随着视角的变化,

航空影像的成像差异引起对空间几何信息不同表达。在这种成像差异中, 有效捕捉几何信息对于增强多视角可靠性至关重要。然而, 经过调研发现, 目前的算法在这方面仍有待深入研究。

传统的卷积神经网络通过堆叠卷积和池化等模块来获得图像的特征描述。然而, 在这些感知过程中, 图像的空间分布信息通常被忽略, 导致在图像度量过程中几何结构的一致性线索未被充分挖掘。采用类似图卷积网络 (Graph convolution network, GCN) 结构对空间分布信息进行感知, 将有望有效建立多视角空-天异源影像的一致性度量关系模型。这种方法能够更好地捕捉影像中的几何信息, 提高度量模型对不同视角下的场景一致性的把握, 从而增强算法在多视角环境中的可靠性。

### 7.2.3 解决思路

在多视角图像匹配中, 提升特征提取的鲁棒性是一个重要的研究问题。对此, 可行的方法包括:

1) 采用多尺度特征、纹理描述子、深度学习方法、几何信息和跨模态融合等方法<sup>[104-105]</sup>, 提高特征对图像的代表能力;

2) 借助于图卷积神经网络, 将几何信息纳入特征提取过程中可以提高鲁棒性;

3) 使用 3D 信息, 如点云数据, 可以帮助捕捉地物的三维结构, 从而在不同视角下实现更好的匹配。

总之, 特征的鲁棒性提升主要体现在两点: 首先, 更充分地挖掘现有的特征; 其次, 补充一些额外的信息作为度量辅助。

## 7.3 度量模型泛化性能不足

度量模型在景象匹配中常用于衡量图像之间的相似性, 从而确定匹配点。然而, 不同场景的图像可能具有多样性, 传统的度量模型在不同场景下的泛化性能可能不佳。模型的泛化性能是评估模型优劣的重要指标之一。当前带标记的无人机视觉定位数据集相对较为有限, 这对于算法模型的泛化能力构成了一个极大的挑战。本节汇总了泛化性能提升的主要待解决问题, 并讨论潜在的解决方法。

### 7.3.1 特征鲁棒性弱

特征在不同场景下的鲁棒性对于泛化性能至关重要, 然而, 目前主流的 CNN 特征在复杂场景中的鲁棒性相对较弱, 对于具有挑战性的场景的特定感知能力存在明显局限。这是因为在网络模型设计中未充分考虑到复杂场景的关键区域感知, 导致传统特征在某些具有显著度量线索的场景中表现良好, 但在复杂场景中效果不佳。

与精度提升类似,融合多种信息对于提升泛化性能同样至关重要.通过模态互补,克服单一模态度量结果的局限性成为一个新的待解决问题<sup>[106]</sup>.为了弥补 CNN 特征鲁棒性较弱的短板,一方面可以建立更细化的特征提取模型,以更好地捕捉复杂场景的关键信息;另一方面,构建多模态融合模型是一种有效的策略,可充分发挥不同模态感知数据的优势.在多变的场景中,利用多模态信息有助于发挥不同算法的长处,规避单一方法的局限性,提高泛化性能.

### 7.3.2 缺少训练样本

训练样本数据为度量模型提供了在已知场景中建立一致性映射关系的知识基础.基于学习的方法的一个典型特点是模型的性能受限于训练数据,这正是庞大规模的预训练模型 ImageNet<sup>[107]</sup> 所试图解决的问题.

为解决这一问题,从问题的本源出发,建立超大规模训练模型,构建多场景、复杂环境的丰富训练样本集,对于克服模型缺少数据的困境至关重要.通过引入更加丰富、多样的训练数据,模型可以更好地捕捉和理解不同场景的特征,提高对新场景的泛化性能.这种方法对于提升景象匹配算法在现实世界多变环境中的适应性具有显著的意义.

### 7.3.3 解决思路

在景象匹配算法中,提升度量模型的泛化性能是一个关键问题.对此,可行的方法包括:

- 1) 提取鲁棒的特征,融合多模态信息,使用数据增强技术和运用迁移学习等方法,可以提高模型在不同场景下的匹配准确度和稳定性.

- 2) 使用多样性的数据集进行训练和测试,可以更好地评估模型的泛化性能.

- 3) 结合多光谱和高光谱数据,或融合图像与激光雷达数据,可以在不同场景下提供更全面的信息,从而提高匹配准确度;通过在训练集中引入不同的变化(如旋转、翻转、遮挡等),来增加模型在不同场景下的适应性,从而提升泛化性能.

