

# 基于光场成像的多线索融合深度估计方法

韩磊<sup>1),2)</sup> 徐梦溪<sup>2)</sup> 王鑫<sup>1)</sup> 王慧斌<sup>1)</sup>

<sup>1)</sup>(河海大学计算机与信息学院 南京 211100)

<sup>2)</sup>(南京工程学院计算机工程学院 南京 211167)

**摘要** 在计算机视觉领域,传统的深度估计方法通常基于单个或多个摄像机拍摄的二维图像实现,忽略了光线的方向信息,获取的深度图精确性较差.以 Raytrix、Lytro 光场相机为代表的新型计算成像设备,能够在一次曝光中同时获取光辐射的空间分布及方向信息,为基于图像的深度估计提供了更多几何依据和线索.近年来,结合微透镜光场相机成像模型的多深度线索融合思想,成为探求深度估计问题的新导向.与基于单一深度线索的深度估计方法相比,当前的多线索融合方法取得了更准确的深度估计效果,然而,尚不能保持对噪声和遮挡的鲁棒性.为此,本文基于光场成像几何提出了一种考虑遮挡并对噪声鲁棒的多线索融合深度估计方法.首先,在借鉴人类视觉多线索优势差异的研究成果基础上,选择对噪声不敏感的聚焦线索和准确度较高的对应线索作为深度估计的依据;接着,针对微透镜光场相机成像模型,剖析聚焦、对应两种深度线索的几何特性,阐明了焦点堆栈轮廓曲线的局部对称性以及 EPI 图像上直线斜率与深度的比例关系,进一步分析并证明了存在遮挡的情况下,分别由焦点堆栈和中心子孔径图像生成的两类轮廓曲线之间仍存在一致性;随后,基于这些几何特性,构建了考虑遮挡和噪声影响的度量标准;然后,建模深度估计为多标签优化问题,将度量标准构造为最小化能量函数,以图割法求解得最优深度估计.在实验环节,本文首先对模型参数做 Ablation 分析:单独考虑对称性、一致性、对应线索的其中之一,或组合考虑两种或全部分别进行实验,结果表明综合考虑多种性质时本文方法效果最好.另外,分别在合成光场数据集和真实场景光场数据集上测试本文及相关文献的深度估计方法,实验结果表明本文深度估计方法准确度较高,对遮挡边缘的细节处理有效,且能够保持对噪声的鲁棒性.

**关键词** 光场;深度估计;多线索融合;微透镜;对称性;极平面图像

**中图法分类号** TP391 **DOI号** 10.11897/SP.J.1016.2020.00107

## Depth Estimation from Multiple Cues Based Light-Field Cameras

HAN Lei<sup>1),2)</sup> XU Meng-Xi<sup>2)</sup> WANG Xin<sup>1)</sup> WANG Hui-Bin<sup>1)</sup>

<sup>1)</sup>(College of Computer and Information Engineering, Hohai University, Nanjing 211100)

<sup>2)</sup>(School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167)

**Abstract** In the field of computer vision, traditional methods of depth estimation are usually based on two-dimensional images captured by single or multiple cameras. The depth maps obtained by those methods are usually inaccurate, since the imaging process ignores the directional information of light. Consumer-level and high-end light-field cameras are now widely available. Light-field cameras, such as Raytrix and Lytro, can record not only the intensity but also the direction of light rays at one exposure. Therefore, a light field provides more geometric constraints and cues to depth estimation based on images. In recent years, depth from combining multiple cues, based on light-field cameras, has become a new direction for the problem of depth estimation. Compared with the methods of depth from a single cue, the current depth estimation methods based on multi-cue fusion achieve more accurate depth maps. However, most previous methods do not

explicitly model occlusions, and cannot capture sharp transitions around object boundaries. On the other hand, conventional methods fail when handling noisy scene. In this paper, a robust multi-cue depth estimation method for occlusion and noise is proposed based on light-field imaging geometry. First of all, inspired by the research of human vision about multiple cues, we select both focus and correspondence depth cues for fusion computation, where the focus cue is insensitive to noise, and the correspondence cue is more accurate. Secondly, according to the imaging model of lenslet light-field camera, we clarify the local symmetry in focal stack profile and the relationship between the depths and the slope of the lines in EPI, and further show that there are consistencies between the two profiles respectively generated by the focal stack images and the center sub-aperture image at the neighborhood of real disparity under occlusion. Thirdly, based on those geometric characteristics, some metrics considering the effects of noise and occlusion are constructed. According to the local symmetry of the focal stack profile, the mirror symmetry metric is established. And the consistency metric is constructed by considering the properties of two kinds of focal stack profile under occlusion. Then the matching metric of disparity cue is designed by using the proportional relationship between the depth value and the slope of line in EPI image. The last but not the least, depth estimation is modeled as a multi-label optimization. We describe the energy function as two parts: the smooth term and the data term consisting of the above metrics. Then we use graph cut to obtain disparity map by minimizing the energy function. In the experiment, Ablation study is executed on the model parameters. We firstly utilize one metric among the symmetry, consistency and correspondence metrics. Then we combine two or all metrics in energy function. Ablation analysis results show that it is the best to consider a variety of metrics synthetically. Additional, we test the depth estimation methods in this paper and the related literatures on the synthesis light-field dataset and real light-field dataset, respectively. The experimental results show that our method can accurately estimate depth, effectively deal with noise and details at occlusion boundary.

**Keywords** light field; depth estimation; multiple cues fusion; lenslet; symmetry; epipolar plane image

## 1 引 言

基于图像分析方法获取物体和场景深度信息(与成像系统距离)是计算机视觉领域的一个研究热点,它在机器人导航、三维重建、摄影测量等方面有着广泛应用<sup>[1]</sup>.在深度信息辅助下,目标分割、识别、跟踪等技术取得较大进展,行为识别、场景理解等技术也有不少突破<sup>[2]</sup>.

然而,在透视投影成像模型下,三维场景向二维图像的映射是一个深度信息丢失的过程,因此,通常需要依据某种深度线索估计出场景深度信息.常用的深度线索有双目线索和单目线索.双目线索多指双目视差线索,单目线索主要有聚焦、散焦、纹理梯度、线条透视、重叠遮挡、运动视差、空气透视和熟悉尺度等<sup>[3]</sup>.这些深度线索揭示了成像模型约束下的

物点与像点的几何关系.

在双目视差线索指引下,基于三角测量原理,可以建立相对成熟的双目立体视觉技术,然而,对应点匹配问题是其难点.基于运动视差线索发展了运动形貌恢复(Shape From Motion, SFM)方法<sup>[4]</sup>.SFM首先提取前一帧图像的特征点,然后在下一帧图像中查找对应特征点以估计视差.当然, SFM也有立体视觉相似的问题,一些特征可能在其他帧图像中丢失,这样就无法计算这些丢失特征点的对应视差,进而相应深度值也无法估计.事实上,形貌恢复(shape from X)与深度估计是同类问题, SFM仅是其中一种,类似的还有分别基于聚焦(focus)、散焦(defocus)、纹理(texture)、阴影(shading)等线索的形貌恢复技术(SFF、SFD、SFT、SFS)<sup>[5]</sup>.其中,基于聚焦线索的 SFF 抗噪声能力较强,但需要在不同距离处采集同一场景的多幅图像,计算每幅图像的聚

焦图。SFD 建立在散焦模型基础上, 算法空间和时间消耗较小, 但可靠度不如 SFF。分别基于阴影和纹理的 SFS 和 SFT 技术, 过度依赖于光照和场景特殊性, 适应范围有限。

由此可见, 基于某种单一线索的深度估计方法各有优势和不足, 各线索间存在一定程度的互补性。人类视觉深度感知的研究也指出: 人类视觉是综合运用多种线索进行深度感知的, 多种深度线索之间存在着不一致性和相互作用<sup>[6]</sup>。受此启发, 多线索融合可能成为提高深度估计准确性、鲁棒性的有效策略。但在传统成像系统中, 同时获取多种深度线索并不方便, 比如单目视觉系统中, 需要移动相机位置或机械改变焦距才能采集多视点图像或焦点堆栈图像, 进而运用视差和聚焦等线索。

光场相机的出现极大地缓解了以上矛盾, 它能同时记录光线的位置和方向, 一次曝光采集的数据即可以计算得到多视角的子孔径图像以及焦点堆栈图像, 为多线索的同时获取创造了便利的物理条件。

近年来, 基于光场相机的深度估计成为 CVPR、ICCV 等顶级会议讨论的话题之一。Tao 等人<sup>[9]</sup>开启了微透镜光场相机成像模型下的多线索融合深度估计方法研究, 该方法综合考虑散焦和对应两种深度线索, 取得了较好效果。随后, Lin、Zhu、Heber、Wang 等也进一步研究了多线索融合的深度估计方法, 效果普遍优于单一线索的深度估计方法<sup>[7-10]</sup>。

当前, 不论是对光场相机这类新型成像设备的认识, 还是对多线索融合深度估计策略的运用, 均处于新兴探索阶段。现有多线索融合深度估计方法提高了深度估计准确度, 但尚不能保持对噪声和遮挡的鲁棒性, 基于的线索不同以及对线索表征的度量标准不同是造成深度估计性能差异的主要原因。为此, 本文在光场成像模型下, 从几何视角剖析深度线索的特征性质, 建立适当的度量标准, 提出了一种考虑遮挡且对噪声鲁棒的多线索融合深度估计方法。本文的主要贡献和创新之处主要体现在以下几个方面:

(1) 运用几何分析的方法, 建立光场成像模型下的深度线索度量标准, 为深度估计提供可靠的量化依据。

(2) 在焦点堆栈轮廓曲线局部对称性的基础上, 进一步指出并证明了遮挡情况下, 两类焦点堆栈轮廓曲线的一致性, 据此提高了遮挡处理能力。

(3) 在对应线索指引下, 运用 EPI 图像中直线的斜率和视差存在比例关系性质, 建立了基于均方

差的度量, 提高了本文方法对噪声的鲁棒性。

后面章节将首先介绍基于光场相机的深度估计相关方法; 其次, 结合光场相机的成像原理, 深入剖析聚焦线索、对应线索在光场成像模型几何约束下的特征性质; 接着, 依据深度线索的这些特征性质, 建立相应的深度度量和置信度度量; 随后, 运用优化技术, 给出本文的深度估计算法; 然后, 在合成光场数据集和真实光场数据集上验证本文算法的有效性。

