

基于参数回归的快速全景图像拼接算法

郭璠, 李小虎, 刘文韬, 唐璠
(中南大学自动化学院, 湖南 长沙 410083)

摘 要: 现实场景中照相机获得的图像视场角范围往往是有限的, 而目前对全景图像的需求日益增大, 因此针对拍摄得到的全景图像序列, 提出了一种基于参数回归的快速全景图像拼接算法。将传统的图像配准任务转化为深度学习结合机器学习的方式, 设计一种基于高斯差分金字塔的多尺度深度卷积神经网络 (MDCNN) 对待拼接图像进行特征提取, 并使用 LightGBM 回归模型对拼接参数进行预测, 获得图像之间的变换矩阵和照相机焦距完成图像对齐, 并设计了一种双曲线图像融合算法消除图像之间的拼接缝。实验结果表明, 所提算法能够实现图像的快速拼接, 获得比已有代表性算法更清晰自然的全景拼接效果, 同时对红外图像也具有很好的适应性。

关键词: 图像拼接; 全景图像; 特征提取; 参数回归; 图像融合

中图分类号: TP391.41

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023182

Fast panoramic image stitching algorithm based on parameter regression

GUO Fan, LI Xiaohu, LIU Wentao, TANG Jin
School of Automation, Central South University, Changsha 410083, China

Abstract: In reality, the field of view of images acquired by cameras was usually limited, and the demand for panoramic images was increasing. Therefore, a fast panoramic image stitching algorithm based on parameter regression was proposed for panoramic image sequences. The traditional image registration task was transformed into deep learning combined with machine learning, a multi-scale deep convolutional neural network (MDCNN) based on Gaussian difference pyramid was designed to extract features of stitching images, and LightGBM regression model was used to predict stitching parameters. The transformation matrix and the focal length of the camera were obtained to align the images, and a hyperbolic image fusion algorithm was designed to eliminate the stitching seam between the images. The experimental results show that the proposed algorithm can quickly mosaic images and obtain clearer and more natural panoramic mosaic effects than the existing representative algorithms. It also has good adaptability for infrared images.

Keywords: image stitching, panoramic image, feature extraction, parameter regression, image fusion

0 引言

为了克服单幅图像的视场角限制, 很多实际应用需要通过全景图像拼接技术快速生成 360°全景图像。目前的图像拼接算法通常包括图像预处理、

图像配准、图像对齐、图像融合等过程, 其中图像配准是整个拼接流程的关键, 是完成后续步骤的基础。尤其对于红外图像来说, 由于红外图像中的特征点往往较少, 难以找到匹配的特征点, 因此图像配准难度较大, 很大程度上会导致拼接失败。同时

收稿日期: 2023-06-08; 修回日期: 2023-09-05

基金项目: 国家自然科学基金资助项目 (No.61502537); 长沙市自然科学基金资助项目 (No.kq2208286); 湖南省自然科学基金资助项目 (No.2023JJ30697)

Foundation Items: The National Natural Science Foundation of China (No.61502537), Changsha Natural Science Foundation (No.kq2208286), The Natural Science Foundation of Hunan Province (No.2023JJ30697)

由于需要对每幅图像都进行特征提取和匹配，这往往需要消耗大量的空间和时间，因此算法抗干扰能力较弱，实时性也较差。

图像配准主要包括特征点的提取和匹配，通过图像配准可计算图像之间的位置变换关系。其原理是基于图像的显著特征完成特征点的提取，并将其表达为特征描述符，然后利用特征描述符的相似性进行关键点的匹配，最后根据匹配点的位置关系来建立两幅图像之间的几何变换关系。在特征点提取的研究中，最先被提出的是 Harris 角点检测算法^[1]，它的核心思想是基于图像局部灰度的变化差异来检测角点。近年来，为了实现多传感器遥感图像的快速自动配准，Shu^[2]结合函数拟合方法对尺度不变特征变换 (SIFT) 算法进行改进和优化，能够有效提高遥感图像配准方法的精度和运行效率。

图像对齐是指根据图像配准得到的图像之间的变换关系，将多幅图像进行映射，构成一幅完整图像的过程。对齐的目的是消除图像拼接时可能出现的不连续、不匹配或重叠等问题，从而使图像拼接的结果更加自然和真实。早期的图像拼接研究往往假设照相机仅做平移和旋转运动，因此可通过全局单应性变换模型来完成图像拼接。Brown 等^[3]提出了经典的全景图像拼接算法，并设计了一款软件 AutoStitch。Zaragoza 等^[4]提出的 APAP (as-projective-as-possible) 算法将图像划分为多个网格，然后分别计算每个网格的局部单应性矩阵。APAP 通过对每个网格进行局部扭曲来对齐重叠区域，但非重叠区域也往往会被影响，产生较严重的畸变。为此，Chang 等^[5]提出了 SPHP (shape-preserving half-projective) 方法，该方法通过将图像划分为重叠区域、过渡区域和非重叠区域，并分别利用不同的变换方式，从而尽可能减少畸变。Lin 等^[6]提出了 AANAP (adaptive-as-natural-as-possible) 算法，它通过结合全局相似性和局部优化的方式，可以有效消除伪影和扭曲，产生高质量的全景图像。Chen 等^[7]提出了一种基于全局相似性先验的图像拼接 (NISWGSP) 方法，其核心思想是利用局部扭曲模型来引导每个网格的变形，通过设计特定的目标函数来尽可能减少扭曲的失真，使拼接后的全景图像看起来较自然。曾国奇等^[8]利用 SURF (speeded up robust feature) 特征描述对连续图像进行匹配，并结合随机抽样一致性算法计算单应性变换矩阵。在此基础上，经过球形变换

校正后利用单应性完成图像拼接。Zhu 等^[9]提出了一种基于渐进变换矩阵的多图像拼接算法 ISGTA (image stitching gradual transformation algorithm) 以消除匹配过程中的形变扭曲。Khamiyev 等^[10]采用多种基于深度神经网络的方式来挖掘单应性矩阵的特征相关性以最终生成全景图。Dai 等^[11]提出了一个边缘引导合成网络 (EGCNet, edge guided composition network)，EGCNet 利用基于卷积神经网络 (CNN) 的 HED (holistically-nested edge detection) 算法来进行显式的边缘引导，给边缘变化的区域分配较大的权重，从而避免出现结构不一致的伪影。

