

控制与决策

Control and Decision

基尼指数成本量指导的抗遮挡光场深度估计方法

张旭东, 董运流, 石梦迪

引用本文:

张旭东, 董运流, 石梦迪. 基尼指数成本量指导的抗遮挡光场深度估计方法[J]. *控制与决策*, 2020, 35(8): 1849–1858.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.1718>

您可能感兴趣的其他文章

Articles you may be interested in

基于焊件识别与位姿估计的焊接机器人视觉引导

Visual guidance of welding robot based on weldment recognition and pose estimation

控制与决策. 2020, 35(8): 1873–1878 <https://doi.org/10.13195/j.kzyjc.2018.1635>

基于时空渐进特征模型的抗遮挡多目标跟踪

Anti-occlusion multi-target tracking with progressive spatio-temporal feature model

控制与决策. 2019, 34(10): 2171–2177 <https://doi.org/10.13195/j.kzyjc.2018.0156>

考虑多位姿估计约束的双目视觉里程计

Stereo visual odometry with multi-pose estimation constraints

控制与决策. 2018, 33(6): 1008–1016 <https://doi.org/10.13195/j.kzyjc.2017.0304>

遮挡场景的光场图像深度估计方法

Light field depth estimation for scene with occlusion

控制与决策. 2018, 33(12): 2122–2130 <https://doi.org/10.13195/j.kzyjc.2017.0936>

分块多特征目标描述子的移动机器人目标跟踪

Person tracking for robot using patches-based-multi-cues representation

控制与决策. 2016(2): 337–342 <https://doi.org/10.13195/j.kzyjc.2014.1822>

基尼指数成本量指导的抗遮挡光场深度估计方法

张旭东[†], 董运流, 石梦迪

(合肥工业大学 计算机与信息学院, 合肥 230009)

摘要: 光场相机能够实现一次拍摄即获得三维场景的多视角信息, 在深度估计领域中具有独特优势. 但是, 当场景中存在复杂遮挡时, 现有深度估计方法提取深度信息的精度会明显降低. 针对该问题, 设计一种基尼指数成本量指导下的抗遮挡光场深度估计方法. 首先, 利用光场重聚焦方法获得焦栈图像; 然后, 构造中心视角与其他视角的基尼指数成本量, 并根据成本最小原则计算得到初始深度图; 最后, 结合彩色图进行联合引导滤波, 获得最终的高精度的深度图像. 实验结果表明, 所提方法对复杂场景更加鲁棒, 能够在较小的算法复杂度下获取更好的深度估计结果. 相比于其他先进方法, 所提方法获取的深度图精度更高, 图像边缘保留效果更好, 在 HCI 数据集上的 MSE_{100} 指标平均降低约 7.8%.

关键词: 光场; 深度估计; 遮挡; 重聚焦; 中心视角; 基尼指数成本量; 联合引导滤波
中图分类号: TP391 **文献标志码:** A

Anti-occlusion light field depth estimation guided by Gini cost volume

ZHANG Xu-dong[†], DONG Yun-liu, SHI Meng-di

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

Abstract: The light field camera can record multi-view information of a three-dimensional scene within one shot, which possesses unique advantage especially in depth estimation. However, the accuracy of the depth information extracted by the existing depth estimation method is significantly reduced when there is complex occlusion in the scene. Aiming at this problem, a method for estimating the depth of anti-occlusion light field based on Gini cost volume is proposed. Firstly, the refocusing images are obtained using the light field refocusing algorithm. Then, the Gini cost volume of the central view and other views are constructed. The initial depth map is calculated according to the principle of minimum cost volume. Finally, the initial depth information is combined with the color map for joint guided filtering, and a high-precision depth map is obtained. The experimental results show that the proposed method is more robust for complex scenes and can obtain better depth estimation results with smaller algorithm complexity. Compared to other advanced methods, the depth images obtained by using the proposed method are more accurate, the edges are clearer, and the MSE_{100} indicator on HCI dataset is reduced by an average of 7.8%.