- 4) 通过在已有数据集上训练模型,然后在目标场景上进行微调,可以提高模型在不同场景下的性能.借助迁移学习可以将在一个场景中训练的模型知识迁移到另一个场景中,从而获得模型的自适应升级效果.

总之,度量模型泛化性能的提升思路主要集中于两点:首先,挖掘变化的环境中位置不变的特征,保持模型的稳定;其次,通过数据增强或采集新数据的方式,使得模型学习更多不同场景的分布知识.

## 7.4 度量模型的实时性不足

在景象匹配算法中,度量模型的实时性提升问题指的是如何在较短的时间内完成图像匹配任务,尤其是在需要即时结果的应用中,如导航、无人机控制和实时监测等.由于无人机移动平台的计算能力有限,因此在设计无人机视觉定位算法模型时,应当保证模型的精简.在此基础上,考虑到地理坐标参考地图的规模问题,需要采取一定的先验位置知识的指导或者采用压缩编码的方式,存储更多的地图数据.本节汇总了实时性提升的主要待解决问题,并讨论潜在的解决方法.

### 7.4.1 算法轻量化

算法轻量化是提升景象匹配实时性的有效策略.通过优化特征提取网络模型的规模,在提取高质量特征的同时最大限度地减少计算量.对于实时性要求高的场景,选择轻量级但能保留有用信息的特征成为一种合理选择<sup>[108]</sup>.此外,通过复杂网络的剪枝,实现网络轻量化同样是一种有效的解决方案.

在对实时性要求较高的情况下,可以考虑采用一些近似算法,例如局部搜索、快速检索等,以降低计算复杂度.这种方式在保持算法性能的同时,通过简化计算过程提高了匹配速度.

### 7.4.2 硬件加速计算

利用专用硬件,如图形处理单元(Graphics processing unit, GPU)进行计算加速是一种有效的手段,可以显著提高计算速度. GPU 在并行处理方面具有出色的性能,适合处理大规模数据集,因此在景象匹配的计算过程中,借助 GPU 能够加速特征提取、配合其他计算密集型任务,提高整体的运算效率.

同时,可将匹配算法分解为多个子任务,并在多个处理单元上并行计算.这样可以充分利用多核处理器或者分布式计算系统,同时处理不同部分的任务,从而加速整个匹配过程,这对于大规模数据的实时匹配尤为重要.

此外,通过预处理数据和结果缓存,将有效减少重复计算.通过提前计算和存储一些计算结果,避免在相同数据上进行重复操作,降低了计算的时间复杂度,提高了整体匹配的速度和效率.

### 7.4.3 解决思路

在景象匹配算法中,提升度量模型的实时性是一个重要的问题.对此,可行的方法包括:

- 1) 通过优化特征提取、特征选择、加速硬件、并行计算、近似算法以及预处理和缓存等方法,可以显著提高算法的实时性能.

2) 在特定的应用中, 需要权衡速度和精度, 选择最适合的算法和优化策略. 基于此, 对特征提取网络进行轻量化设计, 在保证匹配精度不变的前提下, 优化网络规模, 提升特征提取效率.

3) 使用金字塔结构进行预处理, 将匹配结果缓存, 避免重复计算.

度量模型的实时性提升思路主要集中于两点: 首先, 对复杂特征提取模型进行轻量化改造; 其次, 将模型在飞行平台部署的时候进行硬件加速设计.

## 7.5 遥感卫星参考地图管理

由于无人机视觉定位方法的核心在于匹配异源数据, 因此无人机视角与卫星地图的尺度对齐成为实际操作中不可避免的挑战. 卫星地图的获取通常采用滑动窗口在大场景卫星影像中按照无人机影像尺度进行截取. 然而, 这一截取方式导致地图数据库中大量的重叠, 这种冗余数据对于算法的高效运行构成了一定的负担. 本节汇总了遥感卫星地图获取和存储的主要待解决问题, 并讨论潜在的解决方法.

### 7.5.1 数据获取

遥感卫星参考地图是景象匹配算法的基础, 为匹配库提供了对应的地理位置信息, 因此, 高精度地图库直接关系到匹配算法在完成后对地理位置估计的可靠性.

获取参考地图的方法包括实地调查和采集地面控制点等信息. 这些地面控制点通常具有已知的地理坐标, 用于将遥感图像与地球表面的实际位置对应. 除了传统的可视遥感影像作为地图数据库, 现有的地理信息系统数据、数字高程模型等同样可以作为参考地图. 这些数据源通常具有较高的精度和全球覆盖性, 可用于遥感图像的地理定位. 利用丰富模态的地图库参考, 景象匹配的后续地理位置计算将更加准确.

