

DOI:10.19651/j.cnki.emt.2107066

基于3D卷积模块和视差分割的立体匹配方法*

赵倩

(河南科技大学应用工程学院 三门峡 472000)

摘要: 为了提高立体视觉的深度估计准确性和运行效率,提出了基于二元视差分割和3D卷积的立体匹配方法。首先,从立体图像中提取特征,并将特征馈入到分割模块;然后,针对每个视差平面单独运行3D卷积模块,检测目标是否比给定距离更近,或根据某个粗略量级进行深度估计;最后,利用3D卷积层估计二元分割模块的输出,通过精细化处理后,得到最终的视差图。实验结果表明,所提方法在ScencFlow数据集的3-px误差为4.37%,EPE误差为1.06 pixel;在KITTI2015数据集上的小额误差接近深度引导聚合网络(GA-Net)方法。且所提方法在不同深度量化级别的运行效率最高。

关键词: 视差分割;立体视觉;3D卷积;立体匹配;误差;视差图

中图分类号: TP391;TN919.81 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Research of stereo matching method based on 3D convolution module and parallax segmentation

Zhao Qian

(College of Applied Engineering, Henan University of Science and Technology, Sanmenxia 472000, China)

Abstract: To improve the depth estimation accuracy and efficiency of stereo vision, a stereo matching method based on binary disparity segmentation and 3D convolution is proposed. Firstly, features are extracted from the stereo image and fed into the segmentation module. Then, for each parallax plane, the 3D convolution module is run separately to detect whether the target is closer than the given distance, or to estimate the depth according to any rough order. Finally, 3D convolution layer is used to estimate the output of binary segmentation module, and the final disparity map is obtained after fine processing. Experimental results show that the 3-px error of the proposed method is 4.37% and the EPE error is 1.06 pixel in ScencFlow dataset. The small error approximation depth guided aggregation network (GA-Net) method on KITTI2015 dataset. And the proposed method has the highest efficiency in different depth quantization levels.

Keywords: disparity segmentation; stereo vision; 3D convolution; stereo matching; error; disparity map

0 引言

基于立体视觉的深度估计^[1](或称立体匹配)是计算机视觉领域的一个热门研究课题,广泛应用于机器人视觉和仪表检测匹配等任务^[2]。目前,很多深度方法不能提供深度量化级别的灵活性,难以确定目标是否在特定距离内,也不能仅关注特定场景范围^[3]。这是因为当前大部分方法的核心思想是检验大量候选视差,并基于成本函数来选择最可能的视差,以完成深度计算。若目标位于候选视差所跨越的范围之外,则大部分方法会将其映射到成本较低的候选视差上,从而造成错误的深度估计。

目前,立体匹配的大部分研究基于离散成本函数计算或固定视差搜索范围,有些研究值得关注。如文献[4]提出了立体匹配算法和设计的分类标准,即,根据匹配成本、优化视差等方式,对立体视觉方法进行分类。文献[5]在立体匹配任务中,提出深度引导聚合网络(guided aggregation-net, GA-Net),捕捉局部和整体图像的代价依赖关系。第1层是半全局聚集层,用于近似可微;第2层是局部引导聚集层,用于过滤并细化薄结构。文献[6]提出一种多分量权值的最小生成树立体匹配方法。文献[7]提出一种基于简化3D卷积神经网络(3D-convolutional neural network, 3D-CNN)的立体匹配方法,设计了一个融合多尺寸上下文

收稿日期:2021-06-25

* 基金项目:河南省科学规划项目(2019JC008)、河南省科学联合会调研课题项目(SKJL-2019-451)资助

信息的特征提取网络,同时,改进了现有相似度计算步骤,该方法在降低参数数量和匹配精度方面有所提高。文献[8]训练了一个深度神经网络,以计算两个不同图像块的匹配成本。文献[9]提出一种简单而有效的卷积空间传播网络(convolution space propagation network, CSPN)来学习深度立体匹配预测。其中传播采用循环卷积运算的方式。目前很多方法为每个像素指定一个搜索范围内的深度^[10-12]。

