

文章编号 1004-924X(2026)01-0124-15

基于图像匹配的高空大斜视无源目标定位

贾 平*, 李昌灏, 孙 辉, 宋悦铭, 冯卓莹, 徐 芳

(中国科学院 长春光学精密机械与物理研究所

动态光学成像与测量全国重点实验室, 吉林 长春 130033)

摘要:提出一种基于图像匹配的无源定位方法,通过引入基于 Transformer 的特征增强与 MiHo 聚类筛选的两步匹配策略,减轻了高空大斜视条件下传统无源定位算法因微小角度误差导致的定位精度下降程度。根据粗定位结果与飞行参数对航拍图像进行近似正射变换,并截取对应区域的卫星图像。采用 RepVGG 提取图像粗特征,通过互最近邻实现初步匹配,并结合 MiHo 与归一化互相关(Normalized Cross Correlation, NCC)筛选匹配点对。最后,借助 Transformer 模块完成精细化匹配,再根据精匹配结果构建角度误差修正矩阵,多次迭代修正系统误差。实验结果表明,所提方法的定位精度较传统方法有较大幅度提升,在典型应用场景下提升约 70%,在斜距 90 km 的情况下,定位精度可维持在 120 m 左右。该方法突破了传统无源定位对角度精度的高度依赖,验证了基于图像匹配的无源定位路径的可行性与有效性。

关键词:图像匹配;目标定位;航空光电系统;无源定位;大斜视;Transformer

中图分类号:TP75;TP391.41 **文献标识码:**A

doi:10.37188/OPE.20263401.0124 **CSTR:**32169.14.OPE.20263401.0124

Passive target localization under high-altitude large oblique views based on image matching

JIA Ping*, LI Changhao, SUN Hui, SONG Yueming, MA Zhuoluo, XU Fang

(State Key Laboratory of Dynamic Optics and Imaging and Measurement, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China)

* Corresponding author, E-mail: jiap@ciomp.ac.cn

Abstract: The workflow begins with approximate orthorectification of aerial images using coarse positioning results and flight parameters, followed by extraction of the corresponding satellite image regions. RepVGG is employed to extract coarse image features, and initial correspondences are established via nearest-neighbor matching. Candidate pairs are then filtered using MiHo clustering and normalized cross-correlation (NCC). The refined correspondences are further enhanced by a Transformer module to achieve precise alignment. Based on the resulting high-precision matches, an angular error correction matrix is constructed, and iterative refinement is applied to effectively compensate for systematic errors. Experimental results indicate that the proposed method substantially improves positioning accuracy relative to conventional approaches, achieving an approximately 70% enhancement in representative scenarios; even at a slant range of 90 km, the positioning error remains approximately 120 m. By leveraging image match-

收稿日期:2025-09-12;修订日期:2025-10-12.

基金项目:国家自然科学基金资助项目(No. 62405310)

ing under challenging viewing geometries, the proposed passive localization method provides a novel solution to limitations of conventional algorithms. A two-step matching strategy integrating Transformer-based feature enhancement with MiHo clustering-based filtering is introduced, effectively mitigating the accuracy degradation of traditional passive localization under high-altitude, large-oblique viewing conditions in the presence of small angular errors.

Key words: image matching; target location; airborne photoelectric system; passive localization; large strabismus; Transformer.

1 引言

在遥感侦察、精确打击及地理空间测绘等领域,快速、准确地获取战场全地形信息是实现高效指挥与决策支持的基础保障。其中,高空大斜视成像因具备广域覆盖与视差信息丰富等优势,已成为获取复杂地形与动态目标信息的重要手段。相较于传统垂直观测方式,斜视成像不仅提升了探测的实时性与灵活性,还显著增强了三维定位精度,适应性更强,特别适用于复杂作战环境下的目标感知任务。

航空光电系统作为现代空中侦察的核心装备,依托搭载在有人/无人机、飞艇等平台上的EO/IR多光谱相机和激光雷达(LiDAR)等光电载荷,能够克服地形遮挡等限制,实时获取大范围区域的图像与地理信息。相比地基或车/舰载光电系统,航空光电系统在目标识别、测量与定位方面具备更强的空间机动性与数据获取能力,已广泛应用于远程侦察、目标跟踪及情报监视等任务中。随着光学系统与伺服控制技术的不断进步,对航空成像测量系统的要求也不断提升。战场环境瞬息万变,传统基于俯仰角的粗略测量已难以满足实际作战需求,指挥系统迫切需要获取目标的精确地理坐标(包括经纬度与高程),以实现战场态势的精细化感知与地图标注支持。获取目标精确三维坐标的过程即为目标定位,其精度直接关系到后续打击引导与任务规划的有效性。

高空远距离航空相机因其优异的成像分辨率与遥感能力,已在多种作战场景中得到验证。研究适用于该类传感器的大斜视高精度目标定位方法,不仅是实现高可靠战场信息获取的关键环节,也对提升整体作战体系的智能化与自动化

水平具有重要意义。

传统目标定位方法主要依赖于机载角度传感器进行几何反演。然而,在高空大斜视成像条件下,即便是微小的角度误差也会被显著放大,进而导致定位精度的剧烈下降。为此,研究人员围绕误差补偿与系统修正展开了深入探索^[1-6],包括引入数字高程模型(Digital Elevation Model, DEM)^[7]以补偿地形影响,以及采用多无人机协同交叉测量方式提升角度测量精度^[8-9]。尽管相关策略在一定程度上提升了定位准确性,但受限于角度误差在大斜视条件下的高敏感性,现有方法依旧难以从根本上解决定位偏差的问题。

近年来,随着图像匹配技术的持续发展,图像信息辅助定位成为一种颇具前景的研究方向。尤其是在跨视角匹配、不依赖GPS/INS系统的无人机自主定位等方面,相关研究不断取得进展^[10-21]。Yao等提出了一套基于跨视图匹配的无人机地理定位方法,同时发布了面向UAV-satellite配对的基准数据集。该工作把检索、局部匹配与PnP求解视为统一流程,强调跨尺度/跨视角的一致性建模,有利于提高从无人机视角到卫星/航片的配准稳定性。该数据集和方法为后续研究提供了实用的测试平台与基线^[18]。Fan等强调了跨模态/跨源(如无人机图像与高分辨率参考影像)在绝对定位任务中的挑战,提出基于金字塔Transformer或单流金字塔网络(SSPT)等结构以增强跨域语义对齐能力,从而改善检索与粗匹配阶段的稳定性,适应弱纹理与时序变化的情形^[20]。Ren等提出基于注意力机制的图像配准与无人机视觉定位方法,通过引入注意力加权的特征对齐模块来缓解视角、光照与细节差异对匹配的干扰,提高了配准的收敛性与定位精度,在

若干跨视角数据集上验证了有效性^[21]。然而,在高空大斜视场景下,图像间尺度差异大、视角变化剧烈和地物遮挡频繁等问题仍严重制约着匹配精度与稳定性。现有图像匹配定位方法多聚焦于特定子模块(如特征提取、匹配点筛选等)的优化,尚未形成一个完整闭环的、面向实际应用的图像匹配定位流程。因此,本文在兼顾精度、稳定性与实用性的前提下,构建一套适用于高空大斜视成像条件的图像匹配辅助定位机制,依然是当前研究亟待突破的关键问题。

2 定位算法流程

本文提出了一种基于图像匹配的无人机目标精确定位方法,其总体流程如图1所示,分为6个阶段。

在粗定位阶段,根据无人机的飞行数据,利用数字高程模型(Digital Elevation Model, DEM)或地球椭球模型进行初步地理坐标反算,从而获得粗定位结果。该步骤通过建立无人机姿态、航向角、相机内外参数与地理坐标之间的几何关系,得到无人机拍摄影像对应的地面参考区域。

