



引文格式:崔志祥,蓝朝桢,张永显,等.一种热红外与可见光影像深度特征匹配方法[J].武汉大学学报(信息科学版),2023,48(2):316-324.DOI:10.13203/j.whugis20200181

Citation: CUI Zhixiang, LAN Chaozhen, ZHANG Yongxian, et al. A Method Based on Depth Features for Matching Thermal Infrared Images with Visible Images[J]. Geomatics and Information Science of Wuhan University, 2023, 48(2): 316-324. DOI: 10.13203/j.whugis20200181

一种热红外与可见光影像深度特征匹配方法

崔志祥^{1,2} 蓝朝桢² 张永显² 侯慧太² 秦剑琪^{2,3}

¹ 31682部队,甘肃 兰州,730020

² 信息工程大学地理空间信息学院,河南 郑州,450001

³ 63611部队,新疆 库尔勒,841000

摘要:针对无人机热红外影像与光学卫星影像的匹配难题,提出一种基于异源地标数据集学习的深度局部特征匹配方法。首先,利用生成对抗网络学习热红外与可见光影像的灰度分布规律,并进一步合成用于特征提取模型训练的热红外影像地标数据集;然后,联合残差网络和注意力机制模型,从数据集中学习深度不变特征;最后,经过对不变特征的匹配、提纯等处理,获得像对的正确匹配点。试验测试了该方法的性能,并与KAZE、特征检测描述网络和深度局部特征模型进行了对比。结果表明,提出的方法对灰度、纹理、重叠率以及几何变化具有较强的适应性,且匹配效率较高,可为无人机视觉导航提供支撑。

关键词:异源地标数据集;深度局部特征;匹配;热红外;可见光

中图分类号:P237

文献标识码:A

收稿日期:2021-04-14

DOI:10.13203/j.whugis20200181

文章编号:1671-8860(2023)02-0316-09

A Method Based on Depth Features for Matching Thermal Infrared Images with Visible Images

CUI Zhixiang^{1,2} LAN Chaozhen² ZHANG Yongxian² HOU Huitai² QIN Jianqi^{2,3}

¹ Troops 31682, Lanzhou 730020, China

² Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China

³ Troops 63611, Kuerle 841000, China

Abstract: Objectives: Aiming at the matching problem of unmanned aerial vehicle(UAV) thermal infrared images and optical satellite images, a deep local feature matching method based on heterogeneous landmark dataset for learning is proposed. **Methods:** Firstly, the gray distribution law of thermal infrared images and visible images is learned by the generative adversarial network, and the landmark dataset consisting of thermal infrared images for feature extraction model training is synthesized. Secondly, the deep invariant features are learned from the multi-modal landmark dataset by the residual network and attention mechanism. Finally, correct matching points of image pairs are obtained by matching and purifying the invariant features. **Results:** The performance of this method was tested experimentally and compared with KAZE, detect-and-describe network and deep local features. The results show that the adaptability of this method to the grayscale, texture, overlap rate and geometric variations is stronger, and the matching efficiency of this method is higher. **Conclusions:** The effectiveness of this method is proved through multiple sets of experiments. Therefore, the UAV visual navigation is provided support for.

Key words: heterogeneous landmark dataset; deep local feature; match; thermal infrared; visible

基金项目:中原科技创新领军人才计划(194200510023)。

第一作者:崔志祥,硕士,主要从事无人机影像处理的理论与方法研究。1462665541@qq.com

通讯作者:蓝朝桢,博士,副教授。lan_cz@163.com

基于自然路标的无人机视觉导航技术通过在线识别光学卫星影像上的自然特征,解算相机的绝对位姿,可实现全球导航卫星系统(global navigation satellite system, GNSS)拒止下的导航定位^[1-4]。然而,该技术对异源影像匹配方法的依赖性较高。尤其搭载热红外相机在夜间执行任务时,成像原理不同导致热红外影像与光学卫星影像的灰度、纹理等存在显著差异;此外,这两种影像间还可能存在一定的几何畸变,即使通过导航数据预处理仍难以完全消除,从而使得二者的匹配面临巨大挑战,严重制约了无人机视觉导航技术在夜间的应用。

针对热红外与可见光影像的匹配难题,国内外学者进行了大量研究。其中,具有代表性的是基于尺度不变特征变换(scale-invariant feature transform, SIFT)或 KAZE 的改进算法^[5-9],这类算法利用点特征进行匹配,通常具有较高的几何不变性,但描述符受热红外影像梯度平滑和纹理退化的影响较大。为提高点特征的匹配效果,文献[10-11]提出基于弱特征几何关系约束的方法,该方法有效增强了匹配的鲁棒性,但对特征点的检测精度具有依赖性,适用于角点丰富的区域。考虑到边缘特征在热红外影像上的稳定性,文献[12-14]提出点特征与边缘特征联合匹配的方法,提高了特征的可重复性,但边缘图像信息丢失降低了特征点的数量与描述符的独特性。为克服非线性辐射差异对特征检测与描述的制约,文献[15-16]提出基于图像结构与形状相似性度量的模板匹配方法,在几何畸变较小的前提下,实现了多模态遥感影像的快速、精确配准。以上传统匹配方法均基于影像上简单的几何形状或空间关系进行设计,稳定性容易受到影像外观变化的影响。

