



DOI:10.12404/j.issn.1671-1815.2304805

引用格式:郑博文,王琢,曹昕宇.基于复合型2S网络的红外与可见光图像配准研究[J].科学技术与工程,2024,24(16):6783-6791.

Zheng Bowen, Wang Zhuo, Cao Xinyu. Infrared and visible image registration based on compound 2S network[J]. Science Technology and Engineering, 2024, 24(16): 6783-6791.

基于复合型 2S 网络的红外与可见光图像配准研究

郑博文¹, 王琢^{2*}, 曹昕宇¹

(1. 东北林业大学计算机与控制工程学院, 哈尔滨 150040; 2. 东北林业大学机电工程学院, 哈尔滨 150040)

摘 要 针对传统图像配准方法在红外图像与可见光图像配准任务中效果较差的问题。提出一种基于超级点+超级匹配(Superpoint+Superglue, 2S)复合型网络的特征匹配法用于红外与可见光图像配准。方法中首先使用 Superpoint 独特的特征提取方法,充分提取红外图像与可见光图像之间的共性特征。其次利用 Superglue 特征匹配方法中增加匹配约束和使用注意力机制的思想,发挥神经网络的优势,提高匹配效率。在训练阶段通过使用自建数据集的方法,以提高神经网络的泛化性与准确性。结果表明:传统配准方法在 3 组实验图像上的特征点提取重复性评分与准确性评分分别为:(0.006 7, 0.006 1)、(0.001 0, 0.000 8)、(0, 0),特征点正确匹配对数为:7 对、1 对、0 对,平均数量低于估计变换矩阵所需要的最少 4 对匹配点对。而基于 Superpoint+Superglue 的红外与可见光图像配准方法的各项评分为:(0.240 2, 0.262 5)、(0.193 9, 0.172 2)、(0.263 0, 0.264 4),特征点正确匹配对数为:252 对、165 对、252 对,特征点提取评价指标与特征点对正确匹配数量相较于传统方法均大幅度提升,可以较好地完成配准任务。

关键词 图像配准;卷积神经网络(CNN);特征提取;特征匹配

中图分类号 TP391.41;

文献标志码 A

Infrared and Visible Image Registration Based on Compound 2S Network

ZHENG Bo-wen¹, WANG Zhuo^{2*}, CAO Xin-yu¹

(1. College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China;

2. College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China)

[Abstract] Aiming at the problem that traditional image registration methods have poor effect in infrared and visible image registration tasks. A feature matching method based on Superpoint + Superglue (2S) composite network was proposed for infrared and visible image registration. The method first used Superpoint's unique feature extraction method to fully extract common features between infrared and visible light images. Secondly, the idea of adding matching constraints and using attention mechanism in Superglue feature matching method was used to give full play to the advantages of neural network and improve the matching efficiency. In the training phase, the method of using self built datasets was used to improve the generalization and accuracy of the neural network. The results show that the repeatability and accuracy scores of traditional registration methods for feature point extraction on three sets of experimental images are (0.006 7, 0.006 1), (0.001 0, 0.000 8), and (0, 0), respectively. The correct matching logarithms of feature points are 7 pairs, 1 pair, and 0 pairs, with an average number lower than the minimum four matching point pairs required to estimate the transformation matrix. The scores of infrared and visible image registration methods based on Superpoint + Superglue are (0.240 2, 0.262 5), (0.193 9, 0.172 2), (0.263 0, 0.264 4), and the correct matching logarithms of feature points are 252, 165, and 252 pairs. The evaluation index of feature point extraction and the number of correct matching of feature point pairs are significantly increased compared with traditional methods, which can better complete the registration task.

[Keywords] image registration; convolutional neural network(CNN); feature extraction; feature matching

随着成像传感器技术与计算机视觉处理技术的飞速发展,人们逐渐发现单一传感器的信息局限性太大,已经不能满足目前的工作需求,在成像传感器方面,有可见光相机、红外夜视仪、合成孔径雷达(syn-

thetic aperture radar, SAR)、超光谱成像等设备,可以获得相应的可见光图像、红外图像、SAR 图像、超光谱图像等图像,它们都对某一信息具有突出的优势,又因为各自不同的成像原理与成像特点存在相应的缺

收稿日期:2023-06-28 修订日期:2024-03-19

基金项目:中央高校基本科研业务费专项(2572021BF09);黑龙江省自然科学基金(TD2020C001)

第一作者:郑博文(1999—),男,汉族,黑龙江双鸭山人,硕士研究生。研究方向:图像配准、图像融合。E-mail:zbw609676671@163.com。

*通信作者:王琢(1979—),男,汉族,黑龙江哈尔滨人,硕士,副教授。研究方向:林业工程自动化、检测与测试技术。E-mail:wangzhuo@nefu.edu.cn。

点,而如果能将不同类型图像其进行图像融合,就能获取到复杂环境下的全天高质量图像信息。图像融合是一项贴近生活的实用技术,被应用到多个生活领域^[1],在各类型图像融合中,红外图像与可见光图像由于其各自突出的表现形式,被人们针对研究。

将红外与可见光图像融合^[2]的前提是先将其配准,而由于两组图像的成像机理差异,成型设备不同等原因,配准问题一直难以解决。具体原因在于:以往配准方法中用到的特征提取算子,相似度测量方法都是以目标图片和待配准图片之间是同一种传感器为基础设计的,默认两幅图片间的灰度信息是线性变换的,辐射畸变也是一些简单的模糊,噪声等简单形式的图像降质问题。而红外图像是由红外图像传感器捕捉拍摄场景中物体的红外辐射能量,并通过转换将其变为像素灰度信息生成的,可见光图像是在太阳光或人为光源照射下,目标吸收某波长广播,可见光传感器采集目标表面对可见光波段的反射能量生成的,二者间并非简单的辐射畸变,该问题被归类于多源图像配准问题,当试图解决此类多源图像配准问题时,首先使用在同源图像配准上表现良好的基于特征的特征图像配准方法对应的,但效果并不好。

