

高精度视频配准算法中的静态图像配准算法

王苹

引用本文:

王苹. 高精度视频配准算法中的静态图像配准算法[J]. 液晶与显示, 2020, 35(6): 612–618.

WANG Ping. Static image registration algorithm in high-precision video registration algorithm[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6): 612–618.

在线阅读 View online: <https://doi.org/10.3788/YJYXS20203506.0612>

您可能感兴趣的其他文章

Articles you may be interested in

一种基于相位相关与子图像的偏振图像配准方法

Polarized image registration method based on phase correlation and sub-graph

液晶与显示. 2019, 34(5): 530–536 <https://doi.org/10.3788/YJYXS20193405.0530>

基于U-Net的多图谱标签融合算法

Multi-atlaslabel fusion based on U-Net

液晶与显示. 2019, 34(11): 1091–1103 <https://doi.org/10.3788/YJYXS20193411.1091>

一种改进的SURF彩色遥感图像配准算法

Improved SURF algorithm for color remote sensing image registration

液晶与显示. 2017, 32(2): 144–152 <https://doi.org/10.3788/YJYXS20173202.0144>

非凸性优化与动态自适应滤波的湍流退化视频复原

Turbulence-degraded video restoration based on non-convexity optimization and dynamical adaptive filtering

液晶与显示. 2017, 32(6): 482–490 <https://doi.org/10.3788/YJYXS20173206.0482>

多源遥感影像高精度自动配准方法研究

High accuracy-automatic registration method research on multi-source remote sensing image

液晶与显示. 2016, 31(6): 604–612 <https://doi.org/10.3788/YJYXS20163106.0604>

文章编号:1007-2780(2020)06-0612-07

高精度视频配准算法中的静态图像配准算法

王 苹^{1,2*}

(1. 阳光学院 人工智能学院, 福建 福州 350015;

2. 空间数据挖掘与应用福建省高校工程研究中心, 福建 福州 350015)

摘要: 为优化视频配准工作中的静态图像配准算法, 本文应用深度学习的卷积神经网络 VGGNet 设计了一个静态图像配准算法。通过仿真实验结果可知, VGGNet 能够被很好地应用于静态图像配准工作中; 结合均方根误差和 N_{red} 结果可知, 基于 Conv5、FC1 或者 FC2 输出特征的静态图像配准算法具有良好的性能, 基于 FC2 的配准算法性能最好, 其均方根误差值为 0.031 22, N_{red} 值为 74。通过视频缩放和亮度转换后, 静态图像配准算法的性能有所下降, 在亮度变换后基于 FC2 输出特征的配准算法性能较高。与传统 HOG、LBP 特征提取的图像配准算法相比, 本文算法具备较好的配准精度和正确率。研究结果可为当前基于深度学习卷积神经网络的静态图像配准算法提供参考。

关键词: 视频; 图像配准; 深度学习; VGGNet

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.3788/YJYXS20203506.0612

Static image registration algorithm in high-precision video registration algorithm

WANG Ping^{1,2*}

(1. College of Artificial Intelligence, Yango University, Fuzhou 350015, China;

2. Fujian University Engineering Research Center of Spatial Data Mining and Application, Fuzhou 350015, China)

Abstract: In order to optimize the static image registration algorithm in video registration, a static image registration algorithm based on deep learning convolutional neural network VGGNet is proposed. The simulation experiments results show that VGGNet can be well applied to the static image registration work. Combining the root mean square error and N_{red} results, it can be seen that the static image registration algorithm based on the output features of Conv5, FC1, or FC2 has good performance. The algorithm based on FC2 has the best performance with root mean square error value of 0.031 22 and N_{red} of 74. After video scaling and brightness conversion, the performance of the static image registration algorithm decreases. The performance of the registration algorithm based on FC2 is better overall after the brightness conversion. Compared with the traditional image registration algorithm

收稿日期: 2019-12-21; **修订日期:** 2020-02-14.

基金项目: 2019 年福建省属公益类科研院所基本科研专项 (No.2019R1008-9)

Supported by Basic Scientific Research Project of Fujian Provincial Public Welfare Scientific Research Institute in 2019 (No.2019R1008-9)

* 通信联系人, E-mail: wangping11201120@163.com

based on HOG and LBP feature extraction, this algorithm has better registration precision and accuracy. The research results can provide a reference for current static image registration algorithms based on deep learning convolutional neural networks.

