

基于多级网络优化的光场深度值估计

向森^{1,2*}, 黄楠婷^{1,2}, 邓慧萍^{1,2}, 吴谨^{1,2}

¹武汉科技大学信息科学与工程学院, 湖北 武汉 430081;

²武汉科技大学冶金自动化与检测技术教育部工程研究中心, 湖北 武汉 430081

摘要 为了精确而鲁棒地估计光场深度值,提出了一种多级神经网络渐进优化的深度值估计方法。采用四级深度神经网络分别对水平、垂直、对角线和反对角线方向的子孔径图像提取特征并估计中心视点的深度图;各子网络内部利用含跳跃连接的编码器-解码器结构提取全局特征和局部特征;各级子网络之间采用渐进优化的结构和训练策略,即前级子网络生成的深度图作为后级子网络的输入以引导其深度值估计。实验结果表明,所提方法可生成高质量的场景深度图,尤其在物体边界处可得到更加准确的轮廓。此外,该方法对不同分辨率的输入图像均具有良好的鲁棒性,还有高效推理深度值的优点,可更好地满足实际应用的需求。

关键词 图像处理; 深度估计; 光场; 卷积神经网络; 编码器-解码器; 极平面图

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.1010009

Estimation of Light Field Depth Based on Multi-Level Network Optimization

Xiang Sen^{1,2*}, Huang Nanting^{1,2}, Deng Huiping^{1,2}, Wu Jin^{1,2}

¹*School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;*

²*Engineering Research Center for Metallurgical Automation and Measurement Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China*

Abstract This study proposes a depth estimation method based on progressive optimization of a multistage neural network to accurately and robustly estimate the depth of light field. A four-level depth neural network is used to extract features from sub-aperture images in horizontal, vertical, diagonal, and anti-diagonal directions and estimate the depth map of the central viewpoint. In each subnetwork, the encoder-decoder structure having a jump connection is used to extract global and local features. The structure and training strategy of gradual optimization are adopted among subnetworks at all levels, i. e., the depth map generated by the former subnetwork is used as the input of the latter subnetwork to guide its depth estimation. The experimental results demonstrate that the proposed method can generate a high-quality scene depth map, particularly at the object boundary. Moreover, the proposed method has good robustness to input images having different resolutions. It has the advantage of efficient reasoning depth value, which can meet practical application requirements better.

Key words image processing; depth estimation; light field; convolutional neural network; encoder-decoder; epipolar plane image

收稿日期: 2021-07-06; 修回日期: 2021-08-15; 录用日期: 2021-08-25

基金项目: 国家自然科学基金项目(61702384)

通信作者: *xiangsen@wust.edu.cn

1 引言

光场数据同时记录了场景光线的强度信息和方向信息,具有高维度和高灵活性的特点。近年来,光场不但催生了计算成像^[1]等新应用,也给图像超分辨率^[2]、目标检测^[3]、三维重建^[4]等既有的图像处理任务带来了变革。在这些应用和任务中,准确的场景深度值是基础和关键,因此关于深度值估计的研究具有重要的科学意义和实用价值。

现有的光场深度值估计方法可以分为传统方法和基于深度学习的方法。传统方法根据数字重聚焦^[5-6]参数或极平面图像(EPI)中线性结构的最优斜率^[7-8]来获取场景深度。Suzuki等^[9]在结构张量分析之前对EPI进行剪切,将不同剪切值的分析结果集成到最终视差图中,解决了光场视差范围有限的问题。Zhang等^[10]通过最大平行四边形窗口之间的分布距离获取EPI斜率。Sheng等^[11]提出了一种在所有可用方向上提取EPI的方法,并利用结构张量在多个方向上对所有的EPI进行局部深度映射和可靠性估计。Tao等^[12]利用重聚焦算法获取焦堆栈图像后,从中提取散焦和匹配性两种线索进行置信度融合得到深度值。Wang等^[13]和Jia等^[14]进一步在不同的遮挡条件下利用光照一致性的原理将角块划分成不同的区域,并分别计算多个区域的匹配代价。Williem等^[15]利用角块和重聚焦图像分别度量约束角畸代价和约束自适应离焦代价,并综合这两种新的数据代价以降低遮挡的影响。此类方法能够有效计算复杂的遮挡边缘,但其效果受限于场景的复杂程度,并且计算量大,容易受到噪声的影响。

近年来,基于深度神经网络的光场深度值估计方法被陆续提出。相对于传统方法而言,基于深度学习的方法具有强大的特征提取和表征处理能力,能够利用多层神经网络从光场数据中提取深度线索并生成深度值。这些网络利用EPI的线性结构特征^[16]或子孔径图像的相关性特征^[17]获取相应场景的深度。Guo等^[18]设计了一个遮挡感知网络来估计光场图像的深度并对遮挡边缘进行优化;Shi等^[19]利用光流网络获取光场的初始深度图,并利用沙漏网络结构优化深度;Yoon等^[20]使用光场卷积神经网络(LFCNN)来提高光场的角度和空间分辨率。然而现有的光场深度估计方法大部分都采用支路权值共享并且进行全网络整体端到端的训练,未能充分利用光场数据不同方向深度线索的一致性和互

