



DOI: 10.12086/oe.2021.210405

基于边框加权角相关的光场深度估计算法

马 帅^{1,2,3}, 王 宁^{1,2,3}, 朱里程^{1,2},
王 帅^{1,2,3}, 杨 平^{1,2,3}, 许 冰^{1,2,3*}

¹中国科学院自适应光学重点实验室, 四川 成都 610209;

²中国科学院光电技术研究所, 四川 成都 610209;

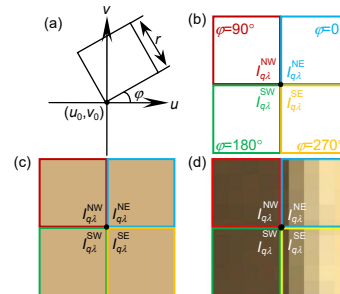
³中国科学院大学, 北京 100049

摘要: 光场成像能同时记录光线的强度信息与方向信息且具备估计场景深度的能力。然而, 深度估计的精度却容易受光场遮挡的影响。因此, 本文提出一种边框加权角相关的深度估计方法来解决该问题。首先, 该方法将光场角度域图像分成四个边框子集并分别度量这些子集中像素的相关性来构建四个代价体积, 以此解决不同类型的遮挡。其次, 该方法提出加权融合策略来融合四个代价体积, 进一步增强算法的鲁棒性, 同时保留算法的抗遮挡能力。最后, 融合后的代价体积利用引导滤波对其进行优化, 以提升深度估计的精度。实验结果表明, 提出的方法在量化指标上优于现有的方法。同时, 在绝对深度测量实验中, 提出的方法能实现高精度的测量。

关键词: 边框角相关; 加权融合; 光场成像; 深度估计

中图分类号: TP391.41; O439

文献标志码: A



马帅, 王宁, 朱里程, 等. 基于边框加权角相关的光场深度估计算法[J]. 光电工程, 2021, 48(12): 210405

Ma S, Wang N, Zhu L C, et al. Light field depth estimation using weighted side window angular coherence[J]. *Opto-Electron Eng.* 2021, 48(12): 210405

Light field depth estimation using weighted side window angular coherence

Ma Shuai^{1,2,3}, Wang Ning^{1,2,3}, Zhu Licheng^{1,2}, Wang Shuai^{1,2,3}, Yang Ping^{1,2,3}, Xu Bing^{1,2,3*}

¹Key Laboratory of Adaptive Optics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

²Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

³University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Light field imaging recodes not only the intensity information of the rays, but also its direction information, and has the ability to estimate the depth of the scene. However, the accuracy of the depth estimation is easily influenced by light field occlusion. This paper proposes a method of weighted side window angular coherence to deal with different types of occlusions. Firstly, the angular patch is divided into four side window subsets, and the coherence of the pixels in these subsets is measured to construct four cost volumes to solve different types of occlusion.

收稿日期: 2021-11-22; 收到修改稿日期: 2021-12-05

基金项目: 四川省科技计划资助(2020JDRC0001); 国家自然科学基金(61805251, 61875203, 62005285); 中国科学院青年创新与促进会(2017429)

作者简介: 马帅(1993-), 男, 博士研究生, 主要从事计算视觉与光场成像的研究。E-mail: msromam@163.com

通信作者: 许冰(1960-), 男, 硕士, 研究员, 主要从事自适应光学与光场成像技术的研究。E-mail: bing_xu_ioe@163.com

版权所有©2021 中国科学院光电技术研究所

Secondly, the weighted fusion strategy is proposed to fuse the four cost volumes to further enhance the robustness of the algorithm and retain the anti-occlusion of the algorithm. Finally, the fused cost volume is optimized by the guided filter to further improve the accuracy of depth estimation. Experimental results show that the proposed method is superior to the existing methods in the quantitative index and can achieve high-precision measurement in the absolute depth measurement experiment.

Keywords: side window angular coherence; weighted fusion; light field imaging; depth estimation

1 引言

随着科技的不断进步, 光场成像技术^[1]获得突飞猛进的发展并受到消费市场与科研领域的青睐。相比于传统成像技术, 光场成像技术能同时记录光线的强度信息与方向信息。正是由于该特性, 光场成像技术具有广泛的应用性, 例如重聚焦^[2]、超分辨^[3]、目标检测^[4]、深度估计^[5-6]等。在这些应用中, 深度估计是光场成像应用于高维视觉的关键一步, 且深度估计的精度将直接影响后续高维视觉任务。

为了从光场成像中获取精确的深度信息, 研究者们提出了许多基于光场成像的深度估计算法, 极大地推动了该领域的发展。Jeon 等人^[7]通过使用亚像素进行匹配来解决子孔径图像的窄基线问题, 使用多标签优化与迭代精化的方法对初始深度图进一步优化, 获得精确的视差估计结果。Johannsen 等人^[8]通过在中心视角中获得更低维度的字典, 然后将中心视角这些字典元素通过生成模型提升至四维光场域的字典元素, 进而实现精确的深度估计。尽管上述方法提出了有效的深度估计方法并获得了具有竞争力的深度估计结果, 但是这些方法没有考虑遮挡对深度估计产生的消极影响, 使得这些方法在遮挡处的深度估计精度受到限制。

为了降低遮挡对深度估计精度的影响, 研究者提出许多经典的方法来解决遮挡问题。Chen 等人^[9]借鉴双边滤波的思想, 将双边一致性度量用于解决光场深度估计中的遮挡问题。同时, 该方法利用引导滤波对代价体积进行优化, 获得高精度且具有锐利遮挡边缘的深度图。Zhang 等人^[10]认为 EPI 中的线特征将平行四边形分成具有差异的两部分且该差异可用于预测 EPI 中线的方向, 同时利用平行四边形算子分别计算水平方向 EPI 与垂直方向 EPI 中线特征的方向以此得到代价值。该方法也提出了有效的加权融合策略对水平与垂直方向的代价值进行融合, 进而处理遮挡与抵抗噪声。实验结果表明, 该方法在合成与真实数据集上均获得了鲁棒性的结果。然而, 仅仅利用直方图距

离去拟合 EPI 中的线特征使得该方法难以处理复杂的遮挡。Wang 等人^[11]提出改进的平行四边形算子, 利用颜色约束与直方图融合的方式来解决上述方法所面临的缺陷。Zhu 等人^[12]通过构建光场遮挡模型发现空域中的遮挡边界与角度域中的边界保持一致, 并提出将空域图像的分割结果映射到角度域图像, 进而解决不同遮挡程度的遮挡问题。该方法能处理不同类型的遮挡问题。然而, 该方法处理遮挡的能力受空域图像的分割精度与光场遮挡模型的限制。Wang 等人^[6]也通过建立光场遮挡模型发现空域与角度域图像中的边界保持一致, 并使用与空域图像中方向相同的直线将角度域图像分成两部分并分别度量这两部分的相关性, 进而处理遮挡问题。该方法提出遮挡感知的深度优化方法, 进一步提升深度估计精度。实验结果表明, 该方法能完美地处理单一遮挡物的遮挡, 但是不能处理多遮挡物的遮挡。