Keywords: light field; depth estimation; occlusion; refocusing; central view; Gini cost volume; joint guided filter

0 引言

在现实生活中, 深度一般是指目标物体到人眼的距离. 在计算机系统中, 深度是指目标物体到相机中心平面的距离. 由于深度信息反映了目标的三维空间信息, 精准的深度信息能够为目标定位^[1]、目标识别^[2]、三维重建^[3]、语义分割^[4]、显著性检测^[5]等计算机视觉任务提供积极的作用. 如今, 深度估计已成为计算视觉领域的重要研究任务之一.

与传统相机不同, 全光场相机^[6]不再捕获像素的累积强度, 而是捕获每个方向的光线强度. 得益于特殊的光学结构^[7], 光场相机能够实现一次拍摄即获得

真实场景的空间信息和角度信息, 渲染生成的四维光场图像包含来自场景的多视角信息, 这使得深度估计成为可能. 光场图像的本质特征是其具有目标物体多个视角的信息, 根据多视角信息不同表征形式, 可以将现有光场深度估计方法分为 4 类^[8-10]: 基于极平面图像 (EPI) 的深度估计^[11-13]、基于数字重聚焦的深度估计^[14-15]、基于多视角立体匹配的深度学习^[16-19]以及基于学习的深度估计^[20-22]. 这些方法可以在一些特定的场景中获取较为准确的深度信息, 但是这些方法都包含了一个重要的隐含前提——成像一致性原则^[23], 即在粗糙的朗伯平面下, 从不同的方

收稿日期: 2018-12-18; 修回日期: 2019-02-28.

基金项目: 国家自然科学基金项目 (61471154).

责任编辑: 侯忠生.

[†]通讯作者. E-mail: xudong@hfut.edu.cn.

向观察场景中同一个空间点,该空间点颜色是相近(相同)的.但是,当场景中存在遮挡点时,成像一致性原则不再成立,导致这些方法不再适用.

为了解决光场深度估计中的遮挡问题,近年来一些学者根据光场图像信息间的几何关系提出了不同的解决方案. Jeon等^[24]引入傅里叶域的相位平移理论,用于表征子孔径图像间的视差变化,有效解决了光场多视角图像视角偏差小的影响.该方法能够处理轻微遮挡,但需要对每个视角分别构造成本函数,时间复杂度较高. Wang等^[25]提出了针对遮挡场景建模的深度计算方法.该方法对于单一遮挡场景效果很好,但依赖于边缘检测的结果,当场景复杂时,边缘检测结果不准确,此方法得到的深度结果较差. Zhang等^[26]在2D EPI中集成旋转平行四边形算子,通过最大化平行四边形窗口的两个部分之间的分布距离来测量EPI的斜率.该方法能够处理遮挡对深度计算的影响,但易受到噪声的干扰,因此对于真实场景的深度估计结果较差.此外,该算法计算量较大,时间复杂度较高,实时性较差. Williem等^[27]引入信息熵作为匹配线索,并结合散焦线索获取深度信息.该方法能够处理较为复杂的遮挡场景,但依赖线索置信度的计算以及融合方法的选择,同时需要对每个空间点计算成本量,时间复杂度较大.综上所述,现有针对遮挡场景的光场深度估计方法主要存在以下两个问题:1)构建的线索或成本函数没有很好地利用光场的多视角信息,不能很好地表征遮挡场景与未遮挡场景的差异;2)优化方法局限于深度图的融合或优化,没有结合彩色图的结构信息.

本文在光场数字重聚焦的框架下设计一种对复杂遮挡有效的光场深度估计方法.通过计算不同焦距处中心视角与其他视角的基尼指数成本量,自适应得到最佳成本量,得到初始深度图;结合彩色图引导滤波初始深度图,根据成本量最小原则,将深度估计表示为能量泛函最小化的过程;引入迭代优化算法,获得最终的全局高精度的深度图像.本文方法简单有效,在保证深度计算精度的同时,降低了时间复杂度.本文的主要贡献归纳如下:1)在光场数字重聚焦的框架下提出一种对复杂遮挡有效的成本函数,该成本函数充分利用光场数据的冗余性,弥补了传统成像一致性的不足,能够在不同遮挡程度的场景中精确获得场景深度,保持场景边缘信息,获得初始深度图;2)结合彩色图的信息,联合优化初始深度图中的匹配误差,在保留深度信息边缘的同时利用彩色图的结构信息控制噪声,进一步提升了深度图精度.