图像融合作为图像拼接的后处理算法，可以有效消除对齐后的拼接缝、重影，使图像呈现出更加自然的效果。最简单的方法是使用直接平均法，即对重叠区域中每个像素取均值。在此基础上，又逐渐演化出了加权平均融合算法。受拉普拉斯金字塔融合算法的启发，Zhang 等^[12]提出了一种密集连接的多流融合网络，能够有效地融合来自不同尺度的前景和背景图像信息。Lu 等^[13]提出了一种双向内容的迁移模块，通过循环神经网络对中间区域的特征表示进行条件预测，除了保证融合过程中的空间和语义一致性外，还采用上下文注意力机制和对抗学习机制来提高融合图像的视觉质量。

总体而言，目前关于全景图像拼接的大部分研究往往集中于传统图像配准算法，导致算法鲁棒性较差。在实际拍摄环境中，由于受到图像噪声、光照、曝光度、模型误差等因素的影响，拼接结果往往会出现鬼影、变形、配准、对齐失败等情况。另外，相关传统图像配准算法的实时性也较差，很难在实际工程系统中得到应用。而目前大部分基于深度学习的拼接算法主要针对两幅图像，针对多幅图像的全景拼接工作尚不多见。因此，迫切需要提出一种快速鲁棒且针对多幅图像的全景图像拼接算法。

1 本文算法

1.1 算法框架

针对上述问题，本文提出了一种基于参数回归的快速全景图像拼接算法，整体框架如图 1 所示。由图 1 可知，本文算法将特征提取和匹配的计算过程，即传统拼接算法图像配准步骤，使用深度学习结合机器学习的技术变成一个黑箱操作，从而简化了计算过程，提高了配准的效率，使算法的实时性

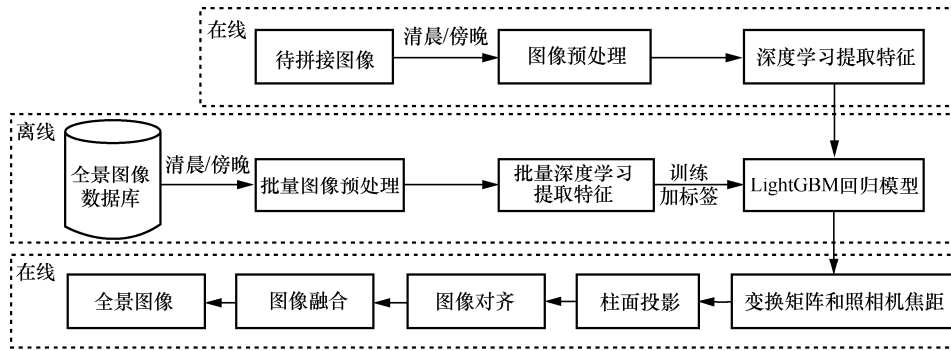


图 1 全景图像拼接算法整体框架

得到了很好的保证。

整个算法分为在线和离线 2 个部分，离线部分主要进行回归模型的训练，回归模型的输出为变换矩阵和照相机焦距。在线部分主要进行图像的预处理和拼接后处理等步骤。其中，图像预处理主要进行图像滤波和光照补偿，拼接后处理部分包括柱面投影、图像对齐和图像融合等过程。在全景图像拼接中，变换矩阵的精确度直接影响图像对齐，它是影响图像拼接效果的关键。为了使模型预测出的变换矩阵参数尽可能准确，本文提出一种多尺度的深度卷积神经网络进行特征提取，并联合 LightGBM 回归模型完成拼接参数的回归学习。

1.2 特征提取网络

针对全景图像拼接场景，本文设计出了一种基于高斯差分金字塔的多尺度深度卷积神经网络 (MDCNN, multi-scale deep convolution neural net-

work)，如图 2 所示。

由图 2 可知，该网络由一个基本骨架 backbone、衍生图 (derived map) 和高斯差分金字塔模块 (DoGPM, difference of Gaussian pyramid module) 组成。backbone 的输入为原图的 7 个衍生图，共包含 6 个 block, block₁ 由卷积层 (Conv)、批归一化层 (BN) 和激活函数 (ReLU) 组成，其余 block 包含一个最大池化层和 2 个卷积层，并且每个 block 的输出会合并 DoGPM 的输出，网络最终会输出一个长度为 25 088 位的一维特征向量。考虑到 VGG (visual geometry group) 网络在特征提取方面的性能，MDCNN 的 backbone 部分和 VGG16 类似，为了在训练过程中加速网络的收敛，对 backbone 中的部分模块使用了 VGG16 的预训练参数。

1) 衍生图

MDCNN 首先会对输入的图像进行不同操作以

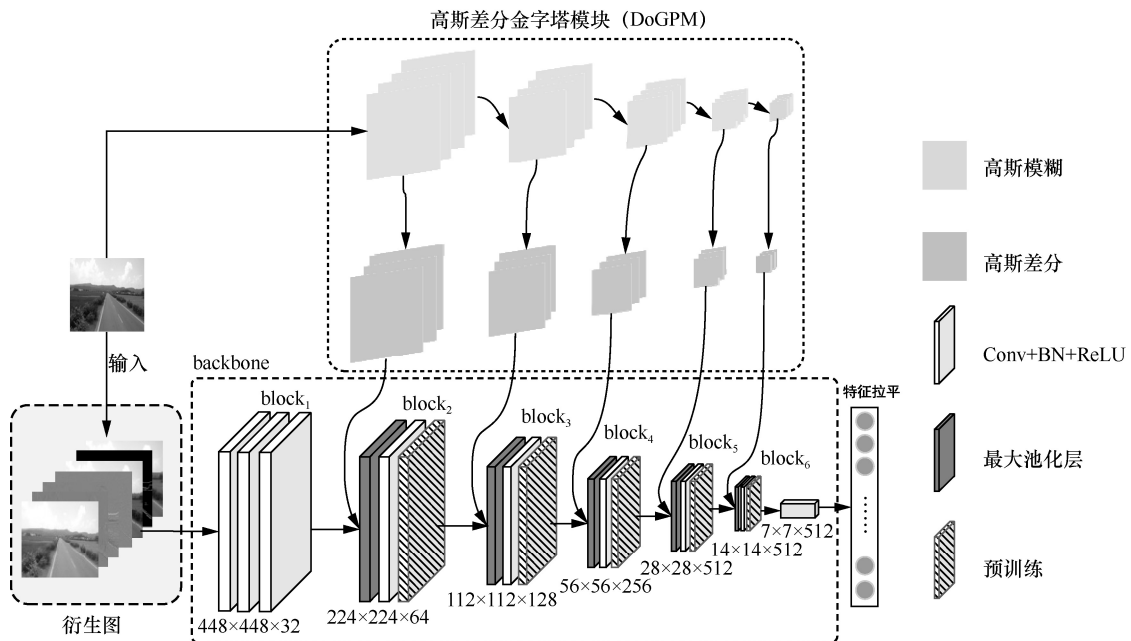
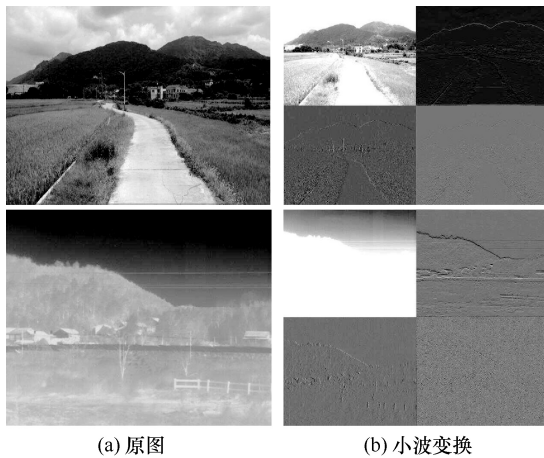