本文提出一种边框加权角度相关的光场深度估计来处理不同类型的遮挡问题。该方法首先将角度域图像分成四个边框子集并分别度量这些子集的角相关性来获得四个代价体积, 进而处理不同类型的遮挡。其次, 应用加权策略将四个代价体积有效地进行融合, 以此增强算法的鲁棒性同时保留算法的抗遮挡能力。最后, 利用引导滤波对加权融合的代价体积进行优化, 进一步提升深度估计精度同时保留深度估计结果在遮挡处的尖锐转接边缘。本文的主要贡献如下:

- 1) 该方法将角度域图像分成四个边框子集并分别度量这些边框子集的角相关性, 进而获取四个代价体积。在正确深度情况下, 仅仅包含来自被遮挡点像素的边框将展现出较强的角相关性, 即该边框中像素的颜色展现出一致性。这样的边框称为“纯物点型边框”。反之, “纯物点型边框”中像素的相关性能直接表征真实深度。

- 2) 该方法利用加权策略得到权重矩阵并将四个代价体积有效地融合成为一个代价体积, 进而提升算法的鲁棒性并保留算法的抗遮挡能力。融合后的代价体积, 利用引导滤波对其每个深度切片进行滤波处理,

进一步提升深度估计精度。

3) 在合成数据集与真实数据集上的实验结果表明, 本文提出的方法能有效地处理不同类型的遮挡问题。同时, 在量化评价中, 本文提出的方法在大部分场景中均获得第一, 超越现有的方法。

2 边框加权角相关的光场深度估计算法

四维光场可被参数化为 $LF(u, v, s, t)$, 如图 1 所示。其中, (u, v) 表示光场角度坐标, 即对物体从不同视角进行成像, (s, t) 表示空间坐标。在目前的研究中, 三个主要的光场特征用于计算代价体积, 即极线平面图 (EPI)、重聚焦序列、角度域图像。本文利用角度域图像来构建代价体积, 以此进行深度估计。为了得到光场角度域图像, 四维光场 $LF(u, v, s, t)$ 需重投影到不同深度标签下的被剪切光场 $LF_\lambda(u, v, s, t)$ [13], 如下所示:

$$LF_\lambda(u, v, s, t) = LF(u, v, s + u(1 - \frac{1}{\lambda}), t + v(1 - \frac{1}{\lambda})) \quad (1)$$

其中: λ 表示深度标签。为了得到空间点对应的角度域图像, 可通过固定空间坐标并从被剪切的光场图像中提取角度像素, 如下所示:

$$I_{q\lambda}(u, v) = LF_\lambda(u, v, s_q, t_q) \quad (2)$$

其中: $I_{q\lambda}$ 表示空间点 $q = (s_q, t_q)$ 在深度标签 λ 下的角度域图像。上述式子可以得到不同深度情况下的角度域图像。在正确深度情况下, 角度域图像中像素的颜

色将展现出一致性, 即像素间具有很强的相关性。反之, 当角度域图像中像素的颜色展示出一致性时, 该角度域图像所对应的深度即为正确深度。利用角度域图像计算深度过程中, 不同深度标签下的角度域图像需被获取, 并度量其中像素的相关性来确定深度。然而, 当遮挡发生时, 角度域图像中的像素将发生变化。正如文献[6]表述的, 当空间点 q 未被遮挡时, 该点对应的角度域图像中的像素将展现出颜色一致性, 如图 2(a)所示。当点 q 被遮挡时, 角度域图像不再具有这样的性质, 如图 2(b)所示。但是, 在角度域图像中来自空间点 q 的像素的颜色依然存在这一致性, 如图 2(b)红色框所示。因此, 处理遮挡问题的关键在于如何挑选来自被遮挡空间点的像素来进行相关性度量, 进而决定正确深度。

2.1 边框角度相关的抗遮挡深度估计方法

为了从角度域图像中提取仅来自被遮挡点的像素, 角度域图像被划分成几个不同模式的边框, 如图 3 所示。文献[14]定义了八个不同模式的边框去划分角度域图像, 其中边框模式包括正方形边框与三角形边框。尽管更多边框会使得算法更加鲁棒以及能够处理不同类型的遮挡, 但是该文献中的正方形边框与三角形边框在划分角度域图像上存在重叠的部分, 进而带来不必要的计算。本文仅定义了正方形边框去划分角度域图像, 如图 3(a)所示。其中, 正方形边框的边长

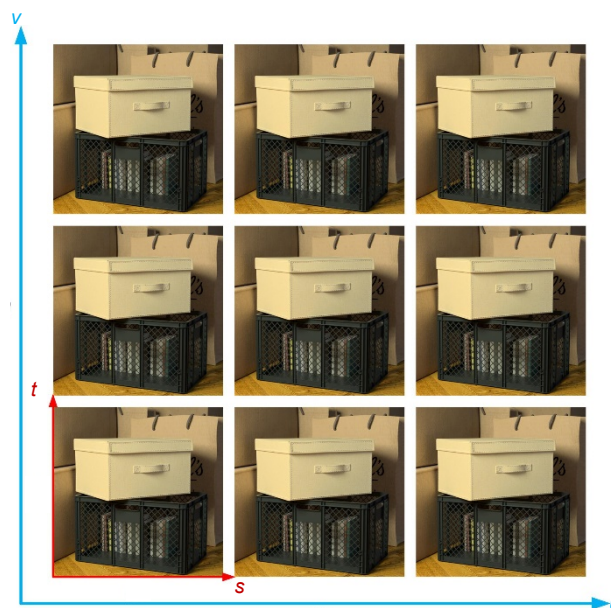


图 1 光场参数化。 (u, v) 表示角度坐标, (s, t) 表示空间坐标

Fig. 1 Light field parametrization. (u, v) represent the angular coordinates, and (s, t) represent the spatial coordinate

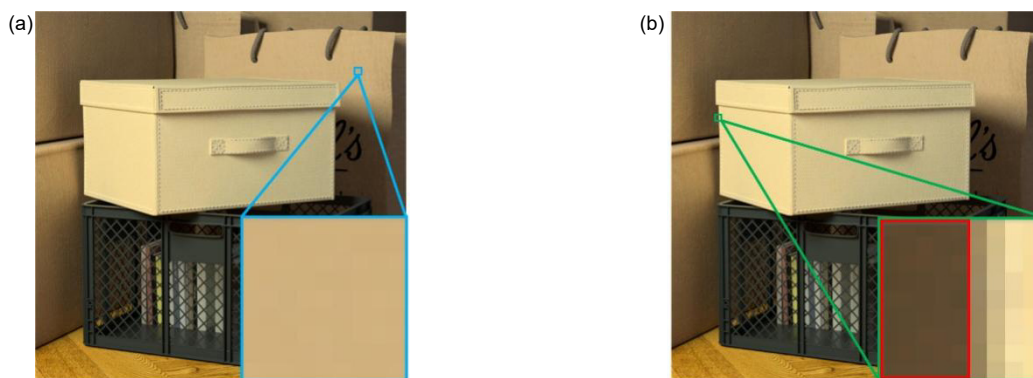


图 2 遮挡与非遮挡情况下的角度域图像。(a) 非遮挡情况下的角度域图像; (b) 遮挡情况下的角度域图像

Fig. 2 The angular image under occlusion and non-occlusion cases.