1 光场成像与遮挡分析

微透镜式光场相机在主透镜与图像传感器之间放置一个微透镜阵列^[7],如图1所示:物体上某一点反射空间中光线经过主透镜聚焦到微透镜阵列所在平面,然后微透镜阵列将光线按照入射方向分散到图像传感器上.因此,图像传感器可以捕捉空间某一点的不同方向的光线,达到记录目标位置和角度信息的目的.获取的4维光场图像用 $L_f(x, y, u, v)$ 表示,其中 (x, y) 表示空间坐标, (u, v) 表示角度坐标.图2所示为光场的多视角图像,它表示在角度坐标 (u, v) 固定的情况下,空间坐标 (x, y) 的变化情况.

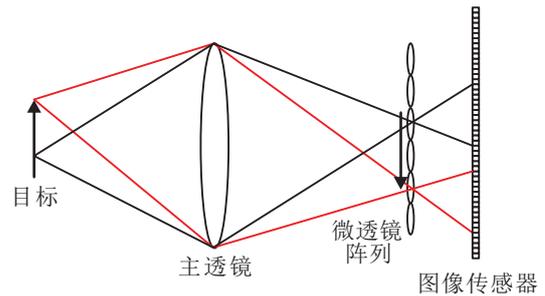


图1 光场相机成像模型

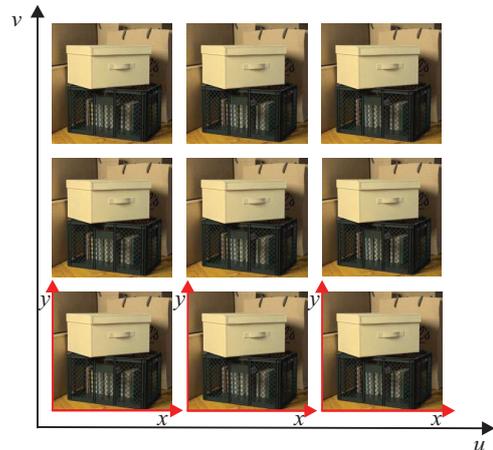


图2 光场多视角图像

光场图像的遮挡状态被定义为:某空间点在多视角图像的中心视角可见,其他视角的部分视角不可见.在非遮挡像素(图3(a)所示)处,所有的视角可见,并且颜色相似度较高;在遮挡像素(图3(b)所示)处,某些视角(非中心视角)被遮挡物所遮挡,形成的多视角块颜色存在差异性.同时,遮挡场景一般被定义成3种类型^[28]:1)非遮挡,场景点没有被其他目标遮挡,满足成像一致性原则(图1(c)所示);2)单遮挡,场景点被一个场景目标遮挡,遮挡视角数目不超过总视角数目的一半,不再满足成像一致性原则(图1(d)所示);3)多遮挡,场景点至少被两个场景目标遮挡,并且遮挡视角数目超过总视角数目的一半,即遮挡视角数目超过了非遮挡视角数目(图1(e)所示).现有针对