与之不同,本文将深度估计构建为一个二元分类问题,将空间中的点分类为平面 π 的“前面”或“后面”。平面 π 可被视为立体相机正前方的围栏,用于检测近于安全距离的物体。通过测试多个这样的平面,可以估计一个像素从“前面”切换到“后面”的深度,也就是该像素的深度。并通过设计一个网络架构计算出立体图像对和视差等级。本文方法的主要优势是:1)支持对延迟和深度量级之间的灵活控制;2)在时间限制较为宽松时,所提方法可计算具有不同量化级别的深度,且执行时间随等级数量增加而线性增加。在所提方法中,每个视差平面单独运行一个轻量级的 3D 卷积模块,利用容器上的单个模块和邻近视差平面的信息能够估计出准确完整深度。该立体匹配方法对立体视觉的研究具有借鉴意义。

1 立体匹配方法的局限

给定一个立体图像对 R 和 S ,通过选择视差范围 $\{d_i\}_{i=0,N}$,建立一个平面扫描容积体(planar swept volume, PSV)。相对于左相机图像,PSV 的每个平面可计算为:

$$\text{PSV}(x, y, d_i) = W(S(x, y), H_{\pi_{d_i}}) \quad (1)$$

其中, $W(\cdot, H)$ 表示单应性 H 的扭曲算子; $H_{\pi_{d_i}}$ 表示与视差 d_i 相对应的深度平面所导致的单应性(为简单起见,下文中 π_{d_i} 表示视差 d_i 处的平面)。

给定匹配成本 C ,则算法中像素的视差为:

$$\bar{d}(x, y) = \operatorname{argmin}_{d_i} C(N_R(x, y), N_{\text{PSV}}(x, y, d_i)) \quad (2)$$

其中, N 为像素邻域;不同算法对成本 C 有着不同选择,可以以中心为 (x, y) 的灰度图像块的归一化互相关,也可以是神经网络的输出^[13]。无论哪种选择方式,都遵循:

$$d(x, y) \in [d_0, d_N] \quad (3)$$

式(3)意味着视差在范围 $D = [d_0, d_N]$ 之外的对象依然会被映射到区间 D 中,这是此类方法的一个重要局限,下文将解释如何解决该局限。

2 提出的方法

2.1 核心网络架构

所提方法核心网络的输入为立体图像对和视差等级 d_i ,并生成基于平面 π_{d_i} 的二元分割。该网络结构如图 1 所示。第 1 个模块特征提取模块从立体图像中提取特征。网络输出为一个 32 通道特征图,分辨率为原始图像的

1/3。无论要检验的深度平面数量多少,特征提取模块仅需要运行 1 次。其后,将左相机影像的特征与扭曲后的右相机影像的特征馈入到分割模块,其中,二元分割模块对输入进行 5 次下采样。在编码器的每个尺度上,均有一个步长为 2 的卷积层,其后为一个步长为 1 的卷积层。解码器遵循相同方法,每个尺度上均有一个步长为 2 的反卷积层,其后为一个步长为 1 的卷积层。最后一个卷积层估计二元分割模块的输出,然后双线性上采样至原始分辨率。最后,通过一个轻量级的 3D 卷积层模块对其进行精细化处理。