## 2 相关工作

早在 1939 年, Gershun 就提出了光场的概念, 定义为光辐射在空间各个位置各个方向的传播<sup>[11]</sup>。他认为, 到达空间不同点处的光辐射量连续变化, 能够通过几何分析进而通过积分的方法来计算像面上每点的光辐射量。Adelson 于 1992 年将光场理论成功运用到机器视觉, 并提出全光场理论 (plenoptic theory)<sup>[12]</sup>。光场理论进一步完善于 1996 年 Levoy 的光场渲染理论 (Light Field Rendering, LFR), 他将光场进行参数化表示, 并提出计算成像公式<sup>[13]</sup>。在此基础上, 2005 年, Ng 发明了第一台手持式光场相机, 将光场成像推向消费市场<sup>[14]</sup>。

光场相机同时记录光线位置和方向的成像特点, 为场景深度估计提供了更多依据。在聚焦深度线索指引下, 针对光场相机成像模型, 学者们探讨了基于焦点堆栈的深度估计方法。这类方法的技术路线是: 由光场相机一次拍摄获取的数据, 运用数字重聚焦技术, 生成一系列焦点堆栈图像; 对焦点堆栈图像进行聚焦区域检测, 估计出场景深度信息。在此技术路线中, 针对聚焦区域检测问题, Kao 运用高频和低频的比率度量聚焦程度, 并采用动态规划法解决区域检测问题, 效率较高<sup>[15]</sup>; Mousnier 使用了梯度检测法, 并利用中值滤波消除区域生长造成的伪影, 进而得到全聚焦图像, 提升了深度精度<sup>[16]</sup>。Li 则克服微透镜渐晕效应对生成全聚焦图像的影响, 建立了高动态范围光场模型<sup>[17]</sup>。尹晓良直接使用 Halcon 软件的聚焦深度函数对焦点堆栈图像进行深度估计, 并应用于 3D 重建<sup>[18]</sup>。Lin 发现对应于非遮挡点的像素轮廓是关于聚焦切片中心对称的<sup>[7]</sup>。利用这一性质, 得到新的聚焦性评价, 辅助提高深度估计鲁棒性。

光场成像模型下, 基于视差线索的深度估计总体上有两种思路: 一是利用光场数据生成子孔径图

像,再基于多视匹配进行深度估计;二是将 4D 光场映射到 2D 的 EPI 图像,在 EPI 图像上利用直线斜率和视差之间的比例关系进行深度估计.针对光场子孔径图像进行多视匹配问题时的微基线问题,Jeon 等人利用傅里叶相移定理求出子透镜图像的亚像素位移,并将中心视角图像和其他视角进行立体匹配,同时引入损耗函数进行多标签优化,然后通过迭代优化将离散视差图转换为连续深度图<sup>[19]</sup>.Zhang 等人同样利用相位平移理论,提出一个基于视差辅助相位的光场合成(DAPS)和深度估计算法框架,由传统立体匹配算法得到初始深度图,再基于相位差和余弦拟合进行迭代地深度优化<sup>[20]</sup>.这两种方法都采用了傅立叶相移理论,通过视差估计精度的提高来降低微基线对误差的放大.

运用光场变换的 EPI 图像进行深度估计的学术团体较多.关于 EPI 图像上的直线斜率估计,很多研究<sup>[21-24]</sup>采用了结构张量法.Wanner 提出全局一致性深度标记算法,在全变分框架下,利用 EPI 图像梯度的结构张量求解纹理斜率,并使用最小化能量函数进行全局约束,实现深度信息的估计<sup>[21]</sup>.Suzuki 就视差范围受限问题,提出在结构张量分析之前先对 EPI 图像进行裁剪,对不同裁剪值的结构张量分析结果进行综合,形成最终视差图<sup>[22]</sup>.Diebold 提出了改进的结构张量法,建立了 EPI 的非恒定灰度模型,可以处理亮度梯度光场、颜色滤波光场的深度估计问题<sup>[23]</sup>.Lüke 充分考虑光场的 4D 本质,将结构张量扩展到 4D 空间,使其具备分析超平面、平面、直线和均质区域能力,使算法更易在并行硬件上实现<sup>[24]</sup>.Lv 从一个预定义集合中选择最优斜率,取代了传统结构张量法<sup>[25]</sup>.综合考虑像素灰度、梯度、空间一致性以及置信度等多种因素作为择优标准,使用匹配代价汇聚和胜者为王的方法进行选择,提高深度估计的鲁棒性,实现了稠密深度估计.Johannsen 另辟蹊径,利用 EPI 图像直线方向与场景深度的关系,设计了稀疏分解子,由中心视图建立低维字典,进而通过稀疏编码稀疏计算朗伯面的可靠视差<sup>[26]</sup>.

近年来,运用光场成像优势,综合考虑多种深度线索成为提高深度估计的有效策略.Tao 使用光场相机融合散焦和对应两种线索进行深度估计,该方法通过裁剪 EPI 图像,发现散焦线索和对应线索分别与 2D EPI 图像的水平方差和垂直方差有一定关联,进而使用马尔科夫模型融合两种深度线索<sup>[6]</sup>.在此基础上,Tao 进一步考虑了阴影线索,运用光场角

度一致性融合了阴影、散焦和对应三种线索进行深度估计<sup>[27]</sup>.Wang 的深度估计融合了遮挡、聚焦等线索<sup>[10]</sup>.杨德刚在相机阵列光场系统下尝试融合散焦和对应两种线索进行深度估计<sup>[3]</sup>.Heber 运用深度神经网络拟合光场数据中蕴含的多种线索进行深度估计<sup>[9]</sup>.

以上方法鉴于传统成像系统深度估计方法的经验,分析光场相机的成像模型下的深度线索新特性,对深度估计问题进行了卓有成效的研究,但在保证准确性,兼顾对噪声和遮挡的鲁棒性方面,仍有很大的提升空间.尤其是不同方法所选择的深度线索不同,建立的深度度量标准不同,深度估计的效果可能也不同.本文正是鉴于这一经验思想,进一步分析深度线索在光场成像模型下的新性质,依次建立新的度量标准,进而提高深度估计准确度和鲁棒性.

### 3 光场深度线索分析

本节介绍微透镜光场相机的成像模型,并在此模型下,寻求深度估计的聚焦线索和对应线索几何特性,以指导遮挡处理,同时为建立相应度量标准奠定理论基础.

#### 3.1 微透镜光场相机成像模型

光场相机<sup>[14]</sup>由三部分构成:主透镜系统、微透镜阵列和光电探测器件,如图 1 所示.成像原理为:物体上一点经过主透镜系统后聚焦于微透镜阵列平面,而后经单个微透镜分散出强度和方向分量,到达探测器的不同面元,此即该物点的光场信息.

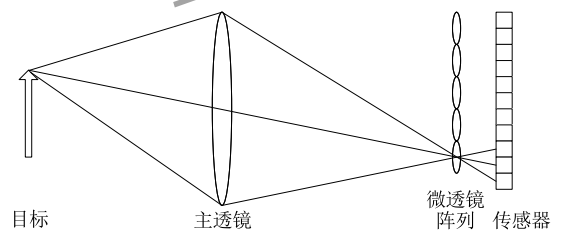


图 1 光场相机的成像示意图

光场相机一次曝光采集的 4D 光场数据经积分运算得到数字重聚焦图像.理论上,由光场数据可以恢复出聚焦于任意深度的图像,也可以恢复出一组多视角图像<sup>[14]</sup>.该成像模型为深度估计提供了多种可靠深度线索.

#### 3.2 聚焦深度线索分析

基于光场数据可以计算生成多组子孔径图像,这里,考虑通过子孔径图像生成焦点堆栈图像的几何原理.将光场数据参数化为 2PP(two Plane

Parametrization)表示,如图 2 所示,平面  $st$  和  $uv$  分别表示相机平面和图像平面,光线通常用向量  $(s, t, u, v)$  表示,辐射强度记作  $r_{st}^{uv}(s, t, u, v)$ ,其中  $r$  的上、下标分别表示参考平面. 输出视差图以中心子孔径图像为参考视图. 建立焦点堆栈时,重聚焦像面以视差为参数由子孔径图像移位叠加而成,即对应视差为  $f$  的重聚焦像面上的  $p$  点灰度辐射强度是由所有子孔径图像上对应视差为  $f$  的像素积分得到的. 据此绘制原理,引入以下符号说明:

$\phi_p(f)$ : 表示由光场数据提取的所有子孔径图像,计算生成的  $p$  点焦点堆栈轮廓曲线. 即  $p$  点对应焦点堆栈(以视差  $f$  为单位)灰度变化曲线.

$\varphi_p(f)$ : 表示由光场数据提取的中心子孔径图像和当前迭代视差图,计算生成的  $p$  点焦点堆栈轮廓曲线.

曲线  $\phi_p(f)$  在真实视差附近表现出对称性(性质 1),其证明见文献[7],本文进一步提炼了两类曲线的一致性(性质 2).

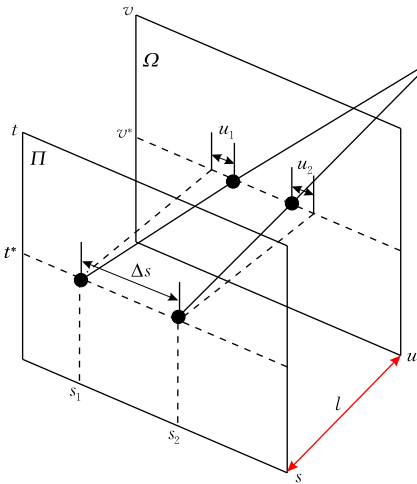


图 2 光场双平面表示

**性质 1.** 在朗伯面假设和非遮挡条件下,某点的焦点堆栈轮廓曲线在真实视差附近呈现镜像对称性.

**性质 2.**  $\varphi_p(f)$  和  $\phi_p(f)$  在真实视差  $d$  的邻域内仍具有一致性,即  $f \in [d - \delta_{\max}, d + \delta_{\max}]$  时,有  $\varphi_p(f)$  和  $\phi_p(f)$  相等或近似相等.