在正投影影像生成阶段,依据粗定位结果确定目标区域范围,从卫星影像数据库中获取对应地区的高分辨率卫星图像,并计算光学载荷图像的正投矩阵。通过该矩阵,可将无人机视角下的倾斜影像转换为近似垂直视角的正投影影像,以减少视角差异对后续匹配精度的影响。

在图像粗匹配阶段,提出一种多阶段、粗到精的匹配策略,以提升匹配精度与稳定性。首先,对卫星图像与正投影影像分别进行1/8下采样以降低计算复杂度。然后,采用轻量级卷积网络可重参数卷积神经网络(Re-parameterizable VGG Network, RepVGG)进行特征提取,获得两幅图像的高层语义特征表示。最后,利用互为最近邻匹配算法(Mutual Nearest Neighbor, MNN)匹配策略实现粗匹配,得到初步对应点对集合。

在匹配点精炼阶段,针对粗匹配结果,进一步采用中间单应性(Middle Homography, MiHo)+归一化互相关(Normalized Cross Correlation, NCC)复合筛选机制进行精炼。MiHo聚类

算法能够根据几何一致性对匹配点进行聚类,而NCC相似度度量则用于评估局部图像块的相关性,从而有效剔除错误匹配点。筛选后,对保留的高置信度匹配点周围区域进行裁剪,提取局部图像块以用于精细匹配。

在精匹配阶段,将裁剪后的图像块输入基于Transformer架构的精匹配模块,通过注意力机制实现局部到全局的特征关联,从而获得高精度的点对匹配结果,由此得到的匹配点坐标关系即为精定位结果。

在误差修正与系统补偿阶段,将精定位结果与初始的粗定位结果进行误差对比,计算系统性偏差所引起的角度误差修正矩阵。该修正矩阵可用于对后续航次的粗定位结果进行校正,实现基于历史匹配误差的自适应系统补偿,从而持续提升定位精度与稳定性。

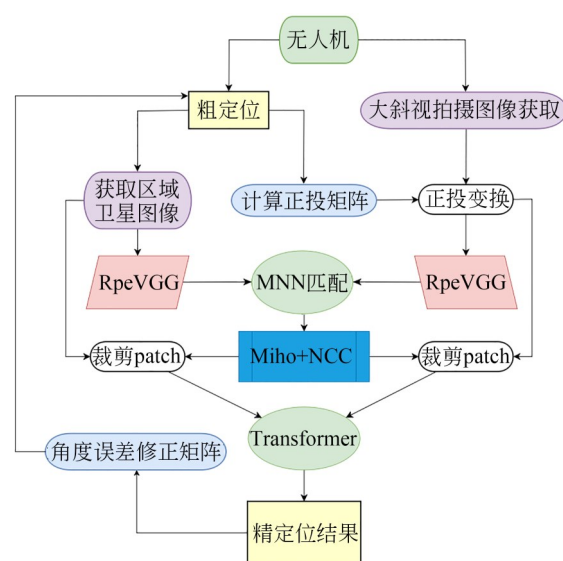


图1 基于图像匹配的无源定位算法流程

Fig. 1 Flow chart of passive positioning algorithm based on image matching

2.1 基于地球椭圆模型的无源定位

目标点 Q 在探测器上的投影点 P 在相机坐标系下的坐标为:

$$P_o = (m, n, -f)^T, \quad (1)$$

其中 f 为成像系统的焦距。

对应的大地坐标系坐标为:

$$P_H = R \cdot P_o + T, \quad (2)$$

其中: R 是飞行参数解算出来的旋转矩阵, T 为无

人机的坐标,大地坐标系指的是 World Geodetic System-1984 Coordinate System (WGS84) 坐标系。

在 WGS84 坐标系中, X 轴指向格林威治子午线与赤道的交点, Y 轴与 X 轴垂直,指向东经 90 度的方向, Z 轴则与地球自转轴重合,指向地理北极。

对于理想光学系统,物点、像点、投影中心三点共线,即目标点在目标投影点与相机坐标系的原点连接的直线上。因此,目标点在大地坐标系下的坐标 Q_H 应该满足:

$$\frac{X_{Q_H} - X_{P_H}}{X_{P_H} - X_T} = \frac{Y_{Q_H} - Y_{P_H}}{Y_{P_H} - Y_T} = \frac{Z_{Q_H} - Z_{P_H}}{Z_{P_H} - Z_T}. \quad (3)$$

又因为目标点位于地球椭圆模型上,所以若目标的大地高为 h ,则目标在大地坐标系下的坐标 Q_H 应满足:

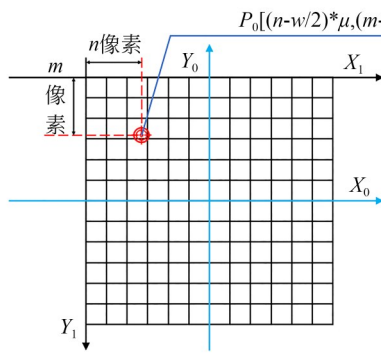
$$\frac{(X_{Q_H})^2}{(R_E)_T^2} + \frac{(Y_{Q_H})^2}{(R_E)_T^2} + \frac{(Z_{Q_H})^2}{(R_P)_T^2} = 1. \quad (4)$$

联立式(1)~式(4)即可求得目标点在大地坐标系下的坐标 Q_H 。

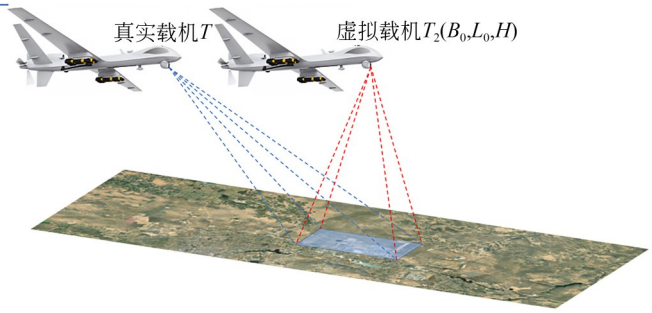
2.2 正投矩阵计算

由图 2(a)可知,图像上一点 $P(n, m)$ 在相机坐标系上的坐标为:

$$P_0[(n - w/2)*\mu, (m - h/2)*\mu, -f], \quad (5)$$



(a) 图像坐标系转成像坐标系
(a) Image to imaging coordinates



(b) 虚拟正投成像载机示意图
(b) Virtual orthographic imaging carrier machine

图 2 正投矩阵计算示意图

Fig. 2 Schematic diagram of front projection matrix calculation

取图像中的 4 个角点,通过式(1)~式(8)求得对应的新图像坐标。通过这 4 对点,即可求得正投矩阵。

其中: w 是图像的像素宽度, h 是图像的像素高度, μ 是像元尺寸, f 是相机焦距。再将其带入式(1)~式(4)中,即可求得图像上点 P 在大地坐标系的坐标 $G_H(x_{G_H}, y_{G_H}, z_{G_H})$ 。

如图 2(b)所示,利用粗定位 (B_0, L_0, H_0) 的结果,可以在目标点上空创建一个虚拟载机 $T_2(B_0, L_0, H)$, H 是原先载机的飞行高度,将其转为大地坐标 $X_{T_2}(x_2, y_2, z_2)$ 。

根据式(2)可得,大地坐标 $G_H(x_{G_H}, y_{G_H}, z_{G_H})$ 对应新载机的导航坐标为:

$$G_F(x_{G_F}, y_{G_F}, z_{G_F}) = R_{F-H}^{-1}(G_H - X_{T_2}). \quad (6)$$

同理,可得大地坐标 $G_H(x_{G_H}, y_{G_H}, z_{G_H})$ 对应的新载机的成像坐标为:

$$G_O(x_{G_O}, y_{G_O}, z_{G_O}) = R^{-1}(G_H - X_{T_2}). \quad (7)$$

由小孔成像原理可知,成像坐标系中的某一点对应的图像坐标 $P(\mu, \vartheta)$ 满足:

$$z_{G_O} * P(\mu, \vartheta, 1) = \begin{bmatrix} \frac{f}{\mu} & 0 & \frac{w}{2} \\ 0 & \frac{f}{\mu} & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} G_O = M * G_O, \quad (8)$$

