

## 遮挡场景的光场图像深度估计方法

张旭东, 李成云<sup>†</sup>, 汪义志, 熊 伟

(合肥工业大学 计算机与信息学院, 合肥 230009)

**摘 要:** 光场相机通过单次拍摄可获取立体空间中的 4 维光场数据, 利用光场的多视角特性可从中提取全光场图像的深度信息. 然而, 现有深度估计方法很少考虑场景中存在遮挡的情况, 当场景中有遮挡时, 提取深度信息的精度会明显降低. 对此, 提出一种新的基于多线索融合的光场图像深度提取方法以获取高精度的深度信息. 首先分别利用自适应散焦算法和自适应匹配算法提取场景的深度信息; 然后用峰值比作为置信以加权融合两种算法获取的深度; 最后, 用具有结构一致性的交互结构联合滤波器对融合深度图进行滤波, 得到高精度深度图. 合成数据集和真实数据集的实验结果表明, 与其他先进算法相比, 所提出的算法获取的深度图精度更高、噪声更少、图像边缘保持效果更好.

**关键词:** 光场相机; 深度估计; 全光场图像; 遮挡; 多线索融合; 联合滤波器  
**中图分类号:** TP391      **文献标志码:** A

### Light field depth estimation for scene with occlusion

ZHANG Xu-dong, LI Cheng-yun<sup>†</sup>, WANG Yi-zhi, XIONG Wei

(School of Computer and Information, Hefei University of Technology, Hefei 230009)

**Abstract:** The light field camera can obtain the four-dimensional light field data from stereoscopic space with one shot. After that, the depth information can be extracted by multiview characteristic of the light field. However, the existing depth estimation method rarely considers the presence of occlusion in the scene. The accuracy of the extracted depth information is significantly reduced, when the scene is blocked. Aiming at this problem, this paper presents a new depth estimation method of light field image based on multi-clues fusion to obtain high-precision depth information. Firstly, the adaptive defocus algorithm and the adaptive matching algorithm are used to extract the depth information of the scene. And then the peak ratio is taken as confidence to synthesize the depth images. Finally, the fusion depth map is filtered by a combined filter with structural consistency, and a high precision depth map is obtained. The experimental results in the virtual data and real data show that, compared with other advanced algorithms, the depth images obtained by the proposed method are more accurate, less noise, and the edges are more clearer.

**Keywords:** light field camera; depth estimation; light field image; occlusion; multi-clue fusion; joint filter

## 0 引 言

光场相机<sup>[1]</sup>是一种新型的多目视觉相机, 巧妙地结合了计算光学理论、计算成像技术及数字图像处理技术等. 微透镜式光场相机在主透镜和图像传感器之间放置了微透镜阵列, 所以单次曝光可以记录 3 维场景的位置、角度等四维光场信息. 由于光场相机具有“先拍摄, 后对焦”的特点, 通过对光场相机获取的光场数据进行分析处理, 可以获得场景的深度信息, 进而在目标识别<sup>[2]</sup>、显著性检测<sup>[3-7]</sup>、超分辨率重建<sup>[8-9]</sup>以及三维表面重建<sup>[10]</sup>等领域都有重要的应用.

目前, 光场的深度估计方法可分为以下 3 类: 基于极平面图像 (EPI) 的深度估计、基于多视角的深度估计及基于重聚焦的深度估计.

1) 基于 EPI 的深度估计利用 EPI 的结构张量得到像素匹配方向的场景深度图. 例如: Chen 等<sup>[11]</sup>利用 EPI 获取深度一致性曲线, 并引入双边一致性度量方法衡量多视角图像匹配的可靠性, 实现深度估计; Li 等<sup>[12]</sup>构造稀疏线性系统, 通过计算 EPI 的结构张量得到局部深度信息, 再利用颜色一致性约束校正错误深度估计区域, 从而获取包含更多细节的连续

收稿日期: 2017-07-13; 修回日期: 2017-11-14.

基金项目: 国家自然科学基金项目 (61403116); 中国博士后基金项目 (2014M560507).

责任编委: 林崇.

作者简介: 张旭东 (1966—), 男, 教授, 博士, 从事智能信息处理、机器视觉等研究; 李成云 (1991—), 女, 硕士生, 从事光场技术、计算机视觉的研究.

<sup>†</sup>通讯作者. E-mail: chengyun@mail.hfut.edu.cn

深度图; Wanner等<sup>[13]</sup>在全变分框架下,将改进的基于约束项的立体匹配算法应用于EPI图像提取深度信息; Zhang等<sup>[14]</sup>对通过EPI图像斜率得到的局部深度信息进行可靠性度量,引入局部线性嵌入方法估计不可靠像素的视差,利用视差标签和传递获得连续一致的深度图. 上述方法EPI的提取过程计算量较大,时间复杂度较高,实时性较差,而且深度估计结果容易受到噪声影响.

2) 基于重聚焦的深度估计充分利用光场的数字重聚焦优势,对焦点堆栈图像进行聚焦性检测,提取深度信息. 例如: Mousnier等<sup>[15]</sup>利用光场数据获取全聚焦图像,之后再利用相机校准参数与图像原始数据之间的映射关系,将全聚焦图像转换为深度图; Lin等<sup>[16]</sup>利用焦点堆栈图像的对称性获取光场深度信息; Lu等<sup>[17]</sup>利用LytroDesk<sup>[18]</sup>获取焦点堆栈图像,进一步通过Laplace算子检测提取深度信息. 上述方法获取的深度图可以保留更多细节,但会引入散焦误差,降低深度图的精确度.