图 2 多尺度深度卷积神经网络结构

获得衍生图，从而利用衍生图来从多个维度获得输入图像的特征，这些图像处理操作包括双边滤波、伽马矫正、边缘检测以及小波变换。通过以上操作可以从多角度获取图像信息，有利于 backbone 更好地提取图像特征。其中，滤波操作可以更好地剔除图像中的噪声，本文主要采用双边滤波^[14]进行图像去噪处理。尤其对于包含色彩信息较少的红外图像，双边滤波能有效去除噪声。其次，使用伽马矫正对图像进行非线性变换，可以改善图像的亮度信息，防止出现局部过明或过暗的情况。经过矫正后的图像能够有效防止光线的影响，在一定程度上保证光照一致性。除了考虑图像的空间域信息，频域信息也值得关注。为此，衍生图还利用了图像小波变换的结果。小波变换是时间（空间）频率的局部化分析，它通过伸缩平移运算对信号（函数）逐步进行多尺度细化，最终达到高频处时间细分、低频处频率细分，从而可聚焦到图像的任意细节，尽可能放大图像中的关键信息，如图3所示。



(a) 原图 (b) 小波变换
图3 可见光图像和红外图像的小波变换

图像的边缘是指图像中局部区域亮度变化显著的部分，反映了灰度值在相邻2个位置的阶跃。这部分往往集中了图像的重要特征，对图像场景的识别和理解十分重要。因此，本文采用 Canny 边缘检测算法的结果作为其中一个衍生图。

2) 高斯差分金字塔模块

为了保留图像的多尺度信息，可通过构建图像金字塔的方式来实现。图像金字塔底部分辨率最高，顶部分辨率最低。本文使用高斯金字塔模型模拟肉眼近距离观察和远距离观察图像的差别。构建图像金字塔可以保证图像特征的尺度不变性。本文所构建的高斯金字塔共5组，每组4层，共20层。

图像金字塔中，每幅图像用 $L(x,y,\sigma)$ 表示，即

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y) \quad (1)$$

其中， $G(x,y,\sigma)$ 为高斯函数； $I(x,y)$ 为输入图像； σ 为高斯函数的标准差，即模糊系数； \otimes 表示卷积操作。每组由4层尺寸相同但模糊系数不同的采样图像组成，模糊系数 σ 的计算式为

$$\sigma(o,r) = \sigma_0 2^{\frac{o+r}{s}} \quad (2)$$

其中， o 为组的索引序号， $o \in [0,1,2,3,4]$ ， r 为每组长层的索引序号， $r \in [0,1,2,3]$ ， s 为高斯金字塔中每组的层数， σ_0 为高斯模糊系数初始值，设置为1.6。仅进行高斯模糊并不能使图像中的关键点信息显露出来，需要进行高斯差分计算出尺度空间的极值，也就是对相邻两层的高斯模糊图像作差，构建高斯差分金字塔。将构建出的高斯差分金字塔依次并入 backbone 每个 block 的池化层之前，从而将图像的多尺度信息并入网络中进行挖掘，以提高网络特征提取的能力。

衍生图和高斯差分金字塔模块的结果均来自对输入图像进行不同的处理操作，将金字塔的不同层级和衍生图作为模型的输入，以获得更具丰富性的特征表示。这种输入方式允许模型同时考虑不同尺度和维度的信息，能够更好地捕捉图像的细节和上下文关系，有利于网络更好地提取特征。

1.3 特征融合策略

对于全景图像拼接场景，由于需要拼接多幅图像，使用 MDCNN 对每幅图像完成特征提取后，需要对多组图像特征进行融合。对于网络输出的特征向量，若直接将每组特征向量依次拼接在一起，这种粗略的数据融合方式没有考虑到图像之间的位置关系，同时存在大量的数据冗余，不利于后续回归模型的学习。因此，需要结合拼接过程中的具体特性，设计相应的特征融合策略。

在实际的 360° 全景拍摄场景下，以图4为例，如采集到 A、B、C 三幅图像，每相邻图像之间均包含重叠区域，A、B、C 三幅图像经过 MDCNN 得到3个特征向量 V_A 、 V_B 、 V_C 。在数据融合时，需要考虑图像之间的位置关系，若将融合后的特征向量直接合并为 $V_A V_B V_C$ ，则忽略了图像 A 和 C 的位置关系。为了解决此问题，以牺牲存储空间为代价，选择将特征向量排列 $V_A V_B V_C V_A$ ，通过重复引入特征向量 V_A ，将图像之间所有的重叠关系均表达出来。

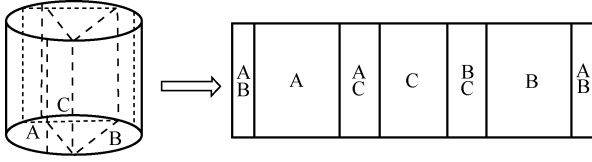


图 4 柱面展开示意

在上述位置关系的基础上, 为了实现细粒度的特征数据融合, 需要对特征向量中每个位置的特征值按照上述融合策略依次进行合并, 对于包含 N 幅图像的特征向量, 特征融合后的结果如图 5 所示。输入图像经过 MDCNN 特征提取后, 每幅图像特征向量的大小为 25 088 位, 特征数据融合后大小为 $(N+1) \times 25\ 088$ 位, 存储这种高维度的特征向量不仅需要极大的存储空间, 而且不利于后续回归模型的学习, 同时发现这种高维度的特征往往存在大量的数据冗余。针对以上问题, 本文利用主成分分析法 (PCA) [15] 进行特征压缩。PCA 是一种维度约减算法, 能够把高维度数据在损失最小的情况下转换为低维度数据。实验结果也证明, 利用 PCA 对特征向量进行压缩, 能够减少信息冗余, 从而在可控的失真范围内提高后续回归模型的运算速度。

