

基于孪生卷积神经网络立体匹配的优化

郭新¹

GUO Xin

摘要

计算机双目视觉从三维世界获取信息,对三维世界进行重构和感知。传统孪生卷积神经网络计算复杂,运算时间长。为满足特征实时提取,提出一种基于孪生卷积神经网络的立体匹配方法。首先通过卷积层和池化层来进行图像的深度特征提取,然后通过点积运算和函数进行分类,最后通过 kitti 数据集训练测试和现有的几种立体匹配的算法进行对比。实验结果表明,所提出的方法具有较好的立体匹配效果和实时性,具有一定的实用性,与 NCC 等匹配代价算法相比, PBM 值提升了 4.53%, 平均视差误差提升了 2.01%。

关键词

双目相机; 立体匹配; 障碍物识别; 卷积神经网络

doi: 10.3969/j.issn.1672-9528.2024.03.029

0 引言

在双目立体视觉系统中,立体匹配是最重要的部分,也是最复杂的部分,其目的就是在采集到的图像对中寻找像素间的一一对应关系。立体匹配的过程是根据相似度算法,在待匹配图像中找到参考图中的匹配对应点,进而建立图像对之间各个像素点的对应关系,以此为基础将二维的图像重建为具有三维特征的深度图^[1]。但现实中容易受到光照、遮挡等外界隐私导致灰度差不同、视差不连续等诸多问题,特征提取的正确率和实时性容易受到制约。本文首先对双目立体匹配进行概述性介绍,然后对目前热门的研究卷积神经网络算法进行研究,最后提出基于孪生卷积神经网络的改进立体匹配算法并进行实验。

1 孪生卷积神经网络立体匹配流程

孪生卷积神经网络立体匹配流程分为两个阶段:网络训练阶段和视差计算阶段。其中,网络训练阶段包括网络结构设计和网络训练,在这一阶段首先针对传统的孪生卷积神经网络结构运算量大、运算时间长这一弊端,将多层卷积替换为池化层进行优化,其次对改进后的网络结构进行立体匹配算法训练,进而更好地应用于障碍物的实时检测^[2-3]。在视差计算阶段应用,通过代价聚合方法对匹配代价进行聚合计算,以减少匹配错误和噪声带来的影响。孪生卷积神经网络立体匹配流程如图 1 所示。

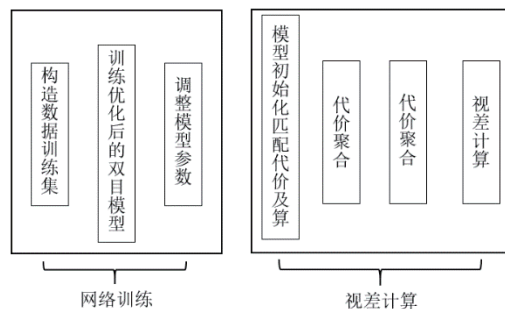
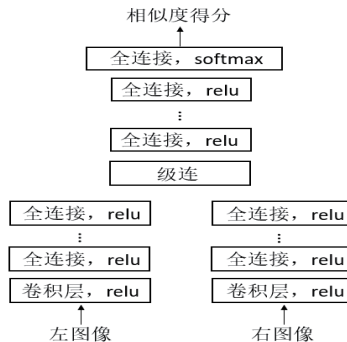


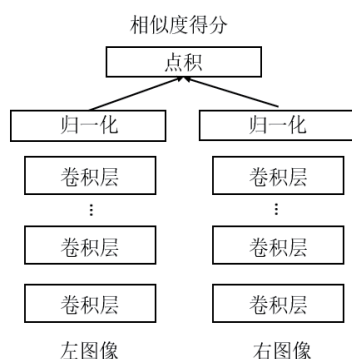
图 1 立体匹配流程

1.1 网络结构设计

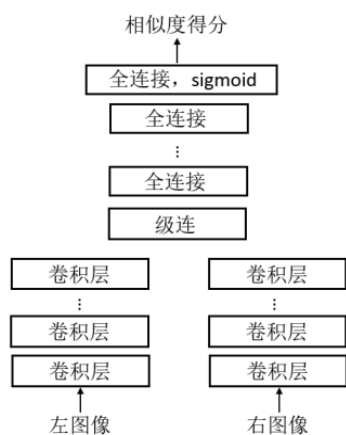
在立体匹配技术上,Zagoruyko 和 Komodakis 首先提出了基于孪生卷积神经网络(MC-CNN)的立体匹配算法,如图 2(a)所示,该网络用于评估两个图像块的匹配代价,通过一个卷积层提取出图像特征,经过若干个全连接层后左右分别输出一对一维特征向量,然后将两个一维特征向量进行融合,最后采用“全连接+softmax”组成分类器,将获取到的特征向量分为相似和不相似两种,但由于该网络中卷积层数较少,只能提取到图像中浅层次的特征。