基于深度学习的匹配方法通过网络模型学习训练,自动挖掘影像中的深度特征和语义信息,能够适应更大的影像差异。如文献[17]提出一种端到端学习的不变特征变换(learned invariant feature transform, LIFT)模型,提取的特征比 SIFT 特征更加稠密,对灰度变化更加鲁棒。文献[18]提出一种基于多通道特征描述符的关键点检测描述网络(detect-and-describe network, D2-Net),实现了宽基线、昼夜影像的精确匹配。文献[19]提出一种基于深度学习的模板匹配方法,实现了尺度、方向基本一致的多模态遥感影像匹配。虽然基于深度神经网络的匹配方法解

决了诸多影像匹配难题,但目前对此类方法的研究主要集中在近景可见光影像和多模态卫星影像领域,针对无人机热红外影像与光学卫星影像匹配的研究还相对较少,因此现有方法难以在效率和鲁棒性方面满足视觉导航应用的需求。

2017 年谷歌提出了一种基于深度局部特征(deep local features, DELF)的大规模图像检索方法^[20]。由于 DELF 特征对影像的灰度与几何变化较为鲁棒,因此常被用作匹配方法的对比基准^[18,21]。DELf 的应用范围集中在可见光影像领域,由于其训练集由简单的影像集合构成,因此便于通过数据增广技术提高数据集的多样性,从而帮助网络模型学习更加稳健的深度特征。

根据 DELF 的特点和无人机视觉导航的应用需求,本文设计了热红外影像合成的方法增广数据集,提高训练集的非线性辐射差异,通过重新训练 DELF 和调整影像金字塔模型,提出一种鲁棒、高效的无人机热红外影像与光学卫星影像匹配方法。

1 深度局部特征提取网络

文献[20]首先利用 ResNet50^[22]残差网络提取影像的多通道特征图,再利用注意力网络在特征图中筛选与地标相关的语义关键点,并用降维的通道向量进行描述,从而得到用于影像匹配的深度局部特征和描述符。

1.1 基于残差网络的特征图提取

特征图是深度局部特征检测与描述符构建的基础。为提取较为鲁棒的深度特征,需要采用层数较多的网络模型,但网络越深,梯度衰减越严重,越容易产生训练误差逐渐增大的病态问题。为此,文献[20]将 ResNet50 作为主干网络提取特征图。ResNet50 由 50 层神经网络组成,结构如表 1 所示,conv1 为 7×7 、64 通道的单卷积层,conv2_x~conv5_x 为 4 个组卷积模块,由若干 BottleNeck 组成,通过在标准前馈卷积网络的基础上增加跳跃连接设计,可实现深层特征与浅层特征的叠加,使得卷积训练只需要拟合残差,剩余梯度可以通过跳跃连接继续反向传播,从而抑制梯度衰减。

ResNet50 中间层输出的计算结果即为影像的特征图。网络层数越浅,深度特征的感受野越小,其在原始影像上的定位精度越高,但特征越简单,主要由点、线等底层特征构成,鲁棒性较差;网络层数越深,深度特征的感受野越大,定位

精度越低,但特征越抽象,主要由深度语义特征构成,鲁棒性较强。为使深度特征同时具有高精度、抽象性的特点,文献[20]将 conv4_x 输出的 1 024 通道的特征图作为深度局部特征检测与描述符构建的基础,每个特征点在原始影像上的位置为局部感受野的中心,通道特征值构成了深度特征的描述向量。虽然与 KAZE 等传统特征相比,深度特征在原始影像上的定位精度较低,但由于后者的重复性较强,匹配点较多,因此通过以多求可靠与几何约束的方法进行筛选,仍然能够获得一定精度范围内的正确匹配点。

表 1 ResNet50 网络结构

Tab.1 Network Architecture of ResNet50

图层名称	输入尺寸/像素	网络结构
conv1	224×224	7×7,64,步长为2
		3×3最大池化,步长为2
conv2_x	112×112	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
conv3_x	56×56	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$
conv4_x	28×28	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 6$
conv5_x	14×14	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$
输出	7×7	7×7均值池化, 2 048×1 000全连接,softmax函数

1.2 基于注意力网络的关键点筛选与描述符生成

特征图的特征点较为密集,用于构建影像的整体表达具有较高的紧致性、完整性和灰度不变性,但用于影像匹配时,相邻特征的相似性会增加大量冗余信息,降低匹配的效率和准确性。为此,文献[20]引入一种注意力网络,通过地标分类监督训练,获取每个特征的注意力得分,根据得分即可在特征图中筛选出判别性高、语义信息丰富的关键点。

图 1 为关键点筛选的网络模型示意图,即去掉 ResNet50 的 conv5_x 和输出层,在 conv4_x 后面增加注意力网络,如图 1 中的橙色部分所示,其由两个 1×1 的卷积层构成,每个卷积层均采用 softplus 函数激活。注意力网络可输出每个特征的相关分数,用于表达特征与地标的相关性。根据特征值与注意力得分,进行加权求和池化降采样处理和全连接层计算,即可对影像地标进行预测,实现监督训练。表达式如下:

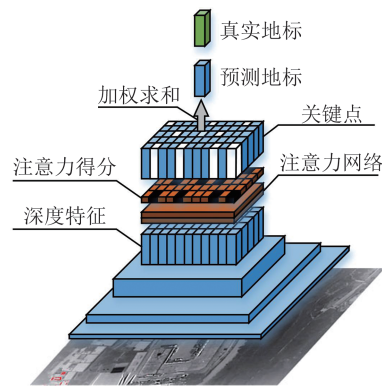


图 1 得分训练

Fig.1 Score Training

$$\mathbf{y} = \mathbf{W} \left(\sum_n \alpha(\mathbf{f}_n; \theta) \cdot \mathbf{f}_n \right) \quad (1)$$