3) 基于多视角的深度估计利用光场多视角图像之间形成的视差提取深度信息. 例如: Yu等<sup>[19]</sup>通过求解不同微透镜后方子透镜图像之间的匹配关系,并采用全局优化方法实现场景的深度估计; Jeon等<sup>[20]</sup>利用傅里叶域的相位平移理论表示子微透镜图像间的亚像素位移,将中心视角与其他视角进行立体匹配,然后利用图割进行多标签优化,最终通过二次函数拟合局部深度图获取连续深度图; Calderon等<sup>[21]</sup>使用分类法完成深度估计,利用损耗函数表示不同深度间的匹配关系,并通过最优方法获得最佳深度估计值; Liu等<sup>[22]</sup>通过渲染增强获得高保真的4D光场,将相同角度维度 $v$ 的中心视角和左视角进行立体匹配,用对应的中心视角和右视角匹配验证,并采用多标签优化方法获取深度信息. 上述方法计算效率较高,但是对匹配噪声难以抑制,此外,当场景中出現遮挡或重复纹理时,视差线索的可靠性会明显降低.

综上,利用单个方法的线索进行深度估计存在精度或鲁棒性不足的问题. 因此,人们提出了利用多线索融合提取深度信息的方法. 根据散焦和匹配线索是场景深度变化在不同维度上的表现, Tao等<sup>[23]</sup>提出了融合散焦(Defocus)和匹配(Correspondence)线索,再利用马尔可夫随机场(MRF)进行全局优化的深度估计算法. 进一步,他们利用阴影和光照线索约束校正深度图,提升了深度估计性能<sup>[24]</sup>. 此类方法降低了散焦误差,且对光线充足的区域效果较好. 然而,当场景中存在遮挡时,对场景深度估计的精确度或鲁棒性

会明显降低. 对此, Wang等<sup>[25]</sup>提出了遮挡知道的深度估计方法,然而此方法仅局限于单个遮挡且高度依赖于边缘检测结果.

为了更好地解决由遮挡引起的场景深度估计精确性和鲁棒性低的问题,本文在融合线索的基础上提出一种对遮挡场景鲁棒的深度估计方法. 通过对各线索的分析,利用对遮挡场景深度估计结果较好的自适应散焦线索和自适应匹配线索提取场景深度信息,然后用峰值比计算它们各自的置信,再结合置信加权求和与MRF对获取的深度信息进行融合,结合各线索优势得到初始深度图,采用具有结构一致性的交互结构联合滤波器对初始深度图进行滤波处理,并引入迭代优化算法进行能量最小化,从而获得全局一致的高精度的深度估计结果.

## 1 散焦与匹配线索获取深度图原理

微透镜式光场相机在主透镜与图像传感器之间放置微透镜阵列,通过一次曝光,可同时获取光辐射的空间分布及辐射传输的方向信息(空间信息和角度信息),如图1所示. 由物体上一点发出的光线会在微透镜所在的像平面上聚集于一点,然后微透镜阵列将光线按照不同的方向分散到传感器上,因此,成像传感器可以记录同一空间点发出的不同方向的光线.

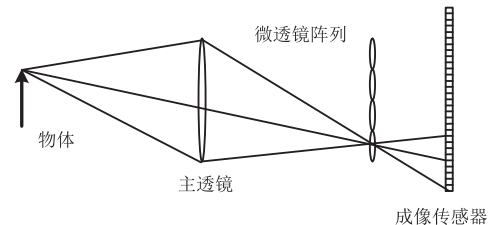


图1 Lytro相机成像模型

图2是光场的多视角图像. 其中: 图2(a)是4D光场示意图,  $(x, y)$ 与 $(u, v)$ 分别表示空间坐标与角度坐标. 固定其中一个角度坐标轴 $u$ ,然后在所有的空间坐标中连续变化另一个角度坐标轴 $v$ ,便可得到光场的一组多视角图像. 连续地变化坐标轴 $u$ ,可以获得该4D光场空间的所有不同视角的图像(多视角图像),如图2(b)所示.

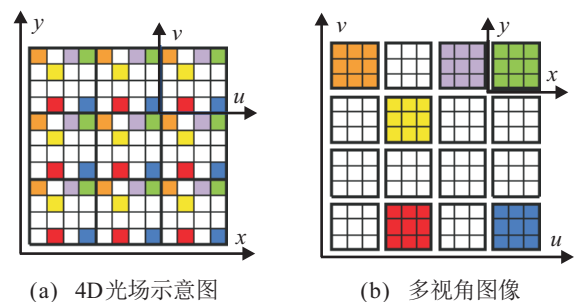


图2 光场的多视角图像

利用光场多视角信息即可进行光场数字重聚焦<sup>[1]</sup>,获得重聚焦图像.图3所示为数字重聚焦原理图.其中: $\Pi$ 是微透镜平面, $\Omega$ 是传感器平面, $\Psi$ 是重聚焦平面.重聚焦公式如下:

$$L_{\alpha F}(x, y, u, v) = L_F\left(x + u\left(1 - \frac{1}{\alpha}\right), y + v\left(1 - \frac{1}{\alpha}\right), u, v\right). \quad (1)$$

其中: $\alpha$ 是聚焦系数, $L_F$ 是输入光场图像, $L_{\alpha F}$ 是重聚焦图像.

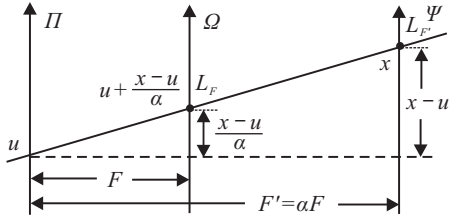


图3 数字重聚焦原理

在获得重聚焦图像和多视角图像之后,可利用散焦线索和匹配线索提取深度信息,具体方法如下.