(a)



(b)



(c)

图2 孪生卷积神经网络 (MC-CNN) 的立体匹配算法

Zbontar 和 LeCun 对该网络结构进行了改进,如图2(b)。为快速孪生卷积神经网络 (MC-CNN-fst), 该网络和 MC-CNN 网络一样参数共享且总体分为左右两个分支: 第一部分为特征提取部分, 通过增加卷积层去除全连接层, 这样网络结构可以获得图像的深度特征, 并将得到的值进行归一化处理, 便于后续操作; 第二部分将两个特征向量进行点积, 对得到的相似度进行评分, 来判断它们的相似性^[4-5]。快速孪生卷积神经网络, 运行速度更快, 从网络结构上看也更加简单, 适用于实时检测, 但获取的视差图精度较低。本文在第二网络结构 (MC-CNN-fst) 的基础上进行改进, 以便更加精确地实现特征提取和实时检测。

传统的孪生卷积神经网络结构中采用多次卷积的方式对输入图像进行缩减, 使该网络卷积核的数量较多, 运算量变大, 运算时间变长, 不适合用于车辆行进过程中实时检测^[6-7]。针对这一点本文在上文所提到的两种网络结构基础上进行优化, 以便更好地应用于障碍物的实时检测。整体网络结构如图3所示。优化后的网络结构同之前两种结构相同, 也分为两个部分: 第一部分为特征提取子网络, 通过卷积层和池化层来进行图像的深度特征提取, 且左右子网络参数共享; 第二部分为特征分类子网络, 通过对提取到的图像特征向量做

点积运算, 得到图像的相似性得分, 然后用 softmax 函数对点积运算结果进行分类。

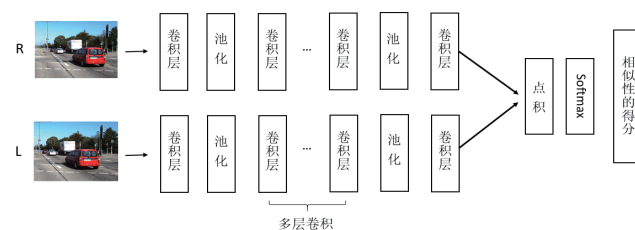


图3 改进的孪生卷积神经网络结构

1.2 网络训练

对改进后的网络结构进行立体匹配算法的训练。本文在 KITTI2015 数据集 200 组图像中, 随机抽取 180 组图像作为训练集, 但是由于 KITTI2015 数据集给定图像数据为完整图像组的数据, 而本文设计的网络结构是在局部立体匹配算法的基础上进行研究, 网络输入的数据为图像对的局部块, 因此需要对图像进行裁剪才能进行网络参数的训练。裁剪时首先需要在参考图像中选取一个像素点 $P(x, y)$ 并以 P 点为中心点裁剪到 $n \times n$ 像素的图像块, 在待搜索图像中以 $p(x-d, y)$, $d \in (0, D)$ 为中心点, 裁剪大小为 $n \times n$ 像素的图像块, 其中 D 为搜索的最大视差范围, 再将得到的一对图像利用公式 (2) 进行标准化处理, 最后作为输入图像对进行网络参数的训练。剩余 20 组图像对作为采集数据, 对网络结构进行验证。

$$U = \frac{X - \bar{x}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (1)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

式中: U 是标准化后的图像灰度矩阵; X 是标准化之前的图像灰度矩阵; x_i 是标准化前图像各像素的灰度值; $\bar{x}$ 是 x_i 的平均值; n 是图像块中像素点的数量。

卷积神经网络的训练需要大量的数据集为基础, ImageNet 数据集是为促进计算机图像识别而设立的一个大型图像数据集。因此, 一般情况下通过 ImageNet 数据集进行训练, 能够获得较好的参数, 加快算法收敛速度, 再通过已获得的少量基础数据对网络进行二次训练, 得到合适的网络参数。

