光场深度估计方法分类及性能比较研究*

苏　博　妮**

(四川文理学院智能制造学院，四川 达州　635000)

摘要:根据不同的光场特征,将现有的光场深度估计方法分为基于极平面、基于多视角、基于焦点堆栈和基于深度学习4类方法.比较和分析这4类方法的特点,阐述了光场参数化表征和光场可视化.在合成和真实光场数据集上,对5种典型深度估计方法分别进行了定量和定性的实验对比分析,研究结果表明,基于深度学习的深度估计方法在运行效率和估计精度上都明显优于传统方法.最后,对光场深度估计方法存在的挑战和未来发展的方向进行展望.

关键词:光场;深度估计;对比分析;深度学习

中图分类号:TP391　　**DOI**:10.19789/j.1004-9398.2021.02.006

0　引　言

深度估计是三维重建过程中的关键问题.常见的深度估计方法包括单目、双目或多目深度估计.单目深度估计是指从单张二维图像中提取场景的深度信息,常用的单目线索有纹理、梯度、散焦/模糊等.双目或多目深度估计(立体视觉)是利用三角测量原理从2张或多张二维图像中获取场景的深度信息.近年来,随着计算摄影技术的不断进步,光场相机(lytro和raytrix)在消费和工业领域都取得了巨大的成功.光场相机具有“先拍照,后聚焦”的特性,使其被广泛应用在三维重建、显著性检测和超分辨率重建等计算机视觉领域.不同于传统的成像设备,光场成像设备可以通过一次曝光、同时记录光线的位置信息和角度信息.正因为这个特点,基于光场图像的深度估计方法成为国内外研究者关注的热点,并取得了一系列研究成果.然而,由于光线角度信息的欠采样、场景存在遮挡、噪声以及非朗伯体等因素的影响,光场深度估计仍然是一个富有挑战性的课题.

光场深度估计方法是利用光场的不同特征获取场景深度信息,且具有自身的特点和适用场景.如:基于极平面的深度估计方法是将三维场景中的点投影到光场不同视角中,形成的直线斜率与视差之间具有比例关系,从而计算出场景深度;基于多视角的深度估计方法,将光场图像转化为多目立体视觉问题,其核心问题是寻找不同视角之间的对应点;基于深度学习的深度估计方法是将光场深度估计看做视差回归问题,利用深度神经网络强大的表征和泛化能力,通过构建四维匹配代价量,搜索不同视角之间的匹配点,从而估计视差图.然而,由于光场相机存在着基线较短,空间和角度分辨率不高的问题,传统立体匹配方法很难直接用于光场深度估计.尽管深度估计方法在运行效率和估计精度方面都取得了较好的效果,但是有监督的深度学习方法需要大量的带标签的训练数据集,而现实场景的视差真值很难获取.由此可见,无论是传统的深度估计方法,还是基于深度学习的深度估计方法都存在着诸多问题和挑战,值得研究者们深入研究和探讨.但是,现有的文献却很少对各种光场深度估计方法进行系统性的比较和分析,鉴于此,本文对目前典型的光场深度估计方法进行定性比较和定量分析,以期为探索光场深度估计未来的研究方向提供一定的参考.

1　光场深度估计方法

传统的深度估计方法主要采用立体匹配的方法,通常是从2幅图像中寻找对应点从而进一步计算视差和深度.光场图像由于包含了丰富的光线角

收稿日期:2020-07-08

* 四川省教育厅科研项目(18ZA0419);四川文理学院智能制造产业技术开发研究专项项目(2017ZZ005Y,2017ZZ002Z);四川文理学院科研项目(2017KZ012Y)

** 通信作者:170972196@qq.com

度信息,可更好地反映场景的深度信息.根据不同的光场特征,将现有的光场深度估计方法分为4类:基于极平面(epipolar plane image,EPI)、基于多视角、基于焦点堆栈和基于深度学习的方法.

1.1 EPI的方法

EPI的方法通过分析光场的结构特征,将四维光场投影到二维EPI图像上,利用直线斜率和深度之间的比例关系,将深度估计转化为求直线斜率问题.Bolles等[1]最早提出EPI的概念和获取方式,并通过在EPI中直线拟合来检测边缘、波峰和波谷,从而重建三维结构;Wanner和Goldluecke[2]提出一种全局一致性标记算法,采用结构张量提取EPI图像中直线的方向,并通过全局优化获取深度.在此方法的启发下,研究者们提出了许多新的方法:Li等[3]采用迭代共轭梯度法构建了一种稀疏线性系统,利用结构张量在EPI中获取局部深度图,然后采用共轭梯度、均值漂移等方法对局部深度图进行全局优化;Kim等[4]提出一种高分辨率图像重建复杂场景的方法,该方法将光场数据进行稀疏表示,且并行处理,采用一种由精到粗深度传播策略提取了高精度的深度信息;Zhang等[5]提出了一种平行四边形算子,通过距离计算EPI中斜线的斜率,该方法对噪声和遮挡都有很好的鲁棒性;Zhang等[6]通过从EPI中选取最优的直线斜率计算局部深度图,利用置信度将像素分为可靠和不可靠2类,对于不可靠的像素采用可靠像素的视差传播来填充以优化深度图.可见,基于EPI的方法可以在一定程度上缓解由光照和镜面反射等因素造成的错误估计,但EPI在提取过程中计算量较大、时间复杂度高,故实时性较差.