目前中外学者对与红外与可见光图像配准算法的研究仍然以传统算法与神经网络为主,即基于区域的配准算法与基于特征的特征配准算法^[3]。前者优点在于方法简单直观,但运算量大,运行时间长,无法满足配准工作的实时性;后者主要利用图像中的特征信息,进行提取、匹配计算变换矩阵,较为效率。但针对红外与可见光图像的特定任务确表现欠佳。

为了解决这一问题,提出超级点+超级匹配(Superpoint+Superglue, 2S)相结合的特征匹配法用于红外热成像与可见光图像配准,利用高性能的神经网络应用于图像配准的特征提取与构建描述子这两个步骤,减少图像配准任务反应时间,提升任务准确性。使后续的检测识别等任务更具时效性,具有重要研究意义。

1 传统算法

1.1 图像配准

多源图像配准^[4]中的多源是指图像成像的传感器不同,而图像配准本身则包含着多重视角、多种时态下对相同物体的两幅或多幅图像,建立变换模型,估算其形变参数,以及像素变换参数。使图像中的共有场景信息达到最大场景的匹配。图1为图像配准流程。假设对于一对红外图像和可将图像,称其为目标图像和浮动图像,其图像上某一点 (x, y)

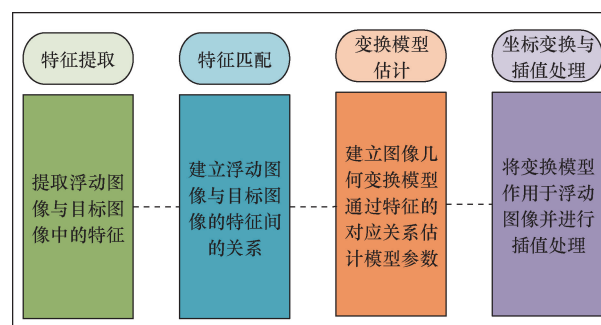


图1 图像配准流程

Fig. 1 Image registration process

的像素值分别为 $I_f(x, y)$ 和 $I_r(x, y)$,则该图像对的变换模型如式(1)所示。

$$I_f(x, y) = G\{I_r[f(x, y)]\} \quad (1)$$

式(1)中: $f(x, y)$ 为图像对的几何变换模型; $G(\cdot)$ 为图像对的像素变换模型。

图像配准的最终任务是利用像素值对应匹配的像素点计算几何对应关系,得到坐标变换矩阵,将浮动图像变换至与目标图像同一位姿,为后续的图像融合等步骤做好前期准备工作。

1.2 配准方法

传统图像配准方法^[5]通过操作步骤中是否涉及特征提取,以及对图像变换模型的处理方式可分为:基于特征的特征配准方法、基于灰度信息的配准方法和基于变换域的配准方法。传统基于特征的特征配准方法中,常用的特征提取方法有:尺度不变特征转换法(scale invariant feature transform, SIFT^[6])、加速稳健特征法(speeded up robust features, SURF^[7])、快速特征点提取和描述法(oriented fast and rotated brief, ORB),表1为3种方法的对比。

在3种传统算法中, SURF算法既有ORB处理速度快的特点,又对图像的旋转和缩放具有一定的不变性。兼顾着速度与性能,因此使用较为广泛。

1.3 方法分析

可见光与红外图像是异源图像,传统的特征配准法虽然可以做到应对灰度图像的线性变换,还有噪声,模糊等简单的辐射畸变形式,但红外与可见光之间的辐射畸变非常严重,而且两种模态图像之间的灰度变化不仅是非线性的^[8],而且是非函数的,非单一映射形式的变化。红外与可见光图像之间特性的差异主要在灰度特性和几何特性两方面,具体来说就是,特征的可重现探测不能保证。以及相似性测度的有效性,合理性不能保证,即匹配策略不合适。

随着神经网络的飞速发展,研究人员发现,其超强的学习能力能够更好的应对不同特征之间差异重复与差异,面对同一特征在不同情况的表现形式,

神经网络通过大量数据的学习,能够做到精准的识别,保证了特征的可重现性,由此基于神经网络的特征提取与匹配方法应运而生,将其应用到图像配准工作上,可以极大提升配准精度。

2 神经网络与图像配准

2.1 神经网络的应用

在卷积神经网络中,独特的多层堆叠组合的结构能够识别到复杂的各类特征,包括特定目标的特征。自2014年以来,图像处理方面的学者们就逐步将神经网络的这一特性应用于图像处理中的特征提取步骤,舍弃原有的传统SIFT方法以及其他类似的算法并初具成效。相较于传统方法,神经网络可以减小不同图像的辐射畸变带来的影响,在保证特征提取准确性的前提下,增加特征匹配时的约束项,获得更好的匹配点对。