(a) The angular image under the non-occlusion case; (b) The angular image under occlusion case

为 r (单位像素), 实验中设置为角度域图像分辨率的一半。为了减少计算量, 本文使用四个离散的边框去划分角度域图像, 如图 3(b)所示。这些离散的边框分别是东北边框($I_{q\lambda}^{NE}$), 西北边框($I_{q\lambda}^{NW}$), 西南边框($I_{q\lambda}^{SW}$), 东南边框($I_{q\lambda}^{SE}$)。当空间点 q 未被遮挡时, 该点对应的角度域图像中正如前面所述, 的所有像素展现出颜色一致性。使用四个边框去划分该角度域图像时, 四个边框中的像素都保持颜色一致性, 如图 3(c)所示。然而, 当空间点 q 被遮挡时, 角度域图像中的所有像素不再具有颜色一致性, 而只有来自空点 q 的像素才保持该特性。当使用四个边框去切分这样的角度域图像时, 只有部分边框中的像素才保持颜色一致性, 如图 3(d)中的 $I_{q\lambda}^{NW}$ 与 $I_{q\lambda}^{SW}$ 边框。那么, 像这样仅包含来自被遮挡点像素的边框称为“纯物点型边框”, 例如 $I_{q\lambda}^{NW}$ 与 $I_{q\lambda}^{SW}$ 。当遮挡发生时, 正确深度可通过“纯物点型边

框”中的像素的颜色一致性, 即角相关性去表征。四个边框的角相关性需被分别度量, 进而计算正确深度。在不同的深度标签下, 每个空间点 q 所对应边框的代价体积可通过以下式子获得:

$$C_{\lambda}^n(q) = \frac{1}{|I_{q\lambda}^n|} \sum_{(u,v) \in I_{q\lambda}^n} \rho(I_{q\lambda}(u,v) - I_{q\lambda}(0,0)), \quad (3)$$

其中: $I_{q\lambda}^n$ 分别表示四个边框的像素集合, $n = \{NE, NW, SW, SE\}$, $|\cdot|$ 表示边框中像素的个数, $I_{q\lambda}(0,0)$ 表示角度域图像的中心像素, $\rho = 1 - \exp(-|v|_2 / (2\sigma^2))$ 是距离度量函数, σ 控制对噪声的敏感度。同时, 得到代价体积 $C_{\lambda}^n(q)$ 后, 初始深度可通过“胜者为王”策略得到, 如下所示:

$$d_{ini}(q) = \arg \min_{\lambda} (\arg \min_c (C_{\lambda}^n(q))), \quad (4)$$

其中 $d_{ini}(q)$ 为初始深度图。

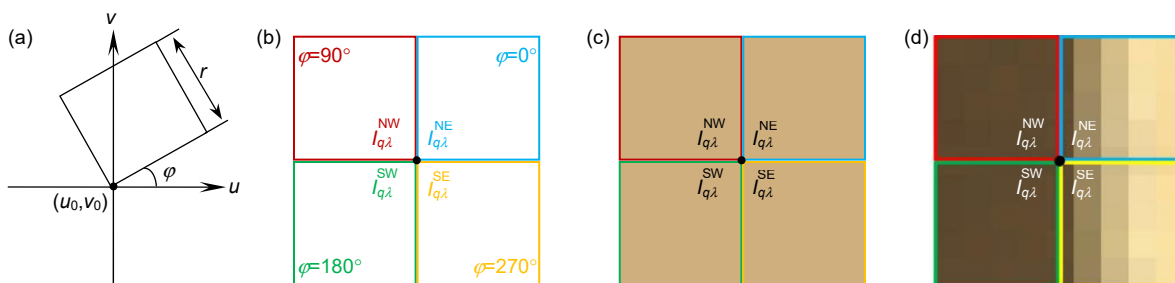


图 3 提出的边框抗遮挡方法。(a) 边框定义; (b) 四个不同的边框, 分别为东北边框, 西北边框, 西南边框, 东南边框; (c) 四个边框划分在非遮挡情况下的角度域图像; (d) 四个边框划分在遮挡情况下的角度域图像

Fig. 3 A side window anti-occlusion method. (a) The definition of the side window; (b) Four side windows, i.e., the northeast, northwest, southwest, and southeast side windows; (c) The side window divides the angular image under the non-occlusion case; (d) The side window divides the angular image under the occlusion case

2.2 加权融合代价体积的方法

在上一小节中,为了解决遮挡问题,角度域图像被分成四个不同模式的边框。这些边框中像素的角相关性被分别度量,并获得四个代价体积,即 $C_{\lambda}^{NE}(q)$, $C_{\lambda}^{NW}(q)$, $C_{\lambda}^{SW}(q)$, $C_{\lambda}^{SE}(q)$ 。正如前面所述,当遮挡发生时,只有“纯物点型边框”才能表征正确深度。换言之,当空间 q 被遮挡时,只有“纯物点型边框”对应的代价体积才能计算正确深度。当空间点 q 未被遮挡时,该点对应的四个边框均为“纯物点型边框”。为了很好地处理遮挡问题,“胜者为王”(WTA)策略通常用来选取“纯物点型边框”对应的代价体积,进而实现具有尖锐转接边缘的深度图。限于“胜者为王”策略的局限,非遮挡点对应的四个“纯物点型边框”的代价体积只能其中一个参与计算。这就意味着其他三个“纯物点型边框”的代价体积不能参与计算深度,使得深度估计精度极大程度上降低。为了提升算法深度估计的精度与抗遮挡性,非遮挡点对应的四个“纯物点型边框”的代价体积需用于计算来提升精度,遮挡点对应的“纯物点型边框”的代价体积用于解决遮挡问题。在深度估计过程中,非遮挡点与遮挡点不能被识别区分,进而不能对这两类点对应的代价体积进行分别处理。为了解决这样的问题,本文提出了加权融合代价体积的方法。受文献[10]启发,为四个边框对应的代价体积分配不同的权重来有效地融合这些代价体积。对于被遮挡的空间点,这些点对应的“纯物点型边框”中的像素在正确深度标签处展现出颜色一致性,即具有较小的代价值。然而其他边框在所有的深度标签下均具有较大的代价值。权重定义来确保“纯物点型边框”在正确深度标签处的代价值更小,即分配更小的权重,而使得其他边框的代价值在该深度标签下的代价值更大,即分配更大的权重。权重定义如下所式:

$$w_n = \exp\left(-\frac{1}{2\alpha^2} \times \frac{C_{\min}^n}{C^n}\right), \quad (5)$$

其中:

$$C_{\min}^n = \min_c(C_{\lambda}^n(q)),$$

$$\overline{C^n} = \frac{1}{n_{\text{label}}} \sum_{\lambda} (C_{\lambda}^n(q)),$$

$n = \{NE, NW, SW, SE\}$, $n_{\text{label}} = 256$ 为深度标签数量, α 控制权重的敏感度。对于被遮挡的点,由于“纯物点型边框”包含的像素都来自被遮挡的空间点,该边框对应的代价体积曲线比较“平坦”,即 C_{\min}^n 与 $\overline{C^n}$ 都同