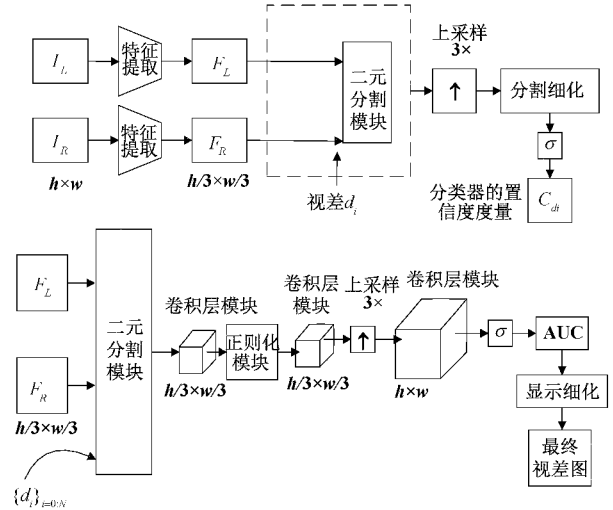


图 1 本文核心网络结构

仅需要多次运行本文网络,即可估计连续或量化分级的深度。为使网络不受视差平面的数量和间隔影响,本文针对每个视差平面单独运行模块 3D 卷积模块。此外,本文应用了一个 3D 编码器-解码器模块(图 1 下方的二元分割模块)以堆叠分割块的输出。利用容器上的单个模块,而非每个置信图上的多个模块,这样就能够利用邻近视差平面的信息,从而提高效率并改进结果。

2.2 对深度的二元分类

一般情况下,视差向量自身的方向已经携带了有价值信息。事实上,在利用式(1)对图像 S 进行扭曲后,视差方向会根据目标在的 $H_{\pi_{d_i}}$ 前方或后方而发生翻转。

这意味着可以训练一个二元分类器,取两个图像 $R(x, y)$ 和 $\text{PSV}(x, y, d) |_{d=d_i}$,并预测 π_{d_i} 后方(或前方)的场景部分。为此,本文利用 CNN 常用的交叉熵损失函数^[11]来训练了一个相对标准的神经网络。

$$L(y, f(x; \theta)) = -y^T \log f(x; \theta) \quad (4)$$

其中, y 为真实值; $f(x; \theta)$ 为模型预测分布。

神经网络将收敛处的分类器输出映射到 $[0, 1]$,以生成如下形式:

$$C_{d_i}(x, y) = \sigma(o(x, y)) \quad (5)$$

其中, o 为网络的输出; $\sigma(\cdot)$ 为 sigmoid 函数; C_{d_i} 可被解读为分类器的置信度量,当 C_{d_i} 接近 1 或 0 时,网络

将确信目标分别处于平面的前方或后方,而 C_{d_i} 的数值接近 0.5 则意味着网络的信任度不足。因此,可将 0.5 作为阈值进行像素分类。

二元深度估计以单个视差为基础,提供了场景相关的有用信息:通过立体相机的周围信息,表明目标与相机是否处于某个特定距离内。当前很多方法无法计算完整深度情况下的二元深度,这是因为这些方法难以测试一组视差集合,并选择最合适的视差。为此,可以针对一组视差平面的集合反复进行这种分类,并在单个容器中不同视差的结果串接在一起,该容器称为置信度容器。

不同位置的视差平面置信度如图 2 所示,其中,对于范围内的目标,在目标的实际视差的置信度将穿过 0.5 水平线;当目标处于预设范围之前或之后时,置信度将分别保持在 1 或 0 附近,且不会越过 0.5 的水平线。假定目标位于视差范围 $[d_0, d_N]$ 之内,如图 2 中的目标 II。对于目标后方较远的视差平面,分类器很可能预测目标 II 处于前方(即 $C=1$)。同理,对于比目标距相机近得多的平面,会预测目标处于后方(即 $C=0$)。但对于靠近目标的视差平面,预测的置信度较低。这是因为平面与目标深度的距离越近,视差向量越小。原则上,会预测出 0.5 的结果。但由于不可避免的分类噪声,仅取曲线相交于 0.5 处的第 1 个视差将导致较大的估计误差。

