



引文格式:秦剑琪,蓝朝桢,崔志祥,等.一种面向无人机绝对定位的卫星基准影像检索方法[J].武汉大学学报(信息科学版), 2023,48(3):368-376.DOI:10.13203/j.whugis.20200229

Citation: QIN Jianqi, LAN Chaozhen, CUI Zhixiang, et al. A Satellite Reference Image Retrieval Method for Unmanned Aerial Vehicle Absolute Positioning[J]. Geomatics and Information Science of Wuhan University, 2023, 48(3): 368-376. DOI: 10.13203/j.whugis.20200229

一种面向无人机绝对定位的卫星基准影像检索方法

秦剑琪^{1,2} 蓝朝桢¹ 崔志祥^{3,1} 张永显¹ 王 岩¹

¹ 信息工程大学地理空间信息学院, 河南 郑州, 450001

² 263611部队, 新疆 库尔勒, 841001

³ 331682部队, 甘肃 兰州, 730020

摘要:针对全球导航卫星系统(global navigation satellite system, GNSS)拒止环境下大范围无人机视觉绝对定位问题,提出了一种聚合深度学习特征的卫星基准影像检索方法。首先,利用预训练的深度学习模型提取无人机与卫星基准影像的局部卷积特征;然后,对局部特征描述符进行聚合,生成影像全局表达;最后,利用影像全局特征进行相似性检索,并采用检索结果精匹配重排序的后处理方法,进一步提高检索准确率。设计了一个新的面向无人机绝对定位的卫星基准影像数据集并进行实验,结果表明,使用所提方法检索无人机影像适配区域的卫星基准影像的准确率达76.07%,可为后续基于视觉的无人机绝对定位提供参考。

关键词:全球导航卫星系统拒止;无人机定位;基准影像;影像检索

中图分类号:P237

文献标识码:A

收稿日期:2020-05-09

DOI:10.13203/j.whugis.20200229

文章编号:1671-8860(2023)03-0368-09

A Satellite Reference Image Retrieval Method for Unmanned Aerial Vehicle Absolute Positioning

QIN Jianqi^{1,2} LAN Chaozhen¹ CUI Zhixiang^{3,1} ZHANG Yongxian¹ WANG Yan¹

¹ Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China

² 263611 Troops, Korla 841001, China

³ 331682 Troops, Lanzhou 730020, China

Abstract: Objectives: In recent years, unmanned aerial vehicle has been widely used and their navigation and positioning rely heavily on global navigation satellite system(GNSS). In the case of GNSS rejection, visual navigation and positioning technology can compensate for this problem, but the technique will also fail to adapt if the approximate location of unmanned aerial vehicle cannot be estimated. To cope with this problem, we propose a reference satellite image retrieval method that aggregates deep learning features to determine the range of unmanned aerial vehicle image adaption region, which can provide reference for the following unmanned aerial vehicle absolute positioning. **Methods:** Firstly, the pre-trained deep learning model is used to extract local convolution features of unmanned aerial vehicle images and satellite images. Secondly, the local aggregation descriptor vector is used to generate the global expression of the images. Finally, the global feature of the image is used to perform similarity retrieval and post-processing method of matching precisely and reranking the retrieval results is used, which further improves the retrieval accuracy. A new satellite reference image data set for absolute positioning of unmanned aerial vehicle is designed and tested.

基金项目:中原科技创新领军人才计划(194200510023)。

第一作者:秦剑琪,硕士,主要从事遥感图像检索研究。179465329@qq.com

通讯作者:蓝朝桢,博士,副教授。lan_cz@163.com

Results: When the queried unmanned aerial vehicle image is similar to the satellite image season in the database, the accuracy of the top 50 candidate images can reach 87.50% using the proposed features for retrieval. Combined with the refined matching re-ranking, the accuracy of the first candidate image can reach up to 76.07%, which satisfies general navigation and positioning applications. **Conclusions:** Although the global descriptor based on deep feature aggregation can effectively represent the images of texture-obvious regions, it is not strong in representing the images of texture-lacking regions and its overlap range between images is high when retrieval is performed. Therefore, the efficiency problem of retrieval and the image feature representation of texture-sparse regions are the directions that need further research.