时较小且差异较小。该边框对应的权重也就越小。由于其他边框包含来自前景遮挡物与被遮挡点的像素,这些边框对应的代价体积曲线趋于“陡峭”,即 C_{\min}^n 与 $\overline{C^n}$ 都较大且差异较大,这些边框对应的权重也就越大。获得每个代价体积的权重后,四个代价体积的权重首先被归一化来保证与其他空间点的关联性,即:

$$w_n = \frac{w_n}{\sum w_n}。$$

四个边框对应的代价体积可通过以下式子进行有效地融合,进而提升算法的鲁棒性,同时具有抗遮挡的能力。

$$C_{\lambda}(q) = w_{NE}C_{\lambda}^{NE}(q) + w_{NW}C_{\lambda}^{NW}(q) + w_{SW}C_{\lambda}^{SW}(q) + w_{SE}C_{\lambda}^{SE}(q), \quad (6)$$

其中: $C_{\lambda}(q)$ 表示融合的代价体积。同时,代价体积融合的深度图可通过以下式子得到:

$$d_{\text{ini_fusion}}(q) = \arg \min_{\lambda} (C_{\lambda}(q)), \quad (7)$$

其中 $d_{\text{ini_fusion}}(q)$ 表示代价体积融合后的深度图。图 4 展示了加权融合代价体积有效性。图 4(b)展示了初始的深度图,即由式(4)得到。该初始深度图尽管有尖锐的转接边界,即遮挡得到很好的处理,但是具有较强的噪声,会影响后期的代价体积优化以及降低最终深度估计的精度。图 4(c)展示了直接将四个代价体积通过求取平均值进行融合后的深度图。在图 4(c)的红色与绿色框中,该深度图尽管具有较弱的噪声,但是得到了模糊的遮挡边缘,使得算法的抗遮挡能力极大地下降同时影响最终深度估计的精度。图 4(d)展示了加权融合四个代价体积后的深度图,即由式(7)得到。在图 4(d)的红色与绿色框中,该深度图在降低噪声的同时获得了尖锐的转接边缘,进而证明该方法在提升算法鲁棒性的同时保留算法的抗遮挡能力。

2.3 引导滤波的代价体积优化方法

融合的代价体积 $C_{\lambda}(q)$ 尽管在一定程度上提升了算法的鲁棒性,但是通过该代价体积获得的深度图依然存在噪声且准确性不高。适当的优化方法应用于代价体积,以进一步提升深度估计精度。常见的优化方法主要有图割^[15]与代价体积滤波等。基于图割的优化方法通常具有庞大的计算量并容易破坏算法的抗遮挡性,即获得模糊的遮挡边界。基于代价体积滤波的方法通常在降低噪声的同时具有优越的保边性,即不会破坏算法的抗遮挡能力。而基于引导滤波的代价体积优化方法获得广泛使用^[16]。然而,利用引导滤波对代

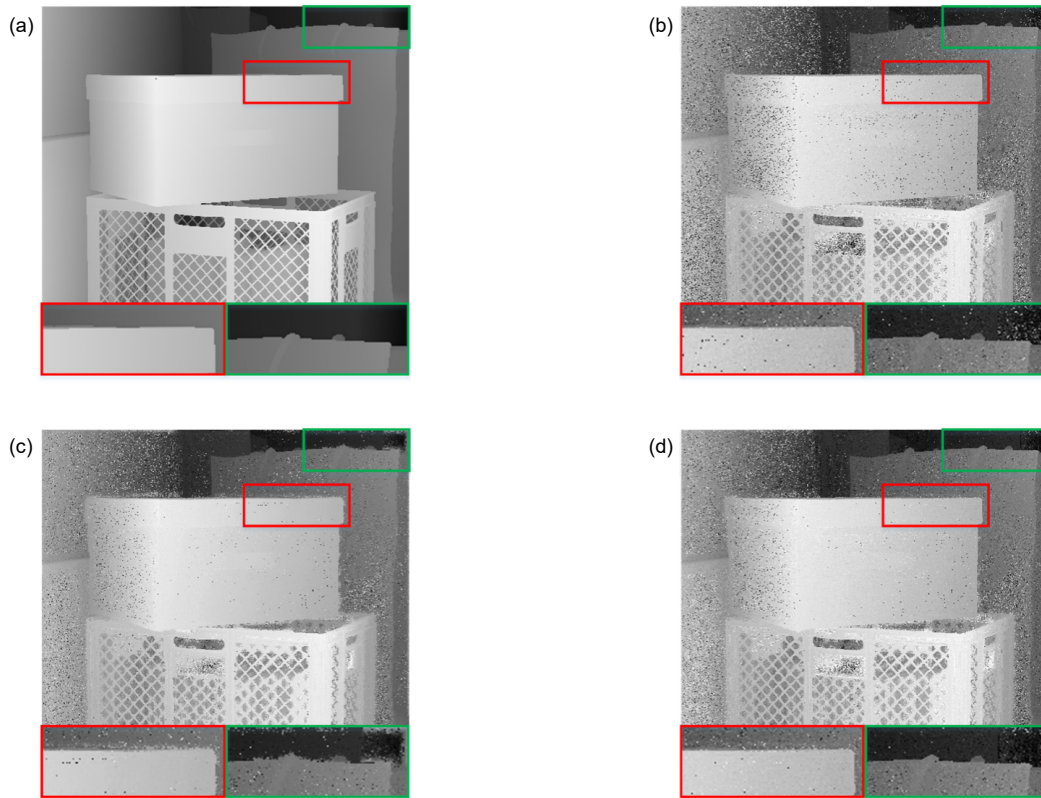


图 4 加权融合代价体的可视化结果。(a) 真值图; (b) 由式(4)获得的初始深度图; (c) 通过求平均融合代价体后的深度图; (d) 通过加权融合代价体体积后的深度图, 即由式(7)获得

Fig. 4 Visualized results of weighted fusion cost volume. (a) The ground truth; (b) The initial depth map obtained by Eq. (4); (c) The depth map obtained by fused cost volume using the average; (d) The depth map obtained by fused cost volume using weight

代价体进行优化的重要前提是很好地处理遮挡, 即获得具有锐利边缘的初始深度图。得益于本文提出的抗遮挡算法, 通过代价体融合后所获得的深度图依然保持尖锐的转接边缘。本文利用光场的中心视角图作为引导图, 对代价体 $C_\lambda(q)$ 的每个深度切片进行引导滤波处理, 如下所示:

$$C_\lambda^{\text{guid}}(q) = \sum_{\hat{q}} F_{q,\hat{q}}(P_{\text{center}}) C_\lambda(\hat{q}), \quad (8)$$

其中: \hat{q} 是以 q 为中心像素的窗口像素坐标, P_{center} 是光场中心视角图, $F_{q,\hat{q}}(\bullet)$ 为滤波核, 可通过下式子获得:

$$F_{q,\hat{q}} = \frac{1}{|\omega|^2} \cdot \sum_{k:(q,\hat{q}) \in \omega_k} (1 + (P_{\text{center}}^q - \mu_k)^T (\psi_k + \xi U)^{-1} (P_{\text{center}}^{\hat{q}} - \mu_k)) \quad (9)$$