1) 利用散焦线索提取深度信息,是根据图像颜色一致性<sup>[24]</sup>计算得到深度散焦响应,从中恢复深度.首先,通过计算一个空间像素的所有角度像素颜色均值与中心视角像素颜色的差,获得散焦深度响应

$$D_{\alpha}(x, y) = \frac{1}{|W_D|} \sum_{(x', y') \in W_D} |\bar{L}_{\alpha}(x', y') - P(x', y')|. \quad (2)$$

其中

$$\bar{L}_{\alpha}(x, y) = \frac{1}{N_{(u, v)}} \sum_{(u', v')} L_{\alpha}(x, y, u', v'),$$

$N_{(u, v)}$ 是角度像素 $(u, v)$ 的数目, $W_D$ 是窗口尺寸(为了增加鲁棒性), $P(x, y)$ 是中心视角像素颜色.然后,在所有剪切图片中取散焦响应的最小值所对应的 $\alpha$ 值,即为深度信息

$$\alpha_D^*(x, y) = \arg \min_{\alpha} D_{\alpha}(x, y). \quad (3)$$

2) 利用匹配线索提取深度信息,是根据立体匹配原理<sup>[26]</sup>,通过计算3D空间中一点在不同视角图像中的视差(方差)捕获场景深度信息.首先,计算得到角度区域的方差

$$V_{\alpha}(x, y) = \frac{1}{N-1} \sum_{u, v} (L_{\alpha}(x, y, u, v) - \bar{L}_{\alpha}(x, y))^2. \quad (4)$$

其中: $N$ 是角度区域的像素个数, $\bar{L}_{\alpha}(x, y)$ 是重聚焦图像在角度像素上的均值.然后,在所有重聚焦图像中取匹配响应最小值所对应的 $\alpha$ 值,即为深度信息

$$\alpha_C^*(x, y) = \arg \min_{\alpha} C_{\alpha}(x, y) =$$

$$\arg \min_{\alpha} \frac{1}{|W_C|} \sum_{(x', y') \in W_C} V_{\alpha}(x', y'), \quad (5)$$

其中 $W_C$ 是以当前像素为中心的窗口尺寸.

## 2 本文方法

由以上分析可知,利用散焦线索和匹配线索可以获得场景的深度信息.然而,当场景中存在遮挡时,散焦和匹配线索不再适用,融合获取的深度图精确度也会明显降低.对此,本文先利用散焦和匹配算法提取深度信息,再融合两种算法提取的深度信息,得到一种对遮挡场景有效的深度估计算法.进一步,为减轻噪声和伪影对深度图精度的影响,本文采用交互结构联合滤波器对深度图进行滤波处理.图4为本文算法的框图.算法主要分为以下几个步骤:1) 对光场数据进行重聚焦处理,得到重聚焦切片;2) 分别采用对遮挡场景深度估计结果较好的自适应散焦算法和自适应匹配算法提取场景深度信息;3) 用峰值比分别计算自适应散焦响应和自适应匹配响应的置信;4) 结合加权置信和MRF融合步骤2)得到的深度信息,获得初始深度信息;5) 用具有结构一致性的交互结构联合滤波器对初始深度信息进行滤波处理,再将滤波后的深度信息用迭代方法进行优化处理,最后得到全局连续的高精度深度图.

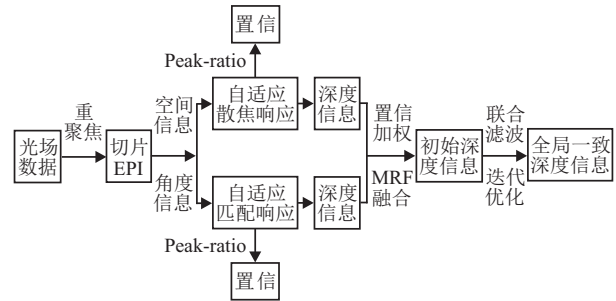


图4 本文算法框图

### 2.1 自适应散焦算法获取深度信息

当场景中存在遮挡时,单个空间像素的整个角度区域 $(9 \times 9)$ 会受遮挡的影响,使得提取的深度信息存在伪影.此时可通过考虑遮挡存在的情况,利用自适应散焦算法<sup>[27]</sup>获取场景深度信息.将单个空间像素的整个角度区域分成9个子区域 $(3 \times 3)$ ,分别计算9个子区域 $N_c(x, y)$ 的散焦响应 $D_{c, \alpha}(x, y)$ 以处理遮挡,如图5所示.

$$D_{c, \alpha}(x, y) = \frac{1}{|N_c(x, y)|} \sum_{(x', y') \in N_c(x, y)} |\bar{L}_{\alpha}(x', y') - P(x', y')|, \quad (6)$$

其中 $c$ 是子区域的索引.

取9个散焦响应中的最小散焦响应作为最终的

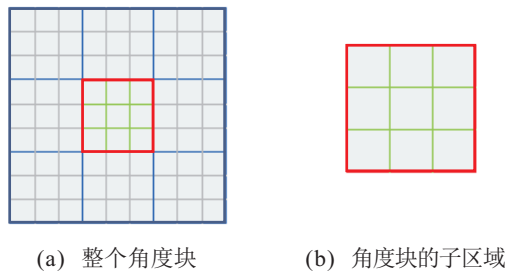


图 5 角度区域示意

散焦响应来实现散焦响应的自适应. 至此, 得到自适应散焦响应  $D_{c^*,\alpha}(x, y)$ , 子区域

$$c^* = \min_c D_{c,\alpha}(x, y).$$

### 2.2 自适应匹配算法获取深度信息

匹配算法假设空间像素的角度块呈现图片一致性, 但是当场景中出現遮挡时, 图片一致性不再满足, 该算法提取的深度信息会产生伪影, 精确度有所降低. 此时可采用基于遮挡原理的自适应匹配算法<sup>[25]</sup>, 用通过中心视角像素的角度边缘将角度块均分成两个区域, 如图 6 所示. 用较小方差的角度区域代替整个角度区域, 计算匹配响应, 实现自适应.