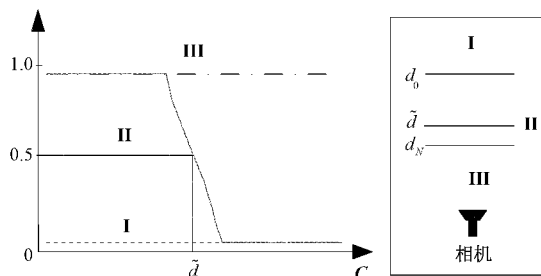


图 2 不同位置的视差平面置信度

为了得到理想视差,可以拟合一个函数,并对视差进行求解,或训练一个交叉点网络。一般,可使用曲线下面积(area under curve, AUC)作为一个简单替代^[15]:

$$d(x, y) = \sum_j C(x, y, d_j) \cdot (d_j - d_{j-1}) \quad (6)$$

式(6)的特点是:首先对错误的置信度数值有一定的容差性;其次估计值可以是连续的值。

2.3 粗略性深度估计

一般情况下,二元深度不足以满足需求,但完全、连续的深度又并非必要。当前方法通过分析所有候选深度,通过 softargmax 等函数得到亚像素视差以进行深度估计。因此,在推理过程中无法改变粒度,或估计任意级别的深度。

对于给定正面平行的平面 π_{d_i} , 其范围分为前后两段,假设置信度 $C_{d_i}(x, y)$ 与一个累积分布函数(cumulative distribution function, CDF)相关:

$$p(d_i \leq d(x, y)) = 1 - p(d_i > d(x, y)) = 1 - C_{d_i}(x, y) \quad (7)$$

其中, $d(x, y)$ 表示像素的视差。由于 $C_{d_i=0} = 1$ 表示所有物体均在零视差平面(即无穷远处的平面)的前方, $C_{d_i=-\infty} = 0$ 表示没有物体在零深度平面的前方。因此,上述 CDF 是合适的。

对于一个较远平面 π_{d_j} , 可表示为:

$$p(d_j < d(x, y) \leq d_i) = C_{d_j}(x, y) - C_{d_i}(x, y) \quad (8)$$

为了得到 L 个量化等级,可利用 $L-1$ 个平面,通过式(8)计算每个量级 bin 的概率,并将像素的视差估计为具有最高概率 bin 的中心 (d_j, d_i) 。这相当于将深度量化分级估计视为硬分割问题。若将其视为软分割问题(即完整深度的情况),可利用这种基于 CDF 的方法简化为第 2.1 节描述的 AUC 方法。

2.4 选择性深度估计

本文假定范围 $[d_0, d_N]$ 覆盖了场景中的所有视差。一般而言,该场景预计 d_0 在 0 处(无穷远处的平面), d_N 在最大视差处。由于无法事先得到该信息,目前很多方法使用 192 个视差等级,其中每个视差等级为 1 pixel 宽, $d_0 = 0$ 。现考虑目标在范围外的情况,例如图 2 中的 III。这种情况下,当前很多方法为了得到最小成本,必须将目标映射到范围内的深度,即使此类方法包含检测范围外目标的策略(例如成本阈值化),但最多也只能发现该目标没有相关深度的信息。

而本文策略可无缝处理范围外的目标,这是由于与目标 III 相关的视差向量方向未发生变化,置信度 C 在整个范围内保持为 1。因此,所提方法不但知道 III 在检验范围之外,还知道其在距离最近的平面前方。若将最远的平面移动到非零视差,则可发现在限制范围之后的对象 I。得益于此,本文能够较好地处理范围选择错误的情况。还可进一步利用该属性,在视差平面的检验存在预算限制的情况下,在整个范围中均匀分布视差平面,其优点是:得到感兴趣区域中的高质量连续深度估计,同时保持相同量级的计算复杂度,对于视差范围 $[d_{\min}, d_{\max}]$, 则可采用 2.2 节方法来处理。

3 实验与分析

3.1 数据集与评价指标

本文使用两个公开数据集进行性能分析。1) SceneFlow 数据集^[16], 包含人工合成的立体匹配数据集,训练图像共 35 454 对,测试图片共 4 370 对。该数据集的标签稠密,视差较为细致,整体质量较为可观。2) KITTI2015 数据集^[17], 该数据包含车辆采集的实景数据,训练图像共 194 对,测试图像 195 对,且均为彩色图像。