**证明.** 本性质的证明基于如下基本概念或观点:光场生成的焦点堆栈是重聚焦图像沿光轴方向排列的图像序列<sup>[15]</sup>,所在序列位置以视差为单位表示;重聚焦图像可以看作是子孔径光线的重积分<sup>[14]</sup>;深度图(视差图)均以中心子孔径图像为参考;不失一般性,证明过程仅以 3 幅子孔径图像的  $s$ - $u$  二维光场为例进行分析.

本性质证明的思路是:(1)证明性质 1 条件下,不仅在真实视差附近存在局部对称性,而且两焦点堆栈曲线相同;(2)指出破坏局部对称性的原因是当前分析点的邻域缺失,概括导致当前分析点邻域缺失的两类情况,并在第(3)、(4)步分别证明分析点邻域缺失情况下,两类曲线仍然保持一致性.

(1)当分析点及其邻域不被遮挡时,如图 3 所示,中心子孔径图像的像素点  $p$ (坐标为  $u_0$ )对应于聚焦平面  $\Pi_2$  上的  $p_0$  点,在上、下子孔径图像上的对应点相对于  $p$  的真实视差为  $d$ .若视差  $f$  在  $d$  邻域偏移  $\delta \in (0, \delta_{\max}]$ ,则视差  $(d - \delta)$  和  $(d + \delta)$  分别对应于聚焦平面  $\Pi_1$  的  $p_3$  和  $\Pi_3$  的  $p_4$  点.聚焦平面  $\Pi_1$  的  $p_3$  点是来自三个子孔径的光线集合  $S_\delta$  的积分,而  $\Pi_3$  的  $p_4$  点是光线集合  $S_{-\delta}$  的积分.  $S_\delta$  和  $S_{-\delta}$  集合如下:

$$S_\delta = \left\{ r_s^{\Pi_2} \left( B, -\frac{B}{d} (u_0 + \delta) \right), \right. \\ \left. r_s^{\Pi_2} \left( -B, -\frac{B}{d} (u_0 - \delta) \right), \right. \\ \left. r_s^{\Pi_2} \left( 0, -\frac{B}{d} u_0 \right) \right\}, \\ S_{-\delta} = \left\{ r_s^{\Pi_2} \left( B, -\frac{B}{d} (u_0 - \delta) \right), \right. \\ \left. r_s^{\Pi_2} \left( -B, -\frac{B}{d} (u_0 + \delta) \right), \right. \\ \left. r_s^{\Pi_2} \left( 0, -\frac{B}{d} u_0 \right) \right\},$$

其中,  $-\frac{B}{d}(u_0 - \delta)$  和  $-\frac{B}{d}(u_0 + \delta)$  分别是  $\Pi_2$  平面上  $p_1$  点和  $p_2$  点的坐标,即  $r_s^{\Pi_2} \left( B, -\frac{B}{d}(u_0 - \delta) \right)$  和  $r_s^{\Pi_2} \left( -B, -\frac{B}{d}(u_0 - \delta) \right)$  是源自同一物点的光线,而

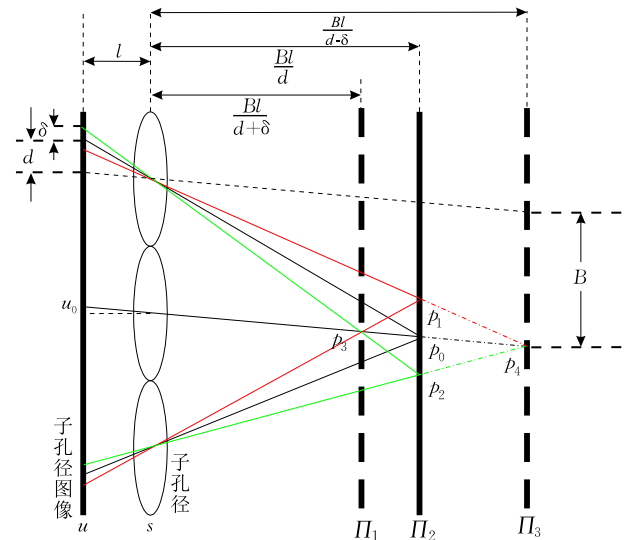


图 3 焦点堆栈局部对称性分析

$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right)$ 和 $r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right)$ 也源自同一物点. 根据朗伯面假设, 有:

$$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 - \delta)\right) = r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 - \delta)\right);$$

$$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right) = r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right).$$

所以,  $S_\delta = S_{-\delta}$ , 对两集合中光线积分后有:  $\phi_p(d-\delta) = \phi_p(d+\delta)$ , 即  $\phi_p(f)$  在真实视差  $d$  的邻域存在镜像对称性.

根据朗伯面假设, 由中心子孔径采集的通过  $p_1$  的光辐射  $r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 - \delta)\right)$  应与  $r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 - \delta)\right)$  和  $r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 - \delta)\right)$  相等; 同理, 在中心子孔径图像中也可根据视差图找到  $p_2$  对应的光辐射. 于是, 可由中心子孔径图像和视差图绘制焦点堆栈图像. 即, 对于  $p_1$  和  $p_2$ , 分别用中心子孔径采集的光辐射取代其他子孔径采集的光辐射, 分别得到汇聚于  $p_3$  和  $p_4$  的光辐射集  $S_\delta^c$  和  $S_{-\delta}^c$ :

$$S_\delta^c = \left\{ r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 + \delta)\right), \right.$$

$$r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 - \delta)\right),$$

$$\left. r_s^{\mu_2}\left(0, -\frac{B}{d}u_0\right)\right\},$$

$$S_{-\delta}^c = \left\{ r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 - \delta)\right), \right.$$

$$r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 + \delta)\right),$$

$$\left. r_s^{\mu_2}\left(0, -\frac{B}{d}u_0\right)\right\}.$$

将  $S_\delta^c$  和  $S_{-\delta}^c$  分别积分, 可以得到  $p$  对应的焦点堆栈曲线上  $(d-\delta)$ 、 $(d+\delta)$  两处的值  $\varphi_p(d-\delta)$ 、 $\varphi_p(d+\delta)$ ; 遍历视差  $d$  的邻域, 即可绘制曲线  $\varphi_p(f)$ .

此时, 因为  $S_\delta^c = S_\delta$ , 且  $S_{-\delta}^c = S_{-\delta}$ , 所以  $\varphi_p(f) = \phi_p(f)$ .

(2) 从以上证明过程不难发现: 焦点堆栈轮廓曲线在真实视差附近存在局部镜像对称性的关键是  $p_0$  的邻域点 ( $p_1$  和  $p_2$ ) 在中心子孔径中可见, 并且是对应物点在聚焦平面  $\Pi_2$  上的真实聚焦. 而当  $p_0$  处于遮挡器的边缘, 或  $p_0$  的邻域点被其他遮挡器遮挡时, 这一关键条件不成立. 下面分析这两类遮挡情况下, 曲线  $\varphi_p(f)$  和  $\phi_p(f)$  之间的关系.

(3) 当  $p_0$  处于遮挡器的边缘时, 如图 4 忽略  $\Pi_{-1}$  焦平面时所示,  $\Pi_2$  是遮挡器对应的聚焦平面,  $p_0$  对应物点位于遮挡器边缘, 邻域  $p_2$  不是物点的真实聚

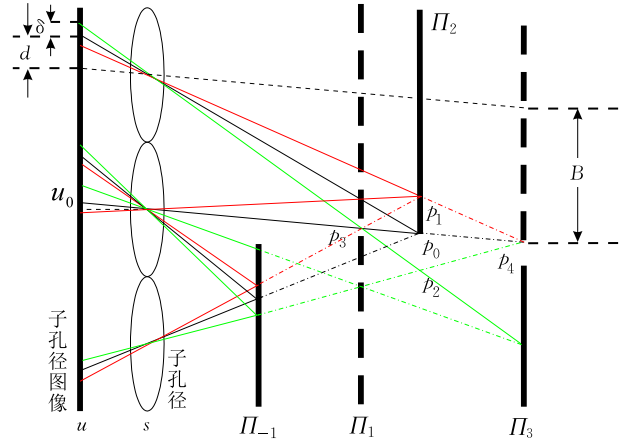


图 4 遮挡下焦点堆栈局部特征分析

焦, 经过它的两条光线并不源自同一物点, 即

$$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right) = r_s^{\mu_3}\left(B, -\frac{B}{d-\delta}(u_0 + 2\delta)\right),$$

$$r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right) = r_s^{\mu_3}\left(-B, -\frac{B}{d-\delta}u_0\right).$$

所以,

$$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right) \neq r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right),$$

进而,  $S_\delta \neq S_{-\delta}$ ,  $\phi_p(f)$  曲线的局部镜像对称性不再满足. 然而,  $S_\sigma$  中的光辐射在中心子孔径图像中均存在对应辐射:

$$r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right) = r_s^{\mu_3}\left(0, -\frac{B}{d-\delta}(u_0 + 2\delta)\right),$$

$$r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right) = r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 - \delta)\right),$$

即

$$S_\delta^c = \left\{ r_s^{\mu_3}\left(0, -\frac{B}{d-\delta}(u_0 + 2\delta)\right), r_s^{\mu_2}\left(0, -\frac{B}{d}(u_0 - \delta)\right), \right.$$

$$\left. r_s^{\mu_2}\left(0, -\frac{B}{d}u_0\right)\right\} = S_\sigma.$$

所以,  $f \in [d - \delta_{\max}, d]$  时,  $\varphi_p(f) = \phi_p(f)$  成立.

$S_{-\sigma}$  中的  $r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right)$  与其中心子孔径图像中对应光辐射  $r_s^{\mu_2}\left(0, -\frac{B}{d}u_0\right)$  不相等, 因为  $p_4$  对应物点受到遮挡, 在中心子孔径图像中不可见. 而  $S_{-\sigma}$  中的  $r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 - \delta)\right)$ 、 $r_s^{\mu_2}\left(0, -\frac{B}{d}u_0\right)$  均在中心子孔径图像中有对应相等的光辐射, 因此, 对  $S_{-\sigma}$  和  $S_{-\delta}^c$  分别积分得到焦点堆栈曲线近似相等, 即  $f \in (d, d + \delta_{\max}]$  时,  $\varphi_p(f) \approx \phi_p(f)$ .