其中: ξ 是惩罚因子, μ_k 与 ψ_k 为中心视角引导图 P_{center} 中窗口大小为 ω_k 内的均值和协方差, U 表示 3×3 的单

位矩阵, P_{center}^q 与 $P_{\text{center}}^{\hat{q}}$ 表示 3×1 的向量。优化的深度图可通过以下式子获得:

$$d_{\text{final}}(q) = \arg \min_{\lambda} (C_\lambda^{\text{guid}}(q)) \quad (10)$$

其中 $d_{\text{final}}(q)$ 为最终的深度图。

3 实验

本文提出的方法在合成数据集与真实数据集上进行实验。合成数据集来源于文献[17]。真实数据集是通过自有的光场相机(Lytro Illum)采集得到。真实数据的视差图通过相机标定得到的内参被换算成为绝对深度。式(3)中的 σ 等于 0.07, 式(5)中的 α 为 0.38, 引导滤波中的半径与步长分别为 3 与 10^{-6} 。本文对提出算法的实验结果进行了量化分析, 并与 Wang 等人[6], Zhu 等人[12], Jeon 等人[7], Strecke 等人[18]的结果进行了对比。量化评价函数使用 BadPix(0.07)与 MSE, 具体定义如下所示:

$$BadPix(0.07) = \frac{1}{|d_{final}(q)|} \cdot |\{q : abs(d_{final}(q) - g(q)) > 0.07\}| \times 100,$$

$$MSE = \frac{1}{|d_{final}(q)|} \sum_q (d_{final}(q) - g(q))^2 \times 100, \quad (11)$$

其中: $d_{final}(q)$ 与 $g(q)$ 分别表示预测的深度图与对应的真值图, $abs(\cdot)$ 表示绝对值函数, $|d_{final}(q)|$ 表示深度图大小, $|\{\cdot\}|$ 表示集合的大小。

3.1 实验结果在合成数据集上的表现

本文提出的方法在合成数据集上的结果与其他方法进行量化对比。实验结果中的遮挡细节也进行了特效放大, 便于直观地证明提出的方法能很好地处理遮挡问题且获得尖锐的遮挡边缘。在合成数据集上, 不同融合策略下的实验结果也被展示, 以此有力地证明本文提出的加权策略能明显提升算法鲁棒性且保留算法的抗遮挡能力。

3.1.1 实验结果与其他方法对比

本文提出的方法与其他方法的对比结果如图 5 所示。这些场景图均包含大量遮挡, 且存在弱纹理。从视觉效果上看, 本文方法能很好地处理遮挡问题, 并获得锐利的遮挡边缘。例如, 场景图 “Backgammon” 包含典型的水平方向的线性遮挡模式。从遮挡边界处的局部放大图看 (图中红色与绿色框), 本文提出的方法获得了清晰的遮挡边缘, 且与真值图的遮挡边缘差异甚微。其他方法却获得不理想的遮挡边缘, 即遮挡处理失败。与 Wang 等人的结果相比, 本文的结果获得了更加清晰的遮挡边界。Wang 等人的方法仅将角度域图像分成两个部分。该方法能够很好处理单一遮挡物的遮挡, 难以胜任多遮挡物的遮挡。在该场景图中 Wang 等人的方法获得了不理想的结果, 其主要原因在于该场景图包含许多细小的遮挡边缘, 超出了算法处理遮挡的能力。场景图 “Stripes” 也包含典型的垂直方向的线性遮挡模式。Jeon 等人, Wang 等人,

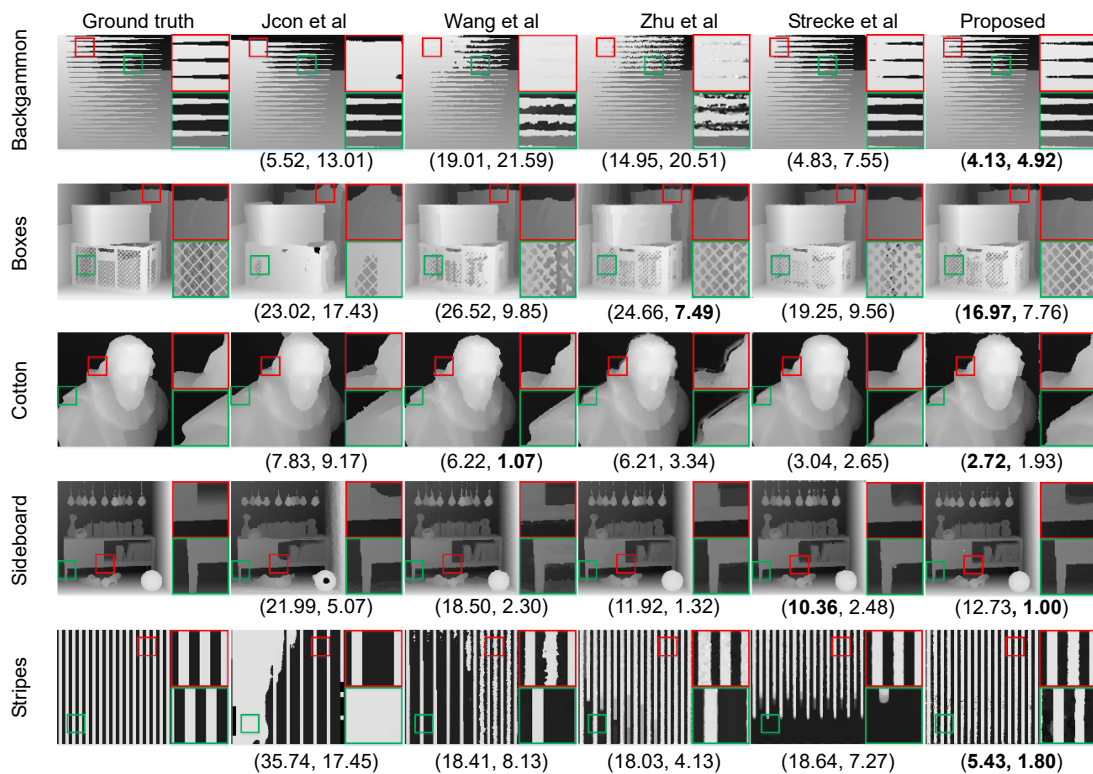


图 5 本文获得的实验结果与其他方法的对比。

场景图像的名称分别是 “Backgammon”, “Boxes”, “Cotton”, “Sideboard”, “Stripes” 场景图下方括号中的数字表示该场景图的量化评价结果, 即分别是 BadPix(0.07)与 MSE

Fig. 5 Comparison of results between the proposed method and other methods.

The names of these scenes are “Backgammon”, “Boxes”, “Cotton”, “Sideboard”, and “Stripes”.