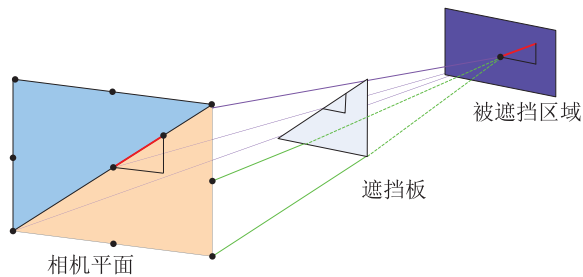


图 6 光场遮挡感知模型

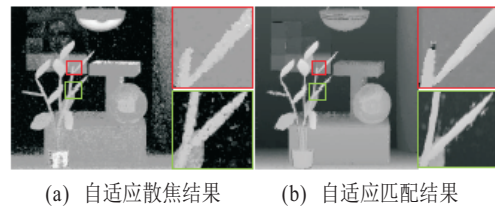
具体做法是, 首先假设  $(u_1, v_1)$  和  $(u_2, v_2)$  分别为两个角度区域的角度像素, 计算两个角度区域的方差

$$V_{\alpha,j}(x, y) = \frac{1}{N_j - 1} \sum_{u_j, v_j} (L_{\alpha}(x, y, u_j, v_j) - \bar{L}_{\alpha,j}(x, y))^2. \quad (7)$$

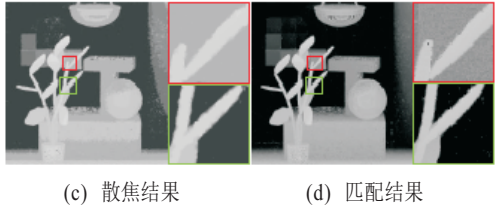
其中:  $N_j$  是角度区域  $j$  ( $j = 1, 2$ ) 里的像素个数,  $\bar{L}_{\alpha,j}(x, y)$  是角度区域的均值. 然后, 令  $i = \arg \min_{j=1,2} \{V_{\alpha,j}(x, y)\}$  为有较小方差角度区域的索引, 可得自适应匹配响应

$$C_{\alpha}(x, y) = V_{\alpha,i}(x, y). \quad (8)$$

图 7 为分别采用自适应散焦算法、自适应匹配算法、散焦算法以及匹配算法获取的深度估计结果. 由图 7 可看出, 自适应散焦算法和自适应匹配算法提取的深度信息遮挡边缘清晰、伪影少, 且能呈现正确的物体形状.



(a) 自适应散焦结果 (b) 自适应匹配结果



(c) 散焦结果 (d) 匹配结果

图 7 不同方法获取的深度估计结果

### 2.3 深度信息融合算法

由图 7 可知: 自适应散焦算法提取的深度信息对遮挡鲁棒, 但对吊灯等明亮的点和区域敏感; 由自适应匹配算法获取的深度信息对遮挡和明亮的区域鲁棒, 但是高度依赖边缘检测结果. 因此, 为了使两种算法的优势互补, 本文结合置信度加权平均与 MRF<sup>[23]</sup> 融合两种算法提取的深度信息, 融合公式为

$$\min_Z \sum_{D \in \{D_d, D_c\}} \lambda_D \sum_i W_i^D |Z_i - Z_i^D| + \lambda_{\text{flat}} \sum_{(x,y)} \left( \left| \frac{\partial Z_i}{\partial x} \right|_{(x,y)} + \left| \frac{\partial Z_i}{\partial y} \right|_{(x,y)} \right). \quad (9)$$

其中:  $Z$  是场景初始深度图;  $\lambda_D$  是自适应散焦和自适应匹配的权重系数;  $\lambda_{\text{flat}}$  是控制输出深度图平整度的拉普拉斯约束参数;  $W$  是深度值置信度, 控制融合过程中两种方法深度值的权重, 有

$$W = R_{\alpha^*}(p) / R_{\alpha^{*2}}(p). \quad (10)$$

$R$  表示深度响应,  $\alpha^*$  和  $\alpha^{*2}$  分别是深度局部最优值 (最大峰或谷) 和下一个局部最优值,  $p$  是空间像素  $(x, y)$  的简写.

图 8 为自适应散焦算法、自适应匹配算法及融合算法得到的深度图. 由图 8 可看出, 融合算法获取的深度图颗粒噪声少, 且保留了自适应匹配深度图与自适应散焦深度图中的遮挡边缘等细节.



(a) 自适应匹配深度图 (b) 自适应散焦深度图 (c) 融合深度图

图 8 融合深度估计结果

### 2.4 融合深度图的交互结构联合滤波与优化

自适应散焦响应对噪声鲁棒, 但是自适应匹配响应对匹配噪声较为敏感, 而且在拍摄真实场景图片

时会引入光照等噪声,因此,会导致融合深度图在一定程度上受噪声影响,使边缘清晰度下降.由图8可看出,融合深度图的边缘存在噪声,此时需对其进行去噪处理.引导滤波<sup>[28-29]</sup>根据引导图像 $P$ 的结构信息保留待滤波深度图 $Z$ 的边缘信息,具有边缘保持特性,所以可用其对深度图进行滤波.然而,当引导图像包含噪声、阴影等额外的结构信息时,滤波后深度图会引入不正确的边缘信息和纹理.为了避免将引导图像 $P$ 中的噪声、阴影等额外结构信息引入滤波后深度图,本文采用具有结构一致性的交互结构联合滤波<sup>[30]</sup>,该方法不是直接用引导图像进行联合滤波,而是先利用引导图像与待滤波深度图的结构交互性对引导图像 $P$ (全聚焦彩色图像)进行处理,得到仅包含图像 $P$ 与图像 $Z$ 一致结构的图像 $P_c$ ,再用图像 $P_c$ 作为引导图像进行滤波.图像 $P_c$ 的获取方法如下.