1.2 基于多视角的方法

根据光场相机的成像原理,可以将光场图像看作虚拟的相机阵列从不同视角拍摄的场景图像集合,所以,可以将多视角立体匹配方法应用于光场图像深度估计.基于多视角的方法,利用各个视角之间颜色一致性等匹配关系估计场景深度.Yu等[7]探索了射线空间中3D线的几何结构,以改善光场的三角测量和立体匹配,由于该方法使用立体匹配估算每个像素的视差,因此,在视差范围较小的数据集上效果不佳;Heber和Pock[8]提出一种主成分分析匹配项,用于多视图立体重建,假设如果要将各个子视图图像投影到同一个投影中心,然后将每个投影的图像视为矩阵中的一行,则矩阵应该是低秩的;Chen等[9]采用表面相机的概念对3D点的角辐射度分布进行建模,并引入了双边一致性度量用于光场深度图估计;Jeon等[10]针对光场相机窄基线的问题,提出基于傅里叶变换的相移理论,创新性地将求子视角之间的亚像素级视差变换为求频域的相移,以中心子视图为参考,对其他子视图进行立体匹配构建深度图,最后采用迭代优化的算法构建了连续的、精度高且平滑性较好的深度图,该方法需要为每个子视角构造代价函数,时间复杂度较高;熊伟等[11]设计了一种自适应成本量的抗遮挡算法,分别为无遮挡、单遮挡和多遮挡类型的场景设置了3种不同的成本量,自适应地选择成本量从而获得局部深度图,最后通过加权中值滤波、图割算法和马尔科夫随机场等方法进一步优化结果;Liu等[12]提出一种基于立体匹配的方法获取光场视频和图像深度图的方法,首先对光场数据进行渲染增强,然后基于傅里叶相移理论求得局部深度图,最后基于图割的多标签优化算法获取精确的深度图.以上基于多视角的深度估计方法虽然在运算效率方面较高,但却很难抑制匹配噪声.

1.3 基于焦点堆栈的方法

数字重聚焦技术是光场相机流行的最大亮点.重聚焦图像是通过平移和叠加各个子视角的图像,使得图像中处于焦点范围内的物体清晰、焦外物体散焦模糊.焦点堆栈是指利用光场重聚焦技术,得到一组聚焦于不同深度的重聚焦图像,其利用光场重聚焦生成焦点堆栈,通过聚焦性检测函数对重聚焦图像进行检测提取深度信息.当场景中的一点在图像中处于最佳聚焦点时,所对应的焦距即为该点的最佳焦距,利用透镜成像公式可计算出深度值.Tao等[13]首次提出将散焦线索和一致性线索相融合获取局部深度图,再采用马尔科夫随机场进行全局优化;Tao等[14]又将阴影线索引入深度估计,利用角度相干性的颜色、深度和阴影一致性等属性来约束和校正深度信息,以此来估计物体形状和深度;Wang等[15]为了更好地处理遮挡问题,提出一种遮挡感知的深度估算方法,通过仅在被遮挡的视图区域中确保成像一致性,该方法可以很好地识别被遮挡的边缘,从而改善深度估计效果;Williem和Park[16]在此遮挡模型的基础上,采用角度熵和自适应散焦相应地提高算法对遮挡的鲁棒性和噪声的敏感性;Zhu等[17]将上述方法扩展到多遮挡情况;杨德刚等[18]在深入分析相机阵列成像时线索互补的特性,提出一种融合模糊和视差双线索的深度估

计方法,该方法通过合成孔径成像方法,获得指定深度上的重聚焦图像,提取不同维度的模糊和视差线索,采用自适应权重融合算法优化融合2种线索值,在马尔科夫随机场上对深度图进行邻域平滑,从而获得全局一致性深度图;胡良梅等[19]针对散焦线索产生误差较大的问题,采用梯度均方差和拉普拉斯算子双聚焦函数检测散焦线索,并通过全聚焦彩色图像对局部深度图进行平滑、填充和滤波,最后在马尔科夫随机场上进行融合优化;Tian等[20]在处理散射媒介中光场图像的恢复和深度估计时,提出了散焦、对应性和传播3种深度线索融合的方法,从而消除了光线反向散射和衰减的影响;韩磊等[21]深入剖析了聚焦线索和对应线索的几何特性,引入了遮挡和噪声的度量标准,将深度估计问题转化为能量最小化问题,采用图割法求得最优的深度图.基于焦点堆栈的方法能够保持更多的图像边缘细节,但对低纹理区鲁棒性不好.

1.4 基于深度学习的方法

随着深度学习在图像分类、分割和识别等高级计算机视觉任务中的广泛应用,基于深度学习的光场深度估计方法也应运而生.Heber和Pock[22]提出一种利用深度学习策略,估计给定光场深度信息的方法,针对每个成像的场景点预测光场域中相应2D超平面的方向,并使用4D正则化来克服无纹理或均匀区域中的预测误差;Johannsen等[23]提出一种采用经过特殊设计的稀疏分解学习方法,通过学习光场中心子视图的结构,并使用此信息来构造一个光场字典用于编码光场的稀疏表示,分析该表示相应原子的差异系数,得出准确而可靠的深度估计;苏钰生等[24]基于改进密集连接型网络估计光场深度,减少了模型的计算量,解决了网络随着层数增加而出现梯度消失的问题;Luo等[25]以十字EPI块为卷积神经网络的输入,并采用图割法全局优化策略对结果进行优化;Peng等[26]提出一种基于卷积神经网络的无监督光场深度估计方法,设计了一种组合损失函数,利用四维光场中的空间和角关联性来约束网络;Shin等[27]设计了一种基于全卷积神经网络的光场深度估计方法,通过数据增强的方法解决了训练数据不足的问题,在效率和精度方面都取得了较好的效果.相比于前3种方法,基于深度学习的光场深度估计方法才刚刚起步,还面临着缺乏训练数据等诸多困难,这方面研究还有很大的提升空间.