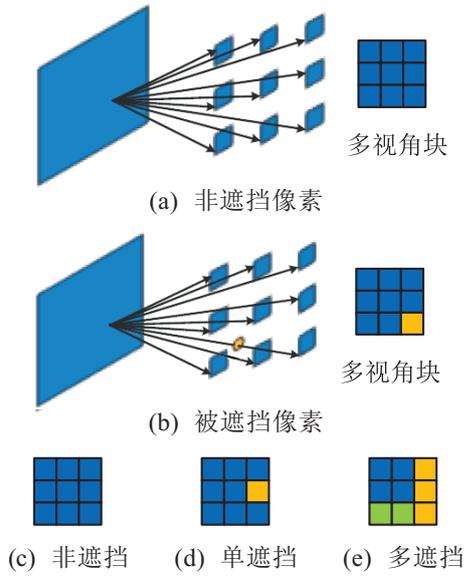


图3 遮挡类型定义

遮挡的光场深度估计方法没有对遮挡视角的数目进行分析,在处理多遮挡场景时不能区分场景点的非遮挡区域和遮挡区域,导致多遮挡场景下深度估计结果较差. 本文构建的成本函数对于多遮挡场景也进行了特定分析,能够有效地处理各种遮挡场景.

2 本文方法

针对复杂遮挡场景下的光场深度估计问题,本文设计一种基尼指数成本量指导下的抗遮挡光场深度估计方法. 主要分为以下几个步骤: 1) 利用数字重聚焦算法获取聚焦到不同深度的焦点堆栈图像; 2) 在焦点堆栈图像上构建自适应基尼指数成本量作为深度成本函数,并通过最小化成本量获取初始深度; 3) 利用彩色图联合引导滤波器迭代优化初始深度,获得全局高精度深度图. 算法框图如图4所示.

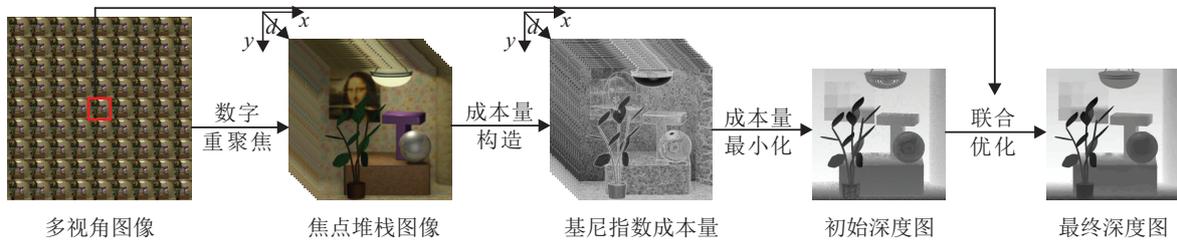


图4 本文算法框图

2.1 光场数字重聚焦

利用光场的数字重聚焦原理^[29],可以将光场的多视角信息重新聚焦,获得聚焦到不同深度平面的重聚焦图像. 图5所示为光场重聚焦原理,其中 F' 是透镜平面到重聚焦平面的距离, F 为焦距,且 $F' = \alpha F$, α 为调焦系数(与深度标签 k 成正比).

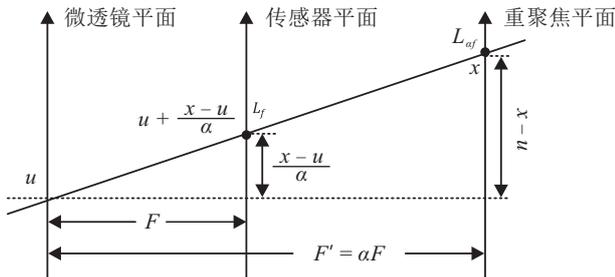


图5 光场数字重聚焦原理

假设光线在重聚焦平面的投影坐标为 $L_{\alpha f}(x, y, u, v)$,根据三角形相似原理,光线在传感器平面的投影坐标为 $L_f\left(u + \frac{x-u}{\alpha}, v + \frac{y-v}{\alpha}, u, v\right)$. 由于光线在自由空间传播时能量不变,有

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

通过不断地改变调焦系数 α 或深度标签 k ,便可利用输入图像 L_f 获得一系列聚焦在不同深度的焦点堆栈图像 $L_{\alpha f}$. 如图6所示,随着深度标签 k 的增大,人物的眼睛(蓝色框部分)逐渐模糊,而绿叶(红色框部分)逐渐清晰.