评价指标采用端点误差(end point error, EPE)与 l 像素误差,这两个指标是立体匹配的常用指标, EPE 表示视

差估计图所有像素点与真实视差值之间的平均欧氏距离, t 像素误差表示 EPE 大于 t 个像素单位的像素点百分比, t 像素误差用 t -px 简化表示。

此外,还采用了交并比^[18](intersection over union, IoU)作为另一个评价指标,IoU 衡量候选框与原标记框的重叠率。在立体匹配中,可以用来评价匹配的准确率。

$$IoU = \frac{\text{area}(H) \cap \text{area}(G)}{\text{area}(H) \cup \text{area}(G)} \quad (9)$$

其中, H 表示候选框; G 表示原目标框;IoU 值越高表示立体匹配的准确率越高。

3.2 训练与参数

首先在 SceneFlow 数据集上训练本文网络。先以 64 个立体图像对组成一个小批,对于其中每对图像,随机选择用于分割的视差平面。使用二进制交叉熵(binary cross entropy, BCE)损失,训练 1 000 代。利用 8 对立体图像组成一个小批,并使用两个不同损失函数:在分割视差平面上使用 BCE 损失函数;在估计出的视差上使用 L1 范式损失函数,两者权重分别为 0.1 和 0.9。将该网络训练 100 代。

对于 KITTI2015 数据集,对两个网络的权重进行微调。对于二元深度和连续深度,批大小均为 8 对图像,但分别对网络进行 5 000 代和 500 代的训练。使用大小为 384×576 的随机图像裁片和 Adam 优化器,且在所有训练中最大视差平面的采样数为 192 个。

3.3 SceneFlow 数据集结果

本文的目的是实现延迟和深度估计准确度(从二元深度到完整连续深度)之间的平衡。为此在 SceneFlow 数据集上与 3 种先进的立体视觉方法进行了比较:文献[5]提出的 GA-Net 方法、文献[7]提出的简化 3D-CNN 方法、文献[9]提出的 CSPN 方法。这些方法并未设计为直接估计深度量级,因此本文用其计算完整深度,然后对结果做适当量化分级。因为量化分级的深度估计是一个分割问题,所以利用平均 IoU 对结果做评估,将深度标签作为类别使用。

从 SceneFlow 数据集中随机选择 1 000 个图像的不同深度量化级别的平均 IoU 结果如表 1 所示。随着深度量化级别的增加,IoU 值都是下降的。虽然所提方法访问的信息较少(即:视差平面较少),但在同一个级别上,平均 IoU 值较高,因此,从重叠的角度看,所提方法取得了较优性能。

表 1 不同深度量化级别的平均 IoU

等级	GA-Net	3D-CNN	CSPN	本文方法
2	0.965 0	0.967 4	0.953 9	0.970 1
4	0.930 1	0.935 5	0.929 8	0.937 3
8	0.877 1	0.882 8	0.877 8	0.890 2
16	0.806 3	0.811 5	0.810 9	0.830 5