综上可知,光场图像由于包含了丰富的场景信息,使得估计深度成为可能.相对于传统的立体视觉深度估计,光场图像深度估计不需要相机标定,从而简化了深度估计过程.光场图像存在着体积大、计算耗时、基线过短和分辨率低等问题,且目前大多数深度估计方法都只适用于朗伯场景下,如何解决这些问题依然是光场深度估计方向的研究热点.为深入比较不同深度估计方法的性能,本文在阐述光场参数化表征和光场可视化方法的基础上,将对基于多视角的光场(light field,LF)[10]、基于焦点堆栈的光场遮挡模型(light field occusion,LF_occ)[15]和代价角度熵(cost augular entropy,CAE)[16]、基于EPI的旋转平行四边形算子(spinning parallelogram operator,SPO)[5]和基于深度学习的极平面网络(epipolar plane image net,EPINET)[27]进行实验对比分析,以期进一步分析不同方法的特性.

2 光场参数化表征和可视化

2.1 光场双平面表征

光场的概念最初是由Gershun[28]提出,用来描述空间中光线在各个位置和方向的辐射.随后,Adelson和Bergen[29]提出用七维全光函数$L=L(x,y,z,\theta,\varphi,\lambda,t)$刻画光线在三维空间中的传播,其中$(x,y,z)$为空间坐标,$(\theta,\varphi)$为光线的方向,$\lambda$和$t$分别表示光的波长和时间.七维全光函数存在着维度高、获取困难等问题.为了进一步简化全光函数,Levoy和Hanrahan[30]于1996年提出光场渲染理论,将全光函数简化为四维,采用双平面表示法将光线用2个平行平面的交点来参数化表征,即$L=L(u,v,s,t)$,如图1所示.聚焦平面$\Pi_{uv}=\{(u,v)\}$描述光场的空间分布,相机平面$\Pi_{st}=\{(s,t)\}$描述光线的角度方向,f为2个平面的距离.光场双平面表征是目前光场最常用的4D光场模型.

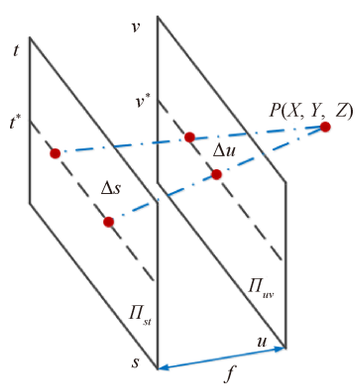


图1 光场双平面参数化表征

2.2　光场可视化

为了获取光场图像,研究者们已经设计了许多成像设备,如编码孔径相机、多自由度相机转台、相机阵列和微透镜阵列光场相机.光场相机是在主透镜和感光器之间插入一个微透镜阵列,用以捕获光线的角度信息.然而,由于受感光器总分辨率的限制,光场相机存在空间和角度分辨率折中的问题.微透镜阵列的个数决定光场图像的空间分辨率,微透镜的直径大小决定光场图像的角度分辨率.每个微透镜对应感光器上的一块区域,称之为宏像素[31].光场相机拍摄的光场图像是由多个宏像素组成,从光场双平面表示的角度解释,即固定了4D光场中的二维空间坐标(u,v),如图2(a)所示.如果取出每个宏像素相同位置的像元组成一幅图,则得到光场的子孔径图像,相当于固定4D光场中的二维角度坐标(s,t),如图2(b)所示.如果分别固定一个空间坐标和一个角度坐标v和t(或者u和s),则可得到光场极平面图像,从图2(c)可知,光场图像拍摄的特殊性使得EPI图像具有线性结构,因此,可根据其线性特性及三角测量原理得到场景中物体的深度.3D空间中某一深度为Z的点P,当改变坐标s时,u坐标会根据公式$\Delta u=\frac{f}{Z}\Delta s$而变化.空间不同深度的点在EPI中形成不同斜率的直线,基于EPI的深度估计方法正是利用这个特点来估计深度.

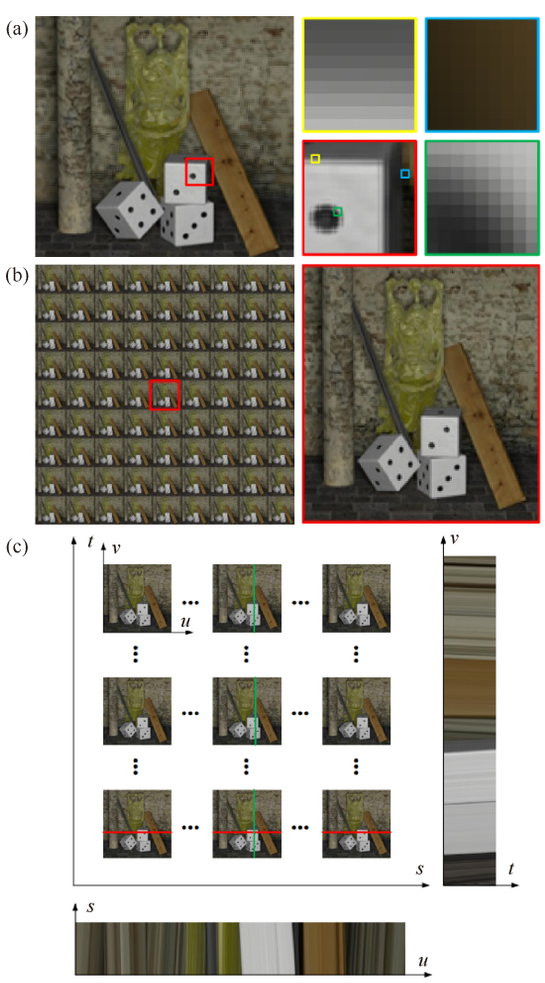


图2　光场图像的可视化[31]

(a)光场宏像素;(b)光场子孔径图像;(c)光场极平面图像

3　实验结果与讨论

3.1　实验环境

本实验是在一台硬件配置为CPU:Intel Core i5-4590 3.3 GHz×4cores,RAM:8 GB和编程环境为Matlab R2014a、TensorFlow2.0的计算机上进行.对LF、CAE、SPO、LF_occ和EPINET 5种典型方法分别在合成和真实数据集进行比较分析.其中,合成数据集采用德国海德堡大学视觉(Heidelberg collaboratory for image,HCI)实验室设计的数据集[32],该数据集由Blender软件渲染而成,包括4个类别共24个场景,光场图像分辨率为9×9×512×512×3,数据集为其中3个类别提供了真实视差图.真实光场数据集是用Lytro Illum相机拍摄的现实场景,其分辨率为14×14×375×541×3.实验中所有参数都是按照文献给出的值进行设置,为比较方便,合成和真实光场图像都选用9×9个视角,深度图层数都设置为100.