(4)  $p_0$  的邻域被其他遮挡器遮挡时, 如图 4 所示,  $S_\sigma$  中  $r_s^{\mu_2}\left(B, -\frac{B}{d}(u_0 + \delta)\right)$  与  $S_\delta^c$  中  $r_s^{\mu_2}\left(-B, -\frac{B}{d}(u_0 + \delta)\right)$  对应, 但两者并不来自同一物点, 因此,

$$r_s^{\pi_2} \left( B, -\frac{B}{d}(u_0 + \delta) \right) \neq r_a^s \left( -(u_0 + d - d_3 + \delta), 0 \right),$$

进而,  $S_\sigma \neq S_\delta^c$ .

但是,  $S_\sigma$  中另两条光线仍与  $S_\delta^c$  的另两条光线辐射对应相等, 即

$$r_s^{\pi_2} \left( -B, -\frac{B}{d}(u_0 - \delta) \right) = r_a^s \left( -(u_0 + d_{-1} - d + \delta), 0 \right),$$

$$r_s^{\pi_2} \left( 0, -\frac{B}{d}u_0 \right) = r_s^{\pi_2} \left( 0, -\frac{B}{d}u_0 \right).$$

所以, 可以认为分别由集合  $S_\delta^c$  和  $S_\sigma$  积分得到的焦点堆栈曲线近似相等, 即  $f \in (d, d + \delta_{\max}]$  时,  $\varphi_p(f) \approx \phi_p(f)$ .

而  $S_{-\sigma}$  集合中的光线与  $S_{-\delta}^c$  中的光线一一对应相等, 即

$$r_s^{\pi_2} \left( B, -\frac{B}{d}(u_0 - \delta) \right) = r_s^{\pi_2} \left( 0, -\frac{B}{d}(u_0 - \delta) \right),$$

$$r_s^{\pi_2} \left( -B, -\frac{B}{d}(u_0 + \delta) \right) =$$

$$r_s^{\pi_{-1}} \left( 0, -\frac{B}{d_{-1}}(u_0 + \delta - d + d_{-1}) \right),$$

$$r_s^{\pi_2} \left( 0, -\frac{B}{d}u_0 \right) = r_s^{\pi_2} \left( 0, -\frac{B}{d}u_0 \right).$$

所以,  $S_{-\sigma} = S_{-\delta}^c$ , 当  $f \in [d - \delta_{\max}, d]$  时, 有  $\varphi_p(f) = \phi_p(f)$  成立.

综上, 存在遮挡情况时, 导致分析点的部分邻域缺失, 在真实视差附近,  $\phi_p(f)$  的局部镜像对称性不再保持, 但是  $\phi_p(f)$  和  $\varphi_p(f)$  仍保持一致性. 证毕.

因此, 对于场景中的非遮挡区域, 可以利用由所有子孔径图像生成的焦点堆栈轮廓曲线在真实视差附近呈现的局部镜像对称性来估计真实视差; 而对于遮挡边缘, 又可以利用两类焦点堆栈轮廓曲线的一致性印证估计视差的可靠性.

### 3.3 对应深度线索分析

如图 2 所示, 采用 2PP 形式表示光场, 由两个平面  $\Pi$  和  $\Omega$  确定光线,  $(s, t) \in \Pi$  表示子孔径的焦点位置,  $(u, v) \in \Omega$  是子孔径图像像素坐标. 光场  $L$  可以看作光线  $(s, t, u, v)$  到灰度值  $L(s, t, u, v)$  的映射:

$$L: (s, t, u, v) \rightarrow L(s, t, u, v).$$

**定义 1.** 同时约束  $\Pi$  和  $\Omega$  的水平维度 ( $s = s^*$ ,  $u = u^*$ ), 得到  $L_{s^*, u^*}: (t, v) \rightarrow L(s^*, t, u^*, v)$ ; 或同时约束  $\Pi$  和  $\Omega$  的垂直维度 ( $t = t^*$ ,  $v = v^*$ ), 得到  $L_{t^*, v^*}: (s, u) \rightarrow L(s, t^*, u, v^*)$ . 由 4D 光场约束得到的 2D 图像  $L_{s^*, u^*}$ 、 $L_{t^*, v^*}$  称为 EPI 图像 (Epipolar Plane Images).

**性质 3.** 在朗伯面假设下, 由同一点发出的光线, 在 4D 光场生成的 EPI 图像中表现为直线段, 直线段的斜率与该点的视差成正比, 与该点的深度成反比.

该性质的证明见文献[21]. 图 5 是典型 EPI 图像的示例, 图中第一行是场景的中心子孔径图像, 第二行是中心子孔径图像中虚线位置处对应的 EPI 图像. 由图 5 可明显看出, 对应同一物点的像素形成一条直线. 直线斜率与物点深度存在比例关系, 因此, EPI 图像中直线斜率分析可以作为深度估计的依据.

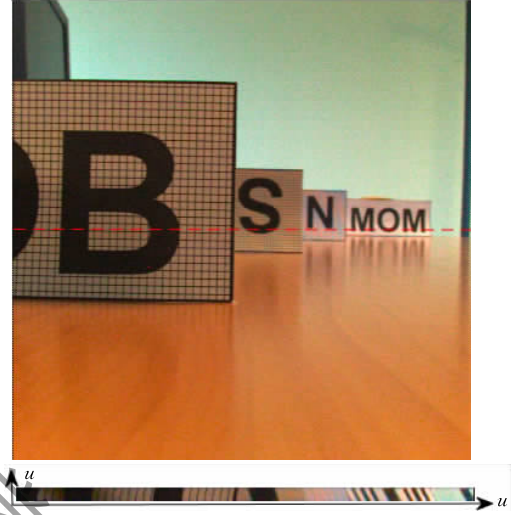


图 5 EPI 图像示例

## 4 深度线索度量

### 4.1 聚焦线索度量

#### 4.1.1 镜像对称性度量

运用性质 1, 建立镜像对称性度量以进行视差估计与优化. 针对中心子孔径图像的每个像素  $u_i$ , 假设视差为  $f$  时, 在焦点堆栈轮廓图像中确立镜像对称度量  $M_p(f)$ :

$$M_p(f) = \int_0^{\delta_{\max}} \rho(\phi_p(f + \delta) - \phi_p(f - \delta)) d\delta \quad (1)$$

式(1)中的函数  $\rho(\cdot) = 1 - e^{-(\cdot)^2 / (2\sigma^2)}$  是一个常见距离函数,  $\sigma$  是调整噪声敏感度的参数.

#### 4.1.2 焦点堆栈曲线一致性度量

运用性质 2, 建立焦点堆栈曲线  $\varphi_p(f)$  和  $\phi_p(f)$  的一致性度量, 可以印证视差估计的准确度, 提升遮挡处理能力.

仍使用距离函数度量两种曲线的一致程度, 匹配度量  $M_p^c(f)$  定义为式(2).

$$M_p^c(f) = \int_{f-\delta_{\max}}^{f+\delta_{\max}} \rho(\phi_p(\delta) - \varphi_p(\delta)) d\delta \quad (2)$$

#### 4.2 对应线索度量

根据性质 3, EPI 图像中某位置像素点对应的视差与其所在直线的斜率成正比. 如果已知某点视差的估计值, 那么可以构造一条直线段, 该直线段上像素与 EPI 图像上对应像素的匹配程度, 反映了视差估计值的准确程度. 由此建立对应线索视差估计的度量标准.

如果已知中心子孔径图像上某像素点  $p$  的视差为  $d$ , 则根据性质 3, 可以计算出  $p$  对应的场景点的深度为  $z = \frac{Bl}{d}$ , 理想情况下,  $p$  在 EPI 图像上对应

直线段  $s$  的斜率为  $\tan\theta = \frac{Z}{l} = \frac{B}{d}$ . 如图 6 所示, 可将直线段  $s$  经过的像素选择出来, 构成集合  $S_p(d)$ .  $S_p(d)$  中像素的灰度值与像素  $p$  的灰度值的相似程度体现了视差  $d$  的可靠度. 因此, 这里通过比较像素  $p$  灰度值和集合  $S_p(d)$  中像素灰度的差异, 得到对应线索一致性度量  $C_p(d)$ , 如式(3).

$$C_p(d) = \frac{1}{|S_p(d)|} \sum_{q \in S_p(d)} (I(q) - I(p))^2 \quad (3)$$

其中,  $|S_p(d)|$  表示集合  $S_p(d)$  的基数,  $I(\cdot)$  表示像素“ $\cdot$ ”的灰度值.

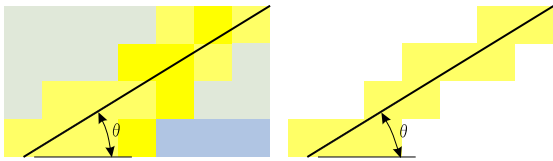


图 6 EPI 图像上候选深度像素集合示意图

### 5 多线索融合深度估计

综合考虑聚焦、对应两类深度线索, 发挥各自对遮挡和噪声的鲁棒性, 进行多线索融合深度估计. 将该问题建模为能量最小化问题, 使能量函数取得最小值的视差图即为估计的视差图, 如式(4)所示. 其中,  $D$  是与中心子孔径图像每个像素一一对应的二维视差图, 矩阵中每个元素的值是对应像素的视差值,  $E(D)$  表示关于视差图  $D$  的能量函数.

能量函数通常由数据项约束和平滑项约束构成(式(5)). 数据项约束主要度量视差值与观测数据的不一致性, 而平滑项约束则体现了区域内的连续性和边界的不连续性.

如式(6), 像素  $p$  对应的数据项约束定义为聚

焦线索的镜像对称度量和匹配度量, 以及对应线索的一致性度量三者之和, 其中,  $D_p$  表示视差图  $D$  中  $p$  处的视差值,  $\lambda_1$  和  $\lambda_2$  表示控制因子.