式中, \mathbf{f}_n 为深度特征; $\alpha(\mathbf{f}_n; \theta)$ 为每个特征的得分函数; θ 为得分函数 $\alpha(\cdot)$ 的训练参数; \mathbf{W} 为全连接层的权重和偏移参数; \mathbf{y} 为地标的 one-hot 预测值。训练过程中,希望逐步建立并优化注意力得分与特征语义之间的关系,即注意力得分高低能够准确评估特征与地标的相关性。为此,文献[20]采用下式所示的交叉熵损失对参数进行惩罚:

$$L = -\mathbf{y}^* \cdot \ln \left(\frac{\mathbf{e}^{\mathbf{y}}}{\mathbf{r}^T \mathbf{e}^{\mathbf{y}}} \right) \quad (2)$$

式中, L 为损失值; \mathbf{y}^* 为地标的 one-hot 真值; \mathbf{r} 为 1 的向量。交叉熵损失函数是对地标预测准确性的度量,损失值越小,地标预测准确性越高;反之则准确性越低。因此,模型训练的最终目标是求解交叉熵损失函数最小时的参数 θ ,从而使得分函数与特征语义密切相关。训练过程中,参数通过梯度反向传播不断更新,并逐渐逼近最优值。损失函数对 θ 的梯度公式为:

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \mathbf{y}} \sum_n \frac{\partial \mathbf{y}}{\partial \alpha_n} \frac{\partial \alpha_n}{\partial \theta} = \frac{\partial L}{\partial \mathbf{y}} \sum_n \mathbf{W} \mathbf{f}_n \frac{\partial \alpha_n}{\partial \theta} \quad (3)$$

网络模型经过地标数据集的弱监督训练,即可获得较为准确的得分函数模型,用于计算每个特征的注意力分数。通过设定得分阈值,将高于阈值的特征予以保留,低于阈值的特征予以剔除,可以筛选出与地标高度相关的语义关键点。

无人机视觉导航过程中,飞行高度和卫星影像的分辨率已知,因此通过无人机影像重采样,可显著降低匹配像对的尺度差异;此外,由于网络模型在训练集中能够学习一定的几何不变特征,因此为提高特征提取的效率,本文舍弃了 DELF 的影像金字塔模型设计,只对原始影像提取关键点,用于匹配。

特征图的描述向量由 1 024 个通道特征值组

成,若直接作为关键点的描述符,维度较高,严重影响无人机热红外影像与光学卫星影像的匹配效率;此外,向量中的大多值判别性较低,影响匹配的准确性。为此,本文对 1 024 维的通道向量进行了两次 L2 正则化、一次主成分分析(principal component analysis, PCA)降维处理^[20],得到 128 维的特征描述符作为关键点的最终表达。

2 基于 GAN 模型的训练集增广

采用传统的色彩增广方法扩充训练集,与热红外影像差别较大,用于 DELF 训练,难以学习可见光与热红外影像上的不变特征。为此,本文采用一种基于生成对抗网络(generative adversarial nets, GAN)^[23]的影像转换方法对训练集进行增广,通过热红外地标影像合成,提高训练集的非线性辐射差异,帮助 DELF 学习两种异源影像上的不变特征。

2.1 热红外影像生成模型

目前,基于 GAN 的神经网络模型在人脸合成、图像增强、艺术效果制作等影像转换领域表现出较为理想的效果,如利用成对数据进行训练的 pix2pix^[24]和 BicycleGAN^[25]以及利用非成对数据进行训练的 CycleGAN^[26]等。本文将该影像转换方法用于异源地标数据集的构建当中^[27]。由于对齐的热红外与可见光影像对获取较为困难,因此本文选择 CycleGAN 作为热红外影像的生成模型。

为解决无成对样本监督下的训练问题,文献[26]在设计 CycleGAN 的过程中采用了一种集合层面的监督训练方法和双生成器、双鉴别器的网络结构。假定 X 域中的影像 x 服从分布 $p_{\text{data}}(x)$, Y' 域中的影像 y' 服从分布 $p_{\text{data}}(y')$ 。生成器 $G: x \rightarrow y'$ 以 x 为基础,用于合成与 y' 相似的目标影像;相反,鉴别器 $D_{Y'}$ 负责识别 $G(x)$ 和样本 y' 的真伪,两者成对立关系。据此,构建损失函数为:

$$L_{\text{GAN}}(G, D_{Y'}, X, Y') = E_{y' \sim p_{\text{data}}(y')} [\ln D_{Y'}(y')] + E_{x \sim p_{\text{data}}(x)} [\ln (1 - D_{Y'}(G(x)))] \quad (4)$$

式中, G 通过最小化 $L_{\text{GAN}}(G, D_{Y'}, X, Y')$ 提高合成影像的真实性; $D_{Y'}$ 通过最大化 $L_{\text{GAN}}(G, D_{Y'}, X, Y')$ 提高对影像真伪的鉴别力;用于训练的对抗性目标损失函数为 $\min_G \max_{D_{Y'}} L_{\text{GAN}}(G, D_{Y'}, X, Y')$ 。与之类似,生成器 $F: y' \rightarrow x$ 和鉴别器 D_X 的对抗性目标损失函数为:

$$\begin{aligned} \min_F \max_{D_X} L_{\text{GAN}}(F, D_X, Y', X) = \\ \min_F \max_{D_X} (E_{x \sim p_{\text{data}}(x)} [\ln D_X(x)] + \\ E_{y' \sim p_{\text{data}}(y')} [\ln (1 - D_X(F(y')))]) \end{aligned} \quad (5)$$