3.2　合成数据集

HCI数据集类别1的光场图像中心视角图和视差图如图3和4所示.图3(a)场景由相互平行且倾斜的背景和前景组成,用来评估算法对精细结构、狭窄间隙和遮挡边缘的处理效果.从图4可知,LF_occ生成的视差图左侧狭窄间隙部分误差较大,EPINET生成的视差图较好地保留了边缘和窄间隙,其他几种方法比LF_occ效果相对较好.图3(b)场景用来评估算法对噪声的鲁棒性,从估计结果可知,受噪声的影响,LF丢失了较多信息,CAE、LF_occ和SPO方法都有不同程度的误差.图3(c)场景包含倾斜、凹凸不平和不精确边缘,可以评估算法在不同几何形状上的性能.从图4场景估计结果可知,LF在平行平面上的估计非常平滑,但在倾斜物体表面上的估计则会产生明显的阶梯式效应,LF_occ估计的结果粗糙不平,而SPO相比LF_occ的结果较为平滑.图3(d)场景评估算法被不同对比度和纹理的影响程度.从图4场景估计结果可知,LF几乎错误估计了

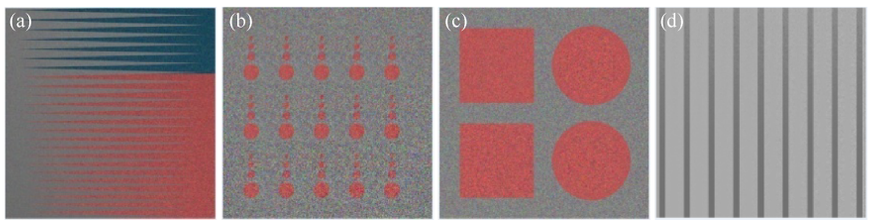


图3　HCI合成光场数据集类别1场景中心视角[32]

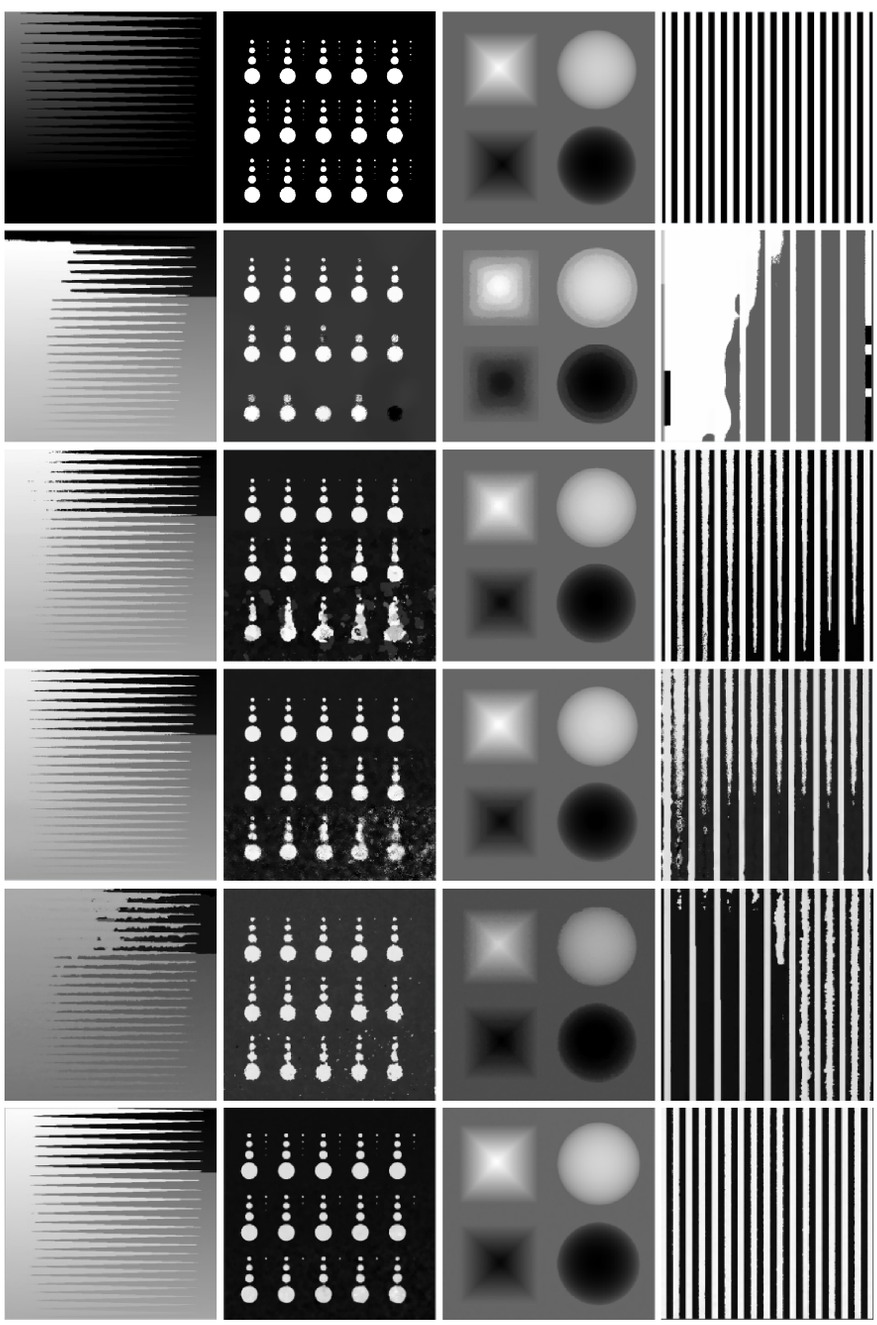


图4　HCI数据集类别1上各方法视差

注:第1~6行分别为真实视差、LF、CAE、SPO、LF_occ和EPINET视差.