Key words: GNSS denied; unmanned aerial vehicle positioning; reference image; image retrieval

近年来,无人机广泛应用于电力巡检、环境监测、精确打击等民用和军事领域,发展前景广阔。无人机的诸多应用都离不开定位与导航技术,高精度的定位与导航能力是无人机遂行各项任务的关键前提。当前,无人机的定位与导航严重依赖全球导航卫星系统(global navigation satellite system, GNSS),GNSS虽然定位精度高,但在无线电信号弱或受到恶意干扰时无法使用,容易出现拒止失效情况,即GNSS无法提供准确有效的定位数据。而基于景象匹配的视觉导航定位技术可以解决此类问题,其被动接收地面反射信号^[1]获取信息的方式不依赖外部系统,具有高自主性,成为当下研究热点。

景象匹配首先应用于巡航导弹末制导阶段,后来逐渐发展为一种视觉导航技术^[2]。在无人机定位导航中,基于景象匹配的视觉导航技术通过获取飞行区附近的实时图像与事先存储的基准图像进行匹配,结合基准图的地理位置信息进行无人机的绝对位姿估计,从而实现无人机的定位与导航。卫星基准影像为地面正射投影,其每个像素都包含绝对平面坐标信息,辅助数字高程模型数据便可获得任意位置的三维地理坐标,因此卫星影像可以作为景象匹配的基准图像。卫星基准影像作为事先获取的先验参考,其区域必须覆盖无人机实时飞行区域,以保证其与实时影像的成功匹配。现有的基于视觉的无人机定位方法^[3]通常假设在GNSS信号突然消失的情况下,辅助惯性导航系统可以估计出无人机概略位置,根据此概略位置选取卫星基准影像区域。但在GNSS完全拒止的情况下,无法估计无人机的大略位置,也就无法确定卫星基准影像大致区域范围。针对此问题,本文提出了在大范围的卫星影像数据库中检索无人机影像适配区域影像的方法,以概略确定无人机影像适配区域范围,为后续无人机绝对定位提供参考。

基于内容的图像检索(content-based image

retrieval, CBIR)^[4]技术为此问题的解决提供了契机,其主要利用图像的可视特征,从数据库中检索出具有相似特征的图像。为了实现图像检索,必须提取图像具有辨别力的特征。无人机影像和卫星影像通常包含了各种各样的语义对象以及相似视觉内容,因此,找到具有足够描述性和鲁棒性的特征是图像检索的关键。

传统的基于人工设计的特征描述符,如尺度不变特征变换^[5]等,依赖底层灰度梯度信息,很难在多源影像上提取出重复特征。近年来,深度学习的方法特别是卷积神经网络(convolutional neural network, CNN)^[6]广泛应用。研究表明,利用CNN连续的层,能够揭示出潜在、复杂、更加丰富且具有代表性的深层图像特征,主要包括学习不变特征变换^[7]、自监督兴趣点特征^[8]、欧氏空间判别斑块深度学习特征^[9]、用于大规模图像检索的深度局部特征(deep local feature, DELF)^[10]等。其中,DELF特征专门为图像检索设计,在全球地标数据集检索中取得了很好的效果。但是,目前已有的检索算法和数据集都是面向地面地标数据分类检索任务的,并不适合无人机影像检索卫星影像任务。因此,在面向无人机自主视觉定位任务中,本文通过引入DELF深度学习特征,结合CBIR技术,提出了一种深度学习特征聚合的无人机影像检索卫星基准影像方法;同时,为验证算法性能,专门针对利用无人机影像检索卫星基准影像的定位任务,设计了一个新的数据集。

1 方法原理

1.1 总体思路

为了在卫星基准影像中检索到与无人机影像适配区域的影像,关键在于提取并描述影像具有辨别力的特征,通过比较特征相似性来完成检索。因此,本文提出了一种深度学习特征聚合的

无人机影像检索卫星基准影像方法,其总体思路如图1所示。和传统面向分类的检索不同,本文主要考虑以下几个方面:(1)为解决不同来源影像的特征表示问题,引入了含有注意力机制的卷积神经网络 DELF 模型提取并描述特征点。(2)针对局部描述符对全局信息描述能力不强、检索效率低的问题,采用了描述符聚合方法,先将局部特征描述符聚合生成紧凑的全局表示向量后再执行检索。(3)针对检索准确率不高的问题,通过多张候选影像精匹配重排序的后处理方法,提高检索的准确率。