### 7.5.2 数据存储

参考地图在存储时需要选择适当的格式, 以确保在算法中能够有效使用. 常见的格式包括栅格数据 (例如 GeoTIFF) 和矢量数据 (例如 Shapefile) 等. 在存储过程中, 需要充分考虑数据的结构、坐标系信息等方面的要素.

选择合理和适当的存储方式对于庞大规模的遥感影像地图库至关重要. 这不仅关系到所需的硬件资源, 同时也对算法的实时性匹配产生直接影响. 有效的存储格式和结构设计能够提高数据的检索效率, 从而保证匹配算法在处理时能够迅速而精准地访问所需信息.

在选择存储格式时, 还应该考虑数据的长期维护和更新. 合理的数据结构和格式不仅有助于提高算法的性能, 同时也为地图库的维护提供了更为便捷的手段. 因此, 在参考地图的存储过程中, 需要综合考虑数据的结构、格式、坐标系等多个因素, 以实现对于算法和系统的高效支持.

### 7.5.3 解决思路

参考地图的精度对于匹配算法的准确性至关重要. 因此, 在获取和存储参考地图数据时, 需要确保数据具有足够的准确性和精度. 对此, 可行的方法包括:

1) 地面控制点的精确测量、精细的几何校正和质量验证等;

2) 借助云存储技术可以方便地存储和共享大规模的遥感数据;

3) 建立基于变化检测的数据更新的机制, 提升数据更新效率, 确保参考地图数据的时效性<sup>[109]</sup>.

总之, 遥感卫星参考地图的获取与存储思路主要有两点: 首先, 提升数据存储与利用的效率; 其次, 面对地物更新时针对性补充待更新部分.

## 7.6 无人机平台差异

不同类型的无人机平台在匹配定位方面表现出不同的特征, 这涉及到它们的机动性、感知能力以及飞行特性等多个方面<sup>[110]</sup>. 在这个广泛的领域中, 可以看到多样化的飞行平台, 每一种都在不同场景下发挥着独特的优势. 例如, 固定翼无人机以其高速巡航和大范围覆盖的特点, 在广阔区域内执行匹配定位任务时具备优势; 而垂直起降的旋翼无人机则适用于狭窄或复杂环境下的匹配定位, 灵活性更高. 混合翼无人机则兼具了固定翼和旋翼的特性, 拓展了适用场景. 在这种多样性中, 学者们致力于进一步提升各类型无人机的定位性能, 以适应不同需求. 因此, 深入了解和研究各种无人机平台对匹配定位的影响, 对于优化系统设计、提高任务执行效率具有重要意义. 通常, 景象匹配算法主要搭载的平台有固定翼、旋翼和混合翼平台, 本节对这些平台的特点进行简要分析.

### 7.6.1 固定翼平台

固定翼无人机通常以其稳定的巡航速度和航程为特点, 为航拍提供了相对固定的视角. 然而, 其在某些方向上的视野可能受到一定限制, 因此需要更全面的传感器集成, 包括惯性导航系统、GPS 和相机等, 以提升对地面目标的感知能力. 由于具备较长的飞行时间, 固定翼无人机有助于覆盖更广泛的区域. 为了在航拍任务中取得更佳效果, 必须考虑

提高定位频率, 以确保数据的准确性. 因此, 对固定翼无人机的性能和感知系统进行进一步的研究和优化, 将为其在各种应用场景中的成功执行任务提供有力支持.

### 7.6.2 旋翼平台

旋翼无人机以其低空飞行的特点, 为感知提供了更高分辨率的可能, 然而却伴随着相对较为剧烈的视角变化. 低空飞行不仅增加了对障碍物检测和避障技术的挑战, 同时也带来了观测影像的透视性, 形成多视角的差异. 然而, 旋翼无人机的飞行时间相对较短, 这要求更为有效的轨迹规划和能源管理. 此外, 负载能力也可能受到限制, 因此在匹配定位时需要权衡负载和飞行性能. 对于旋翼无人机的匹配定位, 需要综合考虑飞行高度、视角变化、飞行时间等多个因素, 以优化系统设计, 确保在特定应用场景中取得最佳性能.

### 7.6.3 混合翼平台

混合翼飞行器因其独特的结构设计而具备多种飞行器的优势, 为匹配定位任务提供了灵活性和多样性. 特别是具备垂直起降能力的混合翼无人机, 在匹配定位时可能需要更加灵活的轨迹规划和动态路径规划. 这是因为其能够从垂直状态转为水平飞行状态, 因此需要考虑更多的飞行动态因素. 在实际操作中, 混合翼无人机可以更自如地适应不同的环境, 从低空到高空的飞行, 使其在匹配定位任务中表现出色. 通过综合考虑其垂直起降能力和灵活性, 算法设计者可以更好地设计匹配定位算法, 以充分发挥混合翼无人机在不同场景下的优越性. 因此, 在匹配定位领域中, 混合翼无人机为实现更为灵活和高效的定位任务提供了潜在的机遇.