补性。同时,神经网络对数据的鲁棒性不足,例如深度值的准确性将随着光场图像分辨率的变化而下降。

针对以上问题,本文提出一种基于多级网络的光场深度值估计方法,以逐级优化的方式获取中心视点的精确深度图。该网络的每一级提取特定方向的深度值线索并估计深度图;同时,以前级网络的深度图作为后级网络的输入,将不同方向深度信息的逐级融合,从而实现深度图的逐级优化,得到精确的深度值。所提方法有以下特点:1) 基于光场不同方向深度线索的互补性和一致性,使用多级子网络结构分别提取水平、垂直、对角线和反对角线方向的深度特征估计深度值,并利用上级深度图引导下级子网络的特征提取,实现逐级融合,得到完整和准确的深度图。2) 各支路网络利用编码器-解码器结构对光场图像进行多尺度的特征提取,同时生成深层语义特征和局部特征,提升在不同分辨率输入下的鲁棒性。3) 基于光场不同方向深度线索的差异性,提出逐级优化的训练策略,以第1支路为起点训练并获取各个支路不同的权值,逐级激活子网络的同时载入上1支路的权重以加快网络的收敛速度。

2 多级网络的深度值估计

2.1 多级网络架构

在Leovy^[21]提出的双平面四维光场模型中,任意光线可利用角度坐标 (u, v) 和空间坐标 (s, t) 来表示,从中提取一组角度坐标和空间坐标 $(s-u$ 或者 $t-v)$ 即可得到EPI。场景点在EPI中的投影为倾斜直线,由于EPI的二维坐标分别是角度和空间坐标,其投影直线的斜率为空间坐标相对角度坐标的变化率,即单位基线的视差,反映的是对应三维场景点的深度。同时,光场包含大量的子孔径视点,从不同方向的基线生成的深度值存在相关性和互补性。基于此,本实验组采用4个方向上的子孔径图像作为网络4支路上的输入,既能减少冗余信息以提高网络训练的效率,又能在获取EPI线性特征的同时捕获多视图之间的相关性特征。受LapSRN^[22]启发,结合光场图像在不同支路具有不同深度信息特征的特点,本实验组在多支路输入的基础上设计一个多级深度神经网络结构,该网络每一级都融合上一级的深度特征,且特征提取模块可以获取对应的支路上的局部特征和上下文信息,如图1所示。

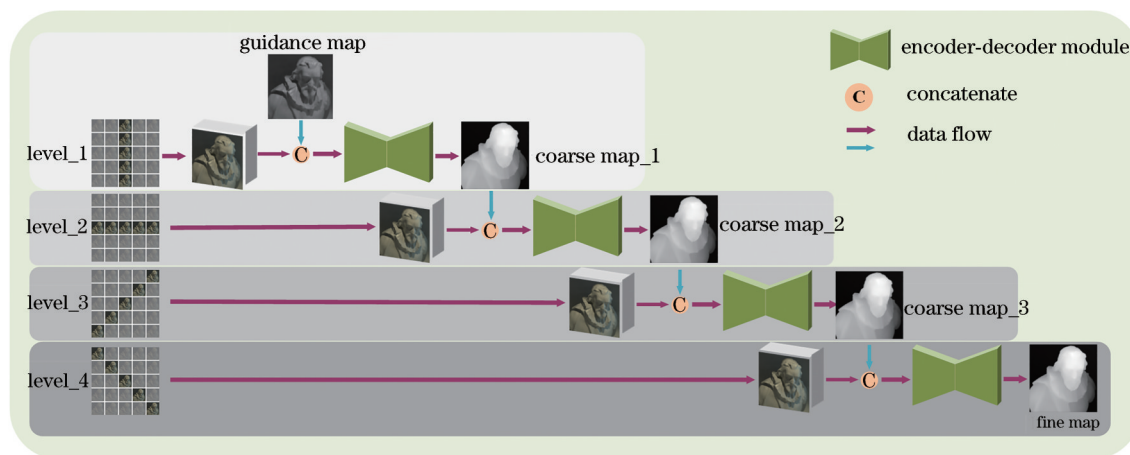


图 1 所提网络架构

Fig. 1 Proposed network architecture

所提网络包含 4 级分支,分别利用水平、垂直、对角线和反对角线方向的子孔径图像来估计中心视点的深度图。考虑支路之间深度线索的相关性和互补性,将前级支路预测的粗糙深度图作为后级支路网络的输入以引导后级支路的训练。所提网络结构具有两方面的优点:一方面,该网络利用各支路深度线索的互补性,渐进地融合了不同方向深度线索,直至最终支路产生完整的深度图;另一方面,该网络利用各方向深度值的一致性,引导和加速了后级网络的训练。需要注意的是,第 1 级支路没有可用的前级深度图进行引导,实验中采用中心子孔径图作为引导图像,以保证各支路结构的一致性。

2.2 多尺度融合特征提取

所提网络中各个支路的基本特征提取单元结构为编码器-解码器模块,如图 2 所示。编码器部分是带有池化层的卷积块,解码器部分是带有上采样层的反卷积块,每个卷积层后面利用 ReLU 激活函数避免线性映射的影响。为了获取更精确的具有

上下文信息的边缘区域特征,编码器中的对应的特征图将通过跳层连接融合到反卷积块中进行进一步的特征提取。此外,通过引入不同尺度的特征,该结构能够提高不同分辨率光场输入情况下的鲁棒性。该模块部分卷积层后添加了 dropout 层以防止过拟合;卷积层使用对称填充,使得网络的输入输出尺寸保持不变,并保留图像边界区域的信息。

2.3 对称填充

卷积神经网络在进行卷积操作时,会根据卷积核的尺寸及卷积核移动的步长改变输出图像的大小。同时,考虑到像素利用率,图像中心部分的像素被利用到的次数更多,而图像边缘的像素利用率低,这就使得经过卷积后的图像丢失边缘信息特征。为了解决这个问题,在卷积层进行卷积前引入了填充操作。一般情况下,神经网络使用零填充的方法解决图像分辨率改变和边缘信息丢失问题,即在对图像进行卷积操作前,在图像周围添加多个值为零的像素。本实验组为了进一步加强边缘像素的使用率,提高物体边缘的精确度,使用对称填充,如图 3 所示。