平滑项约束需要考虑邻域像素的灰度变化和视差变化, 通常认为: 相邻像素若灰度值接近则视差也近似, 灰度值差异较大则视差也较大, 此时, 平滑约束项较小; 否则, 平滑约束值增大. 鉴于此, 式(7)~(8)给出了平滑项约束的定义, 其中,  $I_p$  和  $I_q$  分别表示中心子孔径图像中  $p, q$  处的灰度值,  $\lambda, \lambda_3$  表示控制因子,  $N$  表示相邻像素构成的集合.

$$D^* = \arg \min_D E(D) \quad (4)$$

$$E(D) = E_{\text{data}}(D) + E_{\text{smooth}}(D) \quad (5)$$

$$E_{\text{data}}(D) = \sum_p (M_p(D_p) + \lambda_1 M_p^c(D_p) + \lambda_2 C_p(D_p)) \quad (6)$$

$$E_{\text{smooth}}(D) = \lambda \sum_{\{p,q\} \in N} V_{\{p,q\}}(D_p, D_q) \quad (7)$$

$$V_{\{p,q\}}(D_p, D_q) = 1 - e^{-(I_p - I_q)^2 - (|I_p - I_q| - \lambda_3 |D_p - D_q|)^2 / 2\sigma^2} \quad (8)$$

通过上述分析, 已经把深度(视差)估计建模为一个能量最小化问题, 本质上也是一个多标签优化问题. 假设视差范围为  $[d_{\min}, d_{\max}]$ ,  $d_{\min}, d_{\max}$  分别表示最小、最大视差, 通常, Lytro 微透镜光场相机只有 1 像素的最大视差; 在上述视差范围内, 等步长离散化为若干个视差标签, 构成视差标签集合  $LS$ , 而  $|LS|$  表示标签个数. 深度估计以中心子孔径图像中的每个像素为参考对象, 为其分配一个视差标签, 得到视差图.

本文采用图割算法求解该能量最小化问题, 显然, 式(8)的  $V$  是标签集合  $LS$  的一个度量,  $(LS, V)$  构成一个度量空间. 因此, 可以采用文献[28]提出的 alpha 扩展移动 ( $\alpha$ -expansion moves) 和 alpha-beta 交换移动 ( $\alpha$ - $\beta$  moves) 算法来近似求解. 虽然基于移动算法的图割法可以从任意初始解开始近似计算, 但合适的初始解能够加快算法收敛速度.

文献[28]提出的算法是基于移动规则动态构图、图割求解的迭代过程. 构图时, 需要分别计算数据项约束(式(6))和平滑项约束(式(7))为不同边赋权值. 预先构造一些基础结构有助于提高权值计算速度. 在视差标签集合确定的情况下, 中心子孔径图像上每个像素对应的焦点堆栈曲线  $\phi_p(f)$  在迭代过程中是不变的, 因此, 本文算法首先计算视差标签集合  $LS$  中每个标签对应的数字重聚焦图像, 形成焦点堆栈; 然后, 针对每个像素  $p$ , 以视差标签  $f$  为横坐标, 以  $f$  对应的数字重聚焦图像上  $p$  位置处的灰度值为纵坐标, 绘制焦点堆栈曲线  $\phi_p(f)$ . 另外, 由



性质 1 可知,  $\phi_p(f)$  的局部对称性体现了真实视差分布, 因此, 根据式(1)检测  $\phi_p(f)$  的局部对称性, 以局部对称性最为显著处对应的视差值作为  $p$  的初始视差, 所有像素的初始视差形成中心子孔径图像对应的初始视差图, 以此作为迭代优化的基础。

而由中心子孔径图像和当前视差图绘制的焦点堆栈轮廓曲线  $\varphi_p(f)$  是一个动态过程, 一方面它随着当前视差图而变化, 另一方面, 对每个像素, 只需绘制当前估计视差附近的曲线片段, 计算量较小, 因此, 本文不预先存储  $\varphi_p(f)$  曲线。

对应线索度量所需的 EPI 图像实质上是由 4D 光场裁切而成。4D 光场数据是本文算法的输入, 由其生成 2D 切片的计算较为快速, 另外, 匹配时, 以行或列为单位做临时存储, 不需要预先大规模额外存储。

每次迭代时, alpha 扩展移动完成一次构图, 通过求解该网络图的最小割达到能量最小化的目的, 进而优化视差图。其中, 求解网络图最小割的方法由经典的极大流/最小割算法实现。

基于上述分析, 下面给出本文的深度估计算法。

**算法 1.** 基于光场分析的视差估计。

输入:  $L(s, t, u, v)$  光场数据;  $\delta_{\max}$  最大视差偏移量;

$[d_{\min}, d_{\max}]$  视差范围;  $|LS|$  视差标签个数;

$\lambda, \lambda_1, \lambda_2, \lambda_3, \sigma$  控制参数

输出:  $D$  视差图

1. BEGIN

2. 以视差  $\frac{d_{\max} - d_{\min}}{|LS|}$  为步长, 计算对应的  $|LS|$  幅数字

重聚焦图像, 形成焦点堆栈。

3. 根据以上焦点堆栈数据, 对中心子孔径图像的每个像素  $p$ , 以每幅图像对应的视差为横坐标, 以对应位置灰度值为纵坐标, 绘制焦点堆栈曲线  $\phi_p(f)$ 。

4. 根据式(1), 对中心子孔径图像的每个像素  $p$ , 检测  $\phi_p(f)$  局部对称性, 选取对称性度量最小值对应的视差, 作为像素的初始视差, 形成中心子孔径图像对应的初始视差图  $D$ 。

5. 赋值搜索进展标志变量  $Success=0$ 。

6. 对每一个视差标签  $f \in LS$ :

6.1 对  $D$  进行一次  $\alpha$  扩展移动得  $D'$ , 用图割法求

$\hat{D} = \arg \min E(D')$ , 见文献[28]。

6.2 如果  $E(\hat{D}) < E(D)$ , 则  $D = \hat{D}$ ,  $Success=1$ 。

7. 如果  $Success=1$ , 则转第 5 步。

8. 输出  $D$ , 即为所求的视差图。

9. END

## 6 实验结果与分析

本节将给出本文方法不同阶段的结果, 进而分析算法参数; 然后, 分别在合成光场数据集和真实场景光场数据集上, 对比本文方法及相关方法的性能。

### 6.1 本文方法的参数分析

本文方法所优化的目标函数中包含多个参数, 主要用于控制聚焦对称性度量、焦点堆栈曲线一致性度量、对应线索度量以及平滑特性等因素相对权重。这些参数依据经验值设定。本节实验对参数做 Ablation 分析, 展示单线索深度估计结果和多线索融合深度估计结果。

以 Bedroom 场景为例, 图 7 给出了使用不同深度线索的实验结果。图 7(a) 是场景的中心视图以及局部细节放大。仅考虑聚焦对称性度量和平滑特性时, 即式(6)中参数  $\lambda_1, \lambda_2$  为 0, 本文方法能够估计出场景目标的基本轮廓和深度层次, 但遮挡边界附近存在较大的膨胀或腐蚀现象(图 7(b))。仅考虑对应线索和平滑特性时, 即忽略式(6)中的对称性度量和一致性度量, 本文方法估计的目标轮廓基本清晰, 但均质区域噪声较大(图 7(c))。当式(6)的  $\lambda_2$  为 0 时,

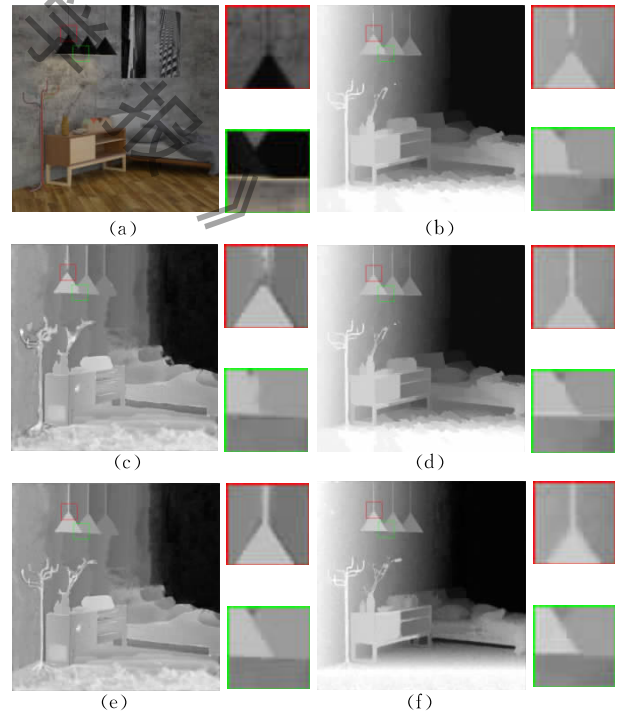


图 7 本文方法使用不同线索(参数)时的深度估计结果((a)中心子孔径图像;(b)对称性+平滑特性;(c)对应线索+平滑特性;(d)对称性+一致性+平滑特性;(e)对应线索+一致性+平滑特性;(f)对称性+对应线索+一致性+平滑特性)

即考虑焦点堆栈曲线的对称性、一致性和平滑特性而忽略对应线索度量,如图 7(d)所示,遮挡边界的膨胀或腐蚀明显减轻.如果考虑对应线索、一致性和平滑特性,而忽略对称性度量,实验结果如图 7(e)所示,灯线的膨胀现象明显减轻,遮挡边界也趋于清晰.同时考虑聚焦对称性度量、焦点堆栈曲线一致性度量、对应线索度量以及平滑性时的实验结果如图 7(f)所示,改善了单一深度线索的不足,深度层次分明,遮挡边界清晰.

## 6.2 实验结果对比分析

分别在合成光场数据集<sup>[29]</sup>和真实场景光场数据集<sup>[16]</sup>上做测试,并与以下方法做性能对比分析:Tao 等人<sup>[6]</sup>的多线索融合的深度估计方法,Jeon 等人<sup>[19]</sup>提出的基于相移理论的对齐匹配方法,Lin 等人<sup>[7]</sup>利用焦点堆栈图像对称特性进行深度估计的方法,Zhu 等人<sup>[8]</sup>利用遮挡器在空间和角度域的一致性进行遮挡建模和深度估计的方法,Heber 等人<sup>[9]</sup>利用卷积神经网络进行深度估计的方法.