Key words: video; image registration; deep learning; VGGNet

1 引言

多视频配准即应用多个相机在同一场景对同一个物体进行拍摄,从而得到多个在时间或者空间上存在某种对应关系的视频,通过配准寻找到这种对应关系的变换参数^[1]。现阶段已经产生了多个视频配准算法^[2-3]。在发展历程上,Caspi^[4]等开展了基于特征和基于区域的视频配准算法研究。Shakil^[5]针对两个拍摄同一物体时自由运动的相机提出了一个新的视频配准算法。Sand^[6]等人提出了一种新的可应用于多个视频配准工作的算法。陈为龙^[7]等提出了视频配准算法中的静态图像配准算法,即通过视频提取出一些图像帧,通过对这些静态的图像帧进行匹配来配准视频,将基于特征的图像配准算法应用到视频配准工作中,取得了有效的结果。深度学习的卷积神经网络因具有模式识别^[8]的优势,已经在图像识别^[9]和分类^[10-11]、目标检测^[12]等领域发挥了重大作用,且已应用于遥感图像配准领域中。

本文应用深度学习的卷积神经网络VGGNet,设计了高精度视频配准算法中的图像配准算法,研究了VGGNet在视频中静态图像配准工作的可行性及其对配准性能的影响,探讨了图像缩放、亮度变换等对视频中图像配准的影响,比较了本配准算法与传统基于方向梯度直方图(HOG)^[13]、局部二值模式(LBP)^[14]特征提取配准算法的性能。

2 基于卷积神经网络的静态图像配准算法

卷积神经网络(CNN)在图像识别和分类问题上能发挥出巨大的优势,因此设计了基于卷积神经网络的静态图像配准算法,研究网络中不同层下图像配准的精度,并阐述了实验的数据源和环境配置。

2.1 多视频配准算法中静态图像配准算法设计

本文的静态图像配准算法及流程如图1所示。首先从视频S1中获取某一帧的图像,并从视频S2中获取某一帧的待配准图像。第二步,通过尺度不变特征变换(SIFT)方法对两张图像进行预处理,降低噪声对图像信息的干扰,获取基本的特征图像,大小为 128×128 。第三步,将特征图像作为CNN模型VGGNet的输入,经过卷积神经网络后,输出抽象化的特征信息。第四步,根据抽象化特征信息获取两幅图像的同名点,从而计算出两幅图像的变换参数,通过图像变换得到最终的配准结果图。

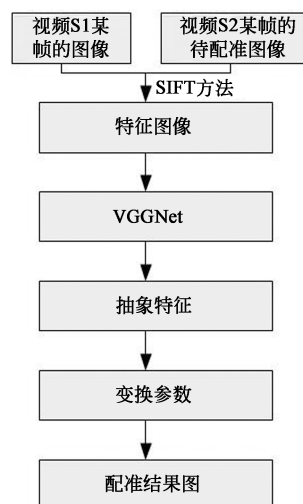


图1 静态图像配准算法

Fig.1 Static image registration algorithm

近几年来,深度学习得到了迅速发展。卷积神经网络在图形分类、目标识别等领域取得了良好成绩。其中,VGGNet模型经过大型数据库ImageNet训练后,可以从图像中提取出更高层次的抽象化特征。仿真实验结果表明16层的VGGNet具有更优异的识别性能。

VGGNet网络结构如图2所示。它由13个卷积层、5个池化层和3个全连接层组成。在卷积层中,它应用了大小为 3×3 的卷积核,这在保证一定范围感受野的条件下减少了网络的参数规