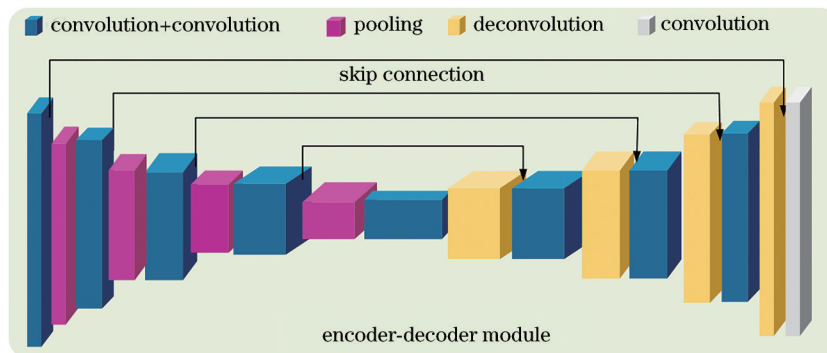


图 2 特征提取模块

Fig. 2 Feature extraction module

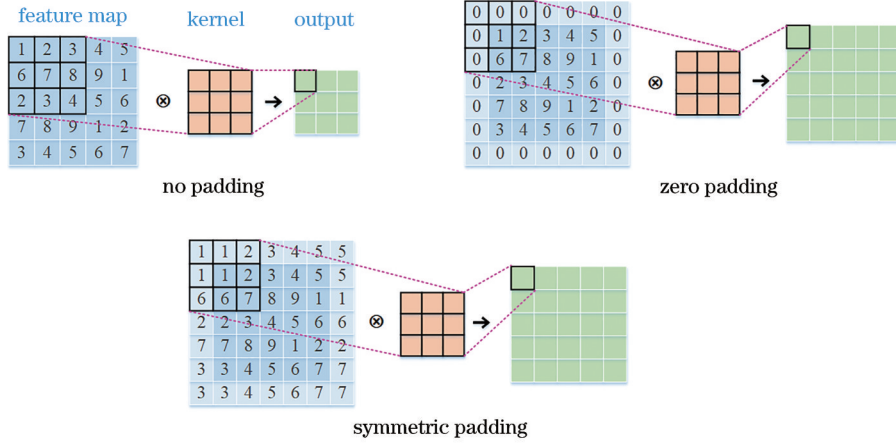


图 3 填充示意图

Fig. 3 Padding diagram

2.4 多级优化的训练策略

光场中不同视图对场景的采集存在差异,这使得不同方向的视图形成空域互补的深度值,将其融合能产生更加准确的深度图。同时,由于各个支路输入的子孔径图像和基线方向均不同,具有不同的特征,若各个支路的权值共享^[19,23],则无法充分挖掘各支路信息的差异性,从而制约网络的性能。因此,本实验组采用多级优化的渐进式训练策略。首先通过随机初始化权重来训练第 1 级网络,待其收敛后,再随机初始化激活第 2 级,继续并同时训练前两级支路直到收敛。然后用同样的方式将第 3 级和第 4 级支路加入训练,直到整个网络完全收敛。值得注意的是,当训练进入到第 K 级 ($K=2,3,4$) 时,第 1 级到第 K 级支路网络的所有参数都在更新。该训练策略可以使 4 个分支的权重参数具有差异性,更好地适应不同方向输入的子孔径图像的采样特征。

3 实验及结果评估

3.1 实施细节

采用 HCI^[24] 和 Wanner^[25] 两个公开光场数据集进行实验,其中训练数据集为 HCI^[24] 中的 16 个综合光场场景,测试数据集有两部分:含有深度真值的 4 个 HCI 场景和 2 个 Wanner 场景;无真值的 4 个 HCI 场景。

由于训练样本只有 16 个角度分辨率为 9×9 、空间分辨率为 512×512 的光场图像,本实验组采用文献^[23]中的方式进行数据扩充,同时去除反射与折射等光照不一致的样本,避免出现深度值估计不稳定的情况,最终将训练样本量增至原始数据的 288 倍。训练中,采用深度值的预测值 y_i 和标注真

值 Y_i 之间的平均绝对误差 (MAE) 作为网络的损失函数。

$$E_{MAE} = \frac{1}{N} \sum_{i=1}^N |Y_i - y_i|, \quad (1)$$

式中: N 表示像素点的个数。训练所采用的优化器为 RMSProp, 批量处理值为 32, 学习率为 10^{-5} 。在处理器为 Intel i7-8700、显卡为 NVIDIA GeForce GTX 1660 的设备上的训练时间为 350 h。

3.2 逐级优化的实验效果

本实验组采用多支路逐级优化的策略来训练所提网络。图 4 为子网络逐级训练后获得的 4 个模型在 HCI 数据集的 4 个场景上进行深度值估计的性能,其中 MSE 为中心视图深度值的均方误差,其值越小代表深度图越准确。从图中可以看出, MSE 随着支路网络的增加而减小,即深度图质量在持续优化。这表明网络级数的增加引入了新的视点信息以产生更准确的深度值;同时网络在前级深度图的深度引导下继续训练,这个过程进一步消除了深度误差,提高了图像质量。