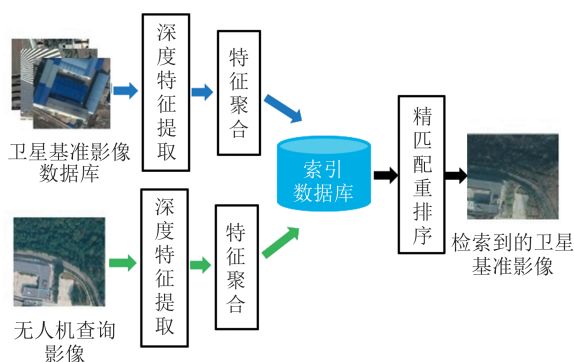


图1 无人机影像检索卫星基准影像流程

Fig.1 Process of Unmanned Aerial Vehicle Image Retrieving Satellite Reference Image

1.2 DELF 特征提取

无人机和卫星影像存在较大的非线性辐射畸变,传统的人工设计的描述符难以准确描述,对此,引入基于注意力机制的 DELF。DELF 选取经典的 ResNet-50^[11]在 ImageNet^[12]上预训练的模型,提取影像深度学习特征,并通过学习得分函数减少背景区域的权重,实现对最显著、最具代表性的区域的局部特征提取。

训练后的注意力机制为每个特征向量 f_n 学习一个得分函数 $\alpha(f_n; \theta)$,其中 θ 是得分函数 $\alpha(\cdot)$ 的参数。为实现该学习,利用了加权求和池化对特征进行处理,其中池化的权重由注意力得分网络获得。训练过程被描述为不断迭代 $f_n \in \mathbf{R}^d$, $n = 1, 2, \dots, N$, 其中 d 表示特征维度并与注意力得分模型一块学习,网络的输出 y 是由特征向量加权求和生成,计算公式为:

$$y = W \left(\sum_n \alpha(f_n; \theta) f_n \right) \quad (1)$$

式中, W 表示 CNN 最终的全连接层权重。

对于训练过程,损失函数 L 采用交叉熵损失的方法,计算公式为:

$$L = -y^* \lg \left(\frac{\exp(y)}{I^T \exp(y)} \right) \quad (2)$$

式中, y^* 对应真值; I 表示单位向量。得分函数 $\alpha(\cdot)$ 的参数通过反向传播算法训练,其梯度计算公式为:

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

式中,输出函数 $\alpha_n = \alpha(f_n; \theta)$ 的反向传播参数 θ 与标准多层感知机参数含义相同。

此外,为应对尺度问题,构建了影像金字塔,对每个尺度的影像独立地进行特征提取,以获得描述不同大小区域的特征。本文构建尺度因子为 $\sqrt{2}$ 的图像金字塔,设置了从 0.25 到 2.0 的尺度范围,分别使用了 0.25、0.356、0.5、0.707 1、1.0、1.414 2、2.0 共 7 种不同的尺度。至此,网络最终产生的为 1 024 维的显著点特征向量,为在特征辨别力和计算复杂度之间取得了良好的平衡,使用主成分分析^[13](principal components analysis, PCA)降低特征向量维度至 40 维。

1.3 VLAD 描述符聚合

在获得 DELF 深层局部特征后,直接使用其进行检索会产生庞大计算量,不利于大规模的无人机影像实时检索卫星影像任务。为此,本文引入描述符聚合方法,先将局部特征描述符聚合生成紧凑的全局表示向量后再执行检索。常用聚合方法有词袋模型^[14](bag of words, BoW)、局部聚合描述符向量^[15](vector of locally aggregated descriptors, VLAD)和费舍尔矢量^[16](Fisher vector, FV)。VLAD 局部聚合描述符相比 FV,其计算量较小,相比 BoW,其码本规模较小,并且能带来更好的检索效果,因此本文使用 VLAD。

VLAD 算法原理是通过 K 均值聚类方法训练一个具有 k 个视觉单词的码本 C_k ,对于每幅图像中的 n 个 d 维局部特征 F ,先找到最近的码本聚类中心,随后所有特征与和它最近的聚类中心的差值做累加,得到 k 个 d 维的残差和向量,串联后生成一个 $k \times d$ 维的 VLAD 向量,表示成 $V = [V_1 V_2 \dots V_k]$ 。经 VLAD 聚合后,用一个向量即可表示影像,检索时直接使用此向量进行比较,而不用对影像中每一个局部特征向量进行比较,提高了检索效率。