首先,计算归一化互相关函数 $\rho(P_p, Z_p)$ ,判断图像 $P$ 和图像 $Z$ 以像素 $p$ 为中心的图像块 $P_p$ 与 $Z_p$ 的相似性,其中

$$\rho(P_p, Z_p) = \frac{\text{cov}(P_p, Z_p)}{\sqrt{\sigma(P_p)\sigma(Z_p)}}. \quad (11)$$

$\text{cov}(P_p, Z_p)$ 是图像块强度的协方差, $\sigma(P_p)$ 和 $\sigma(Z_p)$ 表示方差.当两个图像块有相同边缘时, $|\rho(P_p, Z_p)| = 1$ ;否则 $|\rho(P_p, Z_p)| < 1$ .当两个图像块相似时 $|\rho(P_p, Z_p)|$ 有较大的值.然而 $\rho(P_p, Z_p)$ 是非线性的,在结构优化时不能直接使用,因此,可通过结合局部图像块的最小二乘回归函数 $f(P, Z, a_p^1, a_p^0)$ 和 $f(Z, P, b_p^1, b_p^0)$ 得到两个图像块相似度量函数 $S(P_p, Z_p)$ 来代替 $\rho(P_p, Z_p)$ ,并判断图像块 $P_p$ 与图像块 $Z_p$ 的相似性.

$$\begin{aligned} S(P_p, Z_p) = & \min_{a_p^1, a_p^0} \frac{1}{|N|} f(P, Z, a_p^1, a_p^0) + \\ & \min_{b_p^1, b_p^0} \frac{1}{|N|} f(Z, P, b_p^1, b_p^0) = \\ & (\sigma(P_p)^2 + \sigma(Z_p)^2)(1 - \rho(P_p, Z_p)^2)^2. \end{aligned} \quad (12)$$

其中

$$\begin{aligned} f(P, Z, a_p^1, a_p^0) &= \sum_{q \in N(p)} (a_p^1 P_q + a_p^0 - Z_q)^2, \\ f(Z, P, b_p^1, b_p^0) &= \sum_{q \in N(p)} (b_p^1 Z_q + b_p^0 - P_q)^2, \end{aligned}$$

$a_p^1, a_p^0, b_p^1$ 和 $b_p^0$ 是回归系数, $N(p)$ 表示图像块.

在包含边缘的图像块中,当 $|\rho(Z_p, P_p)|$ 趋于1时, $S(Z_p, P_p)$ 趋于0,表明两个图像块有一致边缘;当 $|\rho(Z_p, P_p)|$ 值较小时, $S(Z_p, P_p)$ 输出较大的值,此时

两个图像块的边缘不一致.当图像块不包含主要边缘时, $S(Z_p, P_p)$ 输出较小的值,此时也可将两个图像块视为一致结构(它们都平滑).所以,最小化式(12)可得图像块一致结构.

最后,基于图像块相似度量,计算得到图像块信息和 $E_s(P, Z, a, b)$ ,即

$$E_s(P, Z, a, b) = \sum_p (f(P, Z, a_p^1, a_p^0) + f(Z, P, b_p^1, b_p^0)), \quad (13)$$

其中 $a, b$ 分别是 $\{a_p^1, a_p^0\}$ 和 $\{b_p^1, b_p^0\}$ 的系数集.根据 $E_s(P, Z, a, b)$ 可以判断整个图像的相似性.通过最小化式(13)可得到仅包含图像 $P$ 与图像 $Z$ 一致结构的图像 $P_c$ .

获得一致结构图像 $P_c$ 后,再用其作为引导图像对融合得到的局部深度图进行边缘保持滤波.之后,希望得到在深度不连续区域有锐利梯度的全局深度图.为此,采用Yang等<sup>[31]</sup>提出的迭代优化算法得到最终的连续深度图 $Z^*$ .本文算法使用4次迭代便能得到较为理想的结果.

图9为引导滤波深度图与交互结构联合滤波深度图对比.由图9可以看出:引导滤波深度图受引导图像中额外结构信息的影响,部分区域深度不连续;交互结构联合滤波深度图遮挡边缘清晰,且噪声颗粒少,说明交互结构联合滤波具有很好的边缘保持特性,且不受引导图片额外结构信息的影响,提高了鲁棒性.



(a) 融合深度图 (b) 引导滤波深度图 (c) 交互结构联合滤波深度图

图9 滤波深度图对比

### 3 实验结果及分析

为了验证所提出算法的有效性,本文分别在合成数据集和真实数据集上进行实验,并与现在流行的深度估计方法进行比较.分别使用文献[23]算法、文献[21]算法、文献[20]算法、文献[25]算法及本文算法,对合成全光场图像与真实场景全光场图像进行深度估计,从定性和定量两个方面对比分析各算法的深度估计结果.

#### 3.1 实验设置

本文实验在配置为Core i7-3770k CPU 3.5 GHz×8cores, RAM 12 GB, Windows7 64位操作系统的台式电脑上完成,编程软件为Matlab 2014a.

实验采用的数据集有合成数据集和真实自然场景数据集. 合成数据集包含Heideberg大学提供的光场基准数据集<sup>[32]</sup>和文献[25]提供的数据集. 合成数据集包含深度图的真值, 可进行定量分析. 真实自然场景数据集采用文献[25]提供的、用Lytro Illum相机拍摄的数据集, 包含一系列的真实遮挡场景, 图像分辨率为 $625 \times 434$ 像素.

在合成数据集上采用均方根误差 (RMSE) 和相对深度误差 ( $B$ ) 作为量化指标评价算法性能, RMSE 和  $B$  取值越小, 表示深度估计性能越好.