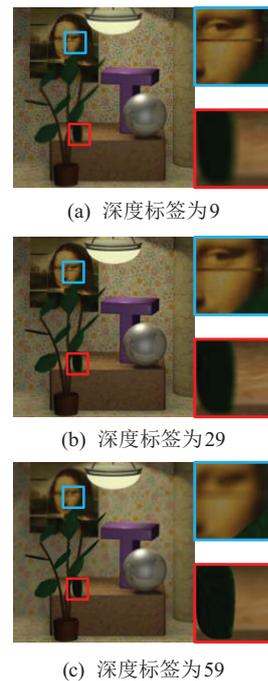


图6 不同深度标签下的焦点堆栈图像

2.2 基尼指数成本量构建

根据成像一致性原则以及重聚焦原理可以知道,在聚焦到正确深度的情况下,非遮挡像素应该具有与中心视角像素相同(或相近)的强度值,被遮挡视角的像素在灰度值上与中心视角像素应该具有较大的差异性.因此,本文利用该性质以及像素强度的权重概率 g 来确定空间像素点的正确深度,如下所示:

$$g(i) = h(i) \times e^{-d(i)}, \quad (2)$$

$$d(i) = \frac{(i - A_k^p(u_c, v_c))^2}{\frac{1}{N} \sum_{n \in N} i_n + 1}. \quad (3)$$

其中: A_k^p 代表图像某空间像素点 $p(x, y)$ 的多视角块 A^p 聚焦到深度 k 时的图像信息(即重聚焦信息); i 代表 A_k^p 中像素点的强度,其范围在 $[0, 255]$; $h(i)$ 代表 A_k^p 中强度值为 i 的概率.由前文分析可知,聚焦到正确深度的多视角块 A^p 应该有着更均匀的颜色,其强度概率 $h(i)$ 的直方图分布更加集中并且波形尖锐、峰值高.反之,在不正确的深度下,多视角块 A^p 的 $h(i)$ 直方图波形平缓、峰值较低.若仅考虑重聚焦图像的 $h(i)$,则只能处理单一遮挡情况,但在多遮挡情况下,遮挡像素多于无遮挡像素,其概率高于无遮挡像素,强度概率直方图不能正确反映场景深度信息的变化规律,因此需要为 $h(i)$ 加入约束(即权重函数 $e^{-d(i)}$). $d(i)$ 为本文设计的强度差异函数,通过分析 $d(i)$ 可以看出,分子 $(i - A_k^p(u_c, v_c))^2$ 表示 A_k^p 中像素 i 与中心像素 $A_k^p(u_c, v_c)$ 强度的差异.为了消除亮度不均衡的影响,本文采用 $\frac{1}{N} \sum_{n \in N} i_n + 1$ 对分子表达的强度差进行归一化, N 表示 A_k^p 中像素点的个数.若 $d(i)$ 趋于 0,表示空间像素点的多视角像素不存在遮挡,权重 $e^{-d(i)}$ 趋于 1,得到的权值概率 $g(i)$ 能够保持 $h(i)$ 峰值和波形;反之,若 $d(i)$ 趋于正无穷,表示空间像素点的多视角像素存在单遮挡或多遮挡,权重 $e^{-d(i)}$ 趋于 0,得到的权值概率 $g(i)$ 也将趋于 0,不能保持 $h(i)$ 峰值和波形.因此,权值概率 $g(i)$ 相较于 $h(i)$ 而言,能够自适应处理各种遮挡环境,可以解决多遮挡时深度计算不精确的问题.

为了度量多视角块 A^p 在不同深度标签 k 下的权值概率 g 分布是否均匀,本文引入基尼指数的概念如下:

$$\text{Gini}(p) = \sum_{n=1}^N p_n(1 - p_n) = 1 - \sum_{n=1}^N p_n^2. \quad (4)$$