The numbers in brackets below the scene maps are the quantitative results of the scene map

Zhu 等人以及 Strecke 等人的方法都获得了不理想的实验结果。从实验结果可以看出, 这些方法的结果获得不完整的竖条纹, 表明这些方法处理遮挡失败。相反, 本文提出的方法获得了与真值图相差甚微的结果, 表明提出的方法能很好处理遮挡问题。场景图 “Boxes” 包含许多复杂的遮挡边界, 例如对应场景图的绿色框中的网孔。该模式也对算法处理遮挡的能力带来更高的挑战。本文提出的方法得到了清晰且完整的网孔, 表明算法处理遮挡的有效性。

从量化结果来看, 本文的量化结果在大部分场景图中都获得了最优得分。与其他方法相比, 本文结果的精度在一些场景图中实现了大幅度提升。尽管一些场景图的量化指标并不是最优的, 但是与最优的指标差距较小。总之, 本文提出的方法能获得锐利的遮挡边缘, 同时实现高精度的深度估计。

3.1.2 加权代价体积融合的有效性

为了证明代价体积融合能提升深度估计的精度, WTA 策略与加权策略处理代价体积的性能对比如表 1 与表 2 所示。WTA 策略处理代价体积表示使用 WTA 策略去选择 “纯物点边框” 对应的代价体积曲线, 即不对四个代价体积进行融合。表 1 展示了两种策略在 BadPix(0.07)指标上的对比。与加权策略相比, WTA 策略在所有的场景图中的性能均下降。一些场景图的指标出现成倍的下降, 例如 “Cotton”, “Pyramid”, “Stripes” 场景图。其中, “Stripes” 场景图的 BadPix(0.07)指标下降近 5 倍。在 MSE 指标上, WTA 策略也在一些场景图中出现不同程度上的下降。这些数据对比表明合适的代价体积融合策略将增强算法的鲁棒性并提升深度估计精度。

其次, 为了证明加权融合代价体积能保留算法的

抗遮挡能力, WTA 策略, 均值融合策略, 加权融合策略融合代价体积的保边性能对比如图 6 所示。为了更好地对比不同策略之间的性能差异, 相应的代价体积没有使用引导滤波对其优化, 即均为初始深度图。均值融合策略表示四个代价体积在深度标签维度上求取平均值。本文对一些遮挡边缘进行了局部放大, 如图 6 中的红色与绿色框所示。理论上, 与其他两种策略相比, WTA 策略融合四个代价体积将获得最佳的遮挡边缘。WTA 策略直接选取 “纯物点型边框” 对应的代价体积来计算深度, 该策略不会降低提出方法的抗遮挡能力, 如图 6 第二列红色与绿色框中的箭头所指的遮挡边界。但是, 该策略仅使用了其中一个代价体积, 导致深度图存在大量噪声。尽管均值融合策略能降低深度图中噪声, 使得算法更加鲁棒, 但是该策略会破坏提出方法的抗遮挡能力。正如图 6 第三列中的红色与绿色框所示, 该策略获得模糊的遮挡边缘, 直接破坏提出方法的抗遮挡能力。这两种方法不能使得算法的鲁棒性与算法的抗遮挡能力同时兼得。本文提出的加权融合代价体积方法既能提升算法的鲁棒性, 也能保留算法的抗遮挡能力, 如图 6 第四列中的红色与绿色框所示。

3.2 绝对深度测量实验

在实验前需要说明的是前面提到的深度估计泛指视差估计。视差与深度之间存在固定转换关系。文献中一般用深度指代视差。光场图像通过自有的光场相机(Lytro Illum)获取, 如图 7(a)所示。拍摄的对象为一块定制的铝材料标准件, 具有三级台阶平面, 即①, ②, ③号平面, 如图 7(b)所示。其中, 相邻两个平面之间的距离为 50 mm, 如图 8 所示。光场相机的物方初始对准深度, 即光场相机与标准件的大致距离为

表 1 使用 WTA 策略与加权融合策略处理代价体积的性能在 BadPix(0.07)指标上的对比

Table 1 Comparison of results in BadPix(0.07) under WTA and weighted strategy

Method	Backgammon	Boxes	Cotton	Sideboard	Pyramids	Stripes
WTA	4.24	19.48	4.63	19.71	2.40	26.43
Proposed	4.13	16.97	2.72	12.73	1.81	5.43

表 2 使用 WTA 策略与加权融合策略处理代价体积的性能在 MSE 指标上的对比

Table 2 Comparison of results in MSE under WTA and weighted strategy

Method	Backgammon	Boxes	Cotton	Sideboard	Pyramids	Stripes
WTA	4.70	7.34	0.99	1.20	0.07	2.53
Proposed	4.92	7.76	1.93	1.00	0.05	1.80

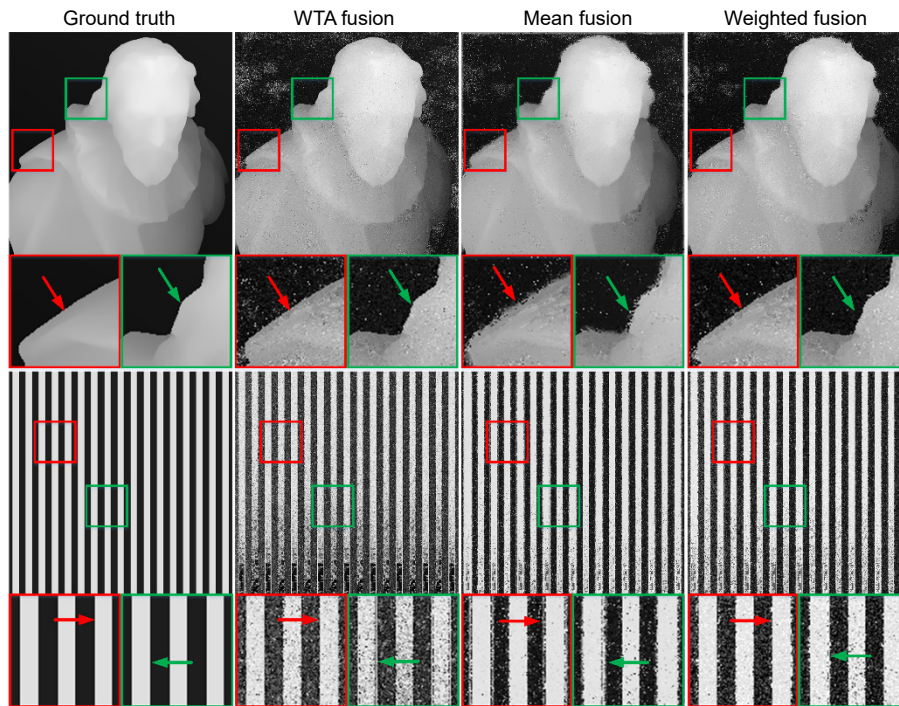


图 6 采用不同策略融合代价体积的性能对比

Fig. 6 Performance comparison of fusion cost volume among different common strategies

450 mm。光场相机采集的光场原始图像需通过解码马赛克^[19]、重采样等一系列操作得到光场子孔径图。本文采用文献^[20]的方法将原始光场图像解码成子孔径图,如图 7(b)的光场中心子孔径图。该场景的视差图可通过本文提出的方法获取,如图 9(a)所示。由于真实场景的光场图像在采集与解码过程中引入大量噪声,本文还使用了文献^[21]中的优化方法对代价体积进行优化,进而获得更加精确的视差图。通过视差图

得到深度图还需对光场相机的内参数进行标定,例如相机主透镜的焦距,主透镜光心到 CCD 的距离,CCD 到微透镜阵列的距离以及微透镜阵列中心等参数。本文通过文献^[22]的方法来标定光场相机,进而获取这些必要的参数。获得光场相机内参数后,视差图可通过这些内参数计算得到绝对深度图,如图 9(b)所示。由于光场相机的光心位置难以确定,相机到标准件之间的真实距离也难以确定。本文测量的是标准件两个