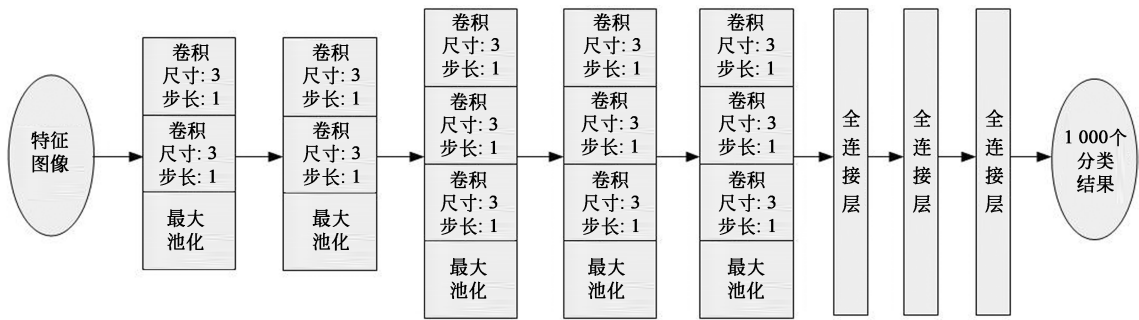


图 2 VGGNet 模型

Fig.2 VGGNet model

模,也引入了更复杂的非线性因素,有效增强了模型的特征表达。最大池化层的窗口大小为 2×2 ,在降采样特征量的同时保证关键特征信息不会丢失。通过反复堆叠小型卷积核和最大池化层,既增强了非线性特征表达能力,又将参数量控制在较低的范围内,具有很好的应用价值。在全连接层中,模型应用多个神经元可有效拟合特征的分布,从而提升分类的准确率。

2.2 实验环境配置

本文选取了官方训练好的 VGGNet 模型。为了使得模型提取出的特征适用于静态图像的配准,本文从视频中提取了 12 000 对静态图像数据对 VGGNet 模型进行了微调,如表 1 所示。其中,2 000 对静态参考图像和待配准图像来自于从多对待配准视频中提取的静态图像,对它们做增强变换,旋转 90° 、旋转 180° 、旋转 270° 、左右变换、上下变换后增大数据集为 12 000 对。随机划分 10 000 对参考图像和待配准图像为训练数据集,剩余的 2 000 对参考图像和待配准图像为测试数据集。

表 1 数据集

Tab.1 Dataset

增加变换方法	参考图像	待配准图像
无	2 000	2 000
旋转 90°	2 000	2 000
旋转 90°	2 000	2 000
旋转 90°	2 000	2 000
左右变换	2 000	2 000
上下变换	2 000	2 000

因此,本文将卷积和最大池化的堆叠操作作为一个网络层,设置一个全连接层为一个网络层,研究它们的输出特征。则此时共有 8 个网络层,命名为 Conv1、Conv2、Conv3、Conv4、Conv5、FC1、FC2、FC3。原 VGGNet 的网络层 FC3 被用于图像的分类工作,不适用于静态图像配准工作,因此舍弃 FC3。

在对 VGGNet 进行微小的训练过程中,设置初始的学习率为 0.003,运行平台是 Ubuntu14.04 服务器,应用的深度学习框架是 Tensorflow。

3 静态图像配准-均方根误差分析

在本章节中,确定了通过卷积网络提取的抽象化特征被用于配准的可行性,并应用均方根误差分析法评估了不同网络层特征的配准精度,最后与传统图像配准算法比较。

3.1 静态图像配准可行性分析

为了验证卷积神经网络提取的抽象化特征对静态图像配准工作的可行性,本文随机选取了一对测试数据集中的参考图像和待配准图像进行实验,结果如图 3 所示。

视频 S1 和视频 S2 是在同一时间段、同一地点、不同角度拍摄的落日图像,分别选取两个视频中的某帧图像作为实验对象。从配准结果图中可知,两幅图像大部分相同,纹理区域均配准成功,因此卷积神经网络可被应用于视频配准中的静态图像配准工作中。