还比较了逐级优化训练和权重共享训练的性能,结果如表 1 所示,其中粗体表示在某一场景中更

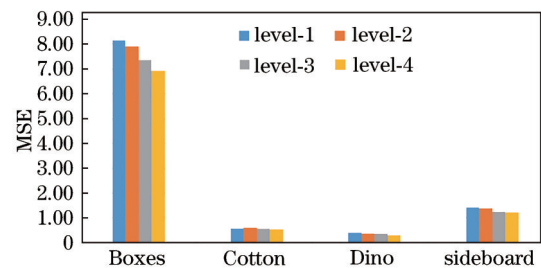


图 4 不同层次训练模型的 MSE

Fig. 4 MSE of different levels of training model

表 1 权重共享策略与所提策略的性能比较

Table 1 Performance comparison between the weight-sharing (WS) strategy and proposed strategy

Scene	BP1 / %		BP3 / %		BP7 / %		MSE × 100	
	WS	Proposed	WS	Proposed	WS	Proposed	WS	Proposed
Boxes	37.52	32.78	23.61	19.47	12.80	13.87	7.43	6.91
Cotton	44.19	30.06	13.16	8.02	2.79	0.99	0.47	0.54
Dino	25.46	19.39	9.44	7.91	3.56	2.60	0.45	0.30
Sideboard	20.98	21.12	10.20	11.96	4.90	6.53	1.26	1.22
Average	32.04	25.84	14.10	11.84	6.01	6.00	2.40	2.24

好的数据,深度图像的质量评估指标为坏点率BP1、BP3、BP7和MSE^[24]，“BPi”表示深度图中绝对误差超过*i*/100的像素点的百分比。从表1可以看出,逐级优化的训练策略效果优于权值共享策略,生成了更加准确的深度图。这是由于各个支路的基线方向不同,所得特征也不相同,权重共享并不能充分利用各支路的互补性;而逐级优化的策略可得到自适应于各个支路特点的权重,从而提高深度值的准确性。

3.3 与现有方法的性能比较

将所提方法与现有主流方法进行了比较,包括:

Shi^[19]、EPINET^[23]、LF^[26]、SPO^[10]、CAE^[15]和LF_PAC^[27]。利用这些方法对HCI数据集^[24]和Wanner数据集^[25]中共6个场景“Boxes”(Bx)、“Cotton”(Co)、“Dino”(Dn)、“Sideboard”(Sd)、“Mona”(Mn)和“Buddha”(Bd)进行深度值估计,并进行定量比较和定性比较以评价其准确性。用于测试的6个场景同时包含了结构简单的平滑区域和结构复杂的遮挡边缘,可全面评估算法性能,其客观质量评估结果如表2所示,其中“Ave”表示Average。从表2可以看出,对BP1指标,所提方法远远优于其他对比方法;随着阈值放松,对比方法的BP3和BP7指标与所提方法逐渐

表 2 不同方法的深度图客观质量对比

Table 2 Objective quality comparison of depth maps with different methods

Method	BP7 / %							MSE × 100						
	Bx	Co	Dn	Sb	Mn	Bd	Ave	Bx	Co	Dn	Sb	Mn	Bd	Ave
LF	23.0	7.8	19.0	22.0	10.7	10.3	15.5	17.43	9.17	1.16	5.07	0.69	1.1	5.77
LF_PAC	22.6	7.4	9.1	10.1	9.4	2.9	10.2	9.92	7.05	0.92	1.10	0.78	0.60	3.40
SPO	14.3	3.2	2.5	7.3	6.5	1.8	5.9	10.88	4.14	0.39	0.99	0.86	0.43	2.95
CAE	17.9	3.4	5.0	9.8	4.7	1.5	7.1	8.42	1.50	0.38	0.88	0.41	0.46	2.01
Shi	22.1	3.3	4.7	10.1	6.4	2.3	8.2	9.16	0.94	0.50	1.37	0.41	0.43	2.14
EPINET	15.1	0.9	1.9	6.6	7.8	1.4	5.6	5.20	0.25	0.19	0.80	0.60	0.26	1.22
Proposed method	13.9	1.0	2.6	6.5	4.9	2.2	5.2	6.91	0.54	0.30	1.22	0.41	0.45	1.64

Method	BP1 / %							BP3 / %						
	Bx	Co	Dn	Sb	Mn	Bd	Ave	Bx	Co	Dn	Sb	Mn	Bd	Ave
LF	65.5	54.2	73.8	59.8	40.4	52.7	57.7	38.4	21.4	45.1	35.4	20.8	23.5	30.8
LF_PAC	79.2	62.1	73.0	72.9	46.3	50.1	63.9	46.0	18.5	31.3	30.1	19.1	11.8	26.1
SPO	69.5	58.3	65.8	70.0	45.6	47.5	59.5	26.7	11.4	15.0	25.3	14.0	8.0	16.7
CAE	72.7	59.2	61.1	56.9	42.2	48.0	56.7	40.4	15.5	21.3	26.8	11.3	8.2	20.6
Shi	68.4	38.2	45.6	63.6	39.6	35.7	48.5	38.5	10.6	14.5	26.3	14.3	8.0	18.7
EPINET	62.4	51.7	41.0	58.6	42.5	39.9	49.2	27.4	4.9	6.6	21.0	14.5	5.2	13.3
Proposed method	32.8	30.1	19.4	21.1	40.9	41.3	30.9	19.5	8.0	7.9	12.0	11.4	7.7	11.1

接近,但仍然存在差距。此外,所提方法在不同场景之间取得的平均坏点率最低,表明所提方法生成了更加准确的深度值。对于MSE,所提方法在大部分场景中取得第1或第2的性能。需要注意的是,MSE计算了误差的平方,因此个别像素点较大的误

差被剧烈放大而产生取值较大的MSE值,故表2中MSE性能与坏点率BPi存在不一致的情况。但总体上看,所提方法产生的误差像素点更少,总体质量更高。[图5(a)]为不同方法的主观结果,[图5(b)]为“Boxes”与“Mona”场景的细节图。