## 8 展望

尽管无人机视觉定位已经取得了一些显著的进展, 但相较于同属于绝对定位范畴的全球定位系统, 仍然存在一定的定位精度差距. 这意味着仍需要不断改进和优化算法, 以适应各种真实世界环境. 在未来, 结合当前人工智能领域最新的“大模型”、“迁移学习”技术<sup>[11]</sup>, 有望为景象匹配带来新的发展契机, 将其推向一个新的高度. 在这一背景下, 加大对无人机视觉定位算法研究的投入显得尤为重要. 通过最新的计算机视觉方法解决当前面临的准确性、稳定性、鲁棒性和实时性等问题, 将为无人机的自主化和智能化发展提供有力支持, 促使其在各个领域发挥更为关键的作用. 因此, 对无人机视觉定位算法的深入研究和技术创新, 将为未来智能化无人机的构建和应用带来积极而深远的影响.

## References

- Li Xue-Long. Vicinagearth security. *Communications of the CCF*, 2022, **18**(11): 44–52  
(李学龙. 临地安防. 中国计算机学会通讯, 2022, **18**(11): 44–52)
- Chen Jie, Fang Hao, Zeng Xian-Lin. On the intelligent development of unmanned platforms in high-risk industries. *Scientia Sinica Informationis*, 2021, **51**(9): 1397–1410  
(陈杰, 方浩, 曾宪琳. 面向高危行业的无人平台智能化发展. 中国科学: 信息科学, 2021, **51**(9): 1397–1410)
- Civil Aviation Administration of China, National Development and Reform Commission, Ministry of Transport. Civil aviation development during the 14th Five Year Plan period exhibition planning [Online], available: <https://www.gov.cn/zhengce/zhengceku/2022-01/07/5667003/files/d12ea75169374a15a742116f7082df85.pdf>, December 31, 2024  
(中国民用航空局, 国家发展改革委员会, 交通运输部. “十四五”民用航空发展规划 [Online], available: <https://www.gov.cn/zhengce/zhengceku/2022-01/07/5667003/files/d12ea75169374a15a742116f7082df85.pdf>, 2024-12-31))
- Zhao Chun-Hui, Zhou Yi-Hui, Lin Zhao, Hu Jin-Wen, Pan Quan. Review of scene matching visual navigation for unmanned aerial vehicles. *Scientia Sinica Informationis*, 2019, **49**(5): 507–519  
(赵春晖, 周映慧, 林钊, 胡劲文, 潘泉. 无人机景象匹配视觉导航技术综述. 中国科学: 信息科学, 2019, **49**(5): 507–519)
- Huang G Q. Visual-inertial navigation: A concise review. In: *Proceedings of the International Conference on Robotics and Automation (ICRA)*. Montreal, Canada: IEEE, 2019. 9572–9582
- Shen L, Stopher P R. Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 2014, **34**(3): 316–334
- Couturier A, Akhloufi M A. A review on absolute visual localization for UAV. *Robotics and Autonomous Systems*, 2021, **135**: Article No. 103666
- Ma J Y, Jiang X Y, Fan A X, Jiang J J, Yan J C. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision*, 2021, **129**(1): 23–79
- Zitová B, Flusser J. Image registration methods: A survey. *Image and Vision Computing*, 2003, **21**(11): 977–1000
- Le Moigne J, Campbell W J, Crompton R F. An automated parallel image registration technique based on the correlation of wavelet features. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, **40**(8): 1849–1864
- Pan W H, Wei S D, Lai S H. Efficient NCC-based image matching in Walsh-Hadamard domain. In: *Proceedings of the 10th European Conference on Computer Vision*. Marseille, France: Springer, 2008. 468–480
- Liu H Z, Guo B L, Feng Z Z. Pseudo-log-polar Fourier transform for image registration. *IEEE Signal Processing Letters*, 2006, **13**(1): 17–20
- Cain S C, Hayat M M, Armstrong E E. Projection-based image registration in the presence of fixed-pattern noise. *IEEE Transactions on Image Processing*, 2001, **10**(12): 1860–1872
- van Dalen G J, Magree D P, Johnson E N. Absolute localization using image alignment and particle filtering. In: *Proceedings of the AIAA Guidance, Navigation, and Control Conference*. San Diego, USA: AIAA, 2016. Article No. 647
- Klein S, Staring M, Pluim J P W. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Transactions on Image Processing*, 2007, **16**(12): 2879–2890
- Loeckx D, Slagmolen P, Maes F, Vandermeulen D, Suetens P. Nonrigid image registration using conditional mutual information. *IEEE Transactions on Medical Imaging*, 2010, **29**(1): 19–29
- Viola P A, Wells W M. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 1997, **24**(2): 137–154
- Yol A, Delabarre B, Dame A, Dartois J E, Marchand E. Vision-based absolute localization for unmanned aerial vehicles. In:

- Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Chicago, USA: IEEE, 2014. 3429–3434
- 19 Fan B J, Du Y K, Zhu L L, Tang Y D. The registration of UAV down-looking aerial images to satellite images with image entropy and edges. In: Proceedings of the 3rd International Conference on Intelligent Robotics and Applications. Shanghai, China: Springer, 2010. 609–617
  - 20 Levin E, Kupiec S A, Forrester T, DeBacker T A, Jamnson T P. GIS-based UAV real-time path planning and navigation. In: Proceedings of the SPIE 4708, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement. Orlando, USA: SPIE, 2002. 296–303
  - 21 Lowe D G. Object recognition from local scale-invariant features. In: Proceedings of the 17th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 1150–1157
  - 22 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
  - 23 Shan M, Wang F, Lin F, Gao Z, Tang Y Z, Chen B M. Google map aided visual navigation for UAVs in GPS-denied environment. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO). Zhuhai, China: IEEE, 2015. 114–119
  - 24 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05). San Diego, USA: IEEE, 2005. 886–893
  - 25 Sun D Q, Roth S, Black M J. Secrets of optical flow estimation and their principles. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 2432–2439
  - 26 Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. In: Proceedings of the International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 2564–2571
  - 27 Masselli A, Hanten R, Zell A. Localization of unmanned aerial vehicles using terrain classification from aerial images. In: Proceedings of the 13th International Conference IAS-13. Cham, Germany: Springer, 2016. 831–842
  - 28 Alcantarilla P F, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: Proceedings of the British Machine Vision Conference. Bristol, UK: BMVA, 2013. 1–11
  - 29 Leutenegger S, Chli M, Siegwart R Y. BRISK: Binary robust invariant scalable keypoints. In: Proceedings of the International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 2548–2555
  - 30 Chiu H P, Das A, Miller P, Samarasekera S, Kumar R. Precise vision-aided aerial navigation. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Chicago, USA: IEEE, 2014. 688–695
  - 31 Mantelli M, Pittol D, Neuland R, Ribacki A, Maffei R, Jorge V, et al. A novel measurement model based on abBRIEF for global localization of a UAV over satellite images. *Robotics and Autonomous Systems*, 2019, **112**: 304–319
  - 32 Wei Q L, Yang Z S, Su H Z, Wang L J. Monte Carlo-based reinforcement learning control for unmanned aerial vehicle systems. *Neurocomputing*, 2022, **507**: 282–291
  - 33 Shan M, Charan A. Google map referenced UAV navigation via simultaneous feature detection and description. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO). 2015. 114–119
  - 34 Woo J, Son K, Li T, Kim G, Kweon I S. Vision-based UAV navigation in mountain area. In: Proceedings of the MVA. Tokyo, Japan: MVA, 2007. 236–239
  - 35 Han X F, Leung T, Jia Y Q, Sukthankar R, Berg A C. MatchNet: Unifying feature and metric learning for patch-based matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 3279–3286
  - 36 Chang J, Lan Z H, Cheng C M, Wei Y C. Data uncertainty learning in face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 5709–5718
  - 37 Tian Y X, Deng X Q, Zhu Y, Newsam S. Cross-time and orientation-invariant overhead image geolocalization using deep local features. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA: IEEE, 2020. 2501–2509
  - 38 Nassar A, Amer K, ElHakim R, ElHelw M. A deep CNN-based framework for enhanced aerial imagery registration with applications to UAV geolocalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018. 1594–1604
  - 39 Noh H, Araujo A, Sim J, Weyand T, Han B. Large-scale image retrieval with attentive deep local features. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3476–3485
  - 40 Teichmann M, Araujo A, Zhu M L, Sim J. Detect-to-retrieve: Efficient regional aggregation for image search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5104–5113
  - 41 Hinzmann T, Siegwart R. Deep UAV localization with reference view rendering. arXiv preprint arXiv: 2008.04619, 2020.
  - 42 Goforth H, Lucey S. GPS-denied UAV localization using pre-existing satellite imagery. In: Proceedings of the International Conference on Robotics and Automation (ICRA). Montreal, Canada: IEEE, 2019. 2974–2980
  - 43 Kinnari J, Verdoja F, Kyrki V. Season-invariant GNSS-denied visual localization for UAVs. *IEEE Robotics and Automation Letters*, 2022, **7**(4): 10232–10239
  - 44 Amer K, Samy M, ElHakim R, Shaker M, ElHelw M. Convolutional neural network-based deep urban signatures with application to drone localization. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW). Venice, Italy: IEEE, 2017. 2138–2145
  - 45 Google map [Online], available: <https://cloud.google.com/map-platform/>, December 31, 2020
  - 46 Bing map [Online], available: <https://www.bing.com/map>, December 31, 2020
  - 47 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015.
  - 48 Nassar A, Lefèvre S, Wegner J. Simultaneous multi-view instance detection with learned geometric soft-constraints. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: 2019. 6558–6567
  - 49 Nassar A, ElHelw M. Aerial imagery registration using deep learning for UAV geolocalization. *Deep Learning in Computer Vision: Principles and Applications*. Boca Raton: CRC Press, 2020. 183–210
  - 50 Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, **24**(6): 381–395
  - 51 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015. 234–241
  - 52 Schleiss M. Translating aerial images into street-map-like representations for visual self-localization of UAVs. In: Proceedings of the XLII-2/W13, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Austria: 2019. 575–580
  - 53 Isola P, Zhu J Y, Zhou T H, Efros A A. Image-to-image translation with conditional adversarial networks. In: Proceedings of