仅利用式(4)和式(5)进行对抗性训练,并不能保证一个单独的输入与输出以有意义的方式成对存在,因为满足 $p_{\text{data}}(x)$ 和 $p_{\text{data}}(y')$ 分布的生成器 G 和 F 有无穷多个,单独优化对抗性损失函数常常导致模式崩溃问题。

为进一步缩小映射空间,文献[26]模仿数学中的可逆函数,引入生成器的循环一致特性,即 $F(G(x)) \approx x$, $G(F(y')) \approx y'$ 。据此,构建循环一致性损失函数为:

$$\begin{aligned} L_{\text{cyc}}(G, F) = E_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] + \\ E_{y' \sim p_{\text{data}}(y')} [\|G(F(y')) - y'\|_1] \end{aligned} \quad (6)$$

综合以上 3 个损失函数,得到最终的生成器函数模型为:

$$\begin{aligned} G^*, F^* = \arg \min_{G, F} \max_{D_X, D_{Y'}} L(G, F, D_X, D_{Y'}) = \\ \arg \min_{G, F} \max_{D_X, D_{Y'}} (L_{\text{GAN}}(G, D_{Y'}, X, Y') + \\ L_{\text{GAN}}(F, D_X, Y', X) + \lambda L_{\text{cyc}}(G, F)) \end{aligned} \quad (7)$$

式中, λ 用来控制 $L_{\text{cyc}}(G, F)$ 在训练中的比重。

2.2 异源地标数据集的构建

为提高 DELF 特征的几何不变性,本文使用谷歌在 2019 年 Kaggle 地标识别与检索挑战赛中公开的影像,构建了近景地标数据集(ground landmark dataset_full, GLD_F),其中包含 335 个地标、21 627 张影像。每个地标数据集由不同距离、不同角度拍摄的近景影像组成,因此有助于提高深度局部特征适应尺度、视角变化的能力。此外,GLD_F 中也包含与实例无关的干扰项,因此本文利用训练好的原始 DELF 网络模型对数据集进行了筛选^[28],得到另一个由 12 060 张与地标密切相关的影像组成的数据集(ground landmark dataset_clear, GLD_C)。

为帮助 DELF 网络模型学习遥感影像上的不变特征,本文利用登封市周边 15 km² 范围内、不同时期的谷歌影像,构建了空中地标数据集(aerial landmark dataset, ALD),其中包含 224 个地标、4 032 张影像。由于 ALD 中既包含建筑物等纹理丰富的目标,也包含农田等纹理贫乏的目标,且影像的时间跨度较大,因此有助于提升深度局部特征对纹理和时相变化的适应性。

为学习热红外与可见光影像上的相似不变特征,本文利用热红外影像生成模型将 GLD_F、GLD_C 和 ALD 数据集的光学影像转换为对应的

热红外影像,从而构成异源地标数据集,记为GLD_F_2、GLD_C_2和ALD_2,用于DELf网络模型重新训练。

3 试验分析

3.1 模型训练

本文对热红外影像生成模型和DELf网络模型进行训练。计算机为联想Y7000P,CPU为i7-8750H,显卡为GeForce GTX 1060,内存为16 GB;实现语言为Python,深度学习框架为PyTorch,操作系统为Ubuntu 16.04。

热红外影像生成模型的训练细节如下:训练集包含658张 640×512 像素的热红外影像和698张 640×512 像素的可见光影像,影像输入前进行均匀分割、尺度缩放和随机裁剪等处理;数据输入的批量大小为1,训练集遍历次数为200,前100次的学习率为0.000 2,后100次的学习率线性衰减至0, λ 为10。

DELf模型的训练细节为:首先,训练ResNet50,为提高模型的学习效率和较小样本下的训练效果,使用GLD_C_2和ALD_2数据集对ImageNet数据集上预训练的ResNet50进行微调,提高特征语义的准确性;其次,进行得分训练,训练集为GLD_F_2和ALD_2。两次训练的超参数一致,数据输入的批量大小为8,初始学习率为0.008,训练集遍历次数为50,每隔10次学习率衰减为原来的0.5。

3.2 匹配试验与结果分析

为检验本文匹配方法的效率和鲁棒性,建立了不同场景下的无人机热红外影像与光学卫星影像测试集,并与两种代表性的匹配方法KAZE和D2-Net进行对比,其中D2-Net为训练好的原始网络模型。试验过程中均采用随机采样一致性(random sample consensus, RANSAC)算法剔除误匹配,阈值设为5像素。试验环境与§3.1中的模型训练环境一致。

3.2.1 试验数据

如图2所示,试验数据为3对异源影像,编号依次为 $D_1 \sim D_6$ 。无人机热红外影像为禅思XT2相机在夜间拍摄,有效波长为 $7.5 \sim 13.5 \mu\text{m}$,焦距为25 mm;可见光卫星影像来源于谷歌地图。

表2列出了试验影像的详细信息。其中,像对1为纹理丰富的城市影像,像对2为纹理稀少的农田影像,二者均无明显的几何变换,主要用于测试匹配方法对非线性辐射差异与稀疏纹理