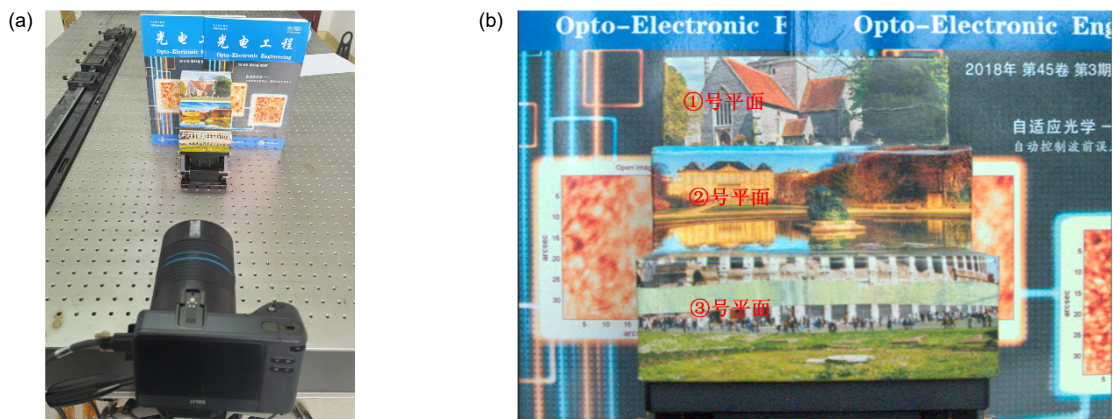


图 7 绝对深度测量的实验图。(a) 光场图像采集; (b) 光场图像的中心视角图

Fig. 7 The experimental image of the absolute depth measure. (a) Light field image capture; (b) The center image of the light field

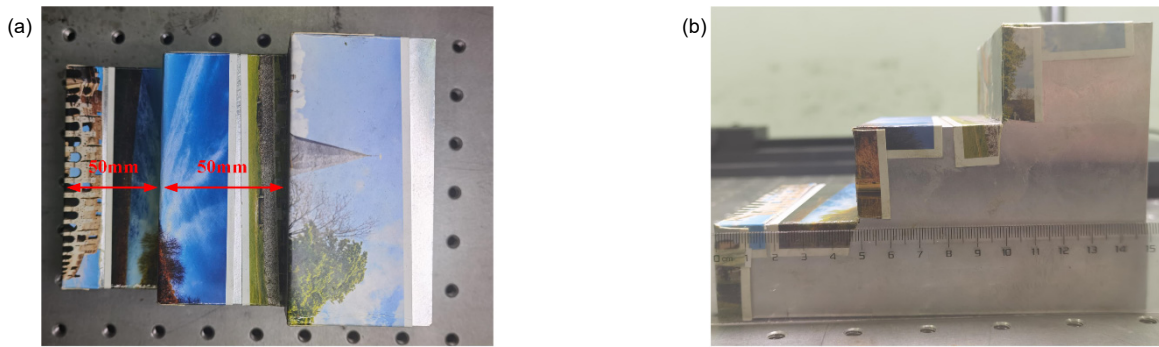


图 8 标准件平面尺寸。(a) 标准件俯视图; (b) 标准件侧视图
 Fig. 8 The plane dimension of the standard part. (a) Top view of standard part; (b) Side view of standard part

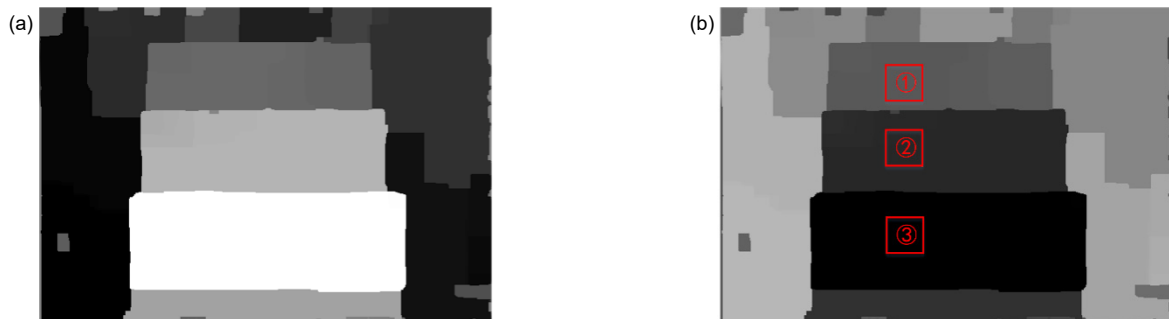


图 9 视差图与深度图。(a) 视差图; (b) 深度图
 Fig. 9 The disparity map and depth map. (a) The disparity map; (b) The depth map

表 3 提出方法绝对深度测量结果

Table 3 The results of absolute depth measurement by the proposed method

Plane	Ground truth/mm	Proposed/mm
①—②	50	58
②—③	50	47

台阶平面之间的距离,即平面①与平面②之间的距离,平面②与平面③之间的距离。求取图 9(b)中红色框中的所有深度值的均值表示相机到该平面的距离,红色框在深度图中的范围分别是(230: 280, 80: 130), (230: 280, 170: 220), (230: 280, 290: 340)。对相机到相邻平面的距离作差即可得到标准件平面之间的距离,如表 3 所示。平面①与②之间的真实间距为 50 mm,实验测量的间距为 58 mm。平面②与③之间的真实间距为 50 mm,实验测量的间距为 47 mm。测量误差主要来源于以下几个方面,即视差图估计不准确,光场相机内参标定误差,光场图像解码误差以及光场相机本身硬件存在的一些误差。这些误差都将引起绝对深度与真实深度之间的误差。

4 结 论

本文提出的边框加权角相关的深度估计算法能处理不同类型的遮挡并实现高精度的深度估计。角度域图像被分成四个边框子集。其中,仅仅包含来自被遮挡点的边框子集在正确深度情况下将展现出像素颜色一致性。通过度量不同边框中像素的一致性可以很好解决遮挡问题。本文提出加权策略去融合四个边框所对应的代价体积。该策略能有效地提升算法的鲁棒性以及深度估计的精度,同时能很好地保留算法的抗遮挡能力。最后,融合后的代价体积通过引导滤波对其进行优化,进一步提升深度估计精度。实验结果表明本文提出的方法能很好地处理遮挡问题并实现精确的

深度估计。在合成数据上, 提出的方法在部分场景中的量化指标得到明显提升。在自有光场相机采集的光场图像上, 提出的方法也精确地估计出视差图, 并通过相机标定的内参数计算出标准的尺寸。尺寸的测量误差保持在较小范围。