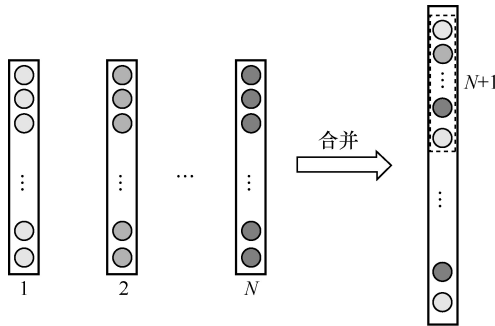


图 5 特征融合示意

1.4 Light GBM 回归模型

在获得全景图像序列的融合特征后, 需要根据特征向量进行拼接参数的预测, 显然这是一个多变量的回归问题, 目前常用的解决方案主要包括机器学习算法和深度学习算法, 考虑到深度学习算法往往需要大规模的数据集训练, 因此选择机器学习中的 GBDT (gradient boosting decision tree) [16] 算法来解决此问题。GBDT 算法的核心在于决策树的迭代训练, 具有容易训练、拟合效果好等优点, 因而得到广泛的应用。传统的 GBDT 算法, 如 XGBoost [17] 在一些高维度、强耦合的领域已经逐渐无法满足要求。在此基础上, LightGBM [18] 在训练效率、准确率和并行化学习等方面都得到了较大的提升, 常被

用来处理大规模数据集和高维度特征, 因此本文使用 LightGBM 算法完成拼接参数的回归学习。

1.5 柱面投影算法

本文的待拼接图像序列在空间上仅存在旋转关系, 因此首先需要先进行柱面投影, 再完成图像对齐。柱面投影是一种将真实世界的图像映射到一个以固定视点为中心的虚拟圆柱体表面上的技术。它通过将图像投影到以像素焦距为半径的圆柱体表面上, 从而保持真实世界中的空间约束关系。在全景图像拼接过程中, 柱面投影算法是不可或缺的步骤, 它在保持图像空间关系方面发挥着重要作用。柱面投影基本原理示意如图 6 所示。

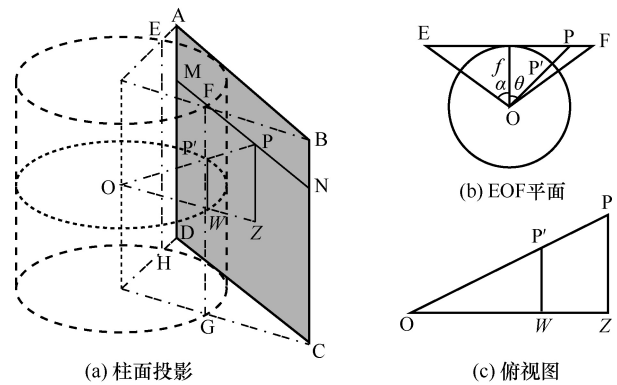


图 6 柱面投影基本原理示意

由图 6 可知, 根据空间立体几何知识, 可以得到投影图像 ABCD 平面到投影柱面 EFGH 之间的空间变换关系。在原图 ABCD 中, 以左上角为坐标原点, 图像的宽和高分别为 W 和 H , 设平面中任意一点 P 的坐标为 (x, y) , 在柱面上的投影点 P' 的坐标为 (x', y') , 从俯视图可以计算出柱面图的宽 W' 和投影横坐标 x' 分别为

$$\alpha = \arctan \frac{W}{2f} \quad (3)$$

$$W' = f\alpha = \arctan \frac{W}{2f} \quad (4)$$

$$\theta = \arctan \frac{x - \frac{W}{2}}{f} \quad (5)$$

$$x' = \frac{W'}{2} + f\theta \quad (6)$$

其中, f 为照相机的像素焦距。进一步可以得到

$$x' = f \arctan \frac{x - \frac{W}{2}}{f} + f \arctan \frac{W}{2f} \quad (7)$$

由相似三角形原理, 可以从侧视图中计算出 y'

$$\frac{f}{\cos \theta} = \frac{\frac{H'}{2} - y'}{\frac{H}{2} - y}, H' = H \quad (8)$$

$$\cos \theta = \frac{f}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \quad (9)$$

$$y' = \frac{H}{2} - \frac{f\left(y - \frac{H}{2}\right)}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \quad (10)$$

其中, (x', y') 为柱面变换后的投影坐标。柱面投影可以将不同角度、不同距离的图像投影到同一个柱面上, 从而保持图像的空间一致性, 同时也能够减少对齐过程中图像的变形和畸变问题, 提高全景拼接图像的质量和稳定性。如图 7(a)所示, 对一张平面图像进行柱面投影, 假设照相机焦距为 500 像素, 柱面投影后的结果如图 7(b)所示。



(a) 原图

(b) 柱面投影

图 7 柱面投影效果

1.6 图像融合算法

理想状况下, 在两幅拼接图像的重叠区域, 每个位置的像素在两幅图像中应该具有相同的像素值, 但实际情况并非如此。即使进行了光照补偿, 但由于光晕、配准偏差以及照相机建模的误差, 在重叠区域仍然存在拼接缝、重影等问题, 这导致图像过渡不太自然, 影响视觉效果, 并且容易影响最终对天际线的检测, 因此图像融合算法十分重要。图像融合算法是图像拼接的后处理方法, 采用合适的图像融合技术, 能够在保留图像信息的同时, 平滑地融合重叠区域, 获得更加自然的视觉效果。

本文在加权平均融合算法的基础上, 提出了一种双曲线融合算法。该算法根据位置和梯度变化曲线计算权值, 使融合效果更加自然, 并保留更多的细节信息。所提算法主要基于这样一个发现: 人们往往容易接受渐入渐出的变化。对于拼接的重叠区域, 使用渐变的融合效果能使过渡区域更加自然。一般来说, 若