本文方法及对比方法均基于 Matlab 软件环境,运行于配置为 Intel Xeon E5 2.3 GHz CPU、64 GB RAM 的硬件平台. Tao 方法的参数沿用文献<sup>[6]</sup>所使用的参数,但将 Lytro 光场数据集视图裁剪为  $9 \times 9$ ; Jeon 方法的参数  $k$  设置取决于数据集类型,代码实现时可选择合成数据集和 Lytro 数据集, Lytro 数据又可选近景和远景,其他参数和文献<sup>[19]</sup>中的设置一致. 以上两种算法分别由作者提供的代码实现,而 Lin 方法无法获得代码,这里根据文献<sup>[7]</sup>的算法思想基于 Matlab 实现,其中,合成光场数据集实验时,  $\lambda_m$ 、 $\lambda_c$ 、 $\lambda_R$  分别设置为 0.5、0.85、0.05,而对于真实场景光场数据,算法参数与文献<sup>[7]</sup>的设置一致. Zhu 方法的  $\gamma_{occ}$ 、 $\gamma_e$  和  $\gamma_c$  三个参数分别设置为 1.6、0.8 和 0.08. Heber 方法在文献<sup>[9]</sup>数据集的基础上增加 HCI 的 Training 类场景作为训练数据集. 实验中,本文方法的参数依经验值设置,合成光场数据集的视差范围为  $[-2, 2]$ ,真实光场的视差范围为  $[-1, 1]$ ,视差标签数为 256,  $\lambda_1$ 、 $\lambda_2$  分别设置为 0.05、0.8,  $\lambda_1$  和  $\lambda_3$  均设置为 0.5,  $\delta_{max}$  和  $\sigma$  分别设置为 0.078 125、0.25.

### 6.2.1 合成光场数据测试

本文实验采用的合成数据集是德国海德堡大学 HCI(Heidelberg Collaboratory for Image Processing)团队提供的标准光场测试数据集<sup>[29]</sup>,该光场测试数据集提供了 24 个精心设计的合成稠密采样 4D 光

场. 每个场景包含四维( $9 \times 9 \times 512 \times 512$ )的光场数据、相机参数以及视差范围等. 其中,针对层叠场景、训练场景两类子集还提供了真实(Ground truth)视差图,为算法性能量化比较提供了方便.

在 HCI 数据集中,Backgammon 光场是专门为测试深度估计算法对存在精细化结构、狭小间隔和遮挡的场景的性能表现而设计的合成场景. 该场景的中心子孔径图像如图 8(a)所示,由两个平行的、倾斜的背景平面和一个前景平面构成,前景平面与背景平面的倾斜方向相反,前景平面是锯齿状的,从而形成了越来越窄的前景结构和越来越精细的背景狭缝.

图 8(b)~(h)分别给出真实视差图(Groundtruth)以及 6 种方法(本文方法、Tao 方法、Jeon 方法、Lin 方法、Zhu 方法和 Heber 方法)估计的视差图. 其中,测试数据集提供的真实视差图和 Jeon 方法的结果以彩色图像表示,本文将其统一转换为灰度图像显示,以便对照分析.

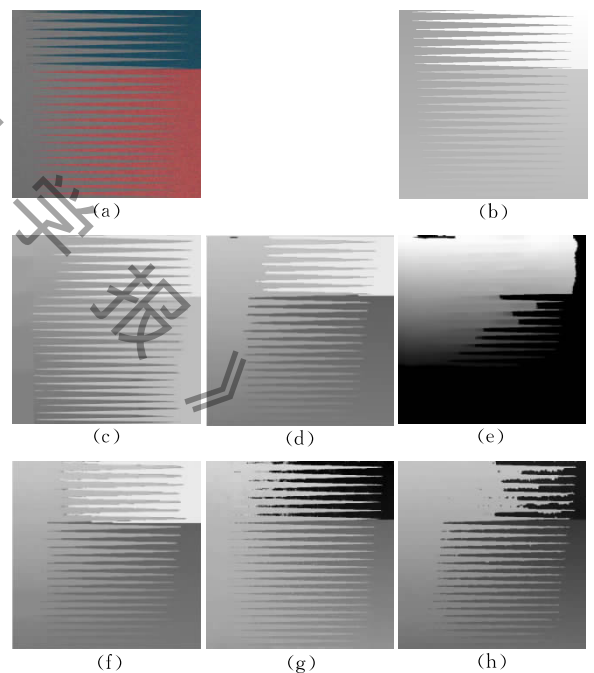


图 8 Backgammon 场景中心子孔径图像及深度估计结果((a)中心子孔径图像;(b)Groundtruth 结果;(c)本文方法结果;(d)Tao 方法结果;(e)Jeon 方法结果;(f)Lin 方法结果;(g)Zhu 方法结果;(h)Heber 方法的结果)

从图 8 的深度估计结果的视觉效果看,Jeon 方法不能清晰区分个别锯齿的前景与背景,Tao 方法和 Lin 方法虽然能够区分锯齿和背景的深度差异,但锯齿边缘毛刺较多,同样,Zhu 方法和 Heber 方法的结果也存在较多毛刺,遮挡边缘不够清晰;而本

文方法的实验结果锯齿边缘较清晰,但相对于真实视差图,对狭缝的深度估计存在偏差。

为了量化比较各算法的性能差异,选取文献[29]的均方误差(MSE)、误识别像素比例(*BadPix*)两个度量指标,分别评价算法的深度估计准确程度和抗遮挡能力.假设估计的视差图为  $D$ ,真实视差图为  $GT$ ,评价区域掩模为  $M$ ,则均方误差度量  $MSE$  定义为

$$MSE_M = \frac{\sum_{x \in M} (D(x) - GT(x))^2}{|M|},$$

误识别像素比例定义为

$$BadPix_M(t) = \frac{|\{x \in M: |D(x) - GT(x)| > t\}|}{|M|}.$$

另外,为了评价算法对噪声的鲁棒性,针对 Backgammon 场景,本文分别在纯净和噪声光场数据上开展实验.表 1 和表 2 给出不同噪声级别时各种算法的均方误差和误识别像素比例,其中,  $\sigma_N = 0$  列是纯净光场数据对应的实验结果,其他列是在光场数据中添加均值为 0、标准差为  $\sigma_N$  的高斯噪声后对应的实验结果.对于 *BadPix* 指标,假定掩模  $M$  为整幅图像,误识别像素的门限参数  $t$  为 0.07,相对于合成光场数据集提供的真实视差数据进行计算.

表 1 不同噪声水平下的均方误差

方法	$\sigma_N = 0$	$\sigma_N = 5$	$\sigma_N = 10$	$\sigma_N = 15$
本文	0.039	0.039	0.043	0.046
Tao	0.042	0.051	0.062	0.070
Jeon	0.144	0.149	0.162	0.203
Lin	0.035	0.041	0.045	0.064
Zhu	0.037	0.040	0.049	0.061
Heber	0.096	0.098	0.107	0.120

表 2 不同噪声水平下的误识别像素比例

方法	$\sigma_N = 0$	$\sigma_N = 5$	$\sigma_N = 10$	$\sigma_N = 15$
本文	11.34%	11.42%	12.51%	15.32%
Tao	12.26%	13.45%	15.18%	17.65%
Jeon	19.48%	23.13%	26.55%	30.24%
Lin	12.15%	13.36%	15.76%	18.14%
Zhu	12.03%	13.95%	15.48%	17.02%
Heber	15.87%	15.93%	18.34%	20.12%

从均方误差(MSE)和像素误识别率(*BadPix*)两项指标看,本文方法、Tao 方法、Lin 方法以及 Zhu 方法基本相当,而 Jeon 方法和 Heber 方法的误差偏大;Jeon 方法和 Tao 方法在原理层面上没有考虑遮挡,但 Tao 方法综合运用了散焦和对应两种深度线索,一定程度上提高了深度估计准确度;本文方法和 Lin 方法考虑了部分遮挡情况下焦点堆栈轮廓曲线与由中心视图生成的焦点轮廓曲线的性质,具有一

定的抗遮挡能力;Zhu 方法建立了遮挡模型,对遮挡的处理较为出色。

如表 1、表 2 所示,当在纯净光场数据加入不同级别的噪声时,本文方法、Tao 方法、Lin 方法和 Heber 方法表现出相对稳定的性能;而 Jeon 方法和 Zhu 方法在噪声增加时,性能下降较为剧烈.这一现象的原因在于:Jeon 方法所基于的相移理论对噪声较为敏感;Zhu 方法所建立的遮挡模型依赖于角块形状,失效于场景纹理与角块形状相似的情况,而噪声会影响局部纹理,造成该方法对噪声不够鲁棒;本文方法、Tao 方法和 Lin 方法使用了聚焦或离焦深度线索,这类线索对噪声不敏感;Heber 方法使用深度神经网络建立 EPI 图像块和深度信息之间的映射,深度神经网络的非线性表达能力使该方法表现出噪声鲁棒性。

### 6.2.2 真实场景光场数据测试

本节在商用光场相机采集的真实场景上进一步测试本文方法对噪声的鲁棒性和遮挡处理能力,由于传感器尺寸小,此类数据通常包含大量噪声.本文使用 Mousnier 等人<sup>[16]</sup>提供的 Lytro 光场数据集进行测试.该数据集包括 30 组 Lytro 摄像数据,其中,25 组是各地的室内外场景,3 组包含运动模糊,1 组包含长时间曝光,1 组是平面摄影.后三类图像不在本文算法的评价范围,本实验主要针对 25 组室内外场景光场图像进行测试,图 9 仅给出其中 5 组典型场景的深度估计结果,第一列是场景的中心子孔径图像,第 2 至 7 列分别是本文方法、Tao 方法、Jeon 方法、Lin 方法、Zhu 方法和 Heber 方法的深度估计结果;上面两行展示两组室内场景、中间一行是局部细节场景、下面两行是典型室外场景。