- the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 5967–5976
- 54 OpenStreetMap [Online], available: <https://www.openstreetmap.org/>, December 31, 2020
- 55 Lin Y P, Medioni G. Map-enhanced UAV image sequence registration and synchronization of multiple image sequences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1–7
- 56 Huang S M, Huang C C, Chou C C. Image registration among UAV image sequence and Google satellite image under quality mismatch. In: Proceedings of the 12th International Conference on ITS Telecommunications. Taipei, China: IEEE, 2012. 311–315
- 57 Wan X, Liu J G, Yan H S, Morgan G L K. Illumination-invariant image matching for autonomous UAV localisation based on optical sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, **119**: 198–213
- 58 Patel B. Visual Localization for UAVs in Outdoor GPS-denied Environments [Master thesis], University of Toronto, Canada, 2019.
- 59 Pluckter K, Scherer S. Precision UAV landing in unstructured environments. In: Proceedings of the International Symposium on Experimental Robotics. Cham, Germany: Springer, 2020. 177–187
- 60 Marcu A, Costea D, Slusanschi E, Leordeanu M. A multi-stage multi-task neural network for aerial scene interpretation and geolocalization. arXiv preprint arXiv: 1804.01322, 2018.
- 61 Pan A N, Yang Y. Remote sensing images registration with different viewpoints. In: Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP). Shanghai, China: IEEE, 2016. 699–704
- 62 Weyand T, Kostrikov I, Philbin J. PlaNet-photo geolocation with convolutional neural networks. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 37–55
- 63 Wu S B, Du C, Chen H, Jing N. Coarse-to-fine UAV image geolocalization using multi-stage Lucas-Kanade networks. In: Proceedings of the 2nd Information Communication Technologies Conference (ICTC). Nanjing, China: IEEE, 2021. 220–224
- 64 Zheng Z D, Wei Y C, Yang Y. University-1652: A multi-view multi-source benchmark for drone-based geo-localization. In: Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM, 2020. 1395–1403
- 65 Workman S, Souvenir R, Jacobs N. Wide-area image geolocalization with aerial reference imagery. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3961–3969
- 66 Hays J, Efros A A. IM2GPS: Estimating geographic information from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 67 Yang H L, Lu X F, Zhu Y Y. Cross-view geo-localization with evolving transformer. arXiv preprint arXiv: 2107.00842, 2021.
- 68 Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. Virtual Event: ICLR, 2021.
- 69 Shi Y J, Liu L, Yu X, Li H D. Spatial-aware feature aggregation for cross-view image based geo-localization. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: ACM, 2019. Article No. 905
- 70 Liu L, Li H D. Lending orientation to neural networks for cross-view geo-localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5617–5626
- 71 Hu S Y, Chang X J. Multi-view drone-based geo-localization via style and spatial alignment. arXiv preprint arXiv: 2006.13681, 2020.
- 72 Jin Y, Li C N, Li Y D, Peng P X, Giannopoulos G A. Model latent views with multi-center metric learning for vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems*, 2021, **22**(3): 1919–1931
- 73 Sun B, Liu G C, Yuan Y. F3-Net: Multiview scene matching for drone-based geo-localization. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: Article No. 5610611
- 74 Ding L R, Zhou J, Meng L X, Long Z Y. A practical cross-view image matching method between UAV and satellite for UAV-based geo-localization. *Remote Sensing*, 2021, **13**(1): Article No. 47
- 75 Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017. 214–223
- 76 Chan E R, Monteiro M, Kellnhofer P, Wu J J, Wetzstein G. piGAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 5795–5805
- 77 Shi Y J, Yu X, Liu L, Zhang T, Li H D. Optimal feature transport for cross-view image geo-localization. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 11990–11997
- 78 Wang T Y, Zheng Z D, Yan C G, Zhang J Y, Sun Y Q, Zheng B L, et al. Each part matters: Local patterns facilitate cross-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(2): 867–879
- 79 Si Shu-Bin, Zhao Da-Wei, Xu Wan-Ying, Zhang Yong-Gang, Dai Bin. Review on visual-inertial navigation and positioning technology. *Journal of Image and Graphics*, 2021, **26**(6): 1470–1482  
(司书斌, 赵大伟, 徐婉莹, 张勇刚, 戴斌. 视觉-惯性导航定位技术研究进展. *中国图象图形学报*, 2021, **26**(6): 1470–1482)
- 80 Castañeda N, Lamy-Perbal S. An improved shoe-mounted inertial navigation system. In: Proceedings of the International Conference on Indoor Positioning and Indoor Navigation. Zurich, Switzerland: IEEE, 2010. 1–6
- 81 Couturier A, Akhloufi M A. Relative visual localization (RVL) for UAV navigation. In: Proceedings of the SPIE 10642, Degraded Environments: Sensing, Processing, and Display. Orlando, USA: SPIE, 2018. Article No. 1064200
- 82 Couturier A, Akhloufi M. Conditional probabilistic relative visual localization for unmanned aerial vehicles. In: Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). London, Canada: IEEE, 2020. 1–4
- 83 Carlone L, Karaman S. Attention and anticipation in fast visual-inertial navigation. *IEEE Transactions on Robotics*, 2019, **35**(1): 1–20
- 84 Tan C W, Park S. Design of accelerometer-based inertial navigation systems. *IEEE Transactions on Instrumentation and Measurement*, 2005, **54**(6): 2520–2530
- 85 Zhao C H, Wang R Z, Zhang T W, Pan Q. Visual odometry and scene matching integrated navigation system in UAV. In: Proceedings of the 17th International Conference on Information Fusion (FUSION). Salamanca, Spain: IEEE, 2014. 1–6
- 86 Li Y J, Pan Q, Zhao C H, Yang F. Scene matching based EKF-SLAM visual navigation. In: Proceedings of the 31st Chinese Control Conference. Hefei, China: IEEE, 2012. 5094–5099
- 87 Liu P D, Heng L, Sattler T, Geiger A, Pollefeys M. Direct visual odometry for a fisheye-stereo camera. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, Canada: IEEE, 2017. 1746–1752
- 88 Babu B W, Kim S, Yan Z X, Ren L.  $\sigma$ -DVO: Sensor noise model meets dense visual odometry. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Merida, Mexico: IEEE, 2016. 18–26
- 89 Buczko M, Willert V. How to distinguish inliers from outliers in visual odometry for high-speed automotive applications. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV). Gothenburg, Sweden: IEEE, 2016. 478–483