将深度学习应用在图像配准领域,也为图像配准任务带来了新的思路和方法^[9]。将神经网络作

用于特征提取与特征匹配工作,可以自动地学习和捕捉图像中的特征和变换关系,再通过固定的单应性矩阵估计公式得到变换矩阵,从而实现更加精确和鲁棒的图像配准。

2.2 复合型2S网络

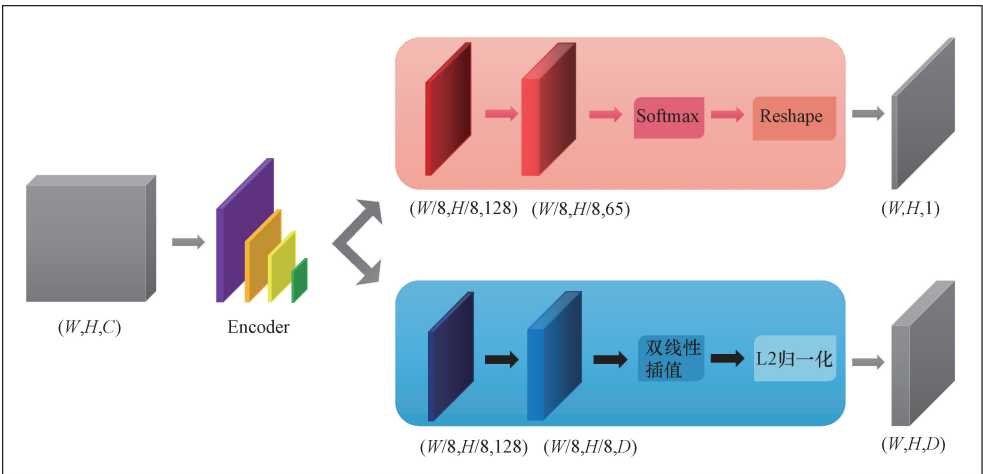
2.2.1 Superpoint 网络

Superpoint^[10]是一个双分支卷积神经网络,可以提取图像中的特征点并构造描述子,其网络结构示意图如图2所示,其网络结构特点如下。

(1)共享编码层(图2结构中有两个网络,但实际上两个网络的前半部分是公用的)。网络的共享部分使用了一种视觉几何组神经网络(visual geometry group,VGG)的下采样结构作为编码器降低图像尺寸,这部分结构由卷积层、最大池化层和非线性激活层组成,通过3个池化层将图像尺寸变为输出的1/8。以往使用神经网络进行特征提取时都是对特征提取和特征匹配分布进行训练,也就是只能单独训练特征点或者描述子的一种,不能用同一个网络

表1 传统方法对比
Table 1 Comparison of traditional methods

方法	SIFT	SURF	ORB
特征检测方法	高斯差分金字塔和尺度空间极值点检测方法来检测极值点;方向梯度直方图HOG描述符来描述关键点的特征	高斯卷积和Hessian矩阵来提取关键点;使用加速的HOG描述符来描述关键点的特征	角点检测算法提取特征点;Brief描述符特征关键点的特征
特征匹配方法	基于最近邻(nearest neighbor)策略的匹配方法	使用暴力匹配法计算特征向量间的欧式距离	利用暴力匹配法,计算特征点之间的汉明距离(Hamming distance)
优点	对于旋转、尺度、光照变化等具有很好的鲁棒性(旋转和尺度方面比SURF好);能够处理非刚性变换	对于旋转、尺度、光照变化等具有很好的鲁棒性(模糊和亮度比SIFT好);速度比SIFT快	计算速度比SIFT和SURF算法都更快;对于图像中存在大量重复或相似区域的情况下,具有较好的鲁棒性
缺点	计算量大,运行速度慢;存在大量重复或相似区域的情况下,可能会出现误匹配的情况	SURF算法的特征描述符可能存在方向性不够明显的问题	鲁棒性不如SIFT和SURF算法;不能处理非刚性变换



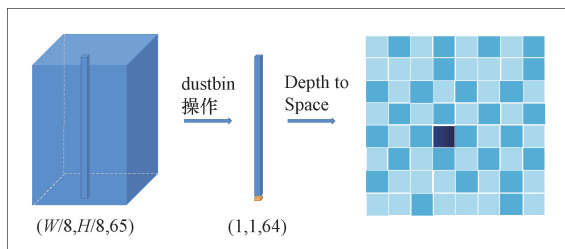
Encoder 为编码器;Softmax 为归一化指数函数;Reshape 为重塑尺度;W 为特征图宽度;H 为特征图高度;
C 为原特征图通道数;D 为描述子特征图通道数

图2 Superpoint 网络结构图
Fig. 2 Superpoint network structure

进行联合训练。这种一前一后的训练方式在特定复杂任务上表现不好,运算资源的浪费,而且网络参数荣誉,训练时间较长。

(2) 特征点解码端。特征点解码端通过卷积神经网络将经过编码的图像 $\left(\frac{W}{8}, \frac{H}{8}, 128\right)$ 处理为符合输出特征维数的大小 $\left(\frac{W}{8}, \frac{H}{8}, 65\right)$, 65 的含义是每个像素点及其所在的各个通道的同一位置共 65 个像素值, 这其中包括对应的 8×8 区域中存在特征点的概率, 与额外的非特征点 dustbin。dustbin 点的加入可以应对不存在特征点的情况, 若不设置 dustbin 点, 当某一 8×8 区域中不存在特征点时, 会出现随机扰动作用于次区域, 以至于后续特征点选择时会将本不是特征点的点, 认定为特征点; 若设置 dustbin 点, 区域中无特征点时, 会将 dustbin 通道的响应值置于最大, 而后抛弃 dustbin 点, 则不会对原区域造成影响。在通道维度上进行 softmax 操作后, 将非特征点去除。之后经过 depth to space 操作, 如图 3 所示。每个像素经过该解码器后会输出一个概率值, 即该点为特征的可能性。最后通过 nms 操作, 即非极大值抑制, 得到响应值较大的位置, 作为特征点。