1.4 检索结果精匹配重排序

由于光照等差异的影响,结构内容相似的不同区域影像的全局特征距离可能较近,影响检索结果。而无人机影像与其适配区域的卫星影像

间存在重叠区域,两者匹配时,理论上存在更多正确匹配的点对。因此,本文采用特征点精匹配的候选影像重排序方法,候选影像依次与查询影像进行特征点匹配,将正确匹配点对数量作为候选影像得分,根据得分对候选影像重新排序,得到更加准确的检索结果。本文对检索得到的候选影像进行精匹配重排序时,查询影像和候选影像间的匹配通常会产生额外时间开销,影响检索效率。但由于本文所用的卫星影像为分块瓦片,尺寸较小,因此,精匹配重排序可以在较短的时间内完成。

2 面向无人机定位的影像检索

2.1 数据集介绍

面向无人机定位的影像检索任务中,无人机影像和卫星基准影像之间通常存在较大的差异,如尺度、光照、旋转等。针对旋转和视角问题,本文假设使用北方向飞行的单目垂直向下摄像头的无人机影像,其与卫星基准影像方向一致;针对尺度问题,本文假设无人机飞行高度和相机参数已知,进一步得出影像分辨率,经重采样后可生成与卫星基准影像尺度相近的影像。

现有的遥感影像数据集多为解决场景分类问题,其组织方式为以语义类别而非地理位置划分,如 UC Merced 土地利用数据集^[17]、NWPU-RESISC45 数据集^[18]等。另外,基于无人机地理定位的多视角多源基准数据集 University-1652^[19],虽然包含了不同平台(无人机与卫星)影像,但其无人机影像为倾斜视角,同样不能适用于本文任务要求。因此,本文面向无人机定位任务的影像检索,提出了数据集应满足如下条件:(1)卫星基准影像覆盖范围较大;(2)卫星基准影像覆盖区域内包含无人机影像;(3)待检索的无人机影像和卫星影像数据具有明确的对应关系,可作为验证检索准确率的依据。

本文获取了中国河南省登封地区约 117 km² 区域的谷歌卫星影像,获取时间为 2019 年 12 月和 2015 年 7 月,分别属于冬季与夏季,两个季节地面植被覆盖具有较大差异。在该区域中选取了典型地貌区域面积约为 15 km² 的无人机影像,作为查询数据。无人机和卫星影像的区域包含关系如图 2 所示。数据集覆盖区域的地貌类型包括山地、丘陵和平原,地物要素丰富,包括城市、农田、水库、森林、道路等地物要素。

合理有效的图像分块组织对遥感影像检索

是很有必要的^[20]。本文借鉴 Nona-tree 思想^[21],为每张谷歌卫星切片影像增加不同覆盖影像。每 4 张相邻瓦片额外增加 5 张瓦片影像,扩充成新的卫星基准影像数据库,其组织方式如图 3 所示。扩张后的卫星瓦片至少可以覆盖查询瓦片面积的 9/16。



图 2 无人机影像与卫星影像的区域位置关系

Fig.2 Area Location Relationship of Unmanned Aerial Vehicle Image and Satellite Image



图 3 卫星影像切片组织方式

Fig.3 Slice Organization of Satellite Image

本文将获取的卫星影像按谷歌地图瓦片模式进行切片,得到的影像分辨率为 1.2 m、大小为 256×256 像素,如图 4 所示,其中图 4(a)表示无人机影像检索 2015 年 7 月卫星影像(在后文中用①表示),图 4(b)表示无人机影像检索 2019 年 12 月卫星影像(在后文中用②表示)。原始数据集中每个年份包含 1 794 张瓦片,共计 3 588 张,按上述组织方式扩充卫星瓦片影像至 6 924 张。

为屏蔽无人机飞行姿态对影像几何特征的影响,使用预处理好的无人机正射影像,将其随机切片成分辨率为 1.2 m、大小为 256×256 像素的瓦片,如图 5 所示。为便于统计结果,相邻 4 张卫星瓦片影像中包含 2 张无人机切片影像,如图 6 所示,底图为数据库中卫星瓦片影像,红框中为无人机瓦片影像,其位置为随机选取。以此方式

随机切片,共产生280张瓦片,扩充后的数据库中卫星影像瓦片与无人机影像瓦片具有重叠区域,它们的重叠区域介于 $9/16 \sim 1$ 之间。每一张无人机瓦片影像均与其周边的9幅卫星瓦片影像(见图3)有重叠区域,检索时,只要检索到其中一张,即为检索正确。