- 90 Steinbrücker F, Sturm J, Cremers D. Real-time visual odometry from dense RGB-D images. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain: IEEE, 2011. 719–722
- 91 Koide K, Yokozuka M, Oishi S, Banno A. Globally consistent 3D LiDAR mapping with GPU-accelerated GICP matching cost factors. *IEEE Robotics and Automation Letters*, 2021, **6**(4): 8591–8598
- 92 Hesch J A, Kottas D G, Bowman S L, Roumeliotis S I. Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics*, 2014, **30**(1): 158–176
- 93 Zhu Qi-Ju, Fan Zhen-Hui, Mei Chun-Bo, Yang Peng-Xiang, Lu Bao-Feng, Yang Chao-Ming. Heading correction and combination algorithm for MEMS with low accuracy based on scene matching. *Journal of Chinese Inertial Technology*, 2023, **31**(9): 870–875  
(朱启举, 樊振辉, 梅春波, 杨鹏翔, 卢宝峰, 杨朝明. 基于景象匹配的低精度 MEMS 航向修正与组合算法. 中国惯性技术学报, 2023, **31**(9): 870–875)
- 94 Yin X C, Wang X W, Du X G, Chen Q J. Scale recovery for monocular visual odometry using depth estimated with deep convolutional neural fields. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 5871–5879
- 95 Ye W C, Lan X Y, Chen S, Ming Y H, Yu X Y, Bao H J, et al. PVO: Panoptic visual odometry. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023. 9579–9589
- 96 Li C, Liu G C, Yuan Y. A multi-source image matching network for UAV visual location. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). Bordeaux, France: IEEE, 2022. 1651–1655
- 97 Kinnari J, Renzulli R, Verdoja F, Kyrki V. LSVL: Large-scale season-invariant visual localization for UAVs. *Robotics and Autonomous Systems*, 2023, **168**: Article No. 104497
- 98 Wen K L, Chu J, Chen J Y, Chen Y, Cai J P. M-O SiamRPN with weight adaptive joint MIoU for UAV visual localization. *Remote Sensing*, 2022, **14**(18): Article No. 4467
- 99 Dai M, Hu J H, Zhuang J D, Zheng E H. A transformer-based feature segmentation and region alignment method for UAV-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(7): 4376–4389
- 100 Tian X Y, Shao J, Ouyang D Q, Shen H T. UAV-satellite view synthesis for cross-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(7): 4804–4815
- 101 Zhu R Z, Yin L, Yang M Z, Wu F, Yang Y C, Hu W B. SUES-200: A multi-height multi-scene cross-view image benchmark across drone and satellite. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, **33**(9): 4825–4839
- 102 Lang C B, Cheng G, Tu B F, Han J W. Global rectification and decoupled registration for few-shot segmentation in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: Article No. 5617211
- 103 Zhang B, Wu Y F, Zhao B Y, Chanussot J, Hong D F, Yao J, et al. Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, **15**: 1814–1822
- 104 Zhang Z Y, Zheng L L, Piao Y, Tao S P, Xu W, Gao T, et al. Blind remote sensing image deblurring using local binary pattern prior. *Remote Sensing*, 2022, **14**(5): Article No. 1276
- 105 Guo J T, Zhang Z R, Mao Y C, Liu S J, Zhu W C, Yang T H. Automatic extraction of discontinuity traces from 3D rock mass point clouds considering the influence of light shadows and color change. *Remote Sensing*, 2022, **14**(21): Article No. 5314
- 106 Yao Y X, Zhang Y J, Wan Y, Liu X Y, Yan X H, Li J Y. Multi-modal remote sensing image matching considering co-occurrence filter. *IEEE Transactions on Image Processing*, 2022, **31**: 2584–2597
- 107 Deng J, Dong W, Socher R, Li L J, Li K, Li F F. ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 248–255
- 108 Wang C H, Huang K Y, Yao Y, Chen J C, Shuai H H, Cheng W H. Lightweight deep learning: An overview. *IEEE Consumer Electronics Magazine*, 2024, **13**(4): 51–64
- 109 Jiang C, Zhao D B, Zhang Q Z, Liu W K. A multi-GNSS/IMU data fusion algorithm based on the mixed norms for land vehicle applications. *Remote Sensing*, 2023, **15**(9): Article No. 2439
- 110 Bappy A M R A, Asfak-Ur-Rafi M, Islam M S, Sajjad A, Imran K N. *Design and Development of Unmanned Aerial Vehicle (Drone) for Civil Applications*. Dhaka: BRAC University Press, 2015.
- 111 Alberts I L, Mercolli L, Pyka T, Prenosil G, Shi K Y, Rominger A, et al. Large language models (LLM) and ChatGPT: What will the impact on nuclear medicine be? *European Journal of Nuclear Medicine and Molecular Imaging*, 2023, **50**(6): 1549–1552