本文方法在 SceneFlow 数据集上的两个测试案例结果如图 3 所示。可以看出,本文输出的图像与真实视差相差

很小,细小的误差集中在左下角区域。直观上看,所提方法的准确性非常高。

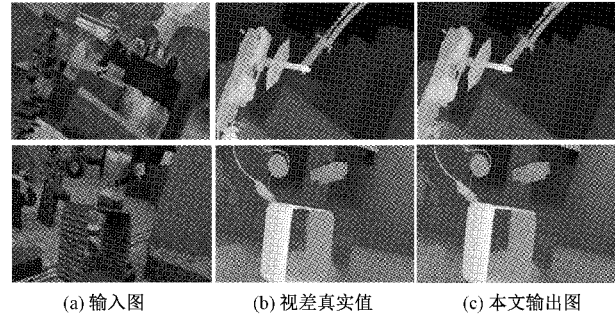


图 3 SceneFlow 数据集的测试结果

在 SceneFlow 数据集上的 t -px(采用 3 pixel 误差)和 EPE 性能对比如表 2 所示。由表 2 可知,与 3 个先进的立体匹配方法相比,本文方法在最小参数数量下,取得了较好的性能。与 GA-Net 相比,本文方法的参数数量明显减少,减少约 28%,EPE 误差增加的很小,但更少的参数数量使得模型更加高效,直接带来运行速度的提升。从表 2 还可以发现,方法的参数数量与立体匹配的精度并不一定成正比关系。例如 CSPN 的参数数量最多,但 EPE 值较大,误差较大。这是因为 CSPN 采用循环卷积运算,这是低效的运算,造成整个模型的臃肿。另外,GA-Net 表现优秀,采用两层结构,第 1 层是半全局聚集层,第 2 层是局部引导聚集层。本文方法的误差与 GA-Net 类似,但所用参数更少。这充分体现所提方法的高效性与精准性。

表 2 在 SceneFlow 测试集上的性能对比

算法	参数数量	3-px/%	EPE/pixel
CSPN	51.13×10^6	6.56	1.29
3D-CNN	3.00×10^6	4.45	1.11
GA-Net	3.90×10^6	3.39	1.03
本文	2.81×10^6	4.37	1.06

3.4 KITTI2015 数据集结果

在 KITTI2015 数据集上的结果对比如图 4 所示。图 4(a)是 KITTI2015 数据集的输入图,这些图包括树木、公路、房屋、天空和车辆等元素。由图 4 可知,本文方法的很多元素视差更准确,比如第 1 行图片的公路路牌视差更加准确,反光的公路面视差更加平滑;第 2 行图片的公路边缘的视差更加准确光滑;第 3 行图片中近点的两个汽车视差更加生动清晰。因此,本文方法的很多细节的深度误差更小,视差更加准确,这主要是因为所提方法可以计算完整深度情况下的二元深度。

各方法在 KITTI2015 数据集上的 t -px 性能如表 3 所示。表中考虑 2-px、3-px 和 4-px,“noc”表示仅非遮挡区域的像素参与计算;“all”表示图中所有像素参与计算。值得一提的是,KITTI2015 数据集在图像数据量上明显少于 SceneFlow

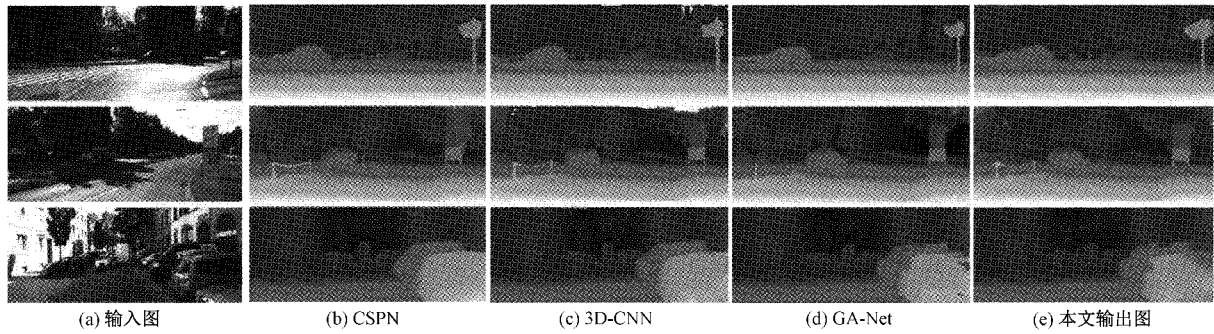


图 4 在 KITTI2015 数据集上的结果对比

数据集,而且 KITTI2015 是自然场景的拍摄图像,而 SceneFlow 多为人工合成图像。由表 3 可知,本文方法依然取得了优秀的匹配精度。总体来看,本文方法的 t -px 性能居第 2 名,明显强于 3D-CNN 和 CSPN,稍微弱于 GA-Net。对此,本文作如下说明:首先,从表 3 中可发现,不管是“noc”还是“all”,从 2-px 误差到 4-px 误差,所提方法与 GA-Net 的性能差距一直在下降。