重叠区域距离第一幅图像越近, 则第一幅图片应该获得较大的权值, 第二幅图片获得较小的权值, 由此计算重叠区域每个像素到图像光心的距离, 如式(11)和式(12)所示。为了使重叠区域图像的细节信息更加丰富, 可以对重叠区域梯度值进行量化, 如式(13)和式(14)所示, 梯度越大说明该位置的细节越丰富, 往往需要更大的权重, 反之更小, 本文使用拉普拉斯算子计算梯度。将位置信息和梯度信息进行加权平均, 最终融合后的图像 $H(x, y)$ 如式(17)所示。

$$d_1(x, y) = \sqrt{(x - x_1)^2 + (y - y_1)^2} \quad (11)$$

$$d_2(x, y) = \sqrt{(x - x_2)^2 + (y - y_2)^2} \quad (12)$$

$$l_1(x, y) = \left\| \sum_{i \in \Omega} I_1(x_i, y_i) - 4I_1(x, y) \right\|^2 \quad (13)$$

$$l_2(x, y) = \left\| \sum_{i \in \Omega} I_2(x_i, y_i) - 4I_2(x, y) \right\|^2 \quad (14)$$

$$w_1(x, y) = \frac{d_1(x, y)}{d_1(x, y) + d_2(x, y)} +$$

$$\frac{l_1(x, y)}{l_1(x, y) + l_2(x, y)} \quad (15)$$

$$w_2(x, y) = \frac{d_2(x, y)}{d_2(x, y) + d_1(x, y)} +$$

$$\frac{l_1(x, y)}{l_1(x, y) + l_2(x, y)} \quad (16)$$

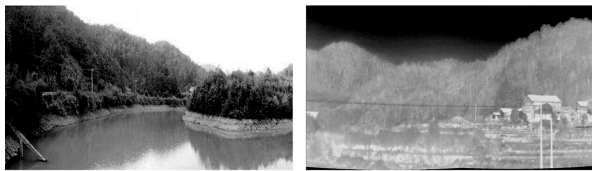
$$H(x, y) =$$

$$\begin{cases} I_1(x, y), (x, y) \in U_1 \\ \lambda_1 I_1(x, y) + \lambda_2 I_2(x, y), (x, y) \in U_1 \cap U_2 \\ I_2(x, y), (x, y) \in U_2 \end{cases} \quad (17)$$

$$\lambda_1 = \frac{w_1(x, y)}{w_1(x, y) + w_2(x, y)}, \lambda_2 = 1 - \lambda_1 \quad (18)$$

其中, d_1 表示融合图像上的一点到待拼接的第一幅图像上一点的欧氏距离, d_2 表示融合图像上的一点到待拼接的第二幅图像上一点的欧氏距离, l_1 表示对待拼接的第一幅图像采用拉普拉斯算子计算梯度后的所得结果, l_2 表示对待拼接的第二幅图像采用拉普拉斯算子计算梯度后的所得结果; λ_1 表示在重叠区域中待拼接的第一幅图像的像素值强度所占的权重系数, λ_2 表示在重叠区域中待拼接的第二幅图像的像素值强度所占的权重系数; U_1 和 U_2 分别表示两幅图像的区域; $w_1(x, y)$ 和 $w_2(x, y)$ 分别表示图像在重叠区域的权重; Ω 表示像素点上、下、

左、右 4 个邻域。将两幅可见光图像和两幅红外图像对齐后进行图像融合实验,融合结果如图 8 所示。



(a) 可见光图像融合效果 (b) 红外图像融合效果

图 8 可见光图像和红外图像的融合效果

2 实验结果与分析

2.1 实验设备和数据集

为了验证本文算法,构建了一套图像采集设备。其中,照相机被固定在一个升降杆上,同时照相机底部包含一个可旋转的云台,升降杆连接到车辆的顶部。当升降杆升起后,离地面高度约 2.5 m,这样可有效防止一些树木、房屋、草丛等物体的干扰,尽可能拍摄到包含清晰天际线的图像。对于特征提取网络部分,往往通过分类任务来训练网络的特征提取能力,因此需要使用图像分类相关的数据集。

本文使用 ILSVRC2012 数据集 (ImageNet 的一个子集),共有 1 000 个类别,训练集包含 130 万幅图片,验证集包含 5 万幅图片。对于回归模型的数据集,由于目前尚未有全景图像拼接相关数据集,因此需要自行进行数据的采集和标注。本文实际拍摄了 64 组图片,每组包含 9 幅可见光图像和 24 幅红外图像。对于可见光图像,采集设备每隔 45° 拍摄一幅,对于红外图像,侦查球每隔 15° 拍摄一幅,其中可见光图像分辨率为 $1\,920$ 像素 \times $1\,080$ 像素,红外图像的分辨率为 $1\,050$ 像素 \times 860 像素。图像采集完之后,需对每组图像进行标注,计算每幅图像对应的变换矩阵及照相机焦距。考虑到照相机在拍摄过程中,可能会进行光学变焦,为了不影响柱面投影的效果,需要增加照相机焦距参数。具体计算过程如下。

- 1) 使用 SIFT 算子对每幅图像进行特征点检测。
- 2) 使用最近邻匹配对每幅图像进行特征点匹配和筛选。
- 3) 使用随机采样一致 (RANSAC) 算法^[19]计算出对应的单应性矩阵,并使用光束平差法对单应性矩阵参数进行捆绑调整,求取精确参数矩阵,并估算照相机焦距。

通过上述步骤,能够完成对大部分数据的标注,但对于少部分匹配错误较大的图像或者特征点

较少的图像,无法计算出对应的参数,因此需要手工标注一些关键点,并计算相应参数。变换矩阵的大小为 3×3 ,包含 9 个参数。对于每组可见光图像,标注了 90 个未知参数。对于红外图像,标注了 240 个未知参数。此外,由于数据集样本较少,可通过数据增强的方式扩充数据集。本文对每组图像数据进行了随机的旋转、平移和变形操作,同时对标签数据进行相应的变换。扩充数据集,不仅有利于降低模型过拟合风险,还可以提高模型鲁棒性和泛化能力。数据集扩充之后,共计 300 组数据,2 700 幅可见光图像,7 200 幅红外图像。

2.2 实验环境与参数设置

本节实验使用 PyTorch 深度学习框架,CPU 为 Intel(R) Xeon(R) CPU E5-2678 v3,显卡为 GeForce RTX 2080Ti 12 GB,操作系统为 Ubuntu 16.04 64 位。对于特征提取网络 MDCNN 的训练,使用随机梯度下降 (SGD) 算法作为网络优化器,学习率调整策略使用余弦退火算法。初始学习率为 0.05,动量为 0.9,权重衰减为 0.000 5,训练周期为 1 000。为了加快模型的收敛,对网络中的部分卷积层选用 VGG16 预训练模型。对于 LightGBM 回归模型的学习,按照 7:3 的比例划分数据集,其中训练集 210 组,测试集 90 组。设置模型的学习率为 0.2,树的最大深度为 20,叶子节点为 20,最大迭代次数为 300。

2.3 实验效果及对比

1) 特征提取网络实验

为了验证本文提出的 MDCNN 特征提取的能力,本节对 AlexNet^[20]、SqueezeNet^[21]、DenseNet^[22]、VGG16、VGG19 和 ResNet^[23]这 6 种分类网络分别进行了实验对比,比较了不同网络在 ILSVRC2012 数据集下的分类性能,分类结果的性能指标如表 1 所示。