对于室内场景,光照水平(大约 400lux)远小于室外光照水平(大约 1000lux),这类场景光照受限,信噪比低.本文方法的深度估计效果较好,尤其对于层次明显的 BSNMOM 场景,本文算法能够清晰确认遮挡边界,且深度层次清楚;而对于 Beers 场景,本文方法虽能区分各啤酒瓶深度信息,但各啤酒瓶间距较小,视差等级接近,因此视差图不够分明. Jeon 方法深度层次分明,但视差标签少,BSNMOM 场景中挡板的深度层次出现重叠现象;Tao 方法融合两种深度线索,深度区分度较好,但噪声较大;Lin 方法也存在视差平滑性不足的问题;Zhu 方法和 Heber 方法的多数遮挡边界清晰,但物体与地面的交界处出现不同程度的过平滑现象。

对于局部细节场景,各方法生成了较理想的视

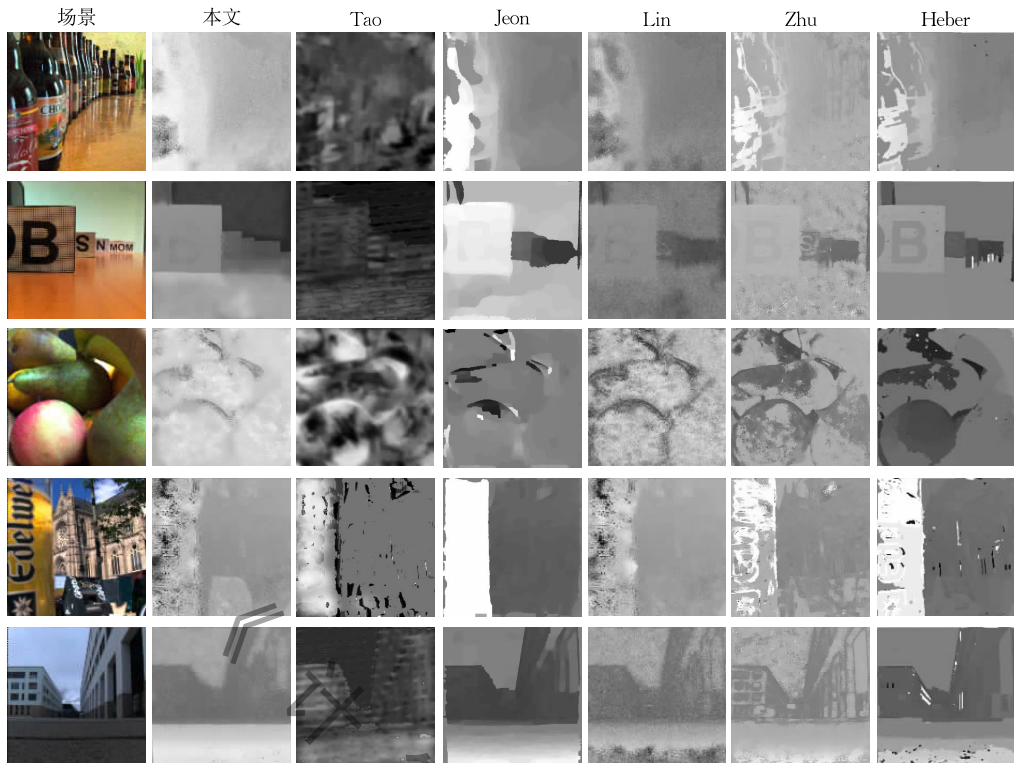


图 9 深度估计结果示例

差图,但均存在不同程度的噪声. Tao 方法在遮挡边界处失真较大. Jeon 方法、Zhu 方法和 Heber 方法目标轮廓非常清晰,但部分区域出现了深度翻转现象. 相比而言,本文方法出现深度翻转的区域最少.

对于室外场景,本文方法优势较为明显,能对远近目标产生层次分明的视差图,但也存在误将远景“天空”作为近景的问题,该问题同样出现在 Lin 方法的实验结果中. 原因可能是这两种方法均运用了焦点堆栈轮廓曲线的局部对称性,而位于无限远处的“天空”在焦点堆栈各图像中趋于一致,焦点堆栈轮廓曲线出现接近水平的直线段,各处对称性度量近似相等,因此容易导致误判. Tao 方法的深度估计结果存在明显噪声,而 Jeon 方法对较远距离景物的区分度不足,存在一定程度的过平滑现象. Zhu 方法和 Heber 方法也有类似现象,这两种方法处理均质区域时出现了不同程度的深度反转问题.

实验发现, Tao 方法和 Jeon 方法在合成光场数据集上的表现尚可接受,但在真实的 Lytro 光场数据集上的深度估计结果却急剧下滑. 对此,以下通过分析源代码来得到该现象的成因.

Tao 方法在计算散焦响应和对应响应之前,将原始光场数据映射到 Tao 定义的格式,该映射依赖于相机,但作者未提供映射定义方法. 因此, Tao 方法当前源代码中的映射未必完全适合 Lytro 光场数

据集,使实验效果受到影响. 但是本文在使用 Tao 方法进行对比实验时,将原光场的  $11 \times 11$  视图减少为  $9 \times 9$  视图,以更符合 Tao 定义的映射,从而改善实验效果.

Jeon 方法的关键步骤之一是将光场的多视图图像拼合成一张 JPEG 图像,后续的亚像素位移计算等深度估计过程基于该 JPEG 图像进行. 图 10(a) 是 Jeon 方法提供的样例图像的拼合效果,而图 10(b) 是由 Lytro 光场测试数据集(图 9 中的第 3 组场景)提供的数据拼合的 JPEG 图像. 不同于 10(a),图 10(b) 的四角为暗区,这是因为该光场测试数据集是由 Lytro 光场相机拍摄的,微透镜成像存在渐晕效应,将传感器数据解码为子孔径图像时,边缘特殊视角对应的图像黑暗模糊,因此,渐晕效应的噪声被传

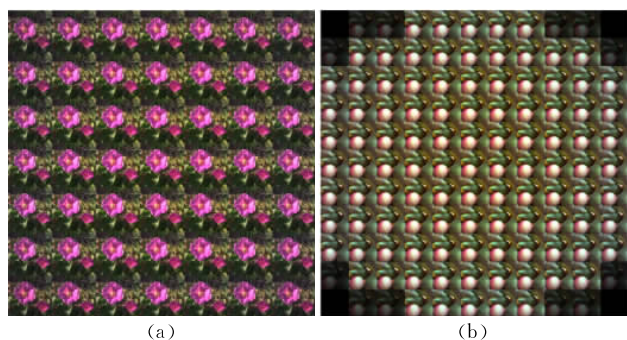


图 10 Jeon 方法图像示例

递到 JPEG 图像中,进而对深度估计的准确度产生了影响。

另一方面,Jeon 方法对其参数  $k$  的选择非常敏感,而 Jeon 提供的代码中仅粗略地依据数据集的类型设置该参数,较难保证能够取得与 Lytro 测试光场数据集相适的参数值,导致其在该数据集上的性能下降。

为了评价算法在计算时间方面的性能,本文以 Lytro 光场数据集的 25 组光场数据为测试对象,统计 6 种方法的各主要计算环节的平均计算时间以及整个深度估计过程的平均计算时间.为了细粒度地分析各方法的时间消耗,这里将参与对比分析的 6 种方法分为 3~4 个关键步骤,如表 3 所示,对每个关键步骤的时间消耗以及所有步骤的总时间消耗进行统计,最后对每种方法在 25 组光场数据的测试时间结果做平均处理,得到图 11 所示的结果.总体上,基于光场的深度估计是计算密集型的,时间消耗普遍较长,但一些研究<sup>[30]</sup>表明,光场计算的多数环节均可用 GPU 加速计算,这也是作者下一阶段将重点研究的内容。

表 3 测试算法的关键步骤

方法	第 1 步	第 2 步	第 3 步	第 4 步
本文	生成焦点堆栈	对称性检测与存储	迭代优化	无
Tao	光场重映射	计算散焦和对应响应	计算深度和置信度	MRF 深度融合
Jeon	成本量构建	成本汇总计算	多标签优化	迭代精细化连续视差图
Lin	生成焦点堆栈	计算聚焦和对应匹配分数	迭代优化	无
Zhu	初始深度估计	遮挡检测	非遮挡视图重选	深度优化
Heber	输入数据处理	超平面方向预测	优化	无

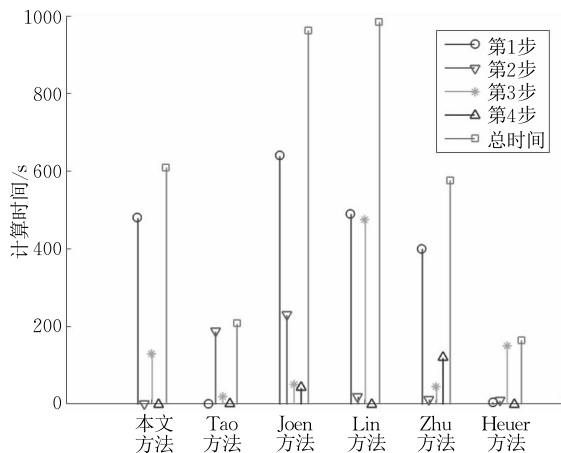


图 11 计算时间

图 11 的 6 种方法的计算时间中,Heber 方法的计算时间最短,但深度神经网络的训练时间未计入.其次,Tao 方法的计算时间较短,主要原因是:一方面,Tao 方法基于光场数据生成的 EPI 图像分析,EPI 图像生成计算代价较小;另一方面,Tao 将散焦重映射等计算负荷重的代码用 C 语言编写(其他环节代码由 Matlab 实现),提高了代码效率,节约了计算时间. Jeon 方法最为耗时,尤其是它第 1 步的成本量构建计算时间较长,这与其依赖于计算量较大的傅里叶相移理论不无关系.本文方法和 Lin 方法都需要生成焦点堆栈,该步骤计算负荷大,是两种方法中最为耗时的环节,另外优化过程中需要动态绘制焦点堆栈轮廓曲线的部分片段,也是运算量较大的环节.总体上,本文方法、Lin 方法和 Zhu 方法计算时间基本相当;但本文方法的平均计算时间少于 Lin 方法,原因在于:第一,本文方法以对称性检测结果作为初始视差图进行迭代优化,有利于减少迭代次数,进而降低时间消耗;第二,本文方法使用两种焦点堆栈轮廓曲线的一致性度量改进了数据项,有利于算法在遮挡边界处的深度估计较快收敛;第三,本文方法使用 EPI 图像上灰度匹配作为对应度量,提高了可靠性,通过临时存储行列 EPI 图像,以空间换取了计算时间。