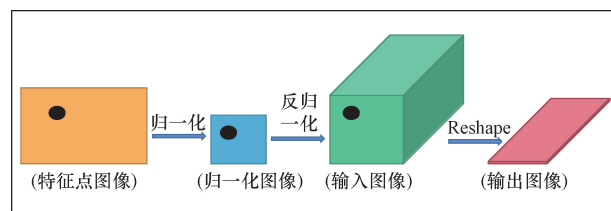
(3) 描述子解码端。通过一个类似于通用对应网络的结构得到描述子, 即将经过编码的图像 $\left(\frac{W}{8}, \frac{H}{8}, 128\right)$ 处理为符合输出特征维数的大小 $\left(\frac{W}{8}, \frac{H}{8}, 256\right)$, 256 是要输出的特征向量的维数。之后对特征点图像进行归一化处理(图 4), 以及根据输入图像大小 $\left(\frac{W}{8}, \frac{H}{8}\right)$ 进行反归一化, 找到对应 $\left(\frac{W}{8}, \frac{H}{8}\right)$ 图像上的特征点位置, 这一步需要插值处理、L2 正则化以及 reshape, 得到大小找到对应 $\left(\frac{W}{8}, \frac{H}{8}\right)$ 图像上的特征点位置, 这一步需要插值处理、L2 正则化以及 reshape, 得到大小为 $(W, H, 256)$ 的特征向量。



W 为特征图宽度; H 为特征图高度; dustbin 为垃圾点;
Depth to Space 为数据及维度转移

图 3 特征点提取示意图

Fig. 3 Feature point extraction diagram



Reshape 为重塑尺度

图 4 归一化操作示意图

Fig. 4 Normalized operation diagram

(4) 损失函数。Superpoint 的损失函数 L 如式(2)所示, 包括特征点检测损失函数 L_p 以及描述子检测损失函数 L_d ,

$$L(X, X', D, D', Y, Y', S) = L_p(X, Y) + L_p(X', Y') + \lambda L_d(D, D', S) \quad (2)$$

式(2)中: X, Y, D 分别为可见光图像中提取的特征点图、特征点真实标签图、描述子; X', Y', D' 分别为红外图像中提取的特征点图、特征点真实标签图、描述子; S 为指示矩阵。

特征点检测采用交叉熵损失函数, 如式(3)所示。

$$L_p(X, Y) = \frac{1}{H_c W_c} \sum_{h,w} \left[-\lg \left(\frac{\exp x_{hwy}}{\sum_{k=1}^{65} \exp x_{hwk}} \right) \right] \quad (3)$$

式(3)中: $H_c = \frac{H}{8}$, 其中 H 为特征图高度; $W_c = \frac{W}{8}$, 其中 W 为特征图宽度; h, w 为元素在选定通道上的横纵坐标位置; x_{hwy} 为选定通道中的各个元素; x_{hwk} 为选定的某一通道。

求出每一条单通道中元素 x_{hwy} 为特征点的概率, 即 softmax 之后的值, 由于规定每 8×8 范围内有一个特征点, 所以整张图像中最多含 $(H \times W)/64$ 有个特征点。

描述子损失函数如式(4)所示。

$$L_d(D, D', S) = \frac{1}{(H_c W_c)^2} \sum_{h,w} \sum_{h',w'} l_d(d_{hw}, d'_{h'w'}, s_{hwh'w'}) \quad (4)$$

式(4)中: D 和 D' 分别为可见光与红外图像中的描述子单元; h, w 为可见光图像中描述子所在坐标点; h', w' 为红外图像中描述子所在坐标点; l_d 为合页函数, 可以保证解的稀疏性, 增强模型的效率及泛化能力; $s_{hwh'w'}$ 为指示函数

$$l_d(d, d', s) = \lambda_d \text{smax}(0, m_p - d^T d') + (1 - s) \max(0, d^T d' - m_n) \quad (5)$$

式(5)中: s 为正确匹配的集合; m_p 为正样本阈值; m_n 为负样本阈值; d 为可见光图像描述子; d' 为红外图像描述子; λ_d 为平衡系数, 进行数量级的平衡。

该损失函数常用于用于学习两个相似度度量

之间的距离,当描述子余项相似度 $d^T d'$ 大于 m_p 时不需要惩罚,反之收到惩罚,即最终目标是尽可能提高描述子的相似程度。

2.2.2 Superglue 网络

Superglue^[11] 是一个特征匹配网络,能够将相似的特征进行匹配并且将不存在匹配点的点滤除,使用时首先将红外与可见光图像输入 Superpoint 网络提取的特征点以及描述子,再将结果直接输入 Superglue 网络,得到目标图像对各特征点之间的匹配关系。网络的核心由注意力图神经网络(graph neural networks, GNN)以及最优匹配层构成。其网络结构特点如下。