其中 M 是相机的内参矩阵。

2.3 基于 Transformer 的图像匹配

如图 3 所示,本方法首先采用 RepVG 网络提取输入图像的下采样粗特征,以获取全局结构信

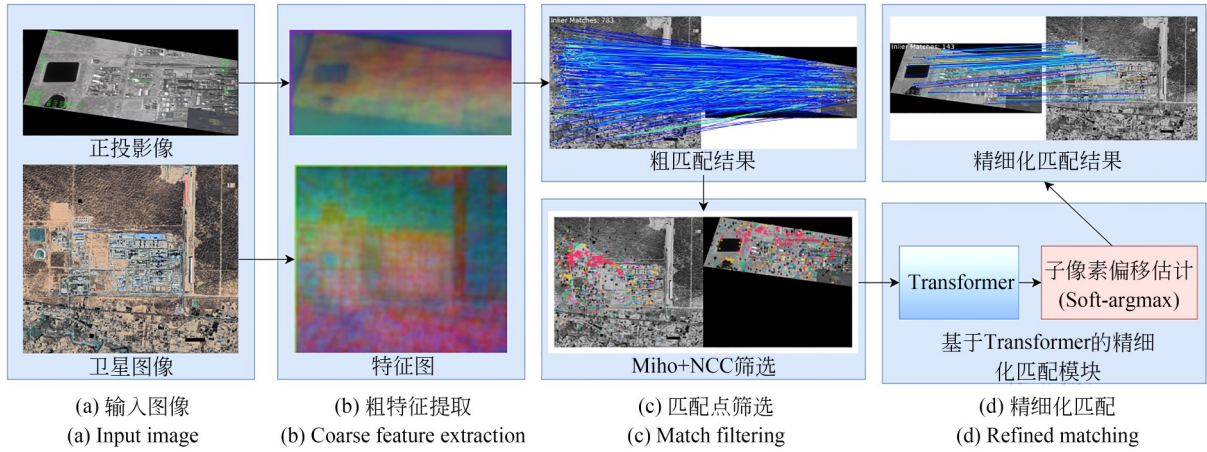


图 3 Transformer-based RP+Miho+NCC 框架

Fig. 3 Transformer-based RP+Miho+NCC framework

息。随后,利用 AffNet 为候选匹配点生成局部仿射帧,从而实现对局部几何变化的建模。接着,采用 MiHo 聚类与 NCC 相似性度量联合筛选出高置信度匹配点对,提升初始匹配的稳定性。在此基础上,再次使用 RepVGG 提取筛选后匹配点周围的局部图像 Patch 特征,以获得更具判别力的特征描述。最后,通过 Transformer 模块对精细特征进行融合匹配,进一步提升匹配精度和稳定性。该方法结合了粗特征的全局表达能力与细特征的局部几何建模优势,具有良好的匹配稳定性和精度表现。

2.3.1 特征粗提取

RepVGG 是一种基于结构重参数化思想的高效卷积神经网络,由 Ding 等于 2021 年提出^[22]。其核心思想是在训练阶段采用多分支结构以提升特征表达能力,在推理阶段则将多分支结构等价转换为单分支卷积形式,从而在保持模型性能的同时显著降低计算量,其训练结构如图 4 所示。

在训练阶段,每个 RepVGG Block 包含以下 3 条并行路径:

$$y = W_{3 \times 3} * x + W_{1 \times 1} * x + I * x + b, \quad (9)$$

其中: $W_{3 \times 3}$ 为标准 3×3 卷积核, $W_{1 \times 1}$ 为 1×1 卷积分支, I 表示恒等映射, b 为偏执项。

训练结束后,3 个分支会通过参数重组(参数折叠)被融合为一个等效的 3×3 卷积:

$$W_{\text{fused}} = W_{3 \times 3} + \text{Pad}(W_{1 \times 1}) + \text{Pad}(I). \quad (10)$$

从而在推理阶段仅需执行一次单分支卷积

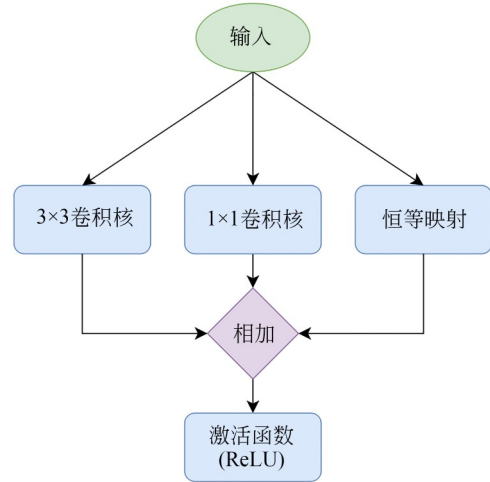


图 4 RepVGG 训练时单个 block 结构

Fig. 4 Single block structure during RepVGG training

运算,大幅减少卷积计算量和显存占用。

每个 RepVGG Block 均由上述多分支卷积模块组成,且在每个卷积层后均采用 ReLU 激活函数,以提高模型的非线性建模能力并缓解梯度消失问题。该设计同时保留了传统 VGG 网络的简洁性,又兼顾了 ResNet 类网络的性能优势。整个网络由多个 Stage 堆叠构成:

Stage1(浅层):侧重于捕获图像的低层纹理与边缘特征;

Stage2(中层):逐步扩展感受野并提取中层语义特征;

Stage3(深层):聚焦于高层语义与全局上下文信息。

在本文的特征提取流程中,输入图像首先经

过1/4下采样后依次通过Stage1至Stage3模块,实现多尺度语义特征建模。随后,通过全局平均池化(Global Average Pooling, GAP)对空间维度进行压缩,最终得到维度为 d 的全局图像特征向量。本文采用的RepVGG网络配置为A0型(RepVGG-A0),其整体通道数配置较轻量,最高输出通道数为1280,在保证特征表达能力的同时具备较高的计算效率。

损失函数采用focal loss函数:

$$\text{FL}(p_i) = -\alpha_i(1-p_i)^\gamma \log(p_i), \quad (11)$$

其中: p_i 是模型对正确类别的预测概率, α_i 是平衡因子, γ 是调节因子。

2.3.2 特征粗匹配

设两幅图像的特征粗匹配特征分别为 $F_0 \in \mathbb{R}^{N \times L \times C}$ 和 $F_1 \in \mathbb{R}^{N \times S \times C}$,其中 N 为批次大小, L, S 分别为两幅特征图的特征点数量, C 为特征维度。为稳定相似度计算的数值范围,先对特征进行归一化处理:

$$\tilde{F}_0 = \frac{F_0}{\sqrt{C}}, \tilde{F}_1 = \frac{F_1}{\sqrt{C}}. \quad (12)$$

随后,计算两幅图像所有特征点对之间的相似度矩阵 $S \in \mathbb{R}^{N \times L \times S}$,其中每个元素通过缩放点积计算得到:

$$S_{b,l,s} = \frac{\tilde{F}_0[b,l] \cdot \tilde{F}_1[b,s]}{T}, \quad (13)$$

其中: $T > 0$ 为温度系数,用以调节相似度分布的平滑程度。

为了将相似度转化为概率分布以便后续匹配,分别沿两个维度对相似度矩阵进行softmax归一化,并对结果逐元素相乘,得到双向softmax归一化的置信度矩阵:

$$P_{b,l,s} = \text{softmax}(S_{b,l,:})_l \times \text{softmax}(S_{b,:,s})_s. \quad (14)$$

最后,基于置信度矩阵 P 进行互为最近邻筛选,即仅保留满足下列条件的匹配对:

$$P_{b,l,s} = \max_{s'} P_{b,l,s'} \text{ 且 } P_{b,l,s} = \max_{l'} P_{b,l',s}. \quad (15)$$