表 3 KITTI2015 数据集上的数值结果 %

方法	2-px		3-px		4-px	
	noc	all	noc	all	noc	all
CSPN	7.81	8.72	5.13	6.02	3.92	4.12
3D-CNN	3.97	4.76	2.52	3.09	1.90	2.35
GA-Net	2.73	3.79	1.91	2.35	1.46	1.85
本文	2.98	3.99	2.13	2.41	1.48	1.89

本文方法与 GA-Net 在 t -px 误差方面的比较如图 5 所示,在 5-px 到 6-px,两者的误差差距接近于 0。由此可以证明,随着 t 的增大,本文方法的 t -px 误差逐渐接近 GA-Net,这也说明了,在 KITTI2015 数据集上,本文方法的小额误差的概率稍高于 GA-Net,但这种区别不明显。

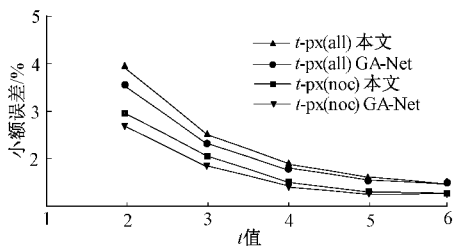


图 5 本文方法与 GA-Net 在 t -px 误差方面的比较

3.5 运行效率

为了深入比较方法在不同时期的运行速度,对不同深度量化级别的运行时间进行统计,结果如表 4 所示,括号中的信息是视频的估计帧率。实验使用 NVIDIA GTX 1050ti,在 2 级量化版本中最快运行速度为 6.9 ms(约 140 fps),在 4 级量化版本中,最快速度为 10.3 ms(约 95 fps),明显低于其他方法。

表 4 不同深度量化级别的运行速度

等级	3D-CNN	GA-Net	CSPN	本文方法
2	10.7 ms	16.7 ms	12.9 ms	6.9 ms
	(约 90 fps)	(约 50 fps)	(约 75 fps)	(约 140 fps)
4	21.1 ms	30.1 ms	20.3 ms	10.3 ms
	(约 45 fps)	(约 30 fps)	(约 45 fps)	(约 95 fps)
8	34.2 ms	46.7 ms	32.1 ms	15.1 ms
	(约 25 fps)	(约 20 fps)	(约 30 fps)	(约 60 fps)
16	40.5 ms	60.2 ms	43.7 ms	20.7 ms
	(约 20 fps)	(约 15 fps)	(约 20 fps)	(约 40 fps)

实验中,4、8 和 16 三个量化等级的深度估计结果如图 6 所示。由图 6 看出,4 级和 8 级一般足以形成场景概况。其中,4 级就可以实现对场景的基本理解,这也从侧面反映了所提方法运行速度非常快。

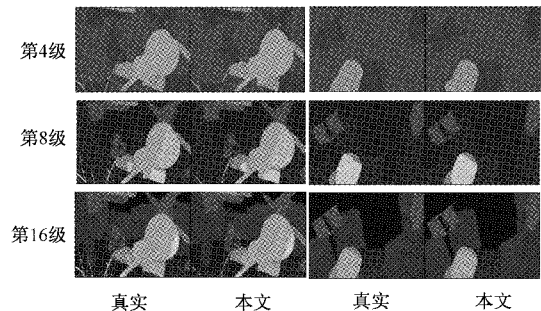


图 6 粗略量级的深度估计结果

4 结 论

当前的立体匹配方法一般对目标是否在特定深度进行测试,本文将该问题分类为两种形式:离该特定深度更近、离该深度更远。为此,提出了用于双目立体视觉深度估计的网络框架,每个视差平面单独运行一个轻量级的 3D 卷积模块,利用容器上的单个模块和邻近视差平面的信息,寻找每个像素在哪个平面上切换标签,从而估计出准确深度。实验结果表明所提方法具有较高的准确性和效率,且效率和准确性非常均衡。未来,本文将研究如何更好区分相似颜色边界,提高边界相似分类的准确性。