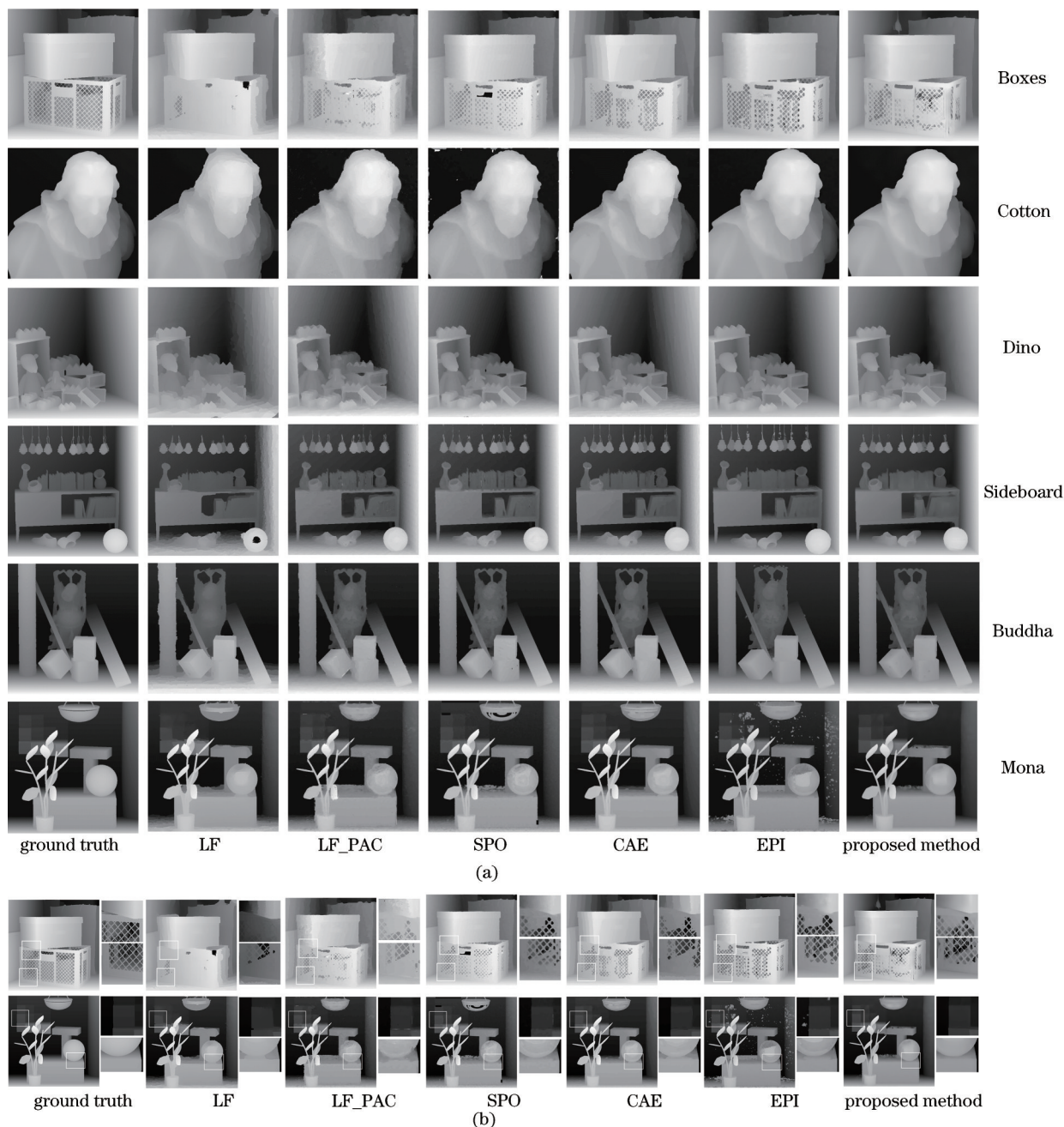


图5 验证数据集的深度图结果比较。(a)主观结果比较;(b)细节比较

Fig. 5 Comparison of depth map results of validation datasets. (a) Comparison of qualitative results; (b) comparison of details

传统方法利用最小化代价函数获取最优的深度值时,通常会面临图像中噪声和物体间的遮挡导致深度值不精确等问题。LF^[26]利用相移定理对子孔径图像进行亚像素偏移,虽然提高了深度估计的精度,

但是该方法更适用于极窄基线情况下的场景,且无法避免噪声和遮挡场景带来问题。LF_PAC^[27]利用多方向上的角相干性质优化了传统方法中的遮挡问题,但是降低了对于噪声的鲁棒性,且仅能处理简单

的单方向遮挡场景,在复杂遮挡下反而引起大量的异常值,如[图 5(a)]中的“Boxes”和“Dino”中多遮挡的效果均较差。SPO^[10]利用旋转平行四边形中左右子窗口的差异最大化准则来提取局部深度值信息,但最佳斜率方向上的差异性容易被忽略,导致遮挡边缘不准确,如[图 5(b)]中“Boxes”的细节图部分,在密集遮挡环境下,SPO的方法在遮挡物边缘表现不佳。CAE^[15]基于颜色相似性对角块中的像素进行加权提高遮挡情况下的鲁棒性,但是该方法过于依赖角块的大小。角块越大,同时包含遮挡物和被遮挡物的像素的可能性就越大,该方法可提升性能;但当场景中无遮挡时或遮挡较小时,该方法反而引起错误的分层,如在[图 5(a)]中“Dino”和“Sideboard”等场景中深度连续的平面上都有明显的分层现象。基于神经网络的EPINET^[23]使用的全卷积网络能够很好地融合不同支路间的特征,但随着网络层次的增加,物体的边缘信息逐渐丢失。所提方法结合了跳层连接,随着网络层次的加深,浅层网络的信息也在深层中进行融合,同时子网络间逐级融入新的特征,在增加网络深度的同时减少了信息的丢失,从图 5 可以看出,所提方法能够