1.3 视差计算

匹配代价计算是要描述两个像素的相似性, 意义是衡量匹配像素与候选像素的相关性, 尽量减小代价, 因为代价越小, 相关性越大。利用设计的卷积神经网络, 可以获取某一点像素点 p 的匹配代价 $C_{CNN}(p, d)$ 。

$$C_{CNN}(p, d) = -(V^l(p) \cdot V^r(p - d)) \quad (3)$$

式中： $V^l(p)$ 、 $V^r(p-d)$ 分别对应左右的子网络输出的特征向量。

但是匹配代价结果不足以生成精度较高的视差图，特别是弱纹理区域，同时为了满足检测的实时性要求，本文采用了 Zhang^[8] 提出的基于十字交叉的代价聚合方法^[8] 对匹配代价进行聚合计算，以减少匹配错误和噪声带来的影响。该代价聚合方法的核心思想就是利用图像的颜色信息和位置关系构建一个合适的支持窗口，其原理如下。

选取图中的任一点 p 为中心，在其相邻的邻域内，在水平方向和竖直方向进行扩展。如图 4 所示，点 p 与 p_l 之间的长度为 PP_l ，并记作左臂 l 。同理可得 $\{u, d, l, r\}$ 分别为上、下、左、右的臂长，且满足两个条件。以左臂为例加以分析。

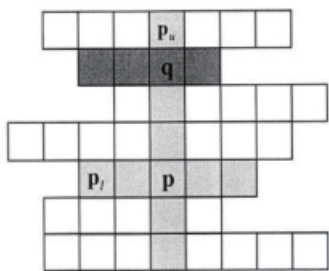


图 4 十字交叉法自适应聚合窗口

(1) P 点与 P_l 之间灰度值的差要小于阈值 t 。

$$|I(p) - I(p_l)| < t \quad (4)$$

(2) P 点与 P_l 点的距离要小于阈值 L_l 。

$$\|PP_l\| = l < L_l \quad (5)$$

基于以上两个条件，可以通过颜色和距离限制臂长的大小，得到最终左右图像的自适应区域 $U(p^l)$ 、 $U(p_d^r)$ ，并通过式 (6) 得到共有区域为最终的聚合窗口 $U(p)$ ：

$$U(p) = \{(x, y) | (x, y) \in U(p^l), (x, y) \in U(p_d^r)\} \quad (6)$$

由公式 (7) 可以得知代价聚合后该区域的代价值：

$$C^i(p, d) = \frac{1}{|U(p)|} \sum_{q \in U(p)} C^{i-1}(q, d) \quad (7)$$

式中： i 为迭代次数，本文的迭代次数为 3，最终代价值为 $C^3(p, d)$ 。

通过上述方法得到聚合代价 $C^3(p, d)$ ，采用赢者通吃的方法进行视察计算，得到点 P 的最终视察 d ，其计算公式为：

$$d = \arg \min_d C^3(p, d) \quad (8)$$

2 试验结果与分析

2.1 运行环境

本文以 Python 为语言基础，以 Pycharm 深度学习框架进行试验，运行环境如表 1 所示。

表 1 运行环境

硬(软)件	名称
处理器	Intel core i5-1035G7
内存	16 GB
操作系统	Win10
语言环境	Python

2.2 视差质量评价

为验证算法的有效性，从数据集中抽取 10 组图像进行测试，同时与一些经典的相似度算法进行比较，采用计算误差匹配像素百分比 (PBM) 和平均视差误差作为量化指标。

2.3 网络结构和参数选择

以 KITTI 数据集为基础进行训练，对三种网络结构进行试验，得到相应的误差匹配百分比和运行时间。

通过表 2 可以得知，随着窗口变大，误匹配像素百分比在减小，而运行时间在增加，当窗口大小 44 像素 × 44 像素时，误匹配像素百分比最小，为 3.89%，此时运行时间最长，为 0.69 s；当窗口大小 18 像素 × 18 像素时，误匹配像素百分比最大，为 4.32%，此时运行时间最长，为 0.38 s。为综合考虑运行时间和误匹配像素百分比，通过引入综合指标 w 值进行比较如式 (9) 所示， w 值越大说明结果越好。因此，窗口宽度为 8 像素 × 18 像素的网络结构作为匹配代驾的网络结构。

$$w = \frac{1 - M}{t} \quad (9)$$

式中： M 为误匹配像素百分比， t 为运行时间。

表 2 实验结果

窗口宽度	18	30	44
误匹配像素百分比 /%	4.32	4.20	3.89
运行时间 /s	0.38	0.47	0.69

2.4 立体匹配算法视差图

以 KITTI2015 数据集的 200 对图像为基础，抛去遮挡等外界因素共提取到 19 986 235 个有效像素点，数据足够多，所以不需要对网络结构进行预处理，通过匹配生成的视差图得到视差图和评测渲染图，如图 5 所示。通过图像可以很清晰地看到采集到的图像大部分区域立体匹配的效果良好，但是在光线较弱的区域，如图 5 (b) 中的右侧区域以及汽车的后挡风玻璃部分误匹配像素较多。