左侧大部分低对比度纹理区域,LF_occ比CAE和SPO估计效果好.EPINET在狭窄间隙、噪声干扰、不同几何形状和重复纹理等场景下都可以精确地估计出场景深度.

HCI数据集类别2的光场图像中心视角图和视差图如图5和6所示.场景估计结果表明:各方法在图5

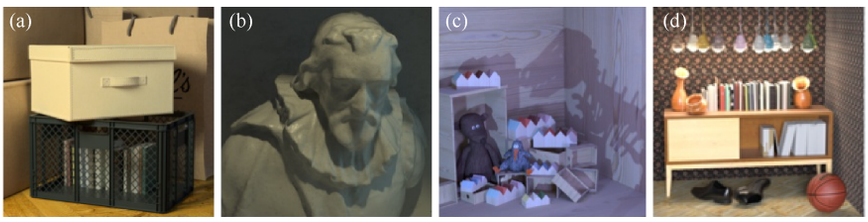


图5 HCI合成光场数据集类别2场景中心视角[32]

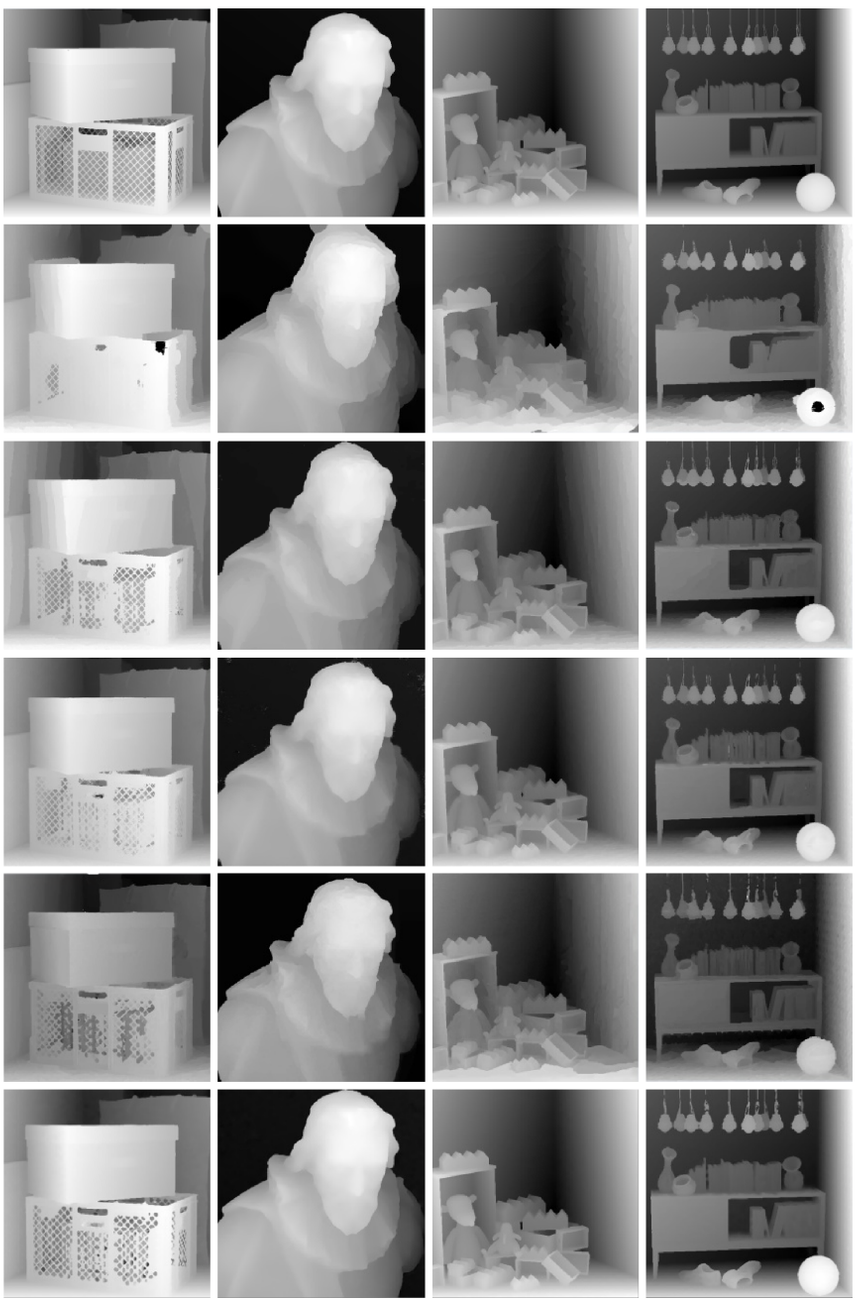


图6 HCI数据集类别2上各方法视差

注:第1~6行分别为真实视差、LF、CAE、SPO、LF_occ和EPINET视差.

(a)场景中都有不同程度的误差,LF几乎误估计了网格箱子的边缘部分;在图5(b)场景中,LF在雕像头顶部存在误匹配,其他方法结果相当;在图5(c)场景下,LF和LF_occ在墙壁上明显粗糙不平;在图5(d)场景中,各个方法在悬挂物体上都有一定程度的误差,其中LF误差最大.与类别1中结果一样,在类别2的4

个逼真场景下,EPINET表现最好.