保证匹配的唯一性与对称性,从而得到可靠的粗匹配结果。

2.3.3 匹配点筛选

为提高图像匹配的稳定性与精度,在粗匹配阶段引入局部几何约束与光度一致性判别机制,构建了一种由粗至精的匹配点筛选流程。具体

而言,在基于互为最近邻的粗匹配之后,引入AffNet估计仿射帧(Local Affine Frame, LAF),再结合基于多个单应模型的几何一致性判别与归一化互相关完成匹配点精筛。

对于每一对粗匹配点 (p_i, q_i) ,在两幅图像中分别以其为中心裁剪固定大小的图像patch P_i^0, P_i^1 。将patch输入仿射网络AffNet,预测其局部几何变换矩阵 $A_i \in \mathbb{R}^{2 \times 2}$ 。结合匹配点坐标,可构造仿射帧:

$$\text{LAF}_i = [A_i | c_i] \in \mathbb{R}^{2 \times 3}, c_i = \text{中心点坐标}. \quad (16)$$

分别可得源图像和目标图像中的LAF表示:

$$\text{LAF}_i^0 = [A_i^0 | p_i], \text{LAF}_i^1 = [A_i^1 | q_i]. \quad (17)$$

在获取所有匹配对的LAF表达后,我们采用MiHo进行几何一致性验证。具体地,将所有匹配对聚类为多个局部平面模型,每个模型对应一个单应变换 H_j 。匹配点是否属于某一单应模型通过投影误差判断:

$$\epsilon_{ij} = \|q_i - \pi(H_j \tilde{p}_i)\|_2. \quad (18)$$

若存在某个 j 使得 $\epsilon_{ij} < \tau$,则该匹配点被视为几何一致。

为进一步增强筛选的判别力,引入基于仿射帧裁剪图像块之间的归一化互相关度量:

$$\text{NCC}(P_i^0, P_i^1) = \frac{\langle P_i^0 - \mu_0, P_i^1 - \mu_1 \rangle}{\sigma_0 \sigma_1}, \quad (19)$$

其中: μ_0, μ_1 为图像块均值, σ_0, σ_1 为标准差。结合几何与光度一致性,筛除“不属于任何平面模型”或“NCC相似度过低”的伪匹配。

2.3.4 基于Transformer的精细匹配

为在粗匹配结果的基础上实现子像素级别的高精度匹配,提出一种基于Transformer的精细化匹配模块(Refinement Transformer, RT)。该模块以MiHo+NCC筛选后的高置信候选点对 (p_1, p_2) 为输入,通过对局部patch的token化、跨图注意力交互、局部相关性建模与可微子像素回归,输出每个候选对的子像素精确位置与匹配置信度掩码。

以每个筛选后的粗匹配点对 (p_1, p_2) 为中心,在RepVGG提取的特征图上裁剪固定大小的局部patch(特征级):

$$P_A \in \mathbb{R}^{w \times w \times C}, P_B \in \mathbb{R}^{w \times w \times C}. \quad (20)$$

将每个 patch 展平为 $N = w^2$ 个 token:

$$\begin{aligned} X_A &= [x_A^1, \dots, x_A^N]^T \in \mathbf{R}^{N \times c}, \\ X_B &= [x_B^1, \dots, x_B^N]^T \in \mathbf{R}^{N \times c}. \end{aligned} \quad (21)$$

对 token 做通道归一化并加入位置编码。

RT 由 4 层同构块堆叠而成, 结构如图 5 所示。

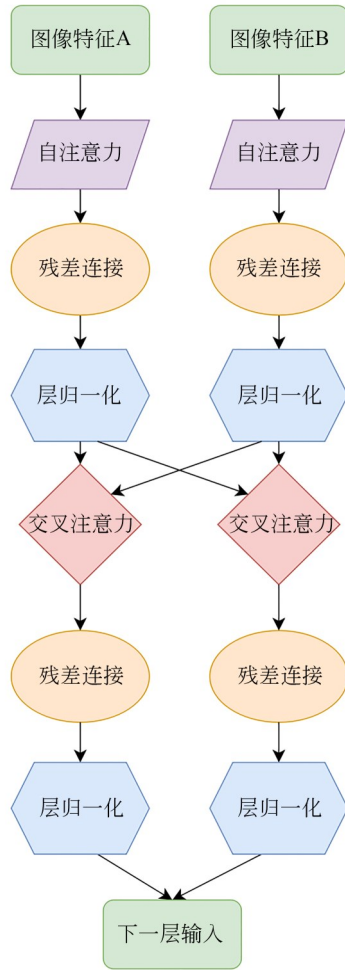


图 5 Transformer 单层结构

Fig. 5 Transformer single-layer structure

第 l 层对 $X_A^{(l)}, X_B^{(l)}$ 的计算流程如下(省略层归一化和残差连接):

(1) 自注意力(Self-Attn)——分别捕获 patch 内部的上下文:

$$\tilde{X}_A = \text{SelfAttn}(X_A), \tilde{X}_B = \text{SelfAttn}(X_B). \quad (22)$$

(2) 交叉注意力(Cross-Attn)——实现 $A \leftrightarrow B$

的信息互传:

$$\begin{aligned} \hat{X}_A &= \text{CrossAttn}(\tilde{X}_A, \tilde{X}_B), \\ \hat{X}_B &= \text{CrossAttn}(\tilde{X}_B, \tilde{X}_A), \end{aligned} \quad (23)$$

其中, 多头注意力的单头计算为:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (24)$$

SelfAttn 与 CrossAttn 按多头并行实现, 且在 $\mathbf{Q}, \mathbf{K}$ 上施加旋转位置编码(Rotary Positional Embedding, RoPE)。

(3) 前馈网络(FFN)——逐 token 的非线性变换与通道融合:

$$X_A^{(l+1)} = \text{FFN}(\hat{X}_A), X_B^{(l+1)} = \text{FFN}(\hat{X}_B). \quad (25)$$

最终输出记为增强特征对:

$$T_A \in \mathbf{R}^{N \times c}, T_B \in \mathbf{R}^{N \times c}. \quad (26)$$

为保留相对几何信息并增强相对位置感知, 采用 RoPE。将每个 token 的位置编码视作对特征向量的二维旋转变换: 对于一个 d 维向量, 可将其分解为若干二维子向量, 对第 k 对子维度施加旋转角度 $\theta_{\text{pos}, k}$, 即:

$$\forall v \in \mathbf{R}^d, v'_{(2k, 2k+1)} = R(\theta_{\text{pos}, k})v_{(2k, 2k+1)}, \quad (27)$$

其中:

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (28)$$

在注意力计算中, 将该旋转作用分别施加于查询 $\mathbf{Q}$ 与键 $\mathbf{K}$, 等价于在内积中引入相对位置信息, 从而使注意力权重自然编码相对位置关系。

利用 Transformer 增强后的 token 表示, 构造局部相似度矩阵(未归一化):

$$\mathbf{S} \in \mathbf{R}^{N \times N}, S_{ij} = \frac{\langle T_A^i, T_B^j \rangle}{\sqrt{C}}. \quad (29)$$

对每个查询 token i (对应 patch A 中位置 i) 沿 j 方向做温度化 softmax 得到概率分布:

$$P_{ij} = \frac{\exp(S_{ij}/\tau)}{\sum_{k=1}^N \exp(S_{ik}/\tau)}, \tau > 0, \quad (30)$$