的鲁棒性;像对3为农村地区的影像,分辨率、方向和视角差异较大,主要用于测试匹配方法对复杂几何变换的适应性。

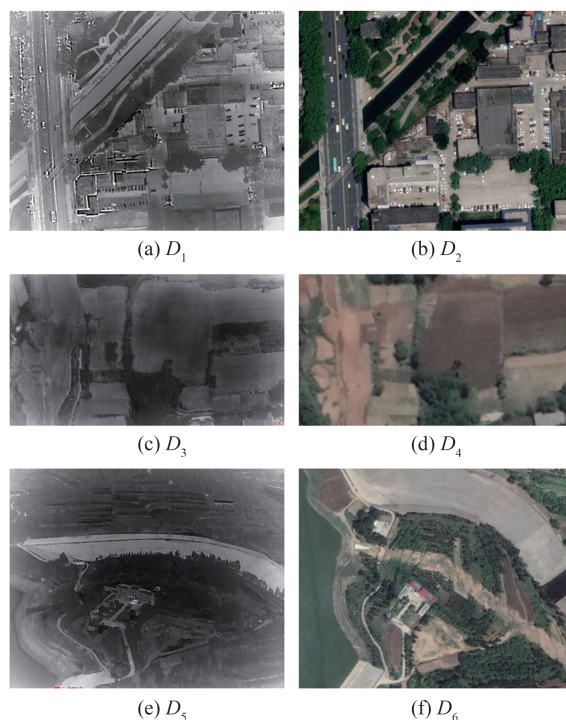


图2 试验影像

Fig.2 Testing Images

表2 异源影像对

Tab.2 Heterogeneous Image Pairs

像对	编号	波段	尺寸/像素	分辨率/m	成像时间	成像地点
1	D_1	热红外	640×512	0.3	2020-03-30	郑州市
	D_2	可见光	640×512	0.3	2019-06-01	
2	D_3	热红外	378×208	0.3	2019-05-16	登封市
	D_4	可见光	378×208	0.3	2018-05-28	
3	D_5	热红外	640×512	—	2019-05-16	登封市
	D_6	可见光	640×512	0.6	2018-05-28	

3.2.2 不同灰度与纹理特征的影像匹配

为测试本文方法用于无人机热红外影像与光学卫星影像匹配时对灰度与纹理变化的鲁棒性,选择几何属性基本一致、具有不同纹理特点的像对1和像对2进行试验。为便于区分,将本文基于异源地标数据集重训练后的DELf记为DELf-VT,DELf则表示训练好的原始网络模型。各方法的匹配结果如图3所示。

图3(a)~3(h)中彩色圆点代表各方法提取的特征,被黄线连接的圆点为RANSAC提纯后的正确匹配点。对比特征点的空间分布可以看出,KAZE提取的特征表现出一定的聚合性,因此在此匹配过程中容易造成相邻特征彼此干扰;而其他3种方法提取的特征都呈现均匀分布的特点,

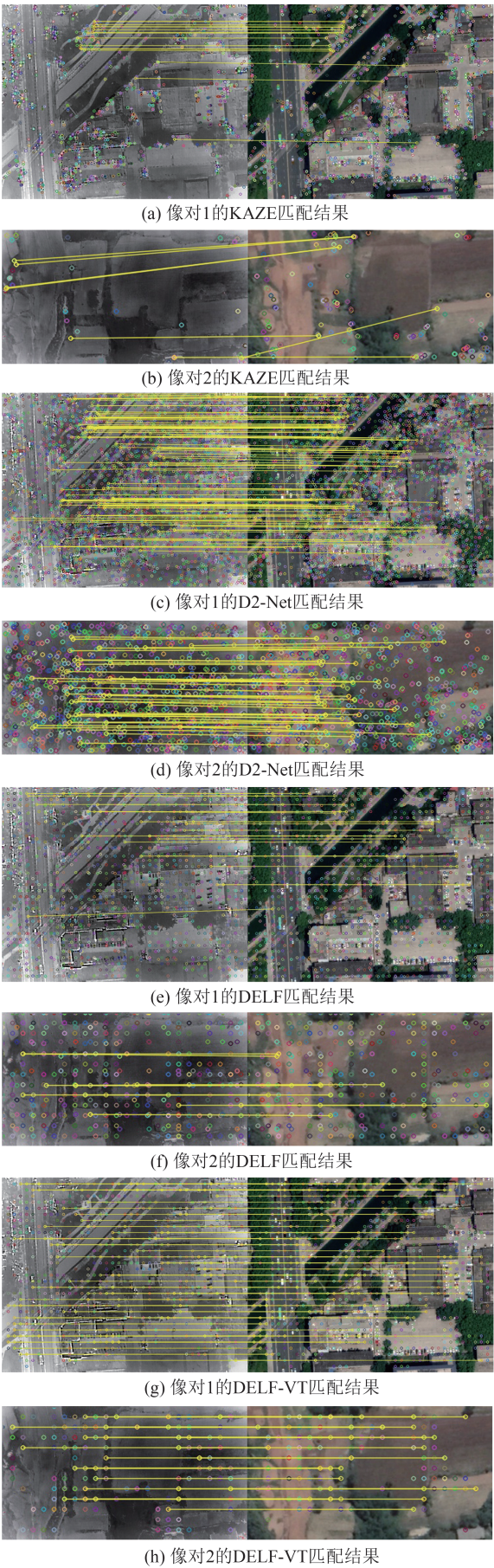


图 3 像对 1、2 的匹配结果

Fig.3 Matching Results of Image Pair 1 and Image Pair 2

因此有利于提高匹配的准确性和几何变换模型的解算精度。

各方法提取的特征数量如表 3 所示,本文计算了可见光影像相对于热红外影像的特征数量变化比例,作为特征稳定性的评价指标之一。由表 3 可知,KAZE 在两种异源影像上提取的特征数量具有显著差异,受影像外观变化的影响较大;而其他 3 种基于深度学习的匹配方法提取的特征数量变化较小,因此受影像非线性辐射差异和纹理退化的影响较小,具有更强的稳定性。