得到更精确的深度图,尤其是产生了更准确的边缘。

为了进一步验证所提网络的有效性,本实验组进一步对HCI中测试数据集下的4个场景“Bicycle”“Bedroom”“Origami”“Herbs”进行深度估计并与上述算法进行对比。这4个场景结构同时包含了多遮挡和复杂遮挡区域以及平滑无纹理区域和反光区域,其中场景“Herbs”中植物间的遮挡,场景“Bicycle”中自行车身的细节的处理难度较大,对算法的要求较高。由于该数据集无Ground truth标注数据,故无法计算客观指标,各方法的主观结果如[图 6(a)]所示。从图中可以看出,不论是传统方法^[10]还是基于深度学习的方法^[23],对于无纹理区域的深度估计均不够精确,如[图 6(a)]第2行场景“Bedroom”中的台灯区域,但所提方法在大部分物体边缘和遮挡区域均取得了更精确的结果。[图 6(b)]展示了“bicycle”和“herbs”细节图。从图中可以看出,所提方法能够获取更加准确的自行车轮廓,并且瓶口边缘和植物的茎叶边缘更加锐利,验证了所提方法对于物体边缘能够获取更加精确的深度值。这是由于所提方法一方面采用了具有

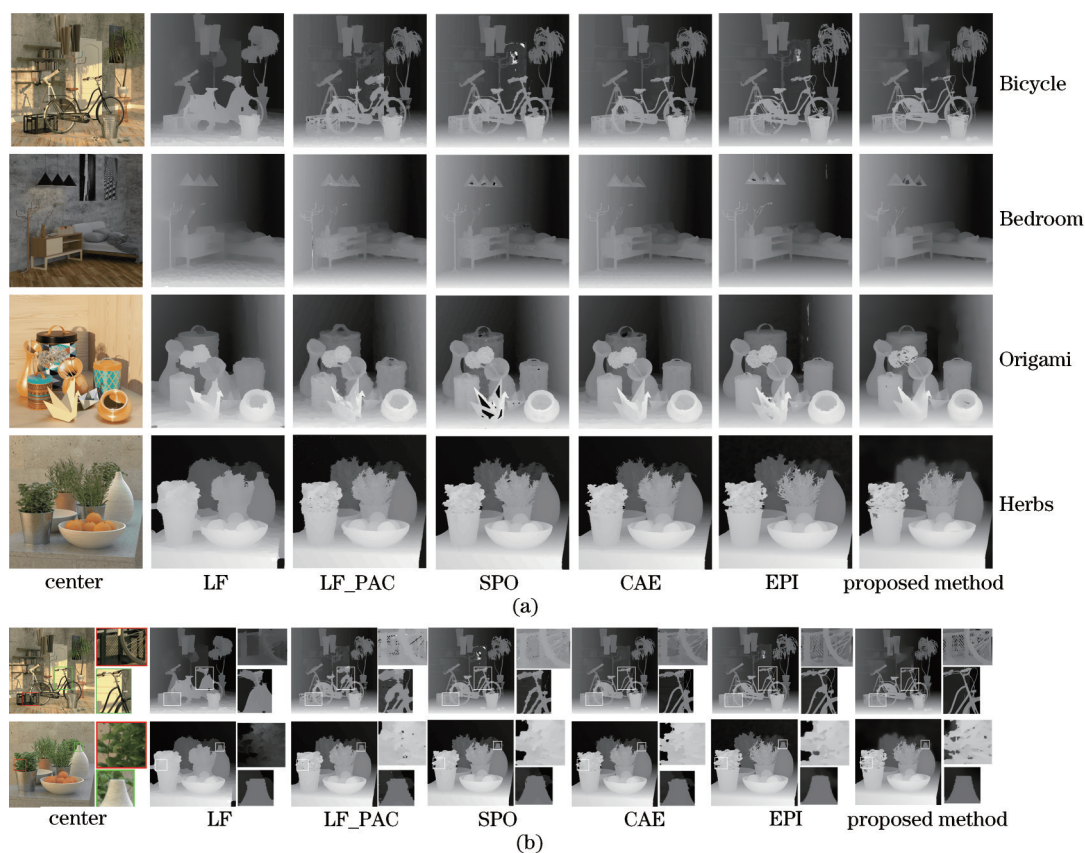


图 6 测试数据集的深度图结果比较。(a) 实验结果图;(b) 细节图

Fig. 6 Comparison of depth map results of test datasets. (a) Experimental results; (b) detail drawing

跨层连接的编码器-解码器进行特征提取,在提取高层语义信息的同时也有效保留了场景的底层特性和细节;另一方面,所提多级网络结构和逐级优化的训练策略能使各个支路子网络充分提取到不同的特征,并逐级对边缘信息进行补充和优化。最后,所采用的对称填充方式也可加强边缘像素的使用率,进一步提高物体边缘的精确度。