其中: τ 是温度参数, 控制 softmax 输出分布的熵, 在本文中等于 1。

定义 patch 内每个候选位置 j 相对于 patch 中心的二维坐标为 $r_j = (x_j, y_j)$ 。则对于查询 token i 的子像素偏移(相对于候选粗位置)可用 soft-arg-

max 求期望:

$$\Delta_i = \sum_{j=1}^N P_{ij} r_j \in \mathbb{R}^2. \quad (31)$$

最终将粗匹配点 p_2 修正为子像素坐标:

$$\hat{p}_2 = p_2 + \Delta_{\text{center}}, \quad (32)$$

其中 Δ_{center} 为对应 patch 中心 token 的偏移。

2.4 角度误差修正矩阵计算

无源定位算法的本质是求解直线与曲线的交点。地球椭圆模型难以修正,通常视为精确的,因此,无源定位的模型可以简化成直线模型。无源定位的本质在于求解观测射线与地球表面模型的交点问题。在多数应用中,地球被建模为椭球体,其几何形状在全球尺度上具有较高精度,因此通常视作已知且无需进一步修正。基于这一前提,可将无源定位问题简化为观测直线与地球椭球面的交点求解问题。

进一步地,在特定分析条件下,若不考虑地球曲率或将局部区域近似为平面,则该模型可进一步简化为线性模型:

$$RX + T = Y, \quad (33)$$

其中: R 表示旋转矩阵, T 为位移矩阵, Y 为目标在地理坐标系下的定位结果, X 为目标点在图像坐标系中的坐标。在实际定位过程中,位移矩阵 T 代表载体平台的位置信息,若假设该位置测量具备较高精度,则主要误差来源可归结为旋转矩阵 R 中的姿态角误差。为补偿该误差,设引入一个修正矩阵 M ,用于修正由姿态误差引起的旋转偏差,则 M 应满足一定的约束关系。

$$MRX + T = Y', \quad (34)$$

其中 Y' 表示目标点的实际定位。

联立式(33)、式(34)可得:

$$M(Y - T) = Y' - T. \quad (35)$$

利用目标点的粗定位与精定位结果,即可求得误差修正矩阵 M 。在一次飞行过程中,可多次进行修正矩阵 M 的迭代,最后获得的修正矩阵 M 代表的是该系统的系统误差。

3 结果与分析

3.1 数据采集

实验数据来自高空大斜视条件下的无人机飞行试验。由于目前尚无针对该应用场景的公

开数据集,为保证实验的真实性与代表性,采用中国科学院长春光学精密机械与物理研究所提供的无人机飞行数据。该数据集包含十余个场景,涵盖无人机斜视成像图像及其对应的飞行参数。具体而言,无人机飞行高度约为 14.6 km,俯仰倾角覆盖 $0^\circ \sim 85^\circ$,目标斜距分布在 20~100 km 之间,能够充分表征高空大斜视成像的典型特征,从而为算法性能验证提供了可靠的数据支撑。

在图像匹配模型的训练阶段,为确保实验的可重复性与结果的可比性,采用与 LoFTR 相同的公开训练数据集。考虑到本研究的应用场景主要为室外地物成像与匹配,因此主要选用 MegaDepth^[23] 和 PhotoTourism^[24] 两个大规模室外数据集。其中,MegaDepth 数据集基于多视图重建生成,涵盖了丰富的地形起伏、建筑结构与视角变化,能够有效提升模型在大尺度、强透视条件下的稳定性;PhotoTourism 数据集则由真实旅游景点照片构成,包含多种光照、天气与拍摄角度差异,有助于增强模型在复杂自然环境下的泛化能力。

实验设备采用一台配备 NVIDIA GeForce RTX 4070 GPU (8 GB 显存) 和 Intel Core i7-11800H 处理器的计算机。

3.2 单次定位结果

图 6 展示了单次定位流程中的中间结果与最终定位效果。为进一步验证本文算法的有效性,对近年来相关研究提出的多种典型目标定位方法进行了对比分析。Du 等提出了一种基于无人机航拍图像的协同目标定位方法^[25]。在飞行高度约 3 000 m、平均倾角约 10° 的条件下,该方法需通过 4 架无人机的协同观测,方能将地面目标的定位误差控制在约 20 m 的水平。相比之下,实验平台的飞行高度达到 14 km,成像倾角远超上述文献所涉场景。在此高空大斜视的实验条件下,由于成像几何放大效应及测量误差的累积影响,推测该方法的定位精度将显著下降,难以达到与本文方法相同的精度水平。

Guo 等开展了基于自适应扩展卡尔曼滤波的机载光电稳定平台目标定位研究^[26]。在飞行高度约 5 000 m、近垂直观测(倾角较小)的条件下进行了目标定位实验。实验结果表明,在不同

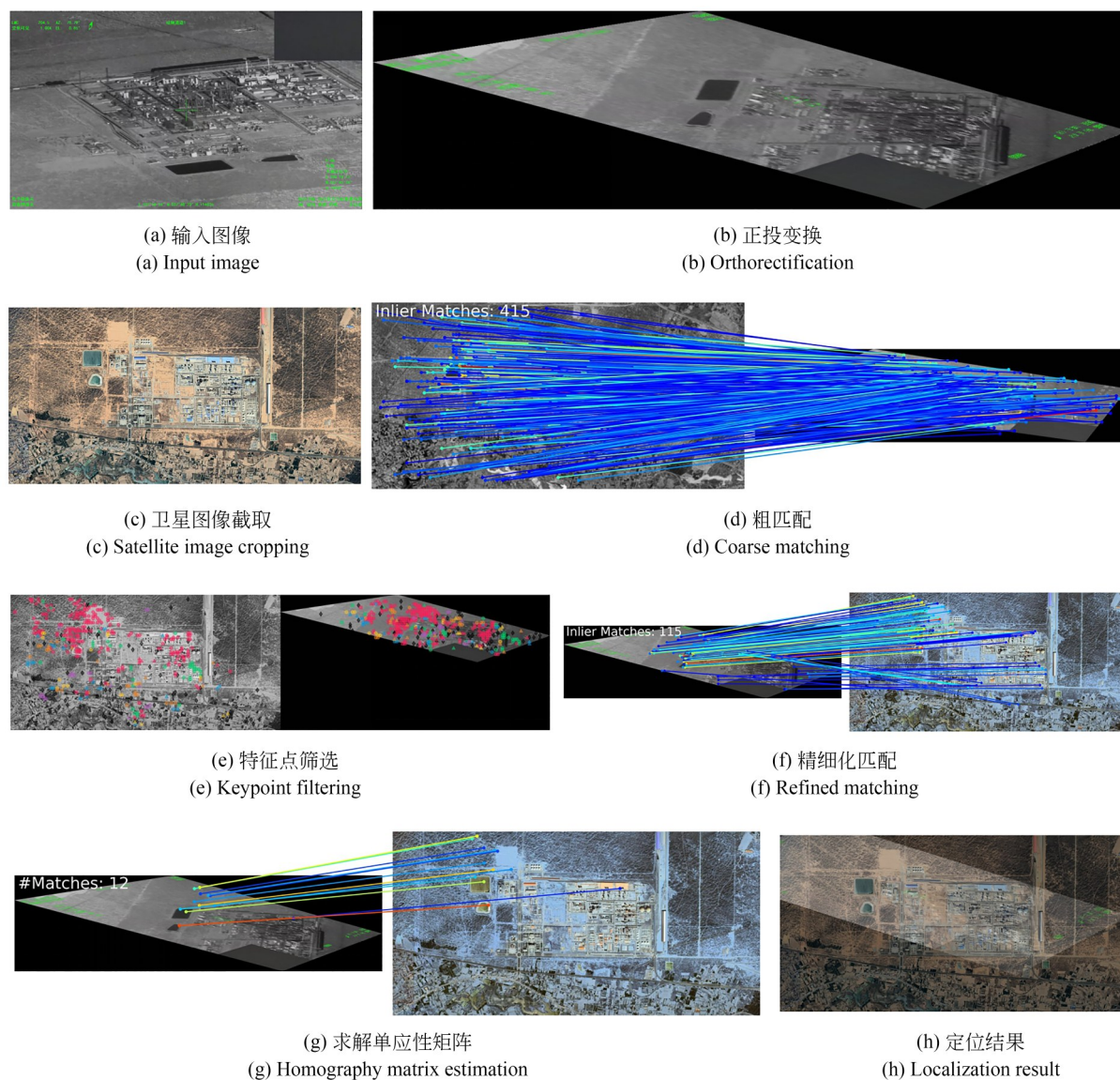


图 6 单次定位结果

Fig. 6 Single positioning results

噪声水平下,该方法的平均定位误差约为 10~11 m,且在 95% 置信水平下的 CEP 半径约为 5.5 m。然而,该方法主要针对低倾角场景设计,且其实验条件(飞行高度、噪声水平)相对理想。相较之下,本文实验场景中飞行高度达到 14 km,目标斜距超过 70 km,属于典型的高空大斜视成像情形。考虑到此类条件下的成像几何畸变、姿态误差放大以及测距不确定性等综合影响,上述方法在本实验条件下的定位性能预计会显著下降,难以优于本文提出的算法。