表 3 特征提取数量

Tab.3 Number of Feature Extraction

算法	D_1 和 D_2 特征			D_3 和 D_4 特征		
	D_1	D_2	变化	D_3	D_4	变化
	特征	特征	比例	特征	特征	比例
KAZE	1 300	1 874	0.442	39	143	2.667
D2-Net	4 380	4 231	0.034	1 151	990	0.140
DELF	1 693	1 638	0.032	403	373	0.074
DELF-VT	685	761	0.111	120	104	0.133

统计图 3 各方法的正确匹配点数量如图 4 所示。

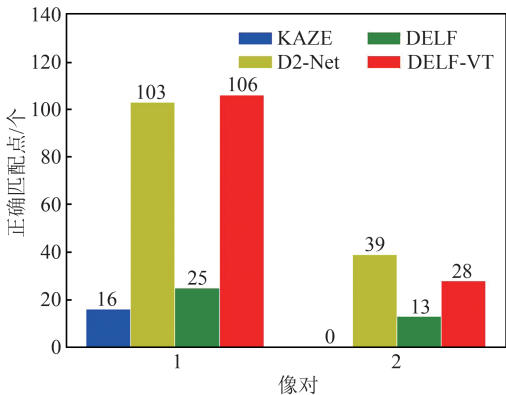


图 4 像对 1、2 的匹配结果统计

Fig.4 Statistics of Matching Results of Image Pair 1 and Image Pair 2

由图 4 结果可知,对于纹理较为丰富、灰度值差异显著的像对 1,各方法均能成功匹配,其中 DELF-VT 的正确匹配点数量最多,且明显多于 DELF,由此证明,本文构建的异源地标数据集可有效提升深度局部特征的灰度不变性。对于纹理较为贫乏、灰度值差异显著的像对 2,KAZE 匹配失败,其他 3 种方法匹配成功,其中 DELF-VT 的正确匹配点数量虽然略少于 D2-Net,但仍多于 DELF,因此再次验证了本文方法设计的合理性。同时值得注意的是,像对 2 的正确匹配点数量明显少于像对 1,由此可见,纹理变化对热红外与可见光影像的匹配具有重大影响。总体上,DELF-VT 和 D2-Net 对灰度与纹理变化的适应性要优

于其他两种方法。

为进一步比较各种特征对灰度与纹理变化的鲁棒性,将特征重复率作为评价指标进行度量,即正确匹配点的数量与2张影像的最小特征点数量的比值。绘制柱形图如图5所示。

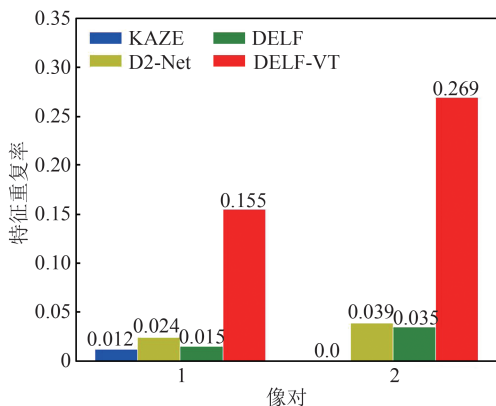


图5 特征重复性测试

Fig.5 Feature Repeatability Test

由图5结果可知,本文方法提取的深度局部特征在热红外与可见光影像上具有更高的重复率,其对灰度与纹理变化的鲁棒性明显优于其他特征。

3.2.3 匹配效率测试

由于无人机热红外影像与光学卫星影像的匹配效率直接关系到视觉导航的位姿优化频率,影响导航精度,因此该指标是评价匹配方法是否适合视觉导航任务的关键。为此,本文统计了各方法的匹配耗时,如表4所示,其中不包含光学卫星影像的特征提取时间,这是由于光学卫星影像的特征在事先构建导航路标库的过程中已经提取,因此并不影响视觉导航的效率。

表4 匹配耗时/s

Tab.4 Matching Time/s

算法	匹配耗时	
	像对1	像对2
KAZE	0.299	0.062
D2-Net	2.159	0.593
DELF	2.315	1.836
DELF-VT	0.286	0.052

由表4可知,DELF-VT的匹配耗时最短,KAZE的匹配耗时略高于DELF-VT,两者对于常规尺寸(640×512 像素)的热红外影像能够达到3 fps的匹配效率,明显快于D2-Net和DELF,因此用于视觉导航中的关键帧位姿解算与优化,基本能够满足需求。尽管DELF-VT和KAZE的匹配效率相当,但前者的鲁棒性更强,正确匹配点更多,因此更适用于无人机视觉导航任务。

3.2.4 不同重叠率的影像匹配

无人机热红外影像与光学卫星影像的重叠率是影响匹配成功率和正确匹配点数量的关键因素之一,直接关系着视觉导航的稳定性。为此,本文测试了匹配方法对影像重叠率的适应性。该试验以 D_1 为基准影像保持不变,通过水平右移 D_2 影像的范围,获取不同重叠率的影像对,其变化步长为0.1。各方法的正确匹配点数量随重叠率变化的曲线如图6所示。