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{(x,y)} (d(p(x,y)) - d_{GT}(p(x,y)))^2}, \quad (14)$$

$$B = \frac{1}{MN} \sum_{(x,y)} (|d(p(x,y)) - d_{GT}(p(x,y))| > \delta_d). \quad (15)$$

其中:  $M$ 、 $N$  表示图像尺寸;  $d(p(x,y))$  表示实验获取的深度图;  $d_{GT}(p(x,y))$  表示深度图真值;  $\delta_d$  表示相对深度允许误差, 实验中取值为0.3.

## 3.2 定性分析

### 3.2.1 合成场景

图10给出了在光场基准数据集<sup>[32]</sup>上文献[23]、文献[21]、文献[20]、文献[25]与本文方法深度估计结果. 由图10可以看出: 文献[23]算法提取的深度图边缘模糊, 且在深度提取时需要使用光场相机的硬件参数; 文献[21]方法获取的深度图不连续, 颗粒噪声较多, 且图像边缘有伪影; 文献[20]方法提取的深度图颗粒噪声较少, 且连续区域精确度较高, 但是边缘区域会被过度平滑, 且光场多视角图像之间存在视角差异, 会导致估计结果存在错误估计区域; 文献[25]方法对边缘和遮挡区域的深度估计效果较好, 但是在一些区域存在噪声; 本文方法融合了对遮挡和边缘区域深度提取效果较好的两个线索, 所以估计的深度图在遮挡区域和物体边缘区域有更少的伪影(如Mona中植物径的形状和叶子与径的遮挡边缘), 同时, 文中用交互结构联合滤波器对融合深度图进行滤波, 所以噪声颗粒较少且能显示物体正确的形状和精确的边缘(如Buddha中佛像和棒状物体的放大区域).

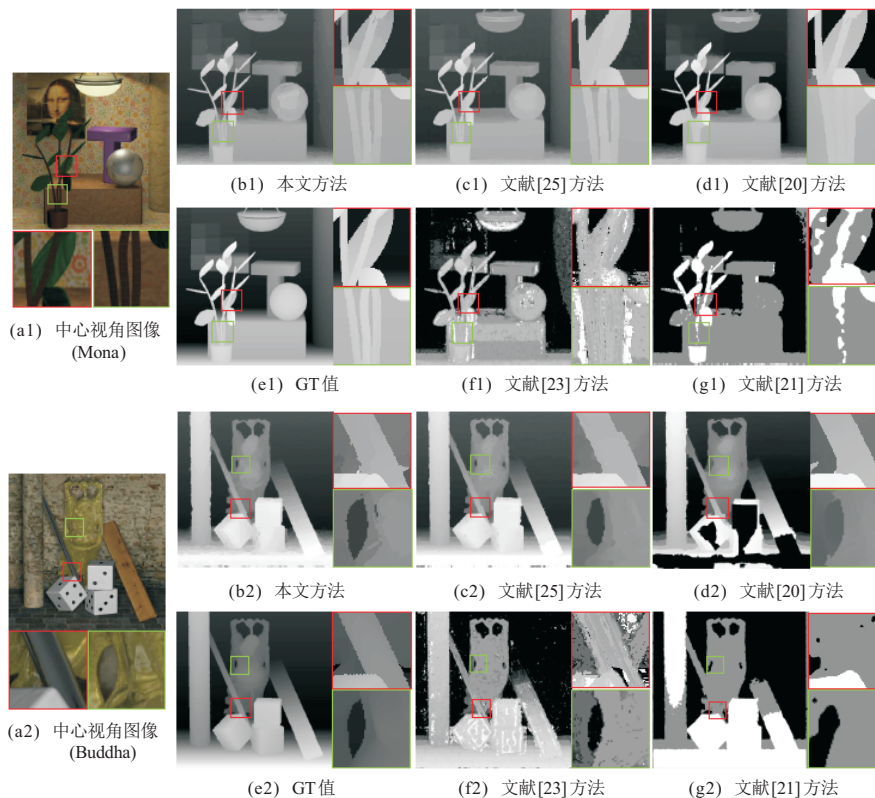


图10 光场基准数据集深度估计结果

图11给出了由文献[25]提供数据集的各方法深度估计结果. 由图11可以看出: 文献[21]方法深度图不连续; 文献[23]方法颗粒噪声较多, 分辨率低, 物体边缘有伪影; 文献[20]方法存在错误估计区域; 本文方法和文献[25]方法在遮挡区域和物体边缘效果较

其他方法好; 但是文献[25]方法会受到颗粒噪声的影响, 遮挡区域和边缘区域深度估计结果比本文方法差; 本文方法更精确(如正确显示了兔子的形状, 兔子耳朵边缘清晰), 几乎无颗粒噪声(如房间里椅子的空洞形状准确, 空洞边缘清晰无伪影). 结果与基准数据