(a) 无人机影像检索2015年7月卫星影像



(b) 无人机影像检索2019年12月卫星影像

图4 本文数据集中的卫星瓦片影像

Fig.4 Satellite Tile Images in the Data Set



图5 本文数据集中的无人机瓦片影像

Fig.5 Unmanned Aerial Vehicle Tile Images in the Data Set



图6 无人机瓦片影像组织方式

Fig.6 Organization of Unmanned Aerial Vehicle Tile Image

2.2 评价指标

无人机影像检索卫星影像任务中,关注的是检索到的前 K 个候选影像中是否包含检索正确影像,即无人机影像适配区域范围的卫星基准影像。本文引入准确率指标 $P@K$,用于统计算法多次查询的前 K 个候选影像的平均准确率,其含义是 N 次检索中,前 K 个候选影像中包含正确影像的次数除以查询总次数 N 得到的均值。 $P@1$ 即表示多次检索中第一张候选影像为正确影像的概率,即第一张候选影像的准确率。对于无人机定位任务,需要将无人机影像检索到的第一张卫星影像作为后续图像配准的绝对参考。因此,无

人机定位任务最关心的便是第一张候选影像的准确率,即 $P@1$ 的值。

3 无人机影像检索实验结果与分析

针对无人机影像检索卫星基准影像任务,在新的数据集上进行实验。首先对视觉词汇数量对检索结果的影响进行实验;然后考察算法对影像季节变化的适应性,同时验证精匹配重排序策略的可行性;最后,验证本文方法在实时状态下获取的无人机影像上的表现。实验使用配置为2.6 GHz Intel CPU、16 GB RAM、NVIDIA GTX1660Ti的笔记本电脑,实现语言为Python,操作系统为Ubuntu 16.04。

3.1 视觉词汇数量实验

视觉词汇的数量决定了VLAD的维数,在满足较高检索正确率的前提下,应尽可能降低词汇量。因此,在提取影像的DELFT特征并使用暴力方法进行检索的情况下,对视觉词汇数量对检索结果的影响进行了实验,结果如图7所示。

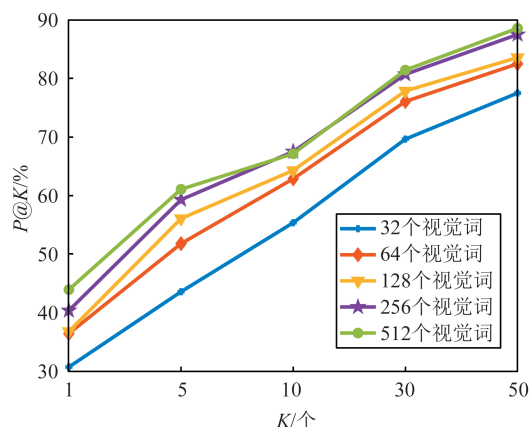


图7 视觉单词数量和 $P@K$ 关系图

Fig.7 Relationship Between the Number of Visual Words and $P@K$

从图7中可以看出,视觉词汇的数量会影响检索准确率,总体而言,随着视觉词汇的增加,准确率在增加,但增幅逐渐减少,尤其当视觉词个数为256时,其 $P@K$ 与512个视觉词的 $P@K$ 愈加接近。主要原因在于对于一个典型区域,地物特征的数量是有限的,因此,词汇量增加到一定数量后,并不增加信息量。相反,VLAD检索的向量维数过高,造成高维向量检索困难,导致检索准确率的下降。因此,本文将视觉词数量确定为256,可以满足一般的地物环境要求。

3.2 检索不同季节卫星影像实验

卫星基准影像所属季节与无人机查询影像

所属季节不一致时,会因植被覆盖的不同造成较大的视觉差异,对这类数据进行实验可以考察算法对影像季节变化的适应性。本文使用无人机影像对两个不同季节的卫星影像进行检索实验,在没有精匹配重排序的情况下,结果见表 1。

表 1 无人机影像检索不同季节卫星影像结果/%
Tab.1 Results of Unmanned Aerial Vehicle Image Retrieving Satellite Image in Different Seasons/%