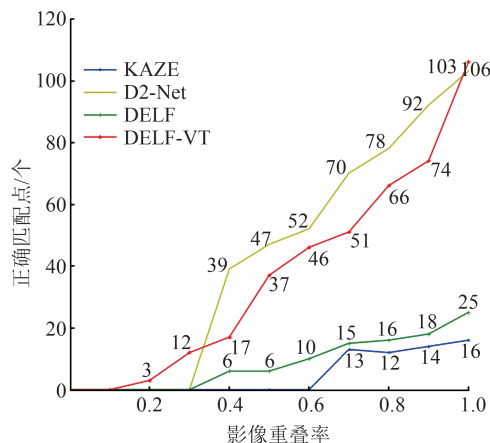


图6 不同重叠率下的正确匹配点数量

Fig.6 Number of Correct Matches Under Different Overlap Rates

从图6可以看出,各方法的正确匹配点数量基本随着影像重叠率的降低而逐渐减少。其中,KAZE虽然在 $[0.7, 1.0]$ 变化缓慢,但在0.6处迅速降为0,这是由于KAZE的正确匹配点分布不均匀导致的,因此其对重叠率变化的适应性有限;DELF和D2-Net对重叠率变化的适应性优于KAZE,能够匹配0.4(含)以上重叠率的影像;DELF-VT虽然在 $[0.4, 0.9]$ 的正确匹配点数量少于D2-Net,但是对0.2(含)以上重叠率的影像均能成功匹配。由此可见,本文方法对影像重叠率的适应范围更广,匹配极端重叠率影像的能力更强,更有助于无人机视觉导航的稳定性。

3.2.5 复杂几何变换下的影像匹配

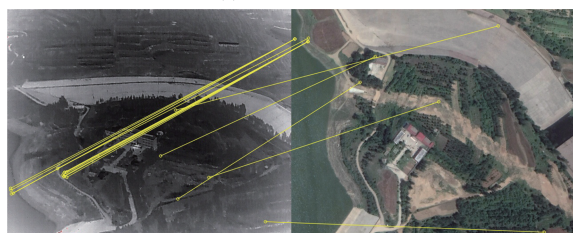
受导航数据精度的制约,预处理后的无人机影像与卫星影像之间可能存在不同程度的几何畸变,因此视觉导航任务要求匹配方法具有一定的几何不变性。为测试各方法对复杂几何变换的适应性,本文选择分辨率、方向和视角差异显著的像对3进行试验,结果如图7所示。

由图7可知,面对复杂的几何变换,KAZE、D2-Net和DELF均匹配失败,只有DELF-VT匹配成功。由此可见,本文方法对无人机热红外影像与光学卫星影像的几何差异具有更强的适应

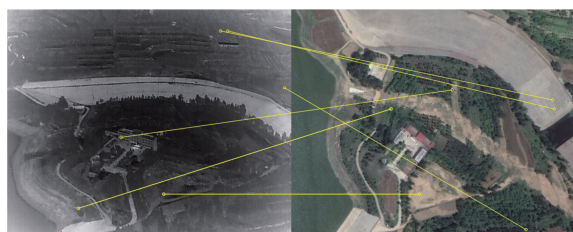
性,更能确保无人机视觉导航的稳定。



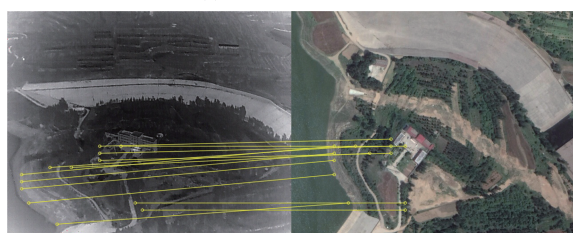
(a) KAZE匹配结果



(b) D2-Net的匹配结果



(c) DELF的匹配结果



(d) DELF-VT的匹配结果

图 7 复杂几何变换下的影像匹配结果

Fig.7 Matching Results Under Complex Geometric Transformation

4 结 语

无人机热红外影像与光学卫星影像的匹配是实现无人机夜间自主绝对定位的关键技术。然而,无人机热红外影像噪声多、对比度低、纹理退化严重,且与光学卫星影像存在显著的非线性辐射差异和几何变换,由此导致两者的匹配及联合应用面临巨大挑战。为此,本文提出一种基于异源地标数据集学习的深度局部特征匹配方法,并在灰度、纹理、效率、重叠率以及几何变换等方面进行了测试,与 KAZE、D2-Net 和 DELF 进行了对比试验。结果表明,本文提出的匹配方法对影像的非线性辐射差异、纹理和重叠率变化以及复杂几何变换具有更强的鲁棒性,且匹配效率较高,适用于无人机视觉导航任务。