袁媛 西北工业大学光电与人工智能研究院教授。主要研究方向为跨域遥感和群体智能决策。

E-mail: [y.yuan@nwpu.edu.cn](mailto:y.yuan@nwpu.edu.cn)

(YUAN Yuan Professor at the School of Artificial Intelligence, Optics and Electronics, Northwestern

Polytechnical University. Her research interest covers cross-domain remote sensing and swarm intelligent decision-making.)



孙柏 西北工业大学计算机学院博士研究生。主要研究方向为度量学习和视觉定位。

E-mail: [sunbo\\_bosun@mail.nwpu.edu.cn](mailto:sunbo_bosun@mail.nwpu.edu.cn)

(SUN Bo Ph.D. candidate at the School of Computer Science, Northwestern Polytechnical University.

His research interest covers metric learning and visual geo-localization.)



刘赶超 西北工业大学光电与人工智能研究院副教授。主要研究方向为跨域遥感和视觉定位。本文通信作者。

E-mail: [liuganchao@nwpu.edu.cn](mailto:liuganchao@nwpu.edu.cn)

(LIU Gan-Chao Associate professor at the School of Artificial Intelligence, Optics and Electronics, North-

western Polytechnical University. His research interest covers cross-domain remote sensing and visual geo-localization. Corresponding author of this paper.)