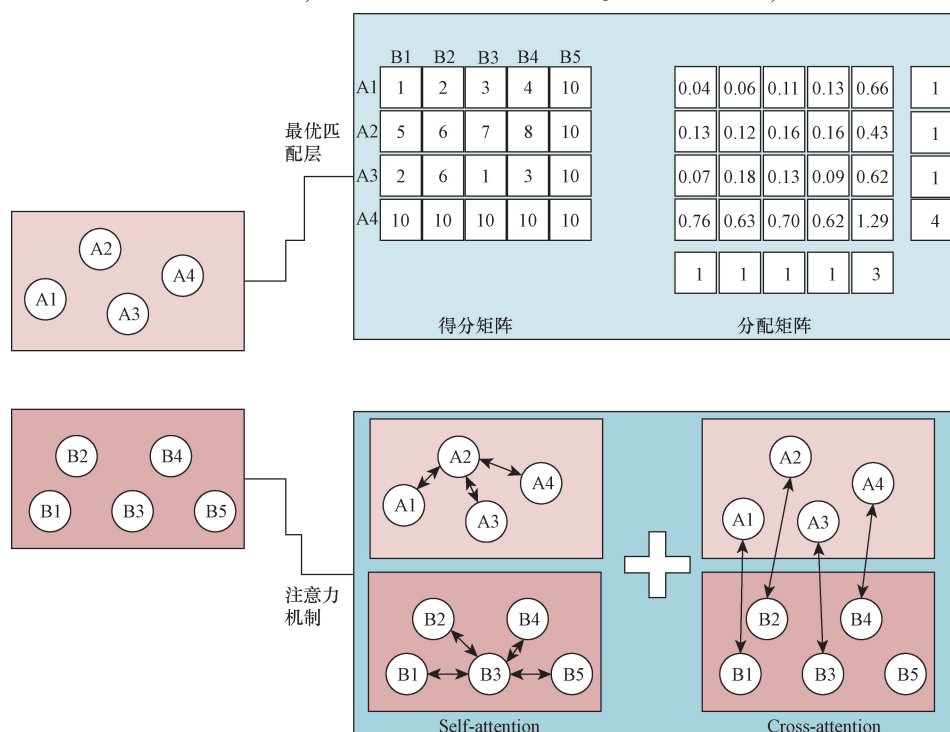
(1)特征点编码器。传统的特征匹配方法只考虑计算特征点周围的梯度值等作为描述子进行匹配,约束项太少导致匹配时无法精准的分辨因图像辐射等因素导致的描述子误匹配,特别是在多源图像特征匹配时表现很差。考虑将低维位置信息作为约束条件引入,与描述子合并后多用层感知机(multi-layer perceptron, MLP)进行升维得到 X_i ,充分利用特征点的位置信息和外观信息。为后续的注意力机制做好准备工作。

(2)注意力 GNN。注意力机制是研究人员在研究人类视觉时受到启发而得到一种思路,人类浏览两个图像“找茬”的过程是有科学性的,具体表现为

人眼来回扫描两幅图像,查找相似点与不同点(如果是相同的点,观察一下该点的周围有没有存在差异的点,直到找到“茬点”/或继续寻找)。在这个过程中人们通过主动寻找上下文来增加特异性,以此找到一些具有奇异性的“茬点”。Superpoint 网络的核心就是在 GNN 中添加了注意力机制,即模拟了人类的“找不同”的过程,只不过变成了“找相同”。在得到特征表示 X_i 后,利用 Self-attention 以及 Cross-attention 来回增强 X_i 的特征匹配性能;然后计算特征向量的内积,得到得分矩阵,再使用 Sinkhorn 算法迭代数次后得到最后匹配得分矩阵。这一步骤在最优匹配层中进行。

(3)最优匹配层。特征的匹配问题属于二次分配问题,求解这类问题需要复杂不切实际的算子。实际处理时学者将其化简为线性分配问题,但只是简单的搭建了一个线性模型,相比之下 SuperGlue 网络利用其深度卷积层,设计了一种更复杂的代价进行求解。

(4)Sinkhorn 算法与分配矩阵 P 。Sinkhorn 算法是对分配矩阵进行迭代的一种方法,用来求解最优传输矩阵; P 的每一行代表来自图 A 中的某个特征点对应图 B 的 N 种匹配的可能性。每一列代表来自图 B 中的某个特征点对应图 A 的 M 种匹配的可能性。如图 5 所示,Self-attention 更加关注在一幅



A1 ~ A4 为可见光图像中的特征点;B1 ~ B5 为红外图像中的特征点;Self-attention 为自注意力机制;
Cross-attention 为跨注意力机制

图5 最优匹配层与注意力机制

Fig.5 Optimal matching layer and attention mechanism

图像中与其他特征点相比更具有特异性的所有点; Cross-attention 则更关注在两幅图像中的外观,即上下位置更相似的点对。

(5) 损失函数。由于 GNN 与最优匹配层均可进行反向传播,且网络训练采用的是有监督训练,即存在真值如式(6)所示。

$$M = \{(i, j)\} \subset A \times B \quad (6)$$

式(6)中: M 为匹配真值即匹配的特征点对; i 为可见光图像中特征点; j 为红外图像中特征点; A 为可见光图像; B 为红外图像。

没有匹配的点 $L \in A, J \in B$ 。可采用负对数似然函数作为损失函数 L_{oss} 进行最小化分配矩阵,如式(7)所示。

$$L_{\text{oss}} = - \sum_{(i,j) \in M} \lg \bar{P}_{i,j} - \sum_{i \in L} \lg \bar{P}_{i,N+1} - \sum_{j \in J} \lg \bar{P}_{M+1,j} \quad (7)$$

式(7)中: $\bar{P}_{i,j}$ 为分配矩阵; $\bar{P}_{i,N+1}$ 、 $\bar{P}_{M+1,j}$ 分别为对分配矩阵进行列求和以及行求和之后构建的新矩阵。

最终目标是最大化精度及召回率。

2.3 复合 2S 型网络的具体应用

2.3.1 共性特征的提取与匹配

对传统特征提取算法进行分析时提出重现性的概念,即特征的可重现探测^[12]。传统方法虽然既能从可见光图像,也能从红外图像上提取到很多的特征点,但点对之间无法匹配,究其原因是在红外与可见光图像提取到的特征点的描述符之间没有建立起联系导致的。

图 6 为传统方法的特征点提取图像,图像中提取出来的特征点数量较多但仅有几对可以建立起正确的联系,有些点对是因为图像的辐射信息导致



(a) 特征点提取结果图像



(b) 特征点匹配结果图像

图 6 特征提取与特征匹配示意图

Fig. 6 Feature extraction and feature matching diagram

位置出现误差,无法匹配,这一问题将由 Superpoint 网络的独特提取方法解决。而有些点对在两幅图中的位置基本相同,只是由于两个特征点的描述子之间没有建立起对应的关系,这一问题将通过 Superglue 的特殊匹配方式解决。