3.4 输入分辨率的鲁棒性能评估

在实际情况中,光场数据的空间分辨率并不恒定,网络的性能也随着光场数据的空间分辨率的变化而变化,因此,网络在不同尺度输入下深度值估计的鲁棒性也十分重要。

基于此,对 HCI 的 4 个合成光场数据进行空间下采样,获得空间分辨率分别为 256、128 和 64 的光场数据,将这些数据送入网络获取对应的深度图,并计算其 MSE 评估质量。同时,采用 EPINET^[23] 作为对比基准,该方法与所提方法具有相似的网络结构,但未考虑多尺度问题,结果如图 7 所示。从图 7 可以看出,对于不同尺寸的输入图像,所提方法对应的柱形条纹图的高度变化更小。进一步地,对图 7 中不同分辨率的输入所得结果的 MSE 进行了方差计算,结果如表 3 所示。从表 3 可以看出,所提方法所得深度图 MSE 的方差远小于 EPINET。这些结果表明,由于采用了带跨层连接的解码器-编码器模块,所提方法能够稳定地对不同空间分辨率的光场数据进行深度值估计,具有良好的鲁棒性。

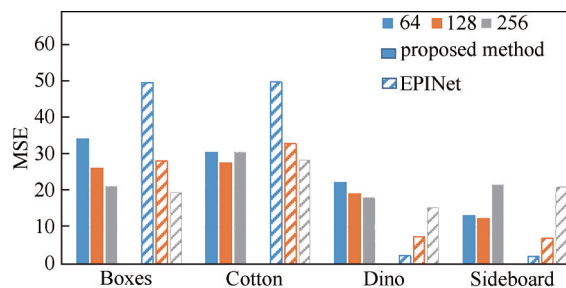


图 7 输入光场空间分辨率不同情况下的 MSE

Fig. 7 MSE under different spatial resolution of input light field

表 3 图 7 场景的 MSE 方差

Table 3 Variance of MSE of scenes in Fig. 7

Method	Box	Cotton	Dino	Sideboard
EPINET	237.33	125.26	44.58	94.40
Proposed method	43.20	2.82	4.86	26.16

为了进一步验证所提方法的鲁棒性,与 EPINET 对 Lytro Illum 相机捕获的真实光场数据^[28] 的测试结果及对不同空间分辨率的真实光场的测试结果进行了比较,结果如图 8 所示。从整体上可以看出,所提方法不仅能够准确估计真实场景的深度图,也能够获取更加精确的物体轮廓。从多尺度上可以看出,当输入光场的空间分辨率改变时,EPINET 受空间分辨率的影响更大,其深度估计结果存在明显的噪声;而所提方法在输入不同尺度的光场情况下,获取深度值更加准确且变化较小,验证了所提方法的鲁棒性。

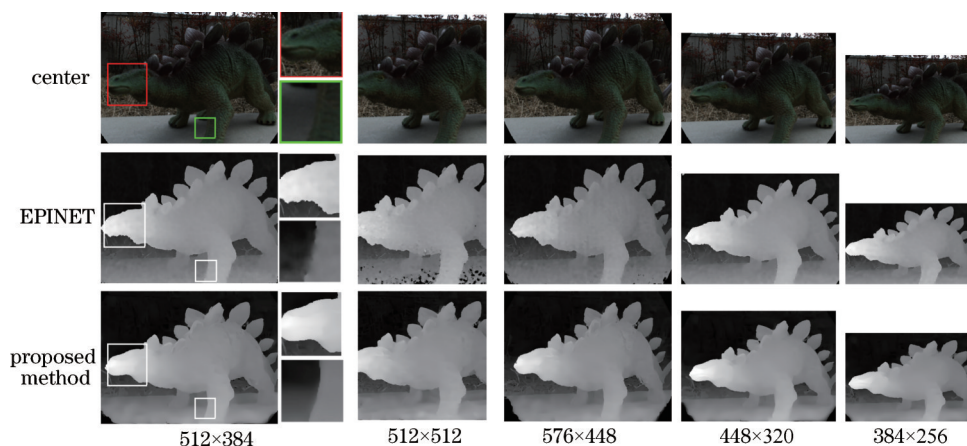


图 8 真实光场实验结果比较

Fig. 8 Comparison of experimental results of real light field

3.5 深度值推理效率对比

除了精确度方面的性能对比,本实验组还在同一硬件设备下对不同方法的深度值计算效率进行了评估。对 HCI 数据集中空间分辨率为 512×

512 的 8 个仿真场景进行了深度值估计,比较了所提方法和 EPINET^[23] 进行深度值获取所耗费的时间,结果如表 4 所示。所用设备的处理器为 Intel i7-8700、显卡为 NVIDIA GeForce GTX 1660。所

提方法和 EPINET 均具有 4 支路结构,但对深度获取模块及融合策略有不同的设计。从表 4 可以看出,所提方法更加高效。这是由于 EPINET 使用全卷积网络,增加了计算的复杂度;而所提方法

在编码器-解码器模块中利用池化层提取多尺度特征的同时,在确保深度图质量的基础上,降低了维度,减少了计算量,从而提高了深度值计算效率。

表 4 测试时间对比

Table 4 Comparison of test time

unit: s

Method	Boxes	Cotton	Dino	Sideboard	Bicycle	Bedroom	Origami	Herbs
EPINET	1.15911	1.16394	1.15558	1.14225	1.07707	1.08461	1.07997	1.08541
Proposed method	0.42078	0.41622	0.42040	0.42168	0.40537	0.40304	0.40388	0.39573

4 结 论

针对光场深度值估计,提出了一个由粗糙到精确逐级优化的深度神经网络,该网络包含 4 级子网络,每级利用特定方向的子孔径图像估计光场中心视点的深度值。子网络采用带有跳层连接的编码器-解码器模块来提取全局和局部特征,进而计算深度值。在各支路之间,前级子网络的深度值将融合后级的输入以引导训练。所提网络采用逐级优化的训练策略,以融合更多的深度线索,提高了深度值的精度。实验结果表明,与现有方法相比,所提方法可以生成更精确的深度图;此外,对输入的分辨率变化具有鲁棒性和更高的效率,因此可以适用于更多的应用。所提方法虽然能够获取准确的深度图并计算出精确的物体边缘,但在弱纹理区域容易产生较大的误差,下一步工作将围绕提高光场场景中弱纹理区域的深度估计以提高整体的光场深度估计精度来展开。