图 11 文献[25]数据集深度估计结果

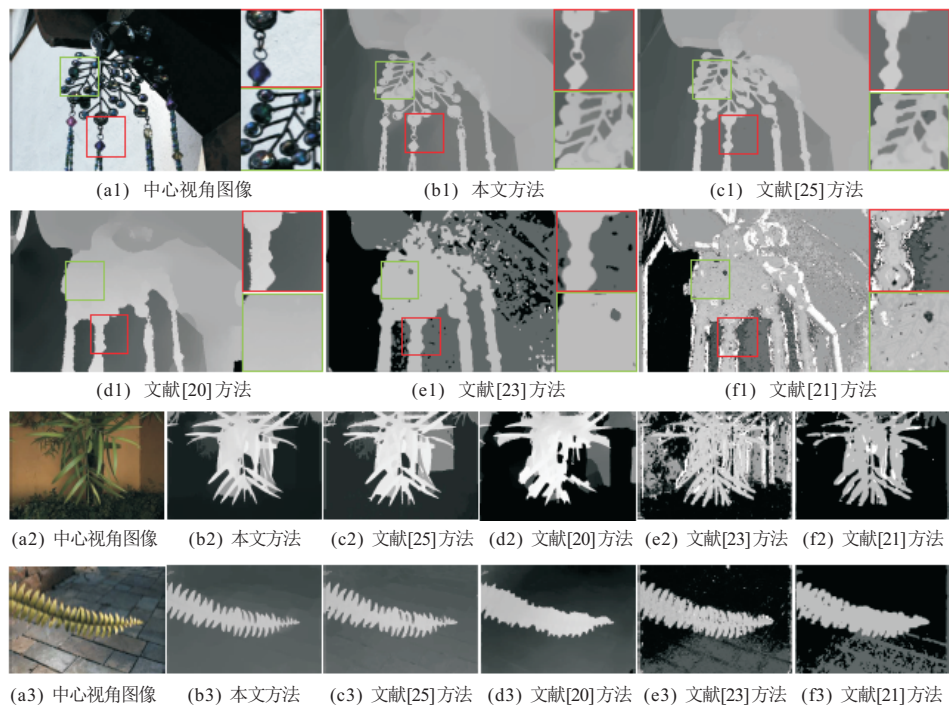


图 12 Lytro Illum 拍摄图像深度估计结果

集保持一致。

### 3.2.2 真实场景

图 12 给出了有精细结构和遮挡的真实场景深度估计结果。可以看出: 本文方法和文献 [25] 方法优于文献 [20]、文献 [21]、文献 [23] 的方法; 本文方法和文献 [25] 方法能正确估计物体形状, 不会过度平滑物体边缘, 边缘保持较好; 本文方法在遮挡区域和物体边缘较文献 [25] 方法性能更好, 尤其是对细小的物体结构(如叶子边缘伪影少、清晰度较高、颗粒噪声少), 且本文是唯一能正确捕获到挂饰各个细小圆环的方法。

### 3.3 定量分析

合成数据集提供了深度图真值, 进而可以进行定量分析。表 1 给出了 5 种方法深度估计结果的评价指

标。文献 [23] 方法由于没有对所使用的硬件参数进行校准, 深度值均方根误差较大。文献 [21] 方法通过局部窗口搜索周围像素作为匹配区域, 并认为匹配区域的深度相同, 因此获取的深度图不连续, 实验数据较差。文献 [20] 方法虽然边缘检测效果较差, 但其充分利用多视角图像间的匹配关系, 并通过权重中值滤波器消除噪声的影响, 所以定量比较时数据较好。文献 [25] 方法边缘保持效果明显优于文献 [20]、文献 [21] 及文献 [23] 的方法, 所以均方根误差与相对深度误差较低。本文方法在遮挡和边缘区域深度估计上较文献 [25] 方法更好, 在边缘保持上具有一定的优势, 实验数据值更小, 与定性评价结果一致。

表 1 各方法在合成数据集上的定量指标

数据集	全光场图像	文献 [23] 方法		文献 [21] 方法		文献 [20] 方法		文献 [25] 方法		本文方法	
		RMSE	<i>B</i>	RMSE	<i>B</i>	RMSE	<i>B</i>	RMSE	<i>B</i>	RMSE	<i>B</i>
HCI 数据集	Mona	4.329	0.335	5.489	0.454	<b>1.749</b>	0.124	2.344	0.198	2.174	<b>0.179</b>
	Buddha	4.365	0.331	5.965	0.474	8.340	0.515	3.861	0.365	<b>3.352</b>	<b>0.308</b>
文献 [25] 数据集	Sculptures	7.134	0.462	7.203	0.497	6.449	0.414	5.215	0.305	<b>3.329</b>	<b>0.271</b>
	Livingroom	8.128	0.914	8.913	0.766	6.848	0.715	7.847	<b>0.544</b>	<b>6.041</b>	0.678
	Plant	7.195	0.639	7.451	0.686	6.336	0.563	<b>1.089</b>	0.564	1.629	<b>0.438</b>

表 2 给出了各方法在不同大小图像上的运行时间, 可以看出, 本文算法的时间复杂度较文献 [23] 方法、文献 [21] 方法、文献 [25] 方法高, 较文献 [20] 方法低。文献 [20] 方法利用图割进行多标签优化后再通过迭代方法优化得到连续深度图, 所以时间复杂度较其他方法高。本文算法时间复杂度相对较高主要是由于在计算自适应散焦响应时每个空间像素需计算 9 次散焦响应再取最小值。

表 2 各方法运行时间对比

方法	HCI 数据集 (9×9×768×768)	文献 [25] 数据集 (9×9×800×600)	Lytro Illum (9×9×625×434)
文献 [23]	825	352	238
文献 [21]	253	176	94
文献 [20]	2582	1874	1401
文献 [25]	264	217	191
本文	946	502	269

## 4 结 论

针对用散焦与匹配方法在提取遮挡场景深度信息过程中会受到散焦伪影及噪声的影响, 且遮挡边界与物体边缘清晰度不高的问题, 本文提出了一种新的多线索融合深度获取方法。将对遮挡场景深度估计效果较好的散焦方法与匹配方法进行融合以提取场景深度信息, 有效地提高了深度图边缘的清晰度。同

时, 为了避免将错误的边缘和纹理从引导图像传递到目标图像, 本文采用具有结构一致性的交互结构联合滤波, 有效地降低了噪声对深度提取的影响。实验结果显示, 本文获取的融合深度图更加平滑, 边缘信息更加精确、丰富, 受遮挡影响有效减弱, 噪声颗粒减少。然而, 本文算法的时间复杂度相对较高, 且本文在自然光照条件下进行深度估计, 没有考虑光照环境对估计结果的影响。接下来的工作是尝试解决高光对深度估计的影响以及通过引入新的优化算法减少算法的时间复杂度。