为了定量比较各方法的性能差异,选取均方误差(MSE×100)、误识别像素率(BadPix0.07)2个度量指标[31],分别评价算法的深度估计准确程度和对噪声的鲁棒性.2个指标值越小,代表方法性能越好.均方误差度量MSE定义为

$$\mathrm{MSE}_M = \frac{\sum_{x \in M} \left(d(x) - gt(x)\right)^2}{|M|} \times 100, \quad (1)$$

式中$d(x)$为估计的视差图,$gt(x)$为真实视差图,M为评估掩膜.

误识别像素率BadPix定义为

$$\mathrm{BadPix}_M(t) = \frac{\left|\left\{x \in M: \left|d(x) - gt(x)\right| > t\right\}\right|}{|M|}, \quad (2)$$

式中t为误识别像素的阈值,本文取值0.07.

5种方法在2个类别8幅场景下的均方误差、误识别率和运行时间见表1.从均方误差和误识别率2个评价指标可知,LF和LF_occ的误差较大,且LF_occ由于考虑了遮挡而误差小于LF. CAE和SPO的误差较小,原因是CAE采用的角度熵和自适应散焦量具有较好的抗遮挡和抗噪能力,SPO是基于EPI的估计方法,具有一定的抗遮挡能力.EPINET误差最小,是因为其采用的深度学习技术具有良好的非线性映射和泛化能力,因而对遮挡和噪声较为鲁棒性.

从运行时间分析,总体上,光场图像由于体积大,数据处理方法大多耗时高.其中LF_occ耗时最多,主要原因是LF_occ需要显式地处理遮挡边缘,采用多标签优化的方法,相对较为耗时; LF与CAE运行时间基本相当,但LF略高是由于需要计算傅里叶变换;SPO基于EPI图像需要计算直线斜率且同样采用多标签优化,耗时比LF和CAE高出一个数量级;基于深度学习的EPINET在不考虑训练时间的情况下,运行效率最高.

3.3 真实数据集

Lytro Illum相机拍摄的真实光场数据集中心视角见图7.图8是不同方法在真实数据集上重建的视差图.真实光场数据包含较大的相机噪声、遮挡、精细结构和高光.由于LF从原理上并未考虑遮挡,与

表1 2个类别经5种方法处理后实验结果

方法	指标	类别1				类别2				平均值
		图3(a)	图3(b)	图3(c)	图3(d)	图5(a)	图5(b)	图5(c)	图5(d)	
LF	均方误差	13.006 8	5.676 2	0.272 9	17.453 7	17.433 6	9.168 4	1.163 7	5.070 9	8.655 8
	误识别率	5.516 4	2.900 2	12.354 2	35.741 3	23.019 1	7.829 1	19.025 5	21.989 1	16.046 9
	运行时间/s	979.87	969.96	929.72	1 093.98	962.18	984.53	1 130.56	987.47	1 004.78
CAE	均方误差	0.224 9	5.082 4	0.753 2	3.555 6	8.424 3	1.505 8	0.382 0	0.875 6	2.600 5
	误识别率	3.923 5	12.401 2	1.680 8	7.871 5	17.884 5	3.368 6	4.968 1	9.844 9	7.742 9
	运行时间/s	811.66	952.02	851.21	758.81	826.97	814.24	832.83	861.56	838.66
SPO	均方误差	4.587 1	5.238 0	0.042 6	6.954 5	9.107 3	1.312 9	0.310 1	1.023 8	3.572 0
	误识别率	3.781 2	16.273 7	0.860 8	14.986 8	15.889 4	2.594 2	2.184 4	9.296 9	8.233 4
	运行时间/s	2 195.0	2 138.0	2 256.0	1 945.0	2 128.0	2 025.0	2 045.0	2 073.0	2 100.63
LF_occ	均方误差	21.586 9	3.301 4	0.097 7	8.130 8	9.850 4	1.067 9	1.136 7	2.303 9	5.934 5
	误识别率	19.006 2	5.821 7	3.172 2	18.408 1	26.520 2	6.218 0	14.912 7	18.494 8	14.069 2
	运行时间/s	5 116.25	10 820.85	11 688.42	19 331.42	10 408.21	6 325.51	10 099.05	13 531.31	10 915.13
EPINET	均方误差	3.628 6	1.634 8	0.008 0	0.950 0	6.239 7	0.190 7	0.167 2	0.827 0	1.705 8
	误识别率	3.579 5	3.183 3	0.192 4	2.461 5	12.838 5	0.508 3	1.285 7	4.800 6	3.606 2
	运行时间/s	1.97	1.96	1.96	1.96	1.97	1.96	1.97	1.97	1.97

注:LF为光场;CAE为代价角度熵;SPO为旋转平行四边形算子;LF_occ为光场遮挡模型;EPINET为极平面网络.