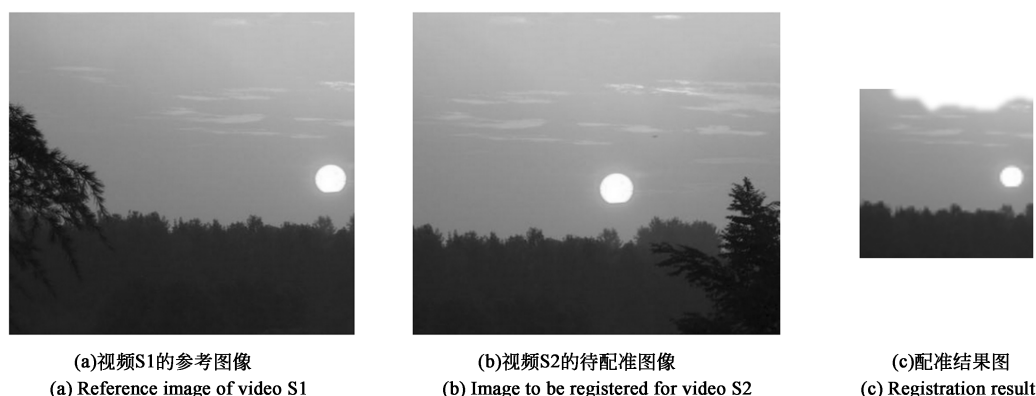


图 3 配准图像

Fig.3 Registration image

3.2 均方根误差分析

采用均方根误差方法来定量分析静态图像配准的精度。均方根误差的计算公式如式(1)所示。

$$y = \sqrt{\frac{1}{N} \sum_{i=1}^N ((X'_i - X''_i)^2 - (Y'_i - Y''_i)^2)} \quad (1)$$

其中, N 为参考图像和待配准图像的对应点对数量, (X'_i, Y'_i) 为参考图像的点 (X, Y) 在待配准图像的对应点坐标, (X''_i, Y''_i) 为对待配准图像进行配准变换后的坐标。均方根误差值越低越好。

对于图 3 中的图像, 卷积神经网络 VGGNet 的各层特征的均方根误差分析值如表 2 所示。由表 2 数据分析可知, 第一个网络层输出特征的配准误差最大; 第二个网络层的配准误差比第一个网络层低约 0.5%; 第三、第四和第五个网络层的特征配准误差较为接近; FC1 特征和第 5 个网络层的输出特征的配准误差较为接近; 经过 FC2 的特征变换后, 输出的特征值的配准误差降低了约 0.4%。

表 2 特征的均方根误差值

Tab.2 Root mean square error for feature

特征源	均方根误差
Conv1	0.045 13
Conv2	0.040 02
Conv3	0.037 59
Conv4	0.037 60
Conv5	0.036 69
FC1	0.036 90
FC2	0.032 94

对于 VGGNet 的各个网络层特征, 计算 2 000

对测试图像的配准均方根误差, 并求取各层配准误差均值, 结果如表 3 所示。由表 3 可知, Conv5、FC1 和 FC2 的输出特征的均方根误差均值较低, FC2 最低为 0.031 22。将包含 FC2 的完整 VGGNet 作为静态图像配准算法开展后续研究。

表 3 特征的均方根误差均值

Tab.3 Mean root mean square error for feature

特征源	均方根误差均值
Conv1	0.049 10
Conv2	0.035 84
Conv3	0.035 92
Conv4	0.034 22
Conv5	0.031 90
FC1	0.033 31
FC2	0.031 22

研究 VGGNet 图像配准算法与基于 HOG、LBP 特征提取的图像配准算法精度, 结果如表 4 所示。

表 4 多个算法的均方根误差均值

Tab.4 Mean root mean square error of different algorithms

算法	均方根误差均值
VGGNet 图像配准算法	0.031 22
基于 HOG 特征提取的图像配准算法	0.040 31
基于 LBP 特征提取的图像配准算法	0.042 29

VGGNet 图像配准算法的均方根误差均值为 0.031 22,明显低于基于 HOG、LBP 特征提取的图像配准算法,具备较高的配准精度。

4 静态图像配准 N_{red} 分析

在本章节,应用 N_{red} 方法研究静态图像配准的正确率,探究了视频缩放和亮度转换对图像配准的影响,最后与传统图像配准算法做比较。

4.1 N_{red} 分析

N_{red} 表示静态图像配准过程中正确同名点的数量,数值越高越好。

根据卷积神经网络层的输出特征,对 2 000 对测试数据集求得图像配准的 N_{red} 值的平均数,结果如表 5 所示。可知第 1~4 个网络层输出特征的配准 N_{red} 均值范围在 60~70 之内;Conv5、FC1 和 FC2 的输出特征的配准 N_{red} 均值分别为 73,73,74,说明这 3 层的输出特征的配准性能较好,在此基础上开展进一步的视频缩放和亮度转换研究。