参 考 文 献

- [1] 方璐,戴琼海. 计算光场成像[J]. 光学学报, 2020, 40(1): 0111001.
Fang L, Dai Q H. Computational light field imaging [J]. Acta Optica Sinica, 2020, 40(1): 0111001.
- [2] Wu G C, Liu Y B, Dai Q H, et al. Learning sheared EPI structure for light field reconstruction[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28(7): 3261-3273.
- [3] 段仲静,李少波,胡建军,等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12): 120005.
Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.
- [4] 殷永凯,于锴,于春展,等. 几何光场三维成像综述[J]. 中国激光, 2021, 48(12): 1209001.
Yin Y K, Yu K, Yu C Z, et al. 3D imaging using geometric light field: a review[J]. Chinese Journal of Lasers, 2021, 48(12): 1209001.
- [5] 武迎春,程星,谢颖贤,等. 边缘复杂光场图像的深度估计散焦响应函数优化[J]. 激光与光电子学进展, 2020, 57(18): 181027.
Wu Y C, Cheng X, Xie Y X, et al. Defocusing response function optimization in depth estimation of boundary complex light field image[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181027.
- [6] 刘晓旻,杜梦珠,马治邦,等. 基于遮挡场景的光场图像深度估计方法[J]. 光学学报, 2020, 40(5): 0510002.
Liu X M, Du M Z, Ma Z B, et al. Depth estimation method of light field image based on occlusion scene [J]. Acta Optica Sinica, 2020, 40(5): 0510002.
- [7] Lü H J, Gu K Y, Zhang Y B, et al. Light field depth estimation exploiting linear structure in EPI[C]// 2015 IEEE International Conference on Multimedia & Expo Workshops, June 29-July 3, 2015, Turin. New York: IEEE Press, 2015: 15330297.
- [8] Neri A, Carli M, Battisti F. A maximum likelihood approach for depth field estimation based on epipolar plane images[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28(2): 827-840.
- [9] Suzuki T, Takahashi K, Fujii T. Disparity estimation from light fields using sheared EPI analysis [C]// 2016 IEEE International Conference on Image Processing, September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 1444-1448.
- [10] Zhang S, Sheng H, Li C, et al. Robust depth estimation for light field via spinning parallelogram operator[J]. Computer Vision and Image Understanding, 2016, 145: 148-159.
- [11] Sheng H, Zhao P, Zhang S, et al. Occlusion-aware

- depth estimation for light field using multi-orientation EPIS[J]. *Pattern Recognition*, 2018, 74: 587-599.
- [12] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 673-680.
- [13] Wang T C, Efros A A, Ramamoorthi R. Occlusion-aware depth estimation using light-field cameras[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 3487-3495.
- [14] Jia Y, Li W H. Multi-occlusion handling in depth estimation of light fields[C]//2017 IEEE International Conference on Multimedia & Expo Workshops, July 10-14, 2017, Hong Kong, China. New York: IEEE Press, 2017: 13-18.
- [15] Williem, Park I K, Lee K M. Robust light field depth estimation using occlusion-noise aware data costs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(10): 2484-2497.
- [16] Heber S, Pock T. Convolutional networks for shape from light field[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3746-3754.
- [17] Zhou W H, Zhou E C, Yan Y X, et al. Learning depth cues from focal stack for light field depth estimation[C]//2019 IEEE International Conference on Image Processing, September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 1074-1078.
- [18] Guo C L, Jin J, Hou J H, et al. Accurate light field depth estimation via an occlusion-aware network[C]//2020 IEEE International Conference on Multimedia and Expo, July 6-10, 2020, London, UK. New York: IEEE Press, 2020: 19870565.
- [19] Shi J L, Jiang X R, Guillemot C. A framework for learning depth from a flexible subset of dense and sparse light field views[J]. *IEEE Transactions on Image Processing*, 2019, 28(12): 5867-5880.
- [20] Yoon Y, Jeon H G, Yoo D, et al. Light-field image super-resolution using convolutional neural network [J]. *IEEE Signal Processing Letters*, 2017, 24(6): 848-852.
- [21] Levoy M, Hanrahan P. Light field rendering[C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques-SIGGRAPH '96, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 31-42.
- [22] Lai W S, Huang J B, Ahuja N, et al. Deep Laplacian pyramid networks for fast and accurate super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5835-5843.
- [23] Shin C, Jeon H G, Yoon Y, et al. EPINET: a fully-convolutional neural network using epipolar geometry for depth from light field images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4748-4757.
- [24] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[M]//Lai S H, Lepetit V, Nishino K, et al. *Computer vision-ACCV 2016. Lecture notes in computer science*. Cham: Springer, 2017, 10113: 19-34.
- [25] Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields[C]//18th International Symposium on Vision, Modeling, and Visualization (VMV), September 11-13, 2013, Lugano, Switzerland. Geneva: Eurographics Association, 13: 225-226.
- [26] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1547-1555.
- [27] Guo Z H, Wu J L, Chen X F, et al. Accurate light field depth estimation using multi-orientation partial angular coherence[J]. *IEEE Access*, 2019, 7: 169123-169132.
- [28] Bok Y, Jeon H G, Kweon I S. Geometric calibration of micro-lens-based light field cameras using line features[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 39(2): 287-300.