总体而言,其余相关算法的适用条件与上述方法相近,均主要面向中低空、低倾角成像环境,

难以在高空大斜视条件下实现优于本文算法的定位精度。

图 7 对不同方法的定位误差进行了可视化,以曲线形式呈现各方法在不同条件下的性能表现(彩图见期刊电子版)。绿色曲线表示在角度误差控制在 0.1° 以内时,通过蒙特卡罗模拟得到的理论最优定位误差;紫色曲线对应实验中实际获得的粗定位误差;黑色曲线为采用数字高程模型方法得到的定位误差;红色曲线表示本文提出算法的定位误差;橙色曲线为经过角度误差修正后的粗定位结果。

从图 7 可进一步看出,本文方法在高空、大斜

视条件下的定位误差明显低于其他方法,其主要原因在于本文在图像匹配与几何校正环节均进行了针对性优化。一方面,基于粗定位结果的正射投影显著减小了影像与地图之间的视角差异,降低了由地形起伏与姿态误差引起的几何失配。另一方面,分层匹配策略有效结合了轻量化特征提取与精细化匹配,使得算法能够在大范围搜索与局部细节保持之间取得平衡。此外,MiHo聚类与NCC相似性度量的结合,在匹配点筛选阶段有效抑制了由噪声与纹理重复引起的误匹配,从而保证了后续单应性估计的稳定性。综上,本文方法在复杂几何条件下依然能够保持较高的定位精度与稳定性。

进一步分析发现,角度误差修正模块在性能提升中起到了关键作用。当姿态角存在微小偏差时,传统DEM方法会放大误差传播,导致位置漂移,而本文通过引入基于影像几何约束的角度误差补偿,有效削弱了这种非线性误差累积,从

而在相同姿态精度条件下实现更优的定位结果。该机制解释了图 7 中红色曲线相对于橙色与黑色曲线的显著下降趋势,也体现出本文算法在误差敏感场景下的优势。表 1 展示了实验中部分定位数据,所涉及的误差均为平均值。

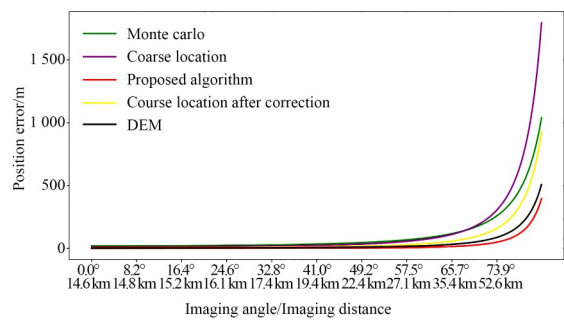


图 7 成像角度与定位误差的关系
Fig. 7 Relationship between imaging angle and positioning error

表 1 传统算法与本文算法的定位误差
Tab. 1 Positioning error between traditional algorithm and proposed algorithm

No.	Flight altitude/km	Target slant range/km	Inclination angle/(°)	Monte carlo error/m	Coarse localization error/m	DEM error/m	Proposed matching error/pixel	Proposed error/m
1	14.7	50~60	72.9~75.8	333.4	370.2	100.2	49.2	52.9
2	14.7	60~70	75.8~77.9	515.2	500.4	150.2	60.8	65.4
3	14.6	70~80	77.9~79.5	719.2	820.8	213.6	71.9	77.3
4	14.6	80~90	79.5~80.7	870.8	1 026.7	330.8	87.7	94.3
5	14.6	90~100	80.7~81.6	1 076.6	1 400.2	500.2	114.6	123.2

3.3 图像匹配算法定位精度对比

为验证所提算法的精度与稳定性,选取多个具有代表性的主流图像匹配算法进行对比实验,包括尺度不变特征变换(Scale-Invariant Feature Transformer, SIFT)^[27]、离散图像空间关键点(Discrete Image Space Keypoints, DISK)^[28]、深度特征网络(Deep Local Feature Network, D2-Net)^[29]、SuperPoint:自监督兴趣点检测与描述子(SuperPoint: Self-Supervised Interest Point Detector and Descriptor)^[30]、基于Transformer的无检测器特征匹配(Detector-Free Local Feature Matching with Transformers, LoFTR)^[31]、LiftFeat:结构化特征学习方法(Lifted Structure Feature, LiftFeat)^[32]、稀疏特征描述子与密集检测器

(Sparse Feature Descriptor with Dense Detector, SFD2)^[33]、最小特征集匹配方法(Matching with INteraction-aware MInimal MATches, MINIMA)^[34]、高效局部特征变换器(Efficient Local Feature Transformer, EfficientLoFTR)^[35],以及本文算法。
在评估过程中发现,SIFT,DISK,D2-Net,SuperPoint和LoFTR等传统或基准方法在本实验场景下的匹配结果不理想,表现为匹配精度低、定位误差大,难以满足精确定位的要求,因此未纳入最终统计分析。
图 8 和表 2 给出了不同飞行高度和斜视条件下,多种图像匹配算法的定位误差对比结果。可以看出,随着目标斜距的增加和倾斜角度的增

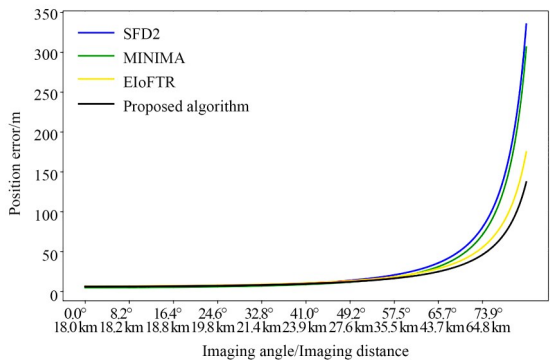


图 8 不同匹配算法定位误差对比

Fig.8 Comparison of positioning errors of different matching algorithms

大,各算法的定位误差均呈上升趋势。在传统方法中,SFD2与MINIMA在低至中等斜视条件下尚能维持百米级精度,但在高斜视(斜距>80 km,倾斜角度>79.5°)时,误差急剧增大,最高超过270 m。相比之下,基于深度特征的ELoFTR在各测试条件下均取得了更优的表现,误差范围维持在60~150 m。与这些方法相比,本文算法

在所有场景下均表现出更高的定位精度,误差稳定保持在50~120 m之间,尤其在大斜距条件下依然保持较好的稳定性,充分体现了它在复杂几何条件下的适用性。

本文算法在所有测试场景中均表现出更高的定位精度与稳定性,主要得益于多阶段、粗到精的匹配策略及MiHo+NCC复合精炼机制。在粗匹配阶段,通过RepVGG提取高层语义特征并结合互为最近邻匹配,快速获得初步匹配点。精炼阶段中,MiHo聚类利用局部几何一致性对匹配点分簇,NCC相似度进一步剔除误匹配,从而形成高置信度匹配集合。该机制能够在拍摄高度不均和大斜距条件下,通过多簇中间单应性的综合利用,实现全局平均定位误差最优化。随后,基于Transformer的精匹配模块通过注意力机制进一步增强局部到全局特征关联,从而保证在复杂几何条件下仍维持较低且稳定的定位误差。由此,本文算法在表2所示的各场景中始终优于SFD2,MINIMA与ELoFTR,充分体现了其稳定性与适用性。