表 5 特征的 N_{red} 均值

Tab.5 N_{red} mean for feature

特征源	N_{red} 均值
Conv1	61
Conv2	66
Conv3	66
Conv4	65
Conv5	73
FC1	74
FC2	74

4.2 视频缩放和亮度转换研究

对视频做缩放操作即对每帧的图像进行了缩放操作。本文对 2 000 对测试数据集采取了不同倍数的缩放操作,并统计在 Conv5、FC1 和 FC2 下的图像配准 N_{red} 均值,结果如图 4 所示。

进行缩放后,图像的配准 N_{red} 均值均不同程度地减小。当采取缩小操作时,配准 N_{red} 均值减小的幅度很大;而当采取放大操作时,配准 N_{red} 均值减小的幅度较小。3 个网络层的输出特征值的配准 N_{red} 均值较为接近。

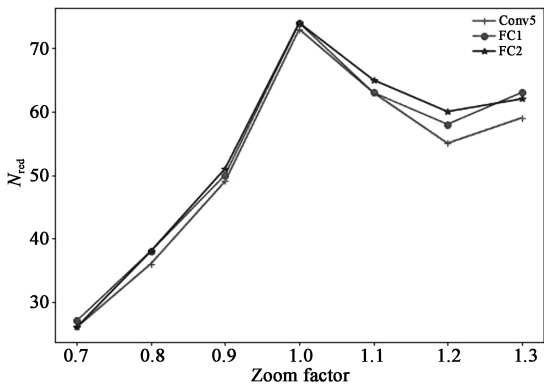


图 4 特征的 N_{red} 均值

Fig.4 N_{red} mean values for feature

在相同的实验环境下,对视频做亮度变换操作,结果如图 5 所示。

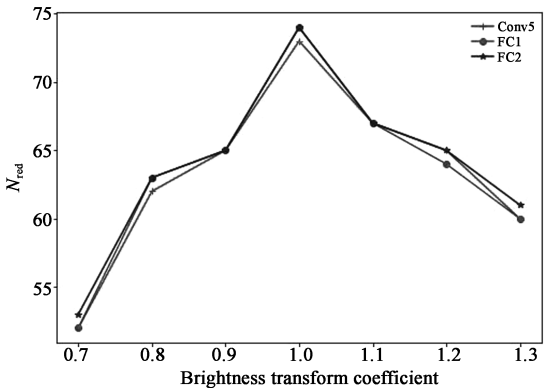


图 5 特征的 N_{red} 均值

Fig.5 N_{red} mean values for feature

采取亮度变换操作后,图像的配准 N_{red} 均值有所减小。减小亮度变换系数的配准 N_{red} 均值减小幅度比增加亮度变换系数大。FC2 的配准 N_{red} 均值总体较高。将包含 FC2 的 VGGNet 作为本图像配准算法开展后续研究。

4.3 图像配准性能分析

统计当前配准算法与经典 HOG、LBP 配准算法的正确同名点数目,结果如表 6 所示。

表 6 多个算法的 N_{red} 均值

Tab.6 N_{red} mean values for feature

算法	N_{red} 均值
VGGNet 图像配准算法	74
基于 HOG 特征提取的图像配准算法	62
基于 LBP 特征提取的图像配准算法	60

VGGNet 图像配准算法的 N_{red} 均值为 74, 远高于基于 HOG、LBP 特征提取的图像配准算法, 具备较高的图像配准正确率。

5 结 论

本文研究了视频配准工作中的静态图像配准算法, 设计了一个由较小尺寸卷积核、非线性激活

操作、池化操作组成的静态图像配准算法模型, 期望捕获一定范围感受野、非线性的图像语义信息。经仿真可知, 算法模型的 FC2 输出特征的配准精度为 0.031 22, 配准 N_{red} 为 74, 具备较好的配准可信度和配准性能, 优于传统的 HOG、LBP 图像特征提取算法。算法模型对于图像的缩放变换和亮度变换操作具备一定的抗干扰能力, FC2 输出特征的综合配准性能较优。

参 考 文 献:

- [1] 熊平, 吕顺华. 多视频多图像实时配准 [J]. 计算机应用与软件, 2014, 31(5): 180-183.
XIONG P, LYU S H. Multi video and multi image real-time registration [J]. *Computer application and software*, 2014, 31 (5): 180-183. (in Chinese)
- [2] 任思阳. 基于 DSP 的多源视频配准算法研究 [D]. 长春: 长春理工大学, 2018.
REN S Y. Research on multi-source video registration algorithm based on DSP [D]. Changchun: Changchun University of technology, 2018. (in Chinese)
- [3] 杨茹. 基于时空相位一致性的多传感器视频配准算法研究与实现 [D]. 西安: 西安电子科技大学, 2013.
YANG R. Research and implementation of multi-sensor video registration algorithm based on temporal and spatial phase consistency [D]. Xi'an: Xi'an University of Electronic Science and technology, 2013. (in Chinese)
- [4] CASPI Y, SIMAKOV D, IRANI M. Feature-based sequence-to-sequence matching [J]. *International Journal of Computer Vision*, 2006, 68(1): 109-116
- [5] SHAKIL O. An efficient video alignment approach for non-overlapping sequences with free camera movement [C]// *Acoustics, Speech and Signal Processing* 2006, Toulouse, France, IEEE, 2006: II 257-II 260.
- [6] SAND P, TELLER S. Video matching [J]. *ACM Transactions on Graphics*, 2004, 22(3): 592-598.
- [7] 陈为龙, 何小海, 郭黎, 等. 高精度多视频配准算法中静态图像配准算法的选取 [J]. 四川大学学报: 自然科学版, 2012, 49(5): 1053-1059.
CHEN W L, HE X H, GUO L, et al. Selection of static image registration algorithm in high-precision multi video registration algorithm [J]. *Journal of Sichuan University: Natural Science Edition*, 2012, 49(5): 1053-1059. (in Chinese)
- [8] 茆旋宇, 郑子扬, 王佩, 等. 基于深度卷积网络的 PRI 调制模式识别方法 [J]. 航天电子对抗, 2019, 35(5): 32-37.
MAO X Y, ZHENG Z Y, WANG P, et al. PRI modulation pattern recognition method based on deep convolution network [J]. *Aerospace electronic warfare*, 2019, 35(5): 32-37. (in Chinese)
- [9] 根绒切机多吉. 基于粗糙集的 CNN 图像识别算法研究 [D]. 成都: 成都理工大学, 2019.
GENRONGQIEJIDUOJI. Research on CNN image recognition algorithm based on rough set [D]. Chengdu: Chengdu University of technology, 2019. (in Chinese)
- [10] 吕朦. 基于多尺度卷积神经网络的图像分类算法研究 [D]. 武汉: 湖北工业大学, 2019.
LYU M. Image classification algorithm based on multi-scale convolution neural network [D]. Wuhan: Hubei University of technology, 2019.
- [11] 赵子豪. 基于卷积神经网络的白细胞图像检测及分类技术 [D]. 西安: 西北大学, 2019.
ZHAO Z H. Detection and classification of leukocyte image based on convolution neural network [D]. Xi'an: Northwest University, 2019. (in Chinese)
- [12] 马得花, 闫宏亮. 全卷积神经网络的车辆点云目标精细化检测 [J]. 测绘科学, 2020(2): 1-9.
MA D H, YAN H L. Vehicle point cloud target fine detection based on full convolution neural network [J]. *Mapping Science*, 2020(2): 1-9. (in Chinese)
- [13] 刘淑萍, 刘羽, 於俊, 等. 结合手指检测和 HOG 特征的分层静态手势识别 [J]. 中国图象图形学报, 2015, 20(6):

781-788.

LIU S P, LIU Y, YU J, *et al.* Hierarchical static gesture recognition combined with finger detection and hog features [J]. *Chinese Journal of image graphics*, 2015, 20(6): 781-788.(in Chinese)

[14] 周汐,曹林.分块 LBP 的素描人脸识别 [J]. 中国图象图形学报, 2015, 20(1):50-58.

ZHOU X, CAO L. sketch face recognition based on block LBP [J]. *Chinese Journal of image graphics*, 2015, 20 (1): 50-58. (in Chinese)

作者简介:



王 苹(1984—),女,福建福州人,硕士,副教授,2012 年于福州大学获得硕士学位,主要从事图像处理、深度学习等方面的研究。E-mail: wangping11201120@163.com