本文随机选取其中 10 组图像与一些经典相似度算法进行对比。



原图 (a)



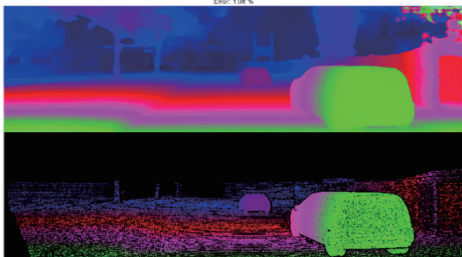
原图 (b)



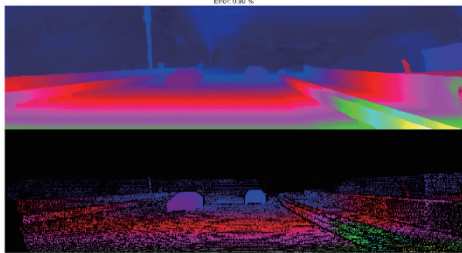
视差图 (a)



视差图 (b)



评测渲染图 (a)



评测渲染图 (b)

图 5 试验结果

表 3 各匹配代价算法的 PBM 和视差

匹配代价方法	PBM 值 /%	平均视差误差 /px
SAD	34.86	23.46
Census	28.43	22.16
AD-Census	20.56	12.78
AD	15.40	9.35
NCC	8.85	5.43
本文算法	4.32	3.42

由表 3 可知，SAD 的算法匹配精度最低误差为 23.46%，AD-Census 算法精度介于 AD 和 Census 两种算法之间的误差为 12.78%，通过结合两种算法，使其具有更强的鲁棒性。NCC 算法相较传统算法匹配精度有一定的提升，但本文算法与之相比，不论是平均视差误差还是误匹配像素百分比，均

有所提升，说明本文提出的基于改进的卷积神经网络立体匹配算法有一定的可行性。

3 结论

本文在卷积神经网络和立体匹配算法的基础下，对卷积神经网络进行深入研究，通过引入池化层对孪生卷积神经网络进行改进，利用 KITTI2015 数据集对网络进行训练，提高匹配正确率。通过测试图像的视差质量，验证了本方法的可靠性。但是在最终的视差图中由于光照、灰度差异等现象，仍存在误匹配的问题，在后续的研究中将对网络继续优化，继续试验，从而达到最优效果。

参考文献：

[1] 马颂德,张正友.计算机视觉:计算理论与算法基础[M].北京:科学出版社,1998.

[2] 陈旭,刘彩霞.双目视觉立体匹配方法研究进展[J].计算机科学,2015,42(10):227-230.

[3] 周芳.双目视觉中立体匹配算法的研究与实现[D].大连:大连理工大学,2013.

[4]BROMLEY J, GUYON I, LECUN Y, et al. Signature verification using a “Siamese” time delay neural network[C]//Proceedings of the 6th International Conference on Neural Information Processing Systems. San Francisco:ICONIP, 1993:737-744.

[5]CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C]//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE, 2005:539-546.

[6] 任珈民,宫宁生,韩镇阳.一种改进的基于孪生卷积神经网络的目标跟踪算法[J].小型微型计算机系统,2019,40(12): 2686-2690.

[7] 梅礼晔,郭晓鹏,郭正红,等.基于空间金字塔池化的深度卷积神经网络多聚焦图像融合[J].云南大学学报(自然科学版),2019(1):18-27.

[8]ZHANG K, LU J, LAFRUIT G. Cross-based local stereo matching using orthogonal integral images[J]. IEEE transactions on circuits & systems for video technology, 2009, 19(7): 1073-1079.

【作者简介】

郭新(1990—),男,江苏徐州人,硕士研究生,助教,研究方向:整车性能评价与试验。

(收稿日期:2023-12-21)