图7 真实光场数据集中心视角

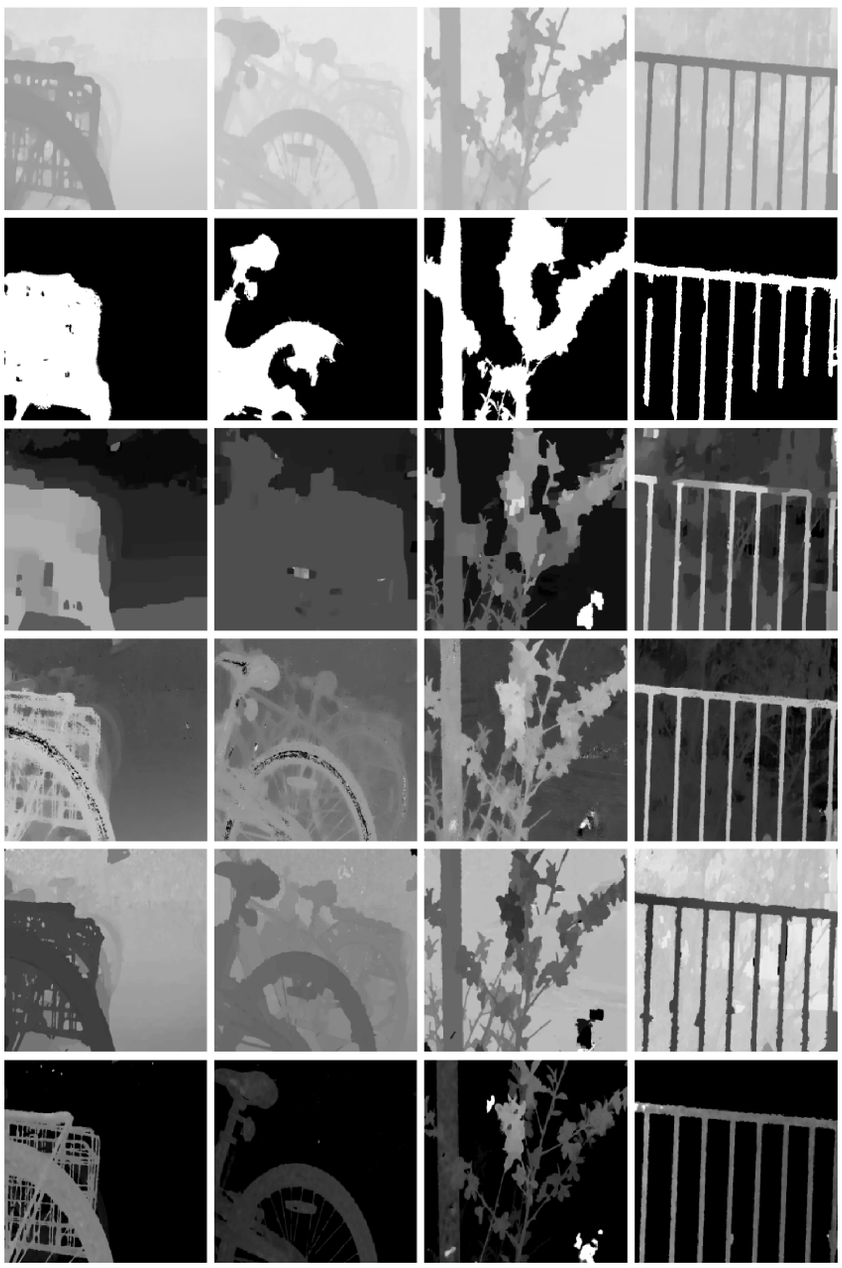


图8 真实光场数据集各方法视差

注:第1~6行分别为真实视差、LF、CAE、SPO、LF_occ和EPINET视差.

合成光场数据集实验结果一样,该方法在真实光场数据上的重建结果误差最大,特别是图7(a)和(b)场景中的自行车辐条边缘几乎全部丢失,在图7(d)幅栅栏上也有部分误估计.CAE估计结果比LF较好,但仍然误差较大.LF_occ和SPO都有较好的抗遮挡和噪声能力,重建的视差图比较清晰和精确.EPINET由于采用了加权中值滤波,故较好地抑制了相机噪声,且对复杂遮挡场景重建结果都非常清晰和精确.但该方法估计结果丢失了远景处的物体(如图7(a)和(b)场景远处的自行车),这与训练缺乏真实数据集有关.需要指出的是,5种方法都是建立在朗伯场景条件下的,不适应于非朗伯场景,如图7(c)高光部分所有方法深度估计结果都存在偏差.

4 总结和展望

作为一种新兴的视觉信息获取技术,光场成像技术已经被广泛应用在三维重建、显著性检测、目标检测和识别等计算机视觉领域.通过梳理和总结光场深度估计方法,并通过实验进一步说明,利用光场结构特征进行场景深度重建,

相比传统的单目和双目深度估计,具有显著优势,但仍然存在挑战,这也是未来的研究趋势:

(1)现有大多数方法都是基于朗伯体假设和简单场景的深度估计,未来还应当对高光或镜面反射的非朗伯体、更精细的结构、倾斜物体表面重建以及低纹理等复杂场景下的深度估计进行深入研究.

(2)与立体匹配相比,由于光场包含密集的视点采样和连续的视差空间,更有利于深度估计.但是,从硬件设计的角度分析,高的角度分辨率会对传感器的性能提出更高的要求.

(3)充分利用光场结构特征,探索多线索融合方法,以进一步提高深度估计的精度和算法效率.已有研究者尝试利用EPI和多视角2种方法进行深度估计,其效果明显优于单一方法.

(4)研究表明,深度学习在计算机视觉领域有着得天独厚的优势.基于深度学习的光场深度估计将会成为未来的主流.然而,深度学习在光场深度估计方面还处于探索阶段,研究者们已经开始关注扩展更多的数据集、设计新的神经网络结构和研究新的损失函数等.

(5)4D光场数据由于包含丰富的场景信息而存在着体积大、计算耗时的问题,未来还需要在提高算法效率,以及开展基于图形处理单元的光场数据处理上进一步研究.

参 考 文 献

[1] BOLLES C R, BAKER H H, MARIMONT H D. Epipolar-plane image analysis: an approach to determining structure from motion[J]. International Journal of Computer Vision, 1987, 1(1): 7-55.

[2] WANNER S, GOLDLUECKE B. Globally consistent depth labeling of 4D light fields[C]//TANG C K, LI D, CHEN Q. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 41-48.

[3] LI J Q, LU M L, LI Z N. Continuous depth map reconstruction from light fields[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3257-3265.

[4] KIM C, ZIMMER H, PRITCH Y, et al. Scene reconstruction from high spatio-angular resolution light fields[J]. ACM Transactions on Graphics, 2013, 32(4): 1-12.