参考文献

- [1] 张喆韬, 万旺根. 基于 LRSDR-Net 的实时单目深度估计[J]. 电子测量技术, 2019, 42(19): 164-169.
- [2] 邓欣, 杨清云, 米建勋, 等. 基于相关滤波的仪表定位方法[J]. 电子测量与仪器学报, 2019, 33(5): 107-115.
- [3] CHEN B, JUN C, ZHANG Z. Variational fusion of time-of-flight and stereo data for depth estimation using edge selective joint filtering [J]. IEEE Transactions on Multimedia, 2018, 20(11): 2882-2890.
- [4] JOHANNSEN O, HONAUER K, GOLDLUECKE B, et al. A taxonomy and evaluation of dense light field depth estimation algorithms[C]. Computer Vision & Pattern Recognition Workshops, Honolulu, HI, USA: IEEE, 2017: 98-107.
- [5] ZHANG F, PRISACARIU V, YANG F, et al. GA-Net: Guided aggregation net for end-to-end stereo matching [C]. Computer Vision and Pattern Recognition (CVPR), Long Beach, USA: IEEE, 2019: 157-168.
- [6] 李亚旋, 谢红薇, 王春丹, 等. 多分量权值的树结构立体匹配算法[J]. 计算机工程与设计, 2020, 41(9): 109-114.
- [7] 张文, 邵小桃, 杨维, 等. 基于卷积神经网络的高效精准立体匹配算法[J]. 计算机辅助设计与图形学学报, 2020, 32(1): 48-56.
- [8] BONTAR J, LECUN Y. Computing the stereo matching cost with a convolutional neural network[J]. Computer Science, 2014, 35(7): 119-127.
- [9] CHENG X J, WANG P, YANG R G. Depth estimation via affinity learned with convolutional spatial propagation network[C]. European Conference on Computer Vision (ECCV), Munich, Germany: IEEE Press, 2018: 103-119.
- [10] 罗久飞, 邱广, 张毅, 等. 基于自适应双阈值的 SURF 双目视觉匹配算法研究[J]. 仪器仪表学报, 2020, 41(3): 242-249.
- [11] 李艳艳, 潘晋孝, 刘宾, 等. 基于相似度匹配的场景深度估计方法[J]. 国外电子测量技术, 2021, 40(3): 37-40.
- [12] 张凝. 立体图像深度估计关键技术研究[D]. 天津: 天津大学, 2019: 15-27.
- [13] GUO W, KOTSIA I, PATRAS I. Tensor learning for regression [J]. IEEE Transactions on Image Processing, 2012, 21(2): 816-827.
- [14] LI A Q, YE D X, CLARENCE W, et al. Convolutional neural network-based leakage detection of crude oil transmission pipes[J]. Instrumentation, 2019, 6(4): 1-10.
- [15] PANG J H, SUN W X, JIMMY S J, et al. Cascade residual learning: A two-stage convolutional neural network for stereo matching [C]. International Conference on Computer Vision (ICCV), Venice, Italy: IEEE Press, 2017: 878-886.
- [16] MENZE M, GEIGER A. Object scene flow for autonomous vehicles[C]. Computer Vision & Pattern Recognition, Las Vegas, USA: IEEE, 2015: 1-7.
- [17] 陈春红. 基于多尺度信息和注意力的立体匹配算法[D]. 大连: 大连理工大学, 2020: 27-36.
- [18] 黄龙, 杨媛, 王庆军, 等. 结合全卷积神经网络的室内场景分割[J]. 中国图象图形学报, 2019, 24(1): 64-72.

作者简介

赵倩, 硕士, 讲师, 主要研究方向为深度学习、数据聚类与分析等。

E-mail: plozard0@163.com