表 2 不同图像匹配算法的定位误差比较

Tab.2 Comparison of positioning error of different image matching algorithms

No.	Flight altitude/km	Target slant range/km	Inclination angle/(°)	SFD2 error/m	MINIMA error/m	ELoFTR error/m	Proposed algorithm error/m
1	14.7	50~60	72.9~75.8	100.8	85.2	60.2	52.9
2	14.7	60~70	75.8~77.9	125.1	100.5	78.7	65.4
3	14.6	70~80	77.9~79.5	147.5	164	100.9	77.3
4	14.6	80~90	79.5~80.7	230.3	190.7	120.3	94.3
5	14.6	90~100	80.7~81.6	275.2	250.8	150.7	123.2

表3总结了各算法在匹配时间、匹配点数量以及内点率等性能指标上的对比结果。从运行

表 3 图像匹配算法性能对比

Tab.3 Performance comparison of image matching algorithms

Algorithm	Matching time/s	Number of matches	Inlier rate/%
SFD2	0.3	570	5.4
MINIMA	0.5	680	5.6
ELOFTR	1.1	1 000	10.8
本文	1.5	360	50.7

效率来看,SFD2和MINIMA的匹配时间较短,分别为0.3 s和0.5 s,而基于深度特征的ELoFTR与本文方法耗时相对较长,分别为1.1 s和1.5 s。在匹配数量方面,ELoFTR能够产生最多的匹配点(约1 000个),而本文方法输出的匹配点数量相对较少(约360个)。值得注意的是,本文方法的内点率高达50.7%,远超SFD2(5.4%),MINIMA(5.6%)和ELoFTR(10.8%)。这一结果表明,虽然本文方法在匹配数量上不占优势,但其输出的匹配点更加可靠,显著提升了有效内点比例,保证了整体定位精度的提升。

理论上,本文方法的高内点率得益于多阶

段、层次化的特征筛选与精匹配机制。首先,采用可重参数卷积神经网络 RepVGG 提取输入图像的下采样粗特征,从全局层面捕获场景的主要结构信息,为后续匹配提供稳定的语义基础。随后,利用 AffNet 为候选匹配点生成局部仿射帧,实现对局部几何变形的建模与归一化处理,从而在大视角差与尺度变化条件下保持特征一致性。

在此基础上,本文引入了中间单应性与归一化互相关相结合的复合筛选机制。其中,MiHo 算法通过对初始匹配点的几何一致性进行聚类,能够有效识别具有相似单应关系的匹配群体,抑制随机误匹配带来的局部不稳定性。而 NCC 相似性度量则从强度域角度评估局部图像块的相关性,对光照变化和纹理噪声具有较强的稳定性。二者结合既兼顾了几何一致性约束,又保留了局部影像内容的相似性特征,从而在理论上显著降低了误匹配概率。

经过 MiHo+NCC 的双重筛选后,保留下来的高置信度匹配点在几何与光度层面均具有一致性。本文进一步以这些点为中心裁剪局部图像块,再次利用 RepVGG 提取高分辨率局部特征,实现更具判别力的特征表达。最后,通过基于 Transformer 的精匹配模块,对局部特征进行全局注意力融合,从多尺度上下文中建立精确的特征关联关系。

整体而言,本文方法在设计上实现了从全局

语义建模→局部几何一致性筛选→局部精细特征融合的逐层优化过程。特别是 MiHo+NCC 复合筛选机制在几何与光度一致性两方面协同约束,使得匹配点集合在理论上更接近真实的空间对应关系,从而解释了本文方法显著高于其他算法的内点率。

实验结果表明,本文算法的定位精度显著优于现有对比方法。同时,该方法在计算效率上亦表现出良好的实用性,尽管单次定位时间略有增加,但仍处于实际应用可接受的范围内,体现了在精度与效率之间的良好平衡。

3.4 消融实验

为验证各模块在算法整体性能中的作用,本文设计了消融实验,对比分析不同模块组合对定位精度的影响。实验方案如下:

方法 a:仅采用粗匹配与精匹配流程,移除 MiHo+NCC 筛选模块;

方法 b:采用粗匹配并引入 MiHo+NCC 筛选机制,但不进行精细化匹配;

方法 c:完整采用“粗匹配+MiHo+NCC 筛选+精细化匹配”的全流程。

各方法的定位误差结果如表 4 所示。从结果可见,方法 c 在各倾角区间均取得最低的定位误差,相较于方法 a 提升约 50%,相较于方法 b 提升约 75%,验证了多阶段匹配策略在高倾角场景下的显著优势。

表 4 消融实验定位误差效果
Tab. 4 Result of positioning error in ablation experiment

No.	Flight altitude/ km	Target slant range/km	Inclination angle/(°)	Error/m		
				Method A	Method B	Method C
1	14.7	50~60	72.9~75.8	97.5	201.4	52.9
2	14.7	60~70	75.8~77.9	117.4	273.8	65.4
3	14.6	70~80	77.9~79.5	145.9	298.4	77.3
4	14.6	80~90	79.5~80.7	225.7	361.9	94.3
5	14.6	90~100	80.7~81.6	250.8	432.5	123.2

由此可知,MiHo+NCC 筛选机制在整体框架中贡献最大。MiHo 模块通过对初始匹配点进行几何一致性聚类,有效抑制了由视差与姿态变化引起的随机误匹配。NCC 模块则在光度层面补偿了纹理差异与光照不一致造成的特征相似度偏差。二者协同作用,显著提升内点比例(由

约 10% 提升至 50% 以上),从而为单应矩阵估计提供了稳定约束。这表明,MiHo+NCC 是算法稳定性与精度提升的核心模块。

精细化匹配模块主要影响定位的亚像素精度。该模块通过局部区域特征重提取与 Transformer 全局优化,有效减小了匹配点间的像素偏

差,使得单应性解算误差显著降低。若省略该阶段(即方法b),定位误差平均增大约60~80 m,表现为系统精度下降而非匹配失败,说明精匹配是提升最终精度的关键。

综上所述,MiHo+NCC模块在匹配稳定性上具有决定性贡献,而精细化匹配模块则对亚像素级精度优化起主导作用。二者相辅相成:前者保障全局匹配的可靠性,后者抑制局部像素偏差。在实时性要求较高的应用场景中,可在保持MiHo+NCC机制的前提下适度降低精细化匹配的计算权重,从而在精度与效率间取得平衡。

4 结 论

本文提出了一种适用于高空大斜视条件的

基于图像匹配的精细化定位方法。该方法在粗匹配基础上引入匹配筛选与精细化模块,实现了更高的几何一致性与匹配精度。实验结果表明,在斜距约90 km的场景中,定位误差可控制在约120 m以内,显著优于传统无源定位算法。该方法在复杂视角条件下具有高精度与稳定性,为高空无源定位提供了可行的技术路径。

作者贡献声明:

贾平:论文构思;
李昌灏:方法论,初稿写作;
孙辉:论文审核与编辑;
宋悦铭:数据管理;
冯卓莹:论文指导;
徐芳:资助获取。

参考文献:

- [1] 李全超. 基于万向节的机载高精度光电平台机构研究[D]. 长春: 中国科学院长春光学精密机械与物理研究所, 2022.
LI Q CH. *Research on Mechanism of Aerial High-Precision Optoelectronic Platform Based on Universal Joint* [D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2022. (in Chinese)
- [2] 周紫君, 郝丰, 吴方, 等. 一种机载惯性/光电组合导航误差修正方法[J]. 中国惯性技术学报, 2024, 32(8): 779-786, 794.
ZHOU Z J, YU F, WU F, *et al.* Error correction method for airborne SINS/EODS integrated navigation [J]. *Journal of Chinese Inertial Technology*, 2024, 32(8): 779-786, 794. (in Chinese)
- [3] 袁美桂, 魏志强, 轩新想. 一种新的机载光电侦察系统目标定位精度分析方法[J]. 光学技术, 2024, 50(2): 215-219.
YUAN M G, WEI ZH Q, XUAN X X. A new precision analysis method for target localization of airborne optoelectronic reconnaissance systems [J]. *Optical Technique*, 2024, 50(2): 215-219. (in Chinese)
- [4] 孙春生, 吴依伦. 无人机载光电吊舱观瞄角误差分配方法研究[J]. 红外与激光工程, 2023, 52(11): 234-244.
SUN CH SH, WU Y L. Research on the error distribution method of unmanned airborne electro-optical pod viewing angle [J]. *Infrared and Laser Engineering*, 2023, 52(11): 234-244. (in Chinese)
- [5] 孙辉. 机载光电载荷对地目标定位技术研究[D]. 长春: 中国科学院长春光学精密机械与物理研究所, 2022.
SUN H. *Research on Ground Target Positioning Technology of Airborne Optoelectronic Payload* [D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2022. (in Chinese)
- [6] ZHANG X F, YUAN G Q, ZHANG H W, *et al.* Precise target geo-location of long-range oblique reconnaissance system for UAVs [J]. *Sensors*, 2022, 22(5): 1903.
- [7] 李梓豪. 基于数字高程模型的大倾角航空相机目标定位研究[D]. 长春: 中国科学院长春光学精密机械与物理研究所, 2023.
LI Z H. *Research on Object Location of High Inclination Aerial Camera Based on Digital Elevation Model* [D]. Changchun: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2023. (in Chinese)
- [8] 彭锐晖, 郭玮, 钱海宁, 等. 多传感器组网的航空目标三维定位算法[J]. 电子学报, 2024, 52(12): 4059-4073.
PENG R H, GUO W, QIAN H N, *et al.* Multi-sensor network-based three-dimensional positioning algorithm for airborne targets [J]. *Acta Electronica Sinica*, 2024, 52(12): 4059-4073. (in Chinese)

- [9] 马德仁,王振桓,张荣兵,等. 基于数据链测距的分布式三维协同定位方法[J]. 遥测遥控, 2025, 46(2): 12-19.
MA D R, WANG ZH H, ZHANG R B, *et al.* Distributed three-dimensional cooperative localization method based on data link ranging [J]. *Journal of Telemetry, Tracking and Command*, 2025, 46(2): 12-19. (in Chinese)
- [10] LIU C Y, DING Y L, ZHANG H W, *et al.* Improving target geolocation accuracy with multi-view aerial images in long-range oblique photography [J]. *Drones*, 2024, 8(5): 177.
- [11] DAI M, HU J H, ZHUANG J D, *et al.* A transformer-based feature segmentation and region alignment method for UAV-view geo-localization [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4376-4389.
- [12] TIAN X Y, SHAO J, OUYANG D Q, *et al.* UAV-satellite view synthesis for cross-view geo-localization [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4804-4815.
- [13] WANG T Y, ZHENG Z D, YAN C G, *et al.* Each part matters: local patterns facilitate cross-view geo-localization [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(2): 867-879.
- [14] WU S B, DU C, CHEN H, *et al.* Coarse-to-fine UAV image geo-localization using multi-stage Lucas-kanade networks [C]. 2021 2nd Information Communication Technologies Conference (ICTC). May 7-9, 2021, Nanjing, China. IEEE, 2021: 220-224.
- [15] ZHENG Z D, WEI Y C, YANG Y. University-1652: a multi-view multi-source benchmark for drone-based geo-localization [C]. *Proceedings of the 28th ACM International Conference on Multimedia*. Seattle WA USA. ACM, 2020: 1395-1403.
- [16] DING L R, ZHOU J, MENG L X, *et al.* A practical cross-view image matching method between UAV and satellite for UAV-based geo-localization [J]. *Remote Sensing*, 2021, 13(1): 47.
- [17] LUO J Q, YE Q, YANG Z X, *et al.* Positioning objects from large oblique ultralong focal length UAV imagery [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6014305.
- [18] YAO Y W, SUN C, WANG T, *et al.* UAV geo-localization dataset and method based on cross-view matching [J]. *Sensors*, 2024, 24(21): 6905.
- [19] FAN J Q, ZHENG E H, HE Y F, *et al.* A cross-view geo-localization algorithm using UAV image and satellite image [J]. *Sensors*, 2024, 24(12): 3719.
- [20] LIU J L, XIAO J C, REN Y F, *et al.* Multi-source image matching algorithms for UAV positioning: benchmarking, innovation, and combined strategies [J]. *Remote Sensing*, 2024, 16(16): 3025.
- [21] REN Y, DONG G H, ZHANG T B, *et al.* UAVs-based visual localization via attention-driven image registration across varying texture levels [J]. *Drones*, 2024, 8(12): 739.
- [22] DING X H, ZHANG X Y, MA N N, *et al.* RepVGG: making VGG-style ConvNets great again [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25, 2021. Nashville, TN, USA. IEEE, 2021: 13728-13737.
- [23] LI Z Q, SNAVELY N. MegaDepth: learning single-view depth prediction from Internet photos [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 2041-2050.
- [24] JIN Y H, MISHKIN D, MISHCHUK A, *et al.* Image matching across wide baselines: from paper to practice [J]. *International Journal of Computer Vision*, 2021, 129(2): 517-547.
- [25] DU M L, ZOU H D, WANG T H, *et al.* A cooperative target localization method based on UAV aerial images [J]. *Aerospace*, 2023, 10(11): 943.
- [26] GUO Z J, ZHAO F, SUN Y, *et al.* Research on target localization in airborne electro-optical stabilized platforms based on adaptive extended Kalman filtering [J]. *Measurement*, 2024, 234: 114794.
- [27] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [28] TYSZKIEWICZ M J, FUA P, TRULLS E. DISK: Learning local features with policy gradient [C]. In *Advances in Neural Information Processing Systems (NeurIPS)*. 2020, 33: 8724-8735.
- [29] DUSMANU M, ROCCO I, PAJDLA T, *et al.*

- D2-Net: A trainable CNN for joint detection and description of local features[C]. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 8092-8101.
- [30] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: self-supervised interest point detection and description [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 18-22, 2018, Salt Lake City, UT, USA. IEEE, 2018: 337-349.
- [31] SUN J M, SHEN Z H, WANG Y A, *et al.* LoFTR: detector-free local feature matching with transformers[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 8918-8927.
- [32] LIU Y P, LAI W P, ZHAO Z, *et al.* LiftFeat: 3D geometry-aware local feature matching [J]. *arXiv preprint arXiv:2505.03422*.
- [33] XUE F, BUDVYTIS I, CIPOLLA R. SFD2: semantic-guided feature detection and description [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 5206-5216.
- [34] REN J W, JIANG X Y, LI Z Z, *et al.* MINIMA: Modality Invariant Image Matching[J]. *arXiv preprint arXiv:2412.19412*.
- [35] WANG Y F, HE X Y, PENG S D, *et al.* Efficient LoFTR: semi-dense local feature matching with sparse-like speed[C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 21666-21675.

作者简介:



贾 平(1964—),男,吉林朝阳川人,博士,研究员,1988年于中国科学技术大学获得硕士学位,2006年于长春光学精密机械与物理研究所获得博士学位,主要从事动态光学成像与测量技术方面的研究。E-mail:jiap@ciomp.ac.cn



李昌灏(2000—),男,山西临汾人,硕士研究生,2023年于浙江大学获得学士学位,主要从事航空图像处理方面的研究。E-mail:lichanghao23@mails.ucas.ac.cn