2.3.2 训练方法

选择使用自建数据集进行训练,自建数据集部分图像如图 7 所示,该数据集由直线、三角形、椭圆、矩形、高斯模糊和噪声等构成的共几万张图像组成。训练思路为:以往监督学习^[13]中输入特征点的位置是人为标定的,即人类认为该点是最具有特征最关注的点,但实际上并不是,人类视觉最关注的点不一定是计算机视觉上能够识别为特征点的点,而对于线段、三角形等标准图像,其特征点位置是确定的,如线段的特征点是两个端点、三角形的特征点是 3 个角点,这样得到的数据集是具有真实的标签信息的,以此来监督训练^[14]。

经过自建数据集训练的网络已经对与特征提取具有良好表现了,再使用 COCO 数据集,进行进一步训练,结果显示,对于形状规矩的图像,网络表现较好,但对于不规则的,目标信息复杂的图像,会出现误识别的现象,针对这一现象使用 Homographic Adaptation 方法,如图 8 所示,对于一张 COCO 数据集的照片,对该图像生成随机单应性矩阵,进行多种主动单应性变换,即单应性矩阵是已知数据,对每一种变换后的图像进行特征提取得到特征点位置,再对特征点施加各自单应性矩阵的逆变换并进行整合,得到这张图像的特征点标签,重复这一过程,进行有监督训练,即可解决该问题^[15]。

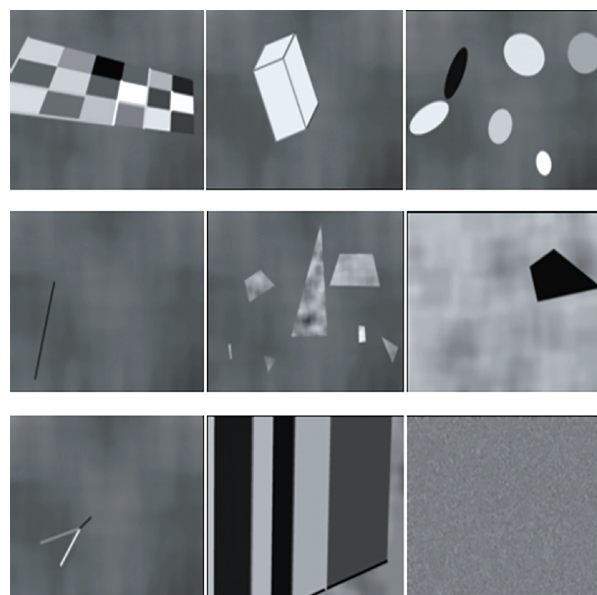
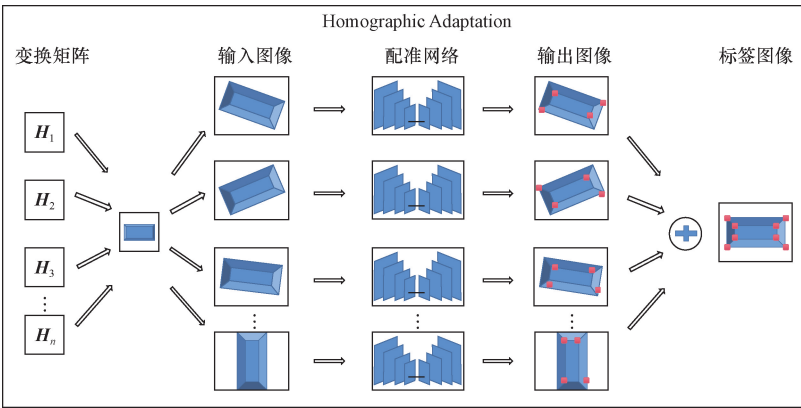


图 7 自建数据集图像

Fig. 7 Self-built data set images



Homographic Adaptation 为单应性变换; $H_1 \sim H_n$ 为自定义变换矩阵

图 8 单应性变换示意图

Fig. 8 Schematic diagram of a homologous transformation

2.3.3 进阶训练

针对红外与可见光图像,为了更好提取共性特征,设计进一步的训练方案。使用已经由科研人员配准好的红外与可见光数据集 RGB-NIR。首先对其中的可见光图像进行特征点提取,此外不优先选择红外图像,原因在于:之前训练的COCO数据集是可见光图像,由于数据集中的图像对是配准好的,因此得到的特征点坐标也是对应于红外图像中特征点的坐标,在之前的 COCO 数据集训练得到的权重的基础上,把坐标标签与红外图像输入网络进行有监督的特征训练得到新的权重。这样得到的网络既具有提取可见光图像特征点的能力,又具有提取红外图像特征点的能力,且由于使用的是可见光图像的特征点进行监督学习,这些点的位置是在两图像中相同的,即达到提取共性特征的目的。

2.3.4 训练结果分析

选择与传统方法进行对比来证明复合型 2S 方法的先进性,选择传统 SURF,改进的 SURF,即增加快速最近邻上搜索(fast library for approximate nearest neighbors,FLANN) + K 邻近匹配算法(K-nearest neighbor batch,KNNbatch) + 随机抽样一致(random sample consensus,RANSAC)等改进辅助匹配方法的作为实验对照组,并且从 Road-scene-infrared-visible-images-master 数据集中提取红外与可见光图像对作为实验对象,该数据集由道路、海洋、森林等包含多目标信息的户外场景可见光与红外图像对组成,从数据集中选择 3 组具有光照噪音明显,复杂树木背景,可识别目标较多的特点的图像对。实验内容为提取红外与可见光图像对间的共性特征并进行特征匹配,使用匹配后的特征点对进行单应性矩阵估计,并作用于红外图像,最终输出原图像对的特征点匹配情况与变换后红外图像,共 3 张图像。表 1 为