方 法	$P@1$	$P@5$	$P@10$	$P@30$	$P@50$
①(本文方法)	40.36	59.29	67.50	80.71	87.50
②(本文方法)	31.79	50.71	57.86	73.57	81.43
①(SIFT)	3.93	12.86	17.86	31.43	37.50
②(SIFT)	2.86	8.57	11.43	21.07	28.57

由表 1 中结果可以看出,本文方法与传统 SIFT 算法相比,检索准确率更高。对于不同季节的卫星影像检索,本文方法都表现出比较高的准确率,尤其是当查询的影像与数据库中卫星影像季节相近时, $P@50$ 可达 87.50%。图 8 为部分不同季节的卫星影像检索结果,每幅无人机查询影像对应的第一、二行分别为检索夏季和冬季卫星影像结果,可以看出,对于受季节影响不大且特征明显的人工地物,检索效果较好;对于受季节影响较大的自然地物,如植被覆盖、河流等区域,本文方法也具有一定适应性。

3.3 候选影像精匹配重排序实验

候选影像精匹配重排序可以进一步提高检索性能,为验证其在本文任务中的可行性,对检

索得到的候选影像进行实验,结果见表 2。由实验结果可以看出,对候选影像进行精匹配重排序后, $P@1$ 即第一张候选影像的准确率大大提高。当前 50 个候选影像精匹配重排序时,单张无人机影像检索的 $P@1$ 最高可以达到 76.07%,但是其精匹配相对耗时较长;当前 30 个候选影像精匹配重排序时, $P@1$ 可以达到 73.21%,平均每张查询无人机影像耗时约为 11 s,满足一般的定位与导航应用要求。



图 8 部分检索结果

Fig.8 Partial Retrieval Results

图 9 所示为重排前后结果对比,第一行和第三行为未精匹配结果;第二行和第四行为精匹配后结果。可以看出,精匹配重排序可以进一步提高检索准确率。

表 2 无人机影像检索卫星影像精匹配重排序前、后结果

Tab.2 Drone Images Retrieval Results of Satellite Images Before and After Reranking the Candidate Images with Precise Matching

不同数据集中检索结果	未重排序前 $P@1/\%$	前 10 个重排 序 $P@1/\%$	前 10 个重排 序 平均用时/s	前 30 个重排 序 $P@1/\%$	前 30 个重排 序 平均 用时/s	前 50 个重排 序 $P@1/\%$	前 50 个重排 序 平均 用时/s
①中检索结果	31.79	62.86	3.89	73.21	11.49	76.07	19.01
②中检索结果	40.36	51.43	3.83	61.07	11.10	66.07	18.26

在检索实验中,由于大量检索不正确发生在水面、大片植被覆盖等无特征或弱特征地区,这也是影响检索准确率 $P@K$ 降低的原因之一,如图 10 所示。

3.4 无人机视频影像实验

为验证本文在实时获取状态下的表现,随机截取了 3 组无人机视频影像,如图 11 所示,分别为平原地区、丘陵地区和几何变形影像,它们包含不同的影像纹理结构特征。平原地区影像包含农田与建筑物区域,可以考察算法对结构性明

显的城市建筑区和边缘特征明显的农田的适应性。丘陵地区影像可以考察算法对不规则纹理地区的适应性。为模拟传感器和环境对成像的影响,第一组与第二组影像在采集过程中添加了降质和偏色的模拟干扰因素,以考察算法抗噪声的能力。无人机广角镜头影像可以考察算法对影像几何变形的适应性。