此外,本文解决异源影像匹配问题的思路为解决其他影像匹配难题提供了参考,即通过目标影像合成、网络模型学习训练的方法进行处理,可能达到令人满意的效果。

参 考 文 献

- [1] Wan Wenhui, Li Yu, Hu Wenmin, et al. Mobile Platform Localization by Integration of Stereo Cameras, IMU and Wheel Qdometer Based on Federated Filter [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(1): 101-106. (万文辉, 李宇, 胡文敏, 等. 基于联邦滤波进行立体相机/IMU/里程计运动平台组合导航定位[J]. 武汉大学学报(信息科学版), 2018, 43(1): 101-106.)
- [2] Goforth H, Lucey S. GPS-Denied UAV Localization Using Pre-existing Satellite Imagery [C]// International Conference on Robotics and Automation (ICRA), Montreal, Canada, 2019.
- [3] Yol A, Delabarre B, Dame A, et al. Vision-Based Absolute Localization for Unmanned Aerial Vehicles [C]//IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, USA, 2014.
- [4] Shan M, Wang F, Lin F, et al. Google Map Aided Visual Navigation for UAVs in GPS-Denied Environment [C]//IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 2015.
- [5] Ye Yuanxin, Shan Jie, Xiong Jinxin, et al. A Matching Method Combining SIFT and Edge Information for Multi-source Remote Sensing Images [J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(10): 1148-1151. (叶沅鑫, 单杰, 熊金鑫, 等. 一种结合 SIFT 和边缘信息的多源遥感影像匹配方法[J]. 武汉大学学报(信息科学版), 2013, 38(10): 1148-1151.)
- [6] Dou J F, Qin Q, Tu Z M, et al. Infrared and Visible Image Registration Based on SIFT and Sparse Representation [C]// Chinese Control and Decision Conference (CCDC), Yinchuan, China, 2016.
- [7] Liu H, Xiao G F. Remote Sensing Image Registration Based on Improved KAZE and BRIEF Descriptor [EB/OL]. [2020-07-13]. <https://link.springer.com/article/10.1007/s11633-019-1218-3>.
- [8] Lü Qianli, Shao Yongshe. Research on Matching Algorithm for Multi-source Remote Sensing Images Based on SIFT Features [J]. *Computer Engineering and Applications*, 2012, 48(36): 171-176. (吕倩利, 邵永社. 基于 SIFT 特征的异源遥感影像匹配方法研究[J]. 计算机工程与应用, 2012, 48(36):

- 171-176.)
- [9] Huang Q Q, Yang J, Wang C Y, et al. Improved Registration Method for Infrared and Visible Remote Sensing Image Using NSCT and SIFT[C]// IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 2012.
 - [10] Hrkac T, Kalafatic Z, Krapac J. Infrared-Visual Image Registration Based on Corners and Hausdorff Distance[M]. Berlin, Germany: Springer, 2007.
 - [11] Feng Xiaolei, Wu Wei, Li Zhi, et al. Hausdorff Distance Using Gradient Orientation Registers Infrared and Visible Image[J]. *Video Engineering*, 2015, 39(3): 5-10. (冯晓磊, 吴炜, 李智, 等. 利用梯度方向的Hausdorff距离配准红外和可见光图像[J]. 电视技术, 2015, 39(3): 5-10.)
 - [12] Ji Li'e, Yang Fengbao, Wang Zhishe, et al. Visible and Infrared Image Matching Algorithm Based on Edge Image and SURF Features[J]. *Infrared Technology*, 2012, 34(11): 629-635. (纪利娥, 杨风暴, 王志社, 等. 基于边缘图像和SURF特征的可见光与红外图像的匹配算法[J]. 红外技术, 2012, 34(11): 629-635.)
 - [13] Yahyanejad S, Rinner B. A Fast and Mobile System for Registration of Low-Altitude Visual and Thermal Aerial Images Using Multiple Small-Scale UAVs[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 104: 189-202.
 - [14] Liu Chang, Cui Tong, He Chenglong, et al. Infrared-Visual Image Registration Based on High-Curvature Feature Matching[J]. *Command Information System and Technology*, 2016, 7(1): 13-17. (刘畅, 崔桐, 贺成龙, 等. 基于高曲率特征点匹配的红外可见光图像配准[J]. 指挥信息系统与技术, 2016, 7(1): 13-17.)
 - [15] Ye Y X, Shan J, Bruzzone L, et al. Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2941-2958.
 - [16] Ye Y X, Bruzzone L, Shan J, et al. Fast and Robust Matching for Multimodal Remote Sensing Image Registration[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 9059-9070.
 - [17] Yi K M, Trulls E, Lepetit V, et al. LIFT: Learned Invariant Feature Transform[C]//The 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016.
 - [18] Dusmanu M, Rocco I, Pajdla T, et al. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019.
 - [19] Nan Ke, Qi Hua, Ye Yuanxin. A Template Matching Method of Multimodal Remote Sensing Images Based on Deep Convolutional Feature Representation[J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(6): 727-736. (南轲, 齐华, 叶沅鑫. 深度卷积特征表达的多模态遥感影像模板匹配方法[J]. 测绘学报, 2019, 48(6): 727-736.)
 - [20] Noh H, Araujo A, Sim J, et al. Large-Scale Image Retrieval with Attentive Deep Local Features[C]// IEEE International Conference on Computer Vision, Venice, Italy, 2017.
 - [21] Revaud J, Weinzaepfel P, Souza C, et al. R2D2: Repeatable and Reliable Detector and Descriptor[EB/OL]. [2020-06-10]. <https://arxiv.org/abs/1906.06195>.
 - [22] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016.
 - [23] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[EB/OL]. [2020-03-23]. <https://arxiv.org/abs/1406.2661>.
 - [24] Isola P, Zhu J Y, Zhou T H, et al. Image-to-Image Translation with Conditional Adversarial Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017.
 - [25] Zhu J Y, Zhang R, Pathak D, et al. Toward Multimodal Image-to-Image Translation[C]// The 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 2017.
 - [26] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks[C]//IEEE International Conference on Computer Vision, Venice, Italy, 2017.
 - [27] Zhang L C, Gonzalez-Garcia A, Weijer J, et al. Synthetic Data Generation for End-to-End Thermal Infrared Tracking[J]. *IEEE Transactions on Image Processing*, 2019, 28(4): 1837-1850.
 - [28] Ozaki K, Yokoo S. Large-Scale Landmark Retrieval/Recognition Under a Noisy and Diverse Dataset[EB/OL]. [2020-10-20]. <https://arxiv.org/abs/1906.04087>.