整理传统 SURF、改进 SURF、本文方法在 3 组实验图片上的表现数据,包括特征点提取数、特征点匹配数、可重复评分(Repeatability)、准确性评分(Accuracy)等,可以看出,本文方法并不是通过增加提取到的特征点基数来提高匹配点对数的,而是提取出更具有共性特征的特征点,运用合适的匹配方法,达到目的,体现的本文方法的优秀性能。

从主观视觉感受与客观评价指标两方面衡量配准性能。主观感受方面,从图 9 可以看出,特征匹配性能远远优于传统方法,匹配点对数量多且正确率高。评价指标方面,以可重复性评分、准确性评分、正确率作为评价标准,具体参数如表 2 所示。

可重复性评分(Repeatability,记为 R_{epe})定义为可见光与红外图像对中检测到的共有特征点数与可见光图像中检测到的总特征点数量的比值,计算如式(8)所示。可重复性指标结果如图 10 所示。

$$R_{epe} = \frac{\{P \mid P \in I_R \cap R_{GB}\}}{\{P \mid P \in R_{GB}\}} \tag{8}$$

式(8)中: P 为特征点个数; I_R 为可见光图像中检测到的特征点; R_{GB} 为红外图像中检测到的特征点。

表 1 对比实验结果				
Table 1 Comparative experimental results				
分组	特征点提取情况	SURF	SURF(KNNbatch) + flann + RANSAC	Superpoint + Superglue
组 1	R_{GB}	1 050	1 050	1 049
	I_R	1 145	1 145	960
	匹配成功点对	7	10	252
组 2	R_{GB}	991	991	851
	I_R	1 189	1 189	958
	匹配成功点对	1	1	165
组 3	R_{GB}	429	429	958
	I_R	1 192	1 192	953
	匹配成功点对	0	2	252

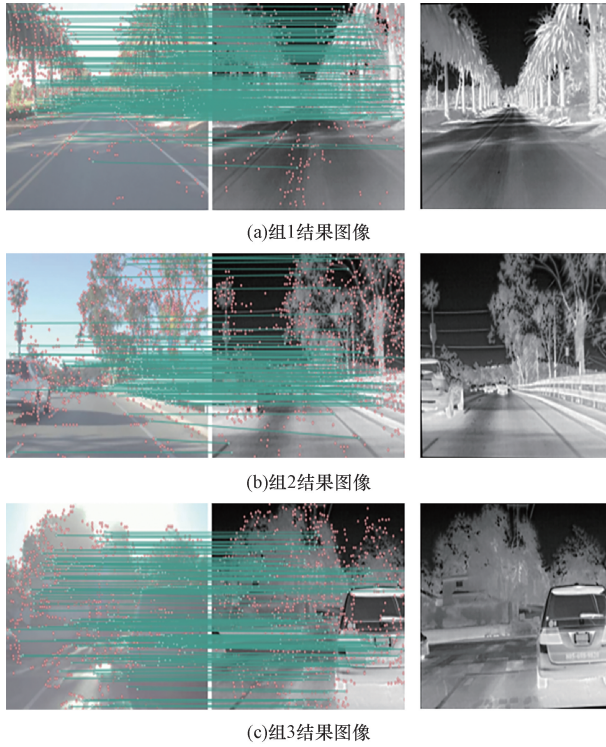


图9 实验结果图像

Fig. 9 Experimental result image

表2 指标参数
Table 2 Index parameter

分组	评价指标/%	SURF	SURF(KNNbatch) + Superpoint + FLANN + RANSAC	Superpoint + Superglue
组1	R_{epe}	0.006 7	0.009 5	0.240 2
	A_{cc}	0.006 1	0.008 7	0.262 5
	C_R	0.025 5	1.000 0	1.000 0
组2	R_{epe}	0.001 0	0.001 0	0.193 9
	A_{cc}	0.000 8	0.000 8	0.172 2
	C_R	0.003 8	0.250 0	0.981 8
组3	R_{epe}	0	0.004 7	0.263 0
	A_{cc}	0	0.001 7	0.264 4
	C_R	0	0.160 0	0.992 1

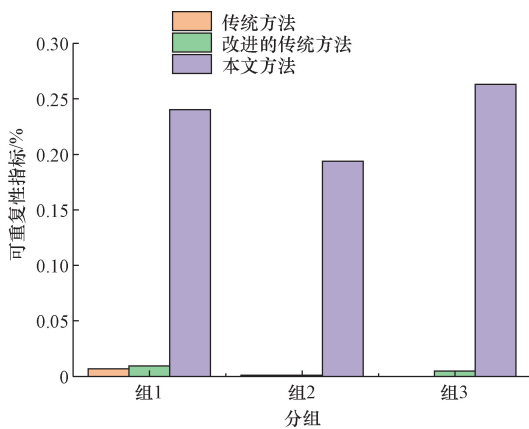


图10 重复性指标

Fig. 10 Repeatability index

准确性评分 (Accuracy, 记为 A_{cc}) 定义为可见光与红外图像对中检测到的共有特征点个数与红外图像中检测到的总特征点数量的比值, 计算公式如式(9)所示。准确性指标结果如图 11 所示。