## 7 结 论

不同于传统相机,微透镜光场相机可以同时记录光辐射强度和方向,使物像之间具有更丰富的几何约束关系,一次曝光所采集的图像中蕴含着多种深度线索.这为借鉴人类视觉多线索融合深度感知机制,解决计算机视觉领域的深度估计问题创造了便利的物理条件.因此,结合微透镜光场相机成像模型的多线索融合策略成为深度估计研究的着力点,也是 ICCV、CVPR 等顶级会议讨论的热点话题之一.本文的研究延续了这一方向,尤其针对当前典型的多线索融合深度估计方法对遮挡和噪声不够鲁棒的问题,采用几何分析的方法,提出了融合聚焦和对应两种线索的深度估计新方法.该方法在分析深度线索的几何性质的基础上,建立适当的度量评价,进而将深度估计建模为能量最小化问题求解.一方面,在焦点堆栈轮廓曲线局部镜像对称性的基础上,进一步指出并证明了遮挡情况下两类焦点堆栈轮廓曲线的一致性,该性质为遮挡处理提供了几何依据.另

一方面,聚焦线索指引下的焦点堆栈生成是子孔径图像移位叠加的计算过程,对噪声具有滤波功能;以均方差为基础,在 EPI 图像上建立的对应线索一致性度量也有一定滤波作用,这些为提升本文方法的噪声鲁棒性奠定了基础.在合成光场数据集上,实验给出了使用不同深度线索度量标准(参数配置)的深度估计效果,另外定量对比了纯净光场数据和添加不同等级高斯噪声的深度估计结果.同时,在真实光场数据集上,定性评价了本文及相关方法在噪声和遮挡情况下的性能.实验结果表明本文方法对噪声和遮挡具有鲁棒性.

本文的研究进一步说明了以下观点:首先,微透镜光场相机是具备同时获取多种深度线索能力的成像设备,为多线索融合深度估计提供了便利条件;其次,融合多种深度线索有利于提高深度估计的准确度;再次,深度线索的选取、特征分析与度量是影响深度估计鲁棒性的关键.研究成果已将聚焦线索拓展应用到遮挡场景,并兼顾对噪声的鲁棒性,是对深度估计问题的有益探索.但是,必须指出包括本文在内的基于光场图像的典型深度估计方法,普遍存在计算量大、运行时间长的问题,一些学者已经开始研究基于 GPU 的光场并行计算,这是非常有实用意义的研究方向.

**致 谢** 感谢国家自然科学基金面上项目和江苏省高校自然科学基金重点项目支持.《计算机学报》编辑部和评委老师给出了宝贵意见,在此一并感谢!

## 参 考 文 献

- [1] Chen Wang, Zhang Mao-Jun, Chong Yang, et al. Omni-directional depth estimation system with single-double mirrors architecture. *Journal of Computer-Aided Design & Computer Graphics*, 2010, 22(6): 1004-1009(in Chinese)  
(陈旺, 张茂军, 崇洋等. 单-双镜面组合结构的全向深度获取系统. *计算机辅助设计与图形学学报*, 2010, 22(6): 1004-1009)
- [2] Li Zhe, Peng Si-Wei. Fingertip tracking and trajectory recognition based on depth image information. *Computer Applications and Software*, 2016, 33(4): 155-159(in Chinese)  
(李哲, 彭四伟. 基于深度图像信息的指尖跟踪与轨迹识别. *计算机应用与软件*, 2016, 33(4): 155-159)
- [3] Yang De-Gang, Xiao Zhao-Lin, Yang Heng, Wang Qing. Depth estimation from light field analysis based multiple cues fusion. *Chinese Journal of Computers*, 2015, 38(12): 2437-2449(in Chinese)

(杨德刚, 肖照林, 杨恒, 王庆. 基于光场分析的多线索融合深度估计方法. *计算机学报*, 2015, 38(12): 2437-2449)

- [4] Lee D J, Merrell P, Wei Z, et al. Two-frame structure from motion using optical flow probability distributions for unmanned air vehicle obstacle avoidance. *Machine Vision & Applications*, 2010, 21(3): 229-240
- [5] Zhang R, Tsai P S, Cryer J E, et al. Shape-from-shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(8): 690-706
- [6] Tao M W, Hadap S, Malik J, Ramamoorthi R. Depth from combining defocus and correspondence using light-field cameras // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Sydney, Australia, 2013: 673-680
- [7] Lin H, Chen C, Bing Kang S, Yu J. Depth recovery from light field using focal stack symmetry // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile, 2015: 3451-3459
- [8] Zhu H, Wang Q, Yu J. Occlusion-model guided antiocclusion depth estimation in light field. *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(7): 965-978
- [9] Heber S, Pock T. Convolutional networks for shape from light field // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 3746-3754
- [10] Wang T C, Efros A A, Ramamoorthi R. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(11): 2170-2181
- [11] Gershun A. The light field. *Studies in Applied Mathematics*, 1939, 18(1-4): 51-151
- [12] Adelson E H, Wang J Y A. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 99-106
- [13] Levoy M, Hanrahan P. Light field rendering // *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. New Orleans, USA, 1996: 31-42
- [14] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report*, 2005, 2(11): 1-11
- [15] Kao Y H, Liang C K, Chang L W, Chen H H. Depth detection of light field // *Proceedings of the IEEE International Conference on Acoustics*. Honolulu, USA, 2007: I-893-I-896
- [16] Mousnier A, Vural E, Guillemot C. Partial light field tomographic reconstruction from a fixed-camera focal stack. *arXiv preprint arXiv: 1503.01903*, 2015
- [17] Li C, Zhang X. High dynamic range and all-focus image from light field // *Proceedings of the 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. Siem Reap, Cambodia, 2015: 7-12

- [18] Yin Xiao-Gen, Zhang Xiao-Fang, Zhang Wei-Chao, et al. Study on 3D reconstruction based on light field digital refocusing. *Journal of Optoelectronics Laser*, 2015, 26(5): 986-991(in Chinese)  
(尹晓良, 张晓芳, 张伟超等. 基于光场数字重聚焦的三维重建方法研究. *光电子·激光*, 2015, 26(5): 986-991)
- [19] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, USA, 2015: 1547-1555
- [20] Zhang Z, Liu Y, Dai Q. Light field from micro-baseline image pair//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, USA, 2015: 3800-3809
- [21] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Providence, USA, 2012: 41-48
- [22] Suzuki T, Takahashi K, Fujii T. Disparity estimation from light fields using sheared EPI analysis//*Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA, 2016: 1444-1448
- [23] Diebold M, Jähne B, Gatto A. Heterogeneous light fields//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA, 2016: 1745-1753
- [24] Lüke J P, Rosa F, MarichalHernández J G, et al. Depth from light fields analyzing 4D local structure. *Journal of Display Technology*, 2014, 11(11): 900-907
- [25] Lv H, Gu K, Zhang Y, Dai Q. Light field depth estimation exploiting linear structure in EPI//*Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. Turin, Italy, 2015: 1-6
- [26] Johannsen O, Sulc A, Goldluecke B. What sparse light field coding reveals about scene structure//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA, 2016: 3262-3270
- [27] Tao M W, Srinivasan P P, Malik J, et al. Depth from shading, defocus, and correspondence using light-field angular coherence//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, USA, 2015: 1940-1948
- [28] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 23(11): 1222-1239
- [29] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields//*Proceedings of the Asian Conference on Computer Vision*. Taipei, China, 2016: 19-34
- [30] Lumsdaine A, Chunev G, Georgiev T. Plenoptic rendering with interactive performance using GPUs. *Proceedings of SPIE-The International Society for Optical Engineering*, 2012, 8295(2): 32



**HAN Lei**, Ph.D. candidate, associate professor. His current research interests include computer vision and digital image processing.

**XU Meng-Xi**, master, associate professor. Her research interests include computer vision and digital image processing.

**WANG Xin**, Ph.D., associate professor. Her research interests include image processing and pattern recognition.

**WANG Hui-Bin**, Ph.D., professor, Ph.D. supervisor. His research interests include electronic imaging, image processing, information acquisition and intelligent system.

## Background

Image-based depth estimation is a common problem in the computer vision community, and it is also the basis of 3D reconstruction, visual navigation and so on. Since normal 2D cameras only record the intensity of light at a certain position on the image sensor, the previous methods usually estimated depth based on a single cue. The resulting depth map was not accurate enough. Fortunately, light-field cameras have recently become available to the consumer market. Unlike traditional photography, light field photography system can record four-dimensional position and direction. Therefore,

much more image information can be obtained from light field photography. Thus, depth cues (such as focus, defocus, correspondence and so on) are available simultaneously in a single capture.

Tao proposed a method of depth estimation which combine both defocus and correspondence depth cues based on light-field cameras. The accuracy of Tao's method was higher than that of the previous single-cue methods. Subsequently, many methods based on multi-cues fusion have been proposed for the problem of depth estimation using light-field cameras.

Those current depth estimation methods based on multi-cues fusion achieve more accurate depth maps, but they can not take into account the robustness of noise and occlusion. In this paper, we present a novel accurate and robust method that computes depth estimation by combining both focus and correspondence depth cues. Experimental results show that our method can accurately estimate depth, effectively deal with noise from the vignetting effect of lenses and details at occlusion boundary.

Our research belongs to the project of the Natural Science Foundation of Jiangsu Higher Education Institutions of China (No. 17KJB520010), whose name is “depth estimation method based on multi-cue fusion in light-field imaging

model”. Geometric pattern analysis for depth cues is one of research contents of this project. In this paper, the feature descriptions of both focus and correspondence depth cues are just a part of the above research content. Besides, this work is supported by the National Natural Science Foundation of China (Nos. 61401195, 61563036).

Our research group has done a lot research on underwater target detection. Related works have been published in international journals and conferences, such as Sensors, Optics & Laser Technology, CVPR etc. Recently, “Focused-Region Segmentation for Refocusing Images from Light Fields” written by Shen Jie, Han Lei, et al. has been published on the Journal of Signal Processing Systems.

计算机学报