### 参考文献 (References)

- [1] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera[J]. Computer Science Technical Report CSTR, 2005, 2(11): 1-11.
- [2] Maeno K, Nagahara H, Shimada A, et al. Light field distortion feature for transparent object recognition[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Portland, 2013: 2786-2793.
- [3] Desingh K, Krishna K M, Rajan D, et al. Depth really Matters: Improving visual salient region detection with depth[C]. British Machine Vision Conf. Bristol, 2013: 98.1-98.11.
- [4] Zhang X, Wang Y, Zhang J, et al. Light field saliency vs. 2D saliency: A comparative study[J]. Neurocomputing,

- 2015, 166(20): 389-396.
- [5] Peng H, Li B, Xiong W, et al. Rgb-d salient object detection: A benchmark and algorithms[C]. European Conf on Computer Vision. Cham: Springer, 2014: 92-109.
- [6] Ju R, Ge L, Geng W, et al. Depth saliency based on anisotropic center-surround difference[C]. IEEE Int Conf on Image Processing, Paris: IEEE, 2014: 1115-1119.
- [7] Zhang J, Wang M, Gao J, et al. Saliency detection with a deeper investigation of light field[C]. Int Conf on Artificial Intelligence. Buenos Aires, 2015: 2212-2218.
- [8] 张旭东, 李梦娜, 张骏, 等. 边缘结构保持的加权BDTV全光场图像超分辨率重建[J]. 中国图象图形学报, 2015, 20(6): 733-739.  
(Zhang X D, Li M N, Zhang J, et al. Edge preserved light field image super-resolution based on weighted BDTV model[J]. J of Image and Graphics, 2015, 20(6): 733-739.)
- [9] Yoon Y, Jeon H G, Yoo D, et al. Learning a deep convolutional network for light-field image super-resolution[C]. Proc of the IEEE Int Conf on Computer Vision Workshops. Santiago, 2015: 24-32.
- [10] Hong W. Light field applications to 3-dimensional surface imaging[D]. Massachusetts Institute of Technology, 2009.
- [11] Chen C, Lin H, Yu Z, et al. Light field stereo matching using bilateral statistics of surface cameras[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Columbus, 2014: 1518-1525.
- [12] Li J, Lu M, Li Z N. Continuous depth map reconstruction from light fields[J]. IEEE Trans on Image Processing, 2015, 24(11): 3257-3265.
- [13] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields[C]. 2012 IEEE Conf on Computer Vision and Pattern Recognition(CVPR). Providence: IEEE, 2012: 41-48.
- [14] Zhang Y, Lv H, Liu Y, et al. Light-field depth estimation via epipolar plane image analysis and locally linear embedding[J]. IEEE Trans on Circuits & Systems for Video Technology, 2017, 27(4): 739-747.
- [15] Mousnier A, Vural E, Guillemot C. Partial light field tomographic reconstruction from a fixed-camera focal stack[EB/OL]. [2017-05-20]. <http://arxiv.org/pdf/1503.01903v1.pdf>.
- [16] Lin H, Chen C, Bing Kang S, et al. Depth recovery from light field using focal stack symmetry[C]. Proc of the IEEE Int Conf on Computer Vision. Santiago, 2015: 3451-3459.
- [17] Lu W, Mok W K, Neiman J, et al. 3D and image stitching with the lytro light-field camera[D]. New York: Department of Computer Science, City University of New York, 2013.
- [18] Cartpauj. Software updates; Lytro illum firmware 2.0, lytro desktop 4.3[EB/OL]. [2017-05-21]. <http://lightfield-forum.com/tag/lytro-desktop-software>.
- [19] Yu Z, Yu J, Lumsdaine A, et al. Plenoptic depth map in the case of occlusions[C]. Multimedia Content and Mobile Devices. Burlingame, 2013: 8667-86671S.
- [20] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Boston, 2015: 1547-1555.
- [21] Calderon F C, Parra C A, Niño C L. Depth map estimation in light fields using an stereo-like taxonomy[C]. Image, Signal Processing and Artificial Vision. Armenia: IEEE, 2014: 1-5.
- [22] Liu F, Hou G, Sun Z, et al. High quality depth map estimation of object surface from light field images[J]. Neurocomputing, 2017, 252: 3-16.
- [23] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]. Proc of the IEEE Int Conf on Computer Vision. Sydney, 2013: 673-680.
- [24] Tao M W, Srinivasan P P, Malik J, et al. Depth from shading, defocus, and correspondence using light-field angular coherence[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Boston: 2015: 1940-1948.
- [25] Wang T C, Efros A A, Ramamoorthi R. Occlusion-aware depth estimation using light-field cameras[C]. Proc of the IEEE Int Conf on Computer Vision. Santiago, 2015: 3487-3495.
- [26] Sun J, Shum H Y, Zheng N N. Stereo matching using belief propagation[C]. European Conf on Computer Vision. Berlin: Springer, 2002: 510-524.
- [27] Williem W, Kyu Park I. Robust light field depth estimation for noisy scene with occlusion[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Las Vegas, 2016: 4396-4404.
- [28] He K, Sun J, Tang X. Guided image filtering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397-1409.
- [29] He K, Sun J. Fast guided filter[EB/OL]. [2017-05-25]. <http://arxiv.org/pdf/1505.00996v1.pdf>.
- [30] Shen X, Zhou C, Xu L, et al. Mutual-structure for joint filtering[C]. Proc of the IEEE Int Conf on Computer Vision. Sontiago, 2015: 3406-3414.
- [31] Yang Q, Yang R, Davis J, et al. Spatial-depth super resolution for range images[C]. IEEE Conf on Computer Vision and Pattern Recognition. Minneapolis: IEEE, 2007: 1-8.
- [32] Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields[C]. Vision, Modeling and Visualization. Geneva: The Eurographics Association, 2013: 225-226.