表 1 分类结果的性能指标

分类网络	Top1 准确率	Top5 准确率
AlexNet	56.522%	79.066%
SqueezeNet	58.092%	80.420%
DenseNet	74.434%	91.972%
VGG16	73.360%	91.516%
VGG19	74.218%	91.842%
ResNet	73.314%	91.420%
MDCNN (不引入 DoGPM)	74.998%	91.807%
MDCNN (不引入衍生图)	75.318%	92.684%
MDCNN (引入 DoGPM 和衍生图)	75.683%	92.728%

由表 1 可知，本文提出的 MDCNN 在 Top1 和 Top5 上的准确率绝大多数优于其他分类网络，网络的分类能力更优。为了验证 MDCNN 中的 DoGPM 模块以及衍生图对网络性能的影响，还进行了相关消融实验。由表 1 可知，MDCNN 不引入 DoGPM 模块时在 Top1 和 Top5 上的准确率分别为 74.998%和 91.807%，而不引入衍生图时在 Top1 和 Top5 上的准确率分别为 75.318%和 92.684%。由此可知，相比于衍生图，DoGPM 模块对模型的影响较大，它对网络性能的提升更加显著。而当同时引入 2 个模块后，MDCNN 的分类准确率达到最佳，在 Top1 和 Top5 上的准确率分别为 75.683%和 92.728%。这证明本文所提 MDCNN 能够有效地提取图像特征，且和目前主流的分类网络相比具有一定的优势。

2) PCA 降维实验

为了验证特征融合过程中 PCA 降维对回归精度的影响，将特征向量分别降维至不同维度，并使用 LightGBM 回归模型计算 MAE (mean absolute error) 值，最终的实验结果如图 9 所示。由图 9 可知，随着特征维度的增加，MAE 值呈减小趋势，即回归精度呈上升趋势。当特征维度增加至 270 维时，回归精度趋于平稳状态，约为 0.024 0；当特征维度为 230 维时，MAE 值最小，约为 0.022 4，此时回归精度最高。因此，综合考虑 PCA 降维对准确率、存储空间以及运行时间的影响，本文选择将特征向量降维至 230 维进行后续实验。

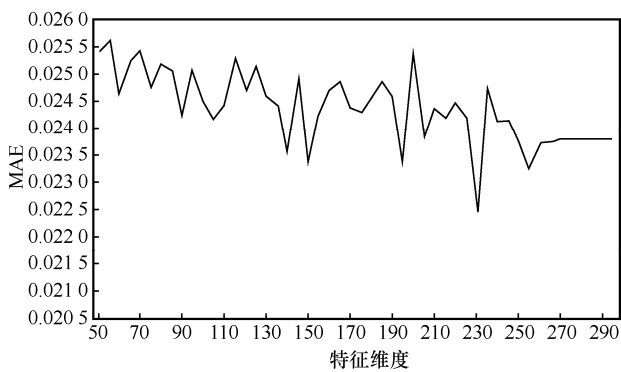


图 9 特征维度与 MAE 值的关系

3) 回归模型实验

针对拼接参数的回归问题，当选用不同的机器学习模型时，回归的准确性会有所差别。本文在测试集上分别使用 Linear、RandomForest、SVM、KNN、XGBoost、BP 和 LightGBM 这 7 种模型进

行实验，实验结果如表 2 所示。其中，MAE 是回归算法度量指标，MAE 值越小，模型的精度越高；均方根误差 (RMSE) 值越小，模型的精度越高； R^2 主要反映了数据之间的拟合度， R^2 值越大，模型拟合度越高、模型效果越好。由表 2 可知，BP 和 Linear 算法的准确率最差，其他算法的准确率都较高。其中，LightGBM 算法的回归精度最高，在 MAE、RMSE 和 R^2 指标上均取得了最优值，模型的预测误差最小，拟合度更高，整体上看，LightGBM 回归模型优于其他算法。

表 2 不同算法的回归性能指标

算法	MAE	RMSE	R^2
Linear	0.098	0.127	0.784
RandomForest	0.028	0.038	0.822
SVM	0.036	0.040	0.835
KNN	0.026	0.037	0.579
XGBoost	0.029	0.043	0.804
BP	0.173	0.200	0.753
LightGBM	0.022	0.035	0.864

4) 全景图像拼接效果

本节实验选取了 6 组可见光图像和 6 组红外图像，分别使用本文算法进行全景图像拼接，实验结果分别如图 10 和图 11 所示。



图 10 可见光图像全景拼接实验结果

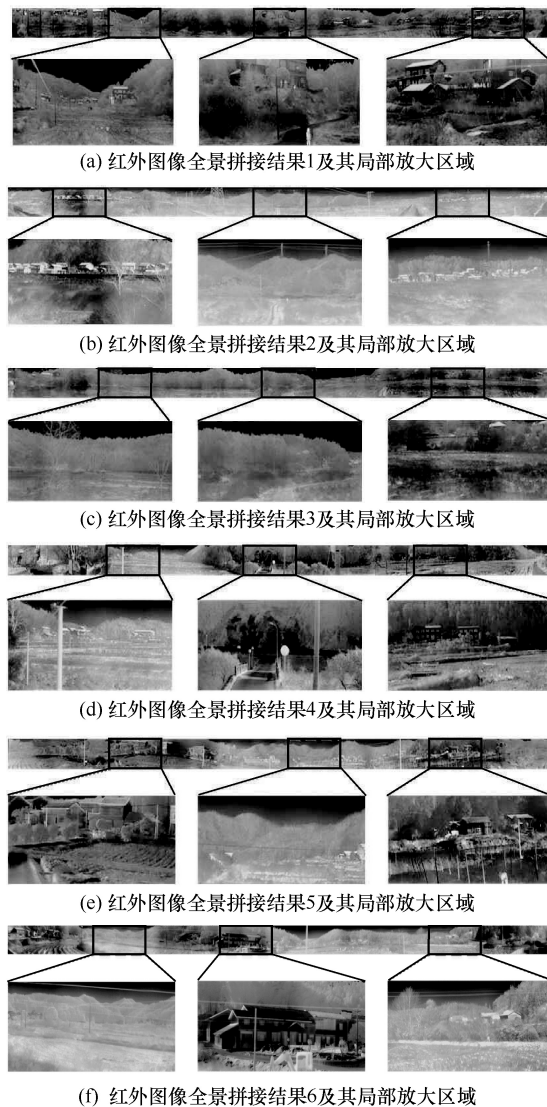


图 11 红外图像全景拼接实验结果

由图 10 可知，可见光图像的全景拼接结果整体视觉效果很自然，图像质量也较高，没有出现明显的畸变、鬼影、拼接缝等问题，信息丢失较少，基本实现了图像的无损拼接。

由图 11 可知，对于红外图像这种分辨率较小、色彩贫乏的伪彩色图像来说，本文算法克服了拼接过程中特征点匮乏、图像之间重叠率较小等问题，同样表现出了较好的拼接效果，局部区域能够很好地对齐和融合。以图 11(f)中的左侧标注的局部区域为例，图 12 为该区域对应的三幅待拼接图像，该区域不存在房屋、树木等包含明显特征的前景对象，并且图像整体都处于过曝光的状态，对比度较低，拼接难度较大，但最终的拼接结果仍然实现了很好的对齐，尤其是对于图像中错落的电网来说，不存在交叉、重叠等问题，证明本文算法对红外图