[5] 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.

[6] ZHANG Y B, LYU H J, LIU Y B, et al. Light-field depth estimation via epipolar plane image analysis and locally linear embedding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(4): 739-747.

[7] YU Z, GUO X Q, LING H B, et al. Line assisted light field triangulation and stereo matching[C]//ORDONEZ N, DENG J, CHOI Y, et al. 2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2013: 2792-2799.

[8] HEBER S, POCK T. Shape from light field meets robust PCA[C]//FLEET D, PAJDLA T, SCHIELE B, et al. European Conference on Computer Vision. Zurich: Springer, 2014: 751-767.

[9] CHEN C, LIN H T, YU Z, et al. Light field stereo matching using bilateral statistics of surface cameras[C]//DANEILIAN M, KHAN F S, FELSBFRG M, et al. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1518-1525.

[10] JEON H G, PARK J, CHOE G, et al. Accurate depth map estimation from a lenslet light field camera[C]//IHAOR, QU Y, LI W L, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1547-1555.

[11] 熊伟,张骏,高欣健,等.自适应成本量的抗遮挡光场深度估计算法[J].中国图象图形学报,2017,22(12):1709-1722.

[12] LIU F, HOU G Q, SUN Z N, et al. High quality depth map estimation of object surface from light-field images[J]. Neurocomputing, 2017, 252: 3-16.

[13] TAO M W, HADAP S, MALIK J, et al. Depth from combining defocus and correspondence using light-field cameras[C]//ORDONEZ N, DENG J, CHOI Y, et al. 2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2013: 673-680.

[14] TAO M W, SRINIVASAN P P, MALIK J, et al. Depth from shading, defocus, and correspondence using light-field angular coherence[C]//ZHAO R, QU Y, LI W L, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1940-1948.

[15] WANG T C, EFROS A A, RAMAMOORTHI R. Depth estimation with occlusion modeling using light-field cameras[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2170-2181.

[16] WILLIEM W, PARK I K. Robust light field depth estimation for noisy scene with occlusion[C]//COHEN N, SHARIR O, SHASHUA A. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4396-4404.

[17] ZHU H,WANG Q,YU J Y. Occlusion-model guided anti-occlusion depth estimation in light field[J]. IEEE Journal of Selected Topics in Signal Processing,2017,11(7):965-978.
[18] 杨德刚,肖照林,杨恒,等.基于光场分析的多线索融合深度估计方法[J].计算机学报,2015,38(12):2437-2449.
[19] 胡良梅,姬长动,张旭东,等.聚焦性检测与彩色信息引导的光场图像深度提取[J].中国图象图形学报,2016,21(2):155-164.
[20] TIAN J D,MUREZ Z,CUI T,et al. Depth and image restoration from light field in a scattering medium[C]//SOCK J,KASAEI S H,LOPES L S,et al. 2017 IEEE International Conference on Computer Vision. Venice:IEEE,2017:2401-2410.
[21] 韩磊,徐梦溪,王鑫,等.基于光场成像的多线索融合深度估计方法[J].计算机学报,2020,43(1):107-122.
[22] HEBER S,POCK T. Convolutional networks for shape from light field[C]//COHEN N,SHARIR O,SHASHUA A. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE,2016:3746-3754.
[23] JOHANNSEN O,SULC A,GOLDLUECKE B. What sparse light field coding reveals about scene structure[C]//KUNDU A,VINEET V,KOLTU V. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE,2016:3262-3270.
[24] 苏钰生,王亚飞,李学华.基于改进密集连接型网络的光场深度估计[J].计算机工程与应用,2020,56(11):142-148.
[25] LUO Y X,ZHOU W H,FANG J P,et al. EPI-patch based convolutional neural network for depth estimation on 4D light field[C]//KING I,WANG J. International Conference on Neural Information Processing. Guangzhou:Springer,2017:642-652.
[26] PENG J Y,XIONG Z W,LIU D,et al. Unsupervised depth estimation from light field using a convolutional neural network[C]//STRUC V,GOMEI F. 2018 International Conference on 3D Vision. Verona:IEEE,2018:295-303.
[27] 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]//MISRA I,GIRSHICK R B,FERGUS R,et al. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake:IEEE,2018:4748-4757.
[28] GERSHUN A. The light field[J]. Journal of Mathematics and Physics,1939,18(1/2/3/4):51-151.
[29] ADELSON E H,BERGEN J R. The plenoptic function and the elements of early vision[M]. Cambridge:MIT Press,1991:46-58.
[30] LEVOY M,HANRAHAN P. Light field rendering[C]//JOHN F. Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New York:ACM,1996:31-42.
[31] WU G,MASIA B,JARABO A,et al. Light field image processing:an overview[J]. IEEE Journal of Selected Topics in Signal Processing,2017,11(7):926-954.
[32] HONAUER K,JOHANNSEN O,KONDERMANN D,et al. A dataset and evaluation methodology for depth estimation on 4D light fields[C]//LAI S H,LEPETIT V,NISHINO K,et al. Asian Conference on Computer Vision. Taiwan:Springer,2016:19-34.

Classification and performance comparison of light field depth estimation methods

SU Boni

(School of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou Sichuan 635000)

Abstract: According to the different characteristics of light field,the existing depth estimation methods are divided into four categories:polar plane-based,multiple view-based,focal stack-based and deep learning-based methods. The characteristics of these four methods are compared and analyzed,and the parametric characterization and visualization of light field are described. On the synthetic and real light field data sets,five typical depth estimation methods are compared and analyzed quantitatively and qualitatively,and the results show that the depth estimation method based on deep learning is obviously superior to the traditional methods in operating efficiency and estimation accuracy. Finally,the challenges and future development directions of optical field depth estimation methods are prospected.

Keywords: light field;depth estimation;comparative analysis;deep learning

(责任编辑:兰丽丽)