$$A_{cc} = \frac{\{P | P \in I_R \cap R_{GB}\}}{\{P | P \in I_R\}} \quad (9)$$

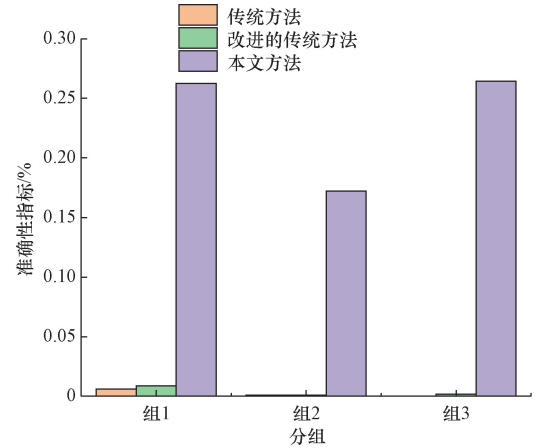


图11 准确性指标

Fig. 11 Accuracy index

正确率评分 C_R 定义为特征点匹配结果中正确匹配点对的数量 (correct matches, 记为 C_M) 与匹配算法建立的总匹配点对数量 (correspondences, 记为 C_{orr}) 的比值, 其计算公式为

$$C_R = \frac{C_M}{C_{orr}} \times 100\% \quad (10)$$

3 结论

(1) 针对红外与可见光图像使用传统方法难以实现配准的问题, 提出一种 Superpoint + Superglue 复合式神经网络配准方法。所提方法使用 Superpoint 网络替换传统方法中的特征提取算子, 实现了图像配准任务中的特征提取。利用增加 dustbin 点, 使用 depth to space, 归一化与反归一化等操作, 解决了以往传统方法针对红外与可见光图像无法提取共性特征、特征数量提取不足等问题。通过复合 Superglue 网络替换传统方法中的暴力匹配策略, 完成特征匹配步骤。利用网络独特的注意力机制以及图匹配策略增加对应点匹配的约束项, 兼顾特征匹配正确率与时间成本。通过采用双分支的网络结构, 进一步提高效率。

(2) 实验证明, 使用本文方法可以完成传统方法无法完成的红外与可见光图像的配准任务, 且各项指标都远高于传统方法, 为后续的图像融合等相关研究奠定了基础。后续研究方向可以增加剪枝、量化等轻量化方法, 或增加预降噪处理等环节, 继续优化网络的参数, 增强网络实用性。

参 考 文 献

- [1] 韩阳, 杨华. 基于非下采样剪切波变换与显著信息加权的图像融合算法[J]. 科学技术与工程, 2021, 21(17): 7224-7229.
Han Yang, Yang Hua. A image fusion algorithm based on nonsubsampling shearlet transform coupled with significant information weighting [J]. Science Technology and Engineering, 2021, 21(17): 7224-7229.
- [2] 白永见, 熊淑华, 吴小强, 等. 基于 FDST 和 MSS 的红外与可见光图像融合[J]. 科学技术与工程, 2017, 17(6): 215-219.
Bai Yongjian, Xiong Shuhua, Wu Xiaoqiang, et al. Infrared and visible images fusion based on FDST and MSS[J]. Science Technology and Engineering, 2017, 17(6): 215-219.
- [3] 徐海洋, 赵伟, 刘建业. 基于边缘结构特征的红外与可见光图像配准算法[J]. 红外技术, 2023, 45(8): 858-862.
Xu Haiyang, Zhao Wei, Liu Jianye. Registration algorithm of infrared and visible light image based on edge structure features [J]. Infrared, 2023, 45(8): 858-862.
- [4] Maes F, Collignon A, Vandermeulen D, et al. Multi-modality image registration maximization of mutual information[C]//Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis. San Francisco: IEEE, 1996: 14-22.
- [5] Barbara Zitová, Flusser J. Image registration methods: a survey [J]. Image and Vision Computing, 2003, 21(11): 977-1000.
- [6] Liu X J, Yang J, Sun J W, et al. Image registration approach based on SIFT[J]. Infrared and Laser Engineering, 2008, 37(1): 156-160.
- [7] Zhang R J, Zhang J Q, Yang C. Image registration approach based on SURF[J]. Infrared and Laser Engineering, 2009, 38(1): 160-165.
- [8] Ni G Q, Li Q. Analysis and prospect of multi-source image registration techniques[J]. Opto-electronic Engineering, 2004, 31(9): 295-302.
- [9] 柳静. 基于模糊数学的对抗网路图像配准建模[J]. 科学技术与工程, 2019, 19(24): 242-247.
Liu Jing. Modeling and analysis of confrontation network image registration based on fuzzy mathematics[J]. Science Technology and Engineering, 2019, 19(24): 242-247.
- [10] Detone D, Malisiewicz T, Rabinovich A. Superpoint: self-supervised interest point detection and description[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2018, 6: 337-349.
- [11] Sarlin P, Detone D, Malisiewicz T, et al. Superglue: learning feature matching with graph neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 4937-4946.
- [12] 李俊廷. 基于共性特征的红外与可见光图像配准算法研究与应用[D]. 成都: 电子科技大学, 2021.
Li Junting. Research and application of infrared and visible image registration algorithm based on common features[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [13] Dosovitskiy A, Fischer P, Springenberg J T, et al. Discriminative unsupervised feature learning with exemplar convolutional neural networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 38(9): 1734-1747.
- [14] Yang Z Q, Dan T T, Yang Y. Multi-temporal remote sensing image registration using deep convolutional features [J]. IEEE Access, 2018, 6: 38544-38555.
- [15] 龙勇志. 红外与可见光图像配准与融合算法研究[D]. 成都: 电子科技大学, 2020.
Long Yongzhi. Research on registration and fusion algorithm of infrared and visible images[D]. Chengdu: University of Electronic Science and Technology of China, 2020.