像同样实现了较好的拼接效果。

5) 拼接算法性能对比

为了验证本文算法在客观性能指标上的优越性，本节采集了 20 组可见光图像和 20 组红外图像进行测试，分别使用不同的全景拼接算法进行拼接实验，并计算拼接结果的自然图像质量评估指标 (NIQE)^[24]、感知因子 (PI)^[25]性能指标，以及拼接成功率和运行时间。不同算法在可见光图像上的性能对比如表 3 所示。其中，对于 NIQE 和 PI 这 2 个指标，其值越小代表性能越好。由表 3 可知，关于全景图像的拼接成功率，APAP 的成功率最低，只有 12 组数据拼接成功，NISWGSP 和 AutoStitch 的成功率较高，而本文算法则达到了 100%，20 组数据全部拼接成功。关于图像质量的客观性能指标，本文算法 NIQE 指标值最小，PI 指标值仅次于 AutoStitch。关于拼接算法的运行速度，本文算法平均拼接仅需要 4.386 s，小于其他算法的运行时间。综合来看，本文算法能够兼顾拼接质量和速度，在拼接性能上具备较好的优势。



图 12 图 11(f)中的左侧标注的局部区域为例的待拼接图像序列

表 3 不同算法在可见光图像上的性能对比

算法	NIQE	PI	拼接成功率	运行时间/s
AutoStitch ^[3]	4.411	4.358	90%	5.146
APAP ^[4]	6.163	7.195	60%	71.632
AANAP ^[6]	5.322	5.976	80%	46.249
SPHP ^[5]	5.675	6.732	85%	62.751
NISWGSP ^[7]	4.831	5.368	95%	12.843
本文算法	4.372	4.373	100%	4.386

同时，为了直观比较 6 种算法的拼接效果，本文选取一组 6 种算法均能成功拼接的图像序列进行实验比较，如图 13 所示。不同算法的拼接结果整体相差不大，但局部位置存在一定差异，尤其对于重叠区域，针对图 13(a)中方框标注的区域，该区域为待拼接图像序列中第 7 幅、第 8 幅图像的重叠区域，下面分别比较 6 种算法在该区域的拼接效果。

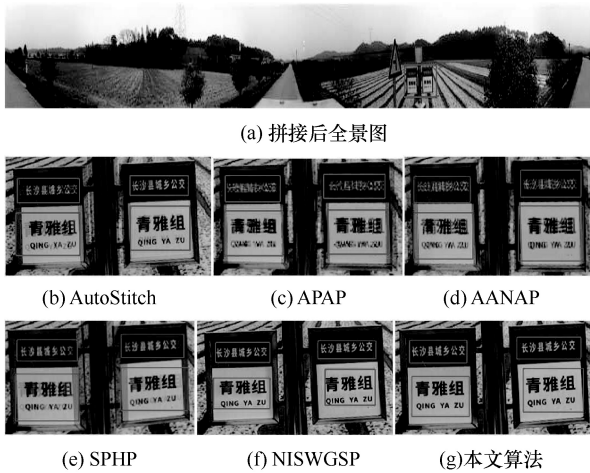


图 13 不同算法拼接效果对比

由图 13 可知，AutoStitch 算法在公交牌区域较模糊，并且左半部分存在重影，右半部分重影较小，而 APAP、AANAP、SPHP 的拼接结果在左右两半部分均存在较严重的重影，尤其 APAP 的结果导致公交牌的字体难以辨认，视觉效果很差。NISWGSP 和本文算法则具有较好的对齐效果，不存在明显的模糊，能够清晰地辨认公交站牌区域的文字，拼接结果质量较好。

对于红外图像而言，由于每组红外图像包含 24 幅分辨率为 1 050 像素×860 像素的图像，拼接结果分辨率通常能够达到 18 000 像素×860 像素。对于超高分辨率的图像而言，计算 NIQE 和 PI 指标较困难，需要耗费大量的时间和内存。对于红外图像来说，由于图像中缺乏色彩信息，进行基于视觉特性的图像质量评估价值不大，因此实验结果仅比较了拼接成功率和运行时间，实验结果如表 4 所示。

表 4 不同算法在红外图像上的性能对比

算法	拼接成功率	运行时间/s
AutoStitch	90%	36.375
APAP	25%	183.276
AANAP	60%	123.164
SPHP	45%	146.527
NISWGSP	80%	80.892
本文算法	100%	8.184

由表 4 可知，本文算法的拼接成功率仍然为 100%。虽然 AtuoStitch 也具有较高的拼接成功率，但运行时间明显上升，达到了 36.375 s，而本文算法在 GPU 上通过神经网络提取特征，利用机器学习模型预测拼接参数，缩短了大量计算时间，最终运行时间仅需要 8.184 s，远少于其他全景拼接算法，原因如下。

相比于可见光图像，红外图像中包含的特征信息较少，细节部分丢失严重；其次红外图像中往往包含大量的干扰噪声，这对于基于特征点提取和匹配的拼接算法来说，增大了拼接难度，很容易出现配准失败的情况。

图 14(a)为其中一组待拼接红外图像序列中的两幅相邻图像。观察两幅图像可知，由于红外图像的视场角较小，拍摄到的大部分为山体，图像之间难以辨认重叠区域，不易得到匹配的特征点。使用 SIFT 进行关键点检测，并使用对特征点进行匹配和筛选，如图 14(b)所示，其中黑色圆圈为检测出的特征点，相连的白色圆圈为匹配的特征点对，可以看出错误匹配的特征点对较多，正常情况下连线应该是平行的，这导致计算出的图像之间的单应性变换矩阵误差较大。AutoStitch、APAP、AANAP、SPHP 和 NISWGSP 算法均依赖上述特征点匹配方法进行拼接，所得到的拼接结果存在明显的畸变。AutoStitch、APAP、SPHP 和 NISWGSP 的拼接结果如图 14(c)~图 14(f)所示，AANPA 由于对齐失败而无法输出结果。