采用本文方法对以上无人机视频影像进行检索实验,由于季节对检索效果有一定影响,因此对于第一组和第二组影像,在冬季数据库中进

行检索;对于第三组影像,在夏季数据库中进行检索,得到的结果如表3所示,其中“正确”的含义

是检索到的卫星基准影像与无人机查询影像属于相同区域范围,影像上存在重叠区域。

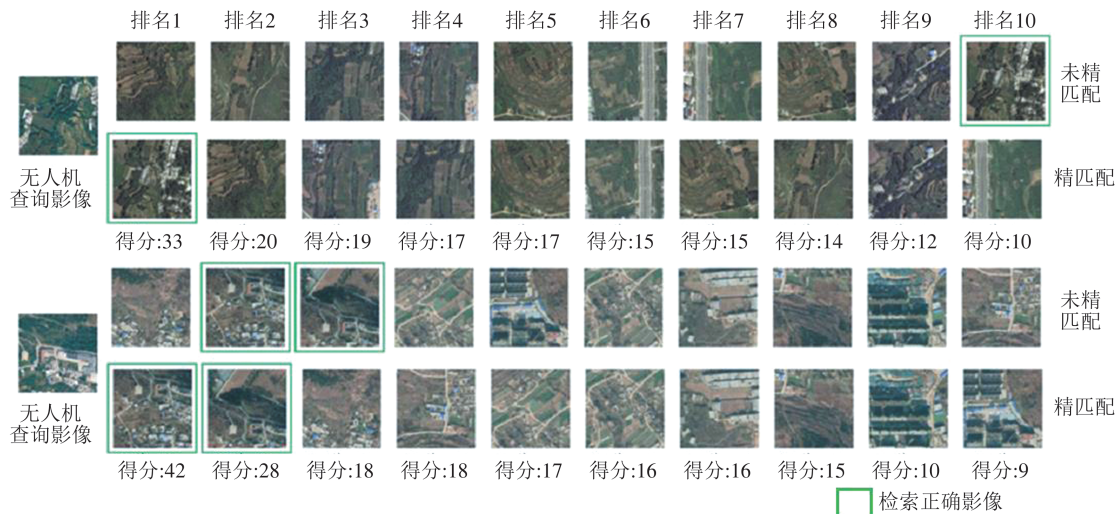


图9 精匹配重排序前、后结果

Fig.9 Results Before and After Reranking with Precise Matching

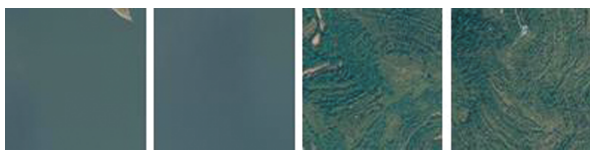


图10 检索失败的区域

Fig.10 Retrieving Failed Areas



图11 无人机视频影像

Fig.11 Unmanned Aerial Vehicle Video Image

由结果分析可知,对第一组和第二组影像,在精匹配重排序前,纹理特征明显的农田和丘陵区域(影像1-1、1-2、2-1、2-2),其排名1的正确率达到75%(影像1-1、1-2、2-2检索正确);对于非密集建筑物区域(影像1-3、2-3)均检索正确;对于密

集建筑物区域(影像1-4、2-4),其效果不佳,但在精匹配重排序后,均得到正确结果。

图12为效果不佳影像精匹配重排序后的结果。对于第三组含有几何形变的影像,本文方法在特征明显的农田区域(影像3-1、3-2)均能正确检索;对于建筑物密集的区域,其检索效果同样不佳,经前30个候选影像精匹配重排序后能得到正确结果(见图12)。实验结果表明,本文方法对实时获取状态下的无人机影像有较好的适应性,尤其是精匹配重排序策略可以进一步提高检索正确率。

4 结 语

针对GNSS完全拒止环境下无人机大范围视觉绝对定位任务中,无法估计大致搜索区域的问题,本文提出的无人机影像检索卫星基准影像的方法,可以较为准确地找到与无人机实时影像适配区域的卫星基准影像,为后续通过影像匹配实现无人机的绝对定位提供范围参考。本文基于深度学习局部特征聚合,生成更具辨别力的影像全局表示特征向量,以此进行相似性检索,并结合精匹配方法对检索到的卫星基准影像进行精匹配重排序,取得了较好的检索效果。

基于深度特征聚合的全局描述符虽然可以有效地表示纹理明显区域影像,但对纹理缺乏区域的影像表示能力不强,检索时,其对影像间的重叠范围要求高。对此问题,本文通过拓展数据库内影像的方式增大目标影像区域被查询影像

覆盖的范围,从而提高检索性能。但检索精度的提高是以存储空间为代价,对于超大范围的影像检索任务,该方法存在效率不高问题。精匹配重排序的后处理方法虽然可以进一步提高检索精度,但其逐个与候选影像匹配的方式同样存在效率问题。在未来的工作中,检索的效率问题和纹理稀少区域的影像特征表示是需要进一步研究的方向。

表 3 无人机视频影像检索卫星影像结果
Tab.3 Results of Unmanned Aerial Vehicle Video Image Retrieval Satellite Image Results