参考文献

- [1] Levoy M, Hanrahan P. Light field rendering[C]//*Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, 1996: 31–42.
- [2] Ng R. Fourier slice photography[J]. *ACM Trans Graph*, 2005, **24**(3): 735–744.
- [3] Zhao Y Y, Shi S X. Light-field image super-resolution based on multi-scale feature fusion[J]. *Opto-Electron Eng*, 2020, **47**(12): 200007.
赵圆圆, 施圣贤. 融合多尺度特征的光场图像超分辨率方法[J]. 光电工程, 2020, **47**(12): 200007.
- [4] Piao Y R, Rong Z K, Xu S, et al. DUT-LFSaliency: versatile dataset and light field-to-RGB saliency detection[Z]. arXiv preprint arXiv:2012.15124, 2020.
- [5] Wu D, Zhang X D, Fan Z G, et al. Depth acquisition of noisy scene based on inline occlusion handling of light field[J]. *Opto-Electron Eng*, 2021, **48**(7): 200422.
吴迪, 张旭东, 范之国, 等. 基于光场内联遮挡处理的噪声场景深度获取[J]. 光电工程, 2021, **48**(7): 200422.
- [6] Wang T C, Efros A A, Ramamoorthi R. Depth estimation with occlusion modeling using light-field cameras[J]. *IEEE Trans Pattern Anal Mach Intell*, 2016, **38**(11): 2170–2181.
- [7] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015: 1547–1555.
- [8] Johannsen O, Sulc A, Goldluecke B. What sparse light field coding reveals about scene structure[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016: 3262–3270.
- [9] Chen C, Lin H T, Yu Z, et al. Light field stereo matching using bilateral statistics of surface cameras[C]//*2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014: 1518–1525.
- [10] Zhang S, Sheng H, Li C, et al. Robust depth estimation for light field via spinning parallelogram operator[J]. *Comput Vis Image Underst*, 2016, **145**: 148–159.
- [11] Wang W K, Lin Y F, Zhang S. Enhanced spinning parallelogram operator combining color constraint and histogram integration for robust light field depth estimation[J]. *IEEE Signal Process Lett*, 2021, **28**: 1080–1084.
- [12] Zhu H, Wang Q, Yu J Y. Occlusion-model guided antiocclusion depth estimation in light field[J]. *IEEE J Sel Top Signal Process*, 2017, **11**(7): 965–978.
- [13] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera[R]. Computer Science Technical Report CSTR, 2005.
- [14] Ma S, Guo Z H, Wu J L, et al. Occlusion-aware light field depth estimation using side window angular coherence[J]. *Appl Opt*, 2021, **60**(2): 392–404.
- [15] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. *IEEE Trans Pattern Anal Mach Intell*, 2001, **23**(11): 1222–1239.
- [16] He K M, Sun J, Tang X O. Guided image filtering[J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, **35**(6): 1397–1409.
- [17] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4d light fields[C]//*Proceedings of the 13th Asian Conference on Computer Vision*, Taiwan, China, 2016: 19–34.
- [18] Strecke M, Alperovich A, Goldluecke B. Accurate depth and normal maps from occlusion-aware focal stack symmetry[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 2017: 2529–2537.
- [19] Shi M D, Zhang X D, Dong Y L, et al. A light field demosaicing method with double guided filtering[J]. *Opto-Electron Eng*, 2019, **46**(12): 180539.
石梦迪, 张旭东, 董运流, 等. 一种双引导滤波的光场去马赛克方法[J]. 光电工程, 2019, **46**(12): 180539.
- [20] Dansereau D G, Pizarro O, Williams S B. Decoding, calibration and rectification for lenslet-based plenoptic cameras[C]//*Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 2013: 1027–1034.
- [21] Williem, Park I K, Lee K M. Robust light field depth estimation using occlusion-noise aware data costs[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **40**(10): 2484–2497.
- [22] Bok Y, Jeon H G, Kweon I S. Geometric calibration of micro-lens-based light field cameras using line features[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(2): 287–300.

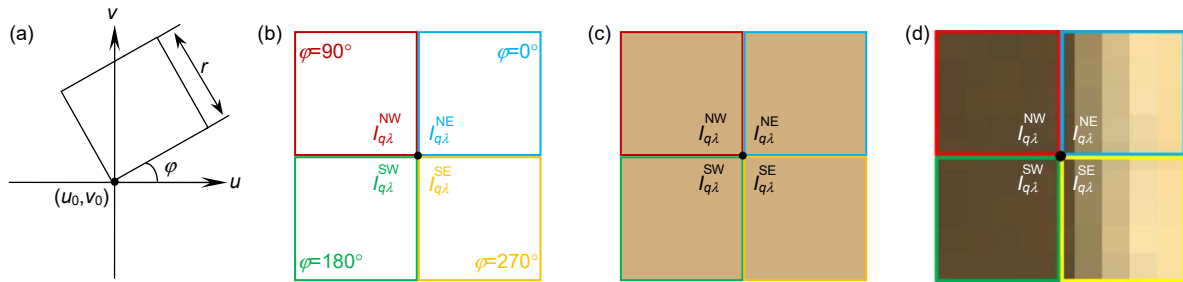
Light field depth estimation using weighted side window angular coherence

Ma Shuai^{1,2,3}, Wang Ning^{1,2,3}, Zhu Licheng^{1,2}, Wang Shuai^{1,2,3}, Yang Ping^{1,2,3}, Xu Bing^{1,2,3*}

¹Key Laboratory of Adaptive Optics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

²Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

³University of Chinese Academy of Sciences, Beijing 100049, China



Anti-occlusion light field depth estimation based on side window angular coherence

Overview: Compared with traditional imaging techniques, the light field imaging technique can recode the intensity information and direction information of rays at the same time, which is favored by both scientific research and commercial fields. Meanwhile, the equipment of recoding light field has also been greatly developed, such as programmable aperture light field, camera array, the gantry system, and micro-lens-array based light field cameras. Due to the characteristics of light field imaging, it is also applied to many fields, such as refocus, 3D reconstruction, super-resolution, object detection, light field edit, and depth estimation. Among them, depth estimation is a key step in the application of light field to high-dimensional vision, such as 3D reconstruction. However, the accuracy of depth estimation is easily influenced by the light field occlusion.

In this paper, we proposed a method of weighted side window angular coherence to solve different types of occlusion problems. Firstly, the angular patch image is divided into four different patterns of side window subsets and the pixels in these subsets are measured respectively to construct four cost volumes. The side window subset which only contains the pixels from the occluded point will exhibit the photo consistency when the depth label represents the true depth label. Then, the true depth can be obtained by the cost volume corresponding to the subsets only containing the occluded pixels, which can deal with the occlusion problem well. Secondly, we proposed a weighted strategy to fuse four cost volumes into one cost volume. According to the characteristics of four side window subsets, the cost volume corresponding to four side window subsets is given by different weighted values, which can enhance the robustness of the proposed algorithm and retain its ability to resist different types of occlusions. Finally, the fused cost volume is optimized by the guided filter to further improve the quality of the depth map. Experimental results in both synthetics and real scenes show the proposed method can handle the occlusion problem well and outperform the existing methods, especially near occlusion boundaries. In synthetic scenes, the proposed method is nearly as good as the other methods in the quantitative index of some scenes, which further proves the effectiveness of our method. In the real scene captured by our light field camera, combined with the calibration parameters of the light field camera, our method can accurately measure the surface size of the standard part, which proves the robustness of our method.

Ma S, Wang N, Zhu L C, *et al.* Light field depth estimation using weighted side window angular coherence[J]. *Opto-Electron Eng*, 2021, 48(12): 210405; DOI: 10.12086/oe.2021.210405

* E-mail: bing_xu_ioe@163.com