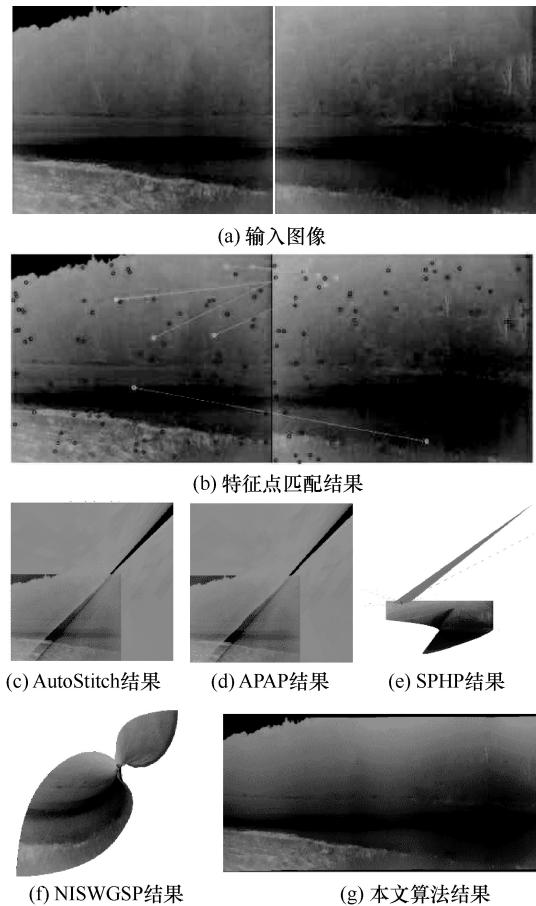


图 14 其他方法拼接失败示例及本文算法结果

本文算法由于不依赖于特征点提取和匹配，而是联合深度学习和机器学习进行端到端的学习，具有较强的准确性和鲁棒性。同时也学习到了更多的先验知识，更加有利于拼接的成功，拼接结果如图 14(g)所示。综上所述，无论是可见光图像还是红外图像，本文算法都具有较强的优势。

3 结束语

本文针对实际拍摄环境，设计了一种基于参数回归的快速全景图像拼接算法。该算法首先通过所提基于高斯差分金字塔的多尺度深度卷积神经网络（MDCNN）来对拼接图像进行特征提取，并使用 LightGBM 回归模型对拼接参数进行预测，获得图像之间的变换矩阵和照相机焦距。然后，根据获得的参数对图像进行柱面投影和扭曲变换，完成图像对齐。最后，利用所设计的双曲线融合算法来完成图像之间重叠区域的融合，并消除图像之间的拼接缝，从而获得更加自然美观的全景图像。实验结果表明，本文算法在实际场景下具有较好的拼接效果，拼接结果图像质量更高，鲁棒性更强，并且拼接速度明显高于其他算法，在实时性要求高的拼接场景下具有明显的优势。

参考文献：

- [1] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//Proceedings of the Alvey Vision Conference 1988. [S.l.]: Alvey Vision Club, 1988: 147-151.
- [2] SHU T. Multi-sensor remote sensing image alignment based on fast algorithms[J]. *Journal of Intelligent Systems*, 2023, 32(1): 1-13.
- [3] BROWN M, LOWE D G. Automatic panoramic image stitching using invariant features[J]. *International Journal of Computer Vision*, 2007, 74(1): 59-73.
- [4] ZARAGOZA J, CHIN T J, BROWN M S, et al. As-projective-as-possible image stitching with moving DLT[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 2339-2346.
- [5] CHANG C H, SATO Y, CHUANG Y Y. Shape-preserving half-projective warps for image stitching[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 3254-3261.
- [6] LIN C C, PANKANTI S U, RAMAMURTHY K N, et al. Adaptive as-natural-as-possible image stitching[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2015: 1155-1163.
- [7] CHEN Y S, CHUANG Y Y. Natural image stitching with the global similarity prior[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 186-201.
- [8] 曾国奇, 牛子凡, 郑丽丽, 等. 基于球形变换的无人机视频图像实时拼接方法[J]. *航空学报*, 2023, 44(23): 1-11.
ZENG G Q, NIU Z F, ZHEN L L, et al. A real time video image stitching method for UAV based on spherical transformation [J]. *Acta Aeronautica et Astronautica Sinica*, 2023, 44(23): 1-11.
- [9] ZHU S D, ZHANG Y Z, ZHANG J, et al. ISGTA: an effective approach for multi-image stitching based on gradual transformation matrix[J]. *Signal, Image and Video Processing*, 2023, 17(7): 3811-3820.
- [10] KHAMIYEV I, ISSA D, AKHTAR Z, et al. Panoramic image generation using deep neural networks[J]. *Soft Computing*, 2023, 27(13): 8679-8695.
- [11] DAI Q Y, FANG F M, LI J C, et al. Edge-guided composition network for image stitching[J]. *Pattern Recognition*, 2021, 118: 108019.
- [12] ZHANG H, ZHANG J M, PERAZZI F, et al. Deep image compositing[C]//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2021: 365-374.
- [13] LU C N, CHANG Y C, CHIU W C. Bridging the visual gap: wide-range image blending[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 843-851.
- [14] PARIS S, KORNPROBST P, TUMBLIN J, et al. Bilateral filtering: theory and applications[J]. *Foundations and Trends in Computer Graphics and Vision*, 2008, 4(1): 1-75.
- [15] SHLENS J. A tutorial on principal component analysis[J]. *arXiv Preprint*, arXiv: 1404.1100, 2014.
- [16] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
- [17] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 785-794.
- [18] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 3149-3157.
- [19] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [21] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model

size[J]. arXiv Preprint, arXiv: 1602.07360, 2016.

- [22] IANDOLA F, MOSKEWICZ M, KARAYEV S, et al. Densenet: Implementing efficient convnet descriptor pyramids[J]. arXiv Preprint, arXiv: 1404.1869, 2014.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [24] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.
- [25] BLAU Y, MECHREZ R, TIMOFTE R, et al. The 2018 PIRM challenge on perceptual image super-resolution[C]//European Conference on Computer Vision. Berlin: Springer, 2019: 334-355.

[作者简介]



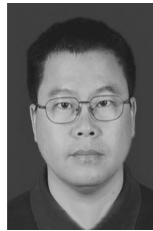
郭璠（1982-），女，湖南临澧人，博士，中南大学副教授、硕士生导师，主要研究方向为图像处理、模式识别、人工智能等。



李小虎（1998-），男，河南安阳人，中南大学硕士生，主要研究方向为计算机视觉、图像处理、模式识别等。



刘文韬（2000-），男，湖北武汉人，中南大学硕士生，主要研究方向为计算机视觉、图像处理、模式识别等。



唐璘（1966-），男，湖南武冈人，博士，中南大学教授、博士生导师，主要研究方向为计算机视觉、人工智能、机器人等。