无人机影像	未精匹配前排名 1	前 10 个候选影像重排序后排名 1	前 30 个候选影像重排序后排名 1	无人机影像	前 10 个候选影像重排序后排名 1	前 30 个候选影像重排序后排名 1
1-1	正确			2-3	正确	
1-2	正确			2-4	错误	正确
1-3	正确			3-1	正确	
1-4	错误	正确		3-2	正确	
2-1	错误	正确		3-3	错误	错误
2-2	正确			3-4	错误	错误

注:“正确”表示检索到的卫星基准影像与无人机查询影像属于相同区域范围,影像上存在重叠区域。



图 12 无人机视频影像检索精匹配重排序后结果
Fig.12 Unmanned Aerial Vehicle Video Image Retrieval Results After Precise Matching and Reranking

参 考 文 献

[1] Zhang Zhanmu, Rui Jie. Fundamentals of Remote Sensing Technology [M]. Beijing: Science Press, 2007. (张占睦, 芮杰. 遥感技术基础[M]. 北京: 科学出版社, 2007.)

[2] Zhao Chunhui, Zhou Yihui, Lin Zhao, et al. Review of Scene Matching Visual Navigation for Unmanned Aerial Vehicles [J]. *Scientia Sinica (Informationis)*, 2019, 49(5): 507-519. (赵春晖, 周映慧, 林钊, 等. 无人机景象匹配视觉导航技术综述 [J]. 中国科学: 信息科学, 2019, 49(5): 507-519.)

[3] Goforth H, Lucey S. GPS-Denied UAV Localization Using Pre-existing Satellite Imagery [C]//2019 International Conference on Robotics and Automation, Montreal, Canada, 2019.

[4] Hirata K, Kato T. Query by Visual Example - Content Based Image Retrieval [C]//The 3rd International Conference on Extending Database Technology, Vienna, Austria, 1992.

[5] Lowe D G. Distinctive Image Features from Scale-

- Invariant Keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [6] Alzu'bi A, Amira A, Ramzan N. Semantic Content-Based Image Retrieval: A Comprehensive Study [J]. *Journal of Visual Communication and Image Representation*, 2015, 32: 20-54.
- [7] Yi K M, Trulls E, Lepetit V, et al. LIFT: Learned Invariant Feature Transform [M]//European Conference on Computer Vision, Amsterdam, Netherlands, 2016.
- [8] de Tone 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, Salt Lake City, USA, 2018.
- [9] Tian Y R, Fan B, Wu F C. L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space [C]//IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017.
- [10] 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.
- [11] 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, USA, 2016.
- [12] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [13] Jégou H, Chum O. Negative Evidences and Co-occurrences in Image Retrieval: The Benefit of PCA and Whitening [C]//The 12th European Conference on Computer Vision, Florence, Italy, 2012.
- [14] Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos [C]//The 9th IEEE International Conference on Computer Vision, Nice, France, 2008.
- [15] Jégou H, Douze M, Schmid C, et al. Aggregating Local Descriptors into a Compact Image Representation [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010.
- [16] Daniilidis K, Maragos P, Paragios N. Improving the Fisher Kernel for Large-Scale Image Classification [C]//European Conference on Computer Vision, Heraklion, Greece, 2010.
- [17] Yang Y, Newsam S. Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification [C]//The 18th SIGSpatial International Conference on Advances in Geographic Information Systems, San Jose, California, 2010.
- [18] Cheng G, Han J W, Lu X Q. Remote Sensing Image Scene Classification: Benchmark and State of the Art [J]. *Proceedings of the IEEE*, 2017, 105(10): 1865-1883.
- [19] Zheng Z D, Wei Y C, Yang Y. University-1652: A Multi-view Multi-source Benchmark for Unmanned Aerial Vehicle-Based Geo-localization [C]//The 28th ACM International Conference on Multimedia, Seattle, WA, USA, 2020.
- [20] Li Deren, Ning Xiaogang. A New Image Decomposition Method for Content-Based Remote Sensing Image Retrieval [J]. *Geomatics and Information Science of Wuhan University*, 2006, 31(8): 659-662. (李德仁, 宁晓刚. 一种新的基于内容遥感图像检索的图像分块策略 [J]. 武汉大学学报(信息科学版), 2006, 31(8): 659-662.)
- [21] Remias E, Sheikholeslami G, Zhang A. Block-Oriented Image Decomposition and Retrieval in Image Database Systems [C]//International Workshop on Multimedia Database Management Systems, Blue-Mountain Lake, USA, 2002.