式(4)表示在样本集合中一个随机样本被分错的概

率,用来评估多视角像素一致性中的随机性.结合式(2)给出的权重概率 g ,可以将基尼指数改写为

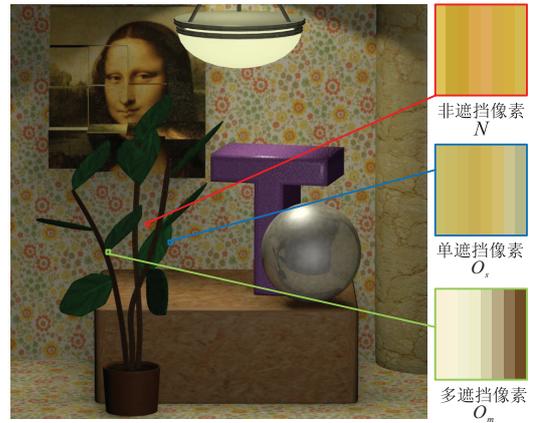
$$\text{Gini}(p, k) = 1 - \sum_i g(i)^2, \quad (5)$$

其中 $\text{Gini}(p, k)$ 为 A_k^p 中的基尼指数.算法先计算多视角块中 $\{R, G, B\}$ 三个通道的基尼指数,再加权求和得到基尼指数成本量 $C(p, k)$,公式如下:

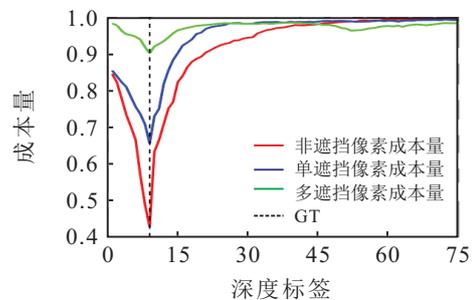
$$C(p, k) = \frac{\text{Gini}_R(p, k) + \text{Gini}_G(p, k) + \text{Gini}_B(p, k)}{3}. \quad (6)$$

图7展示了同一真实深度的3种遮挡类型空间点的基尼指数成本量曲线,在正确的深度下,3种遮挡点的基尼指数成本量均能够获得最小的值(图7(b),GT为9).由式6和图7(b)可知,成本量是关于图像像素空间坐标 p 和深度标签 k 的函数,反映了不同像素点在不同深度标签上的成本大小,成本越小表明该像素点取该深度值越可靠.因此,求解深度的问题便转化为提取每个空间点最小成本所对应的深度标签 k .从而,获得初始深度图 $d_i(p)$ 为

$$d_i(p) = \arg \min_k C(p, k). \quad (7)$$



(a) 中心视角图像



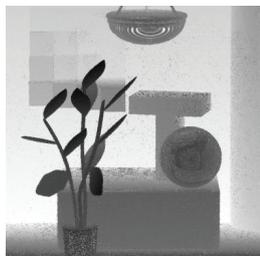
(b) 成本量曲线

图7 不同区域基尼指数成本曲线

由图8(b)可以看出,由基尼指数成本量提取的深度信息对遮挡场景鲁棒,遮挡边缘清晰,伪影少,能够正确呈现物体形状.



(a) 中心视角图像



(b) 初始深度图



(c) 最终深度图

图8 本文方法不同步骤深度图对比

2.3 彩色与深度信息联合引导的深度图优化

通过自适应基尼指数成本量可以有效地处理复杂遮挡场景,但是该成本函数获得的深度图可能存在空间不连续性,这是因为计算多视角基尼指数时没有考虑空间点的邻域信息,割裂了像素的空间相关性(图9(a)所示),因此需要对初始深度进行平滑约束和优化.而现有深度图优化方法主要基于深度图的优化或融合,而深度图自身所含信息的局限性制约了最终的优化结果,由图9(b)红框部分可看出优化后的深度图依然含有噪声^[30].传统的彩色图引导滤波在利用彩色图的结构信息保留深度图边缘的同时,可能会将彩色图的边缘或阴影等额外的结构信息引入深度图,导致深度图引入不正确的边缘信息(图9(c)蓝色框部分所示)^[31].为了更好地平滑深度图并且不引入彩色图的边缘信息,本文采用彩色图和深度图联合引导的方法^[32]进行深度图优化,将优化问题表示为能量泛函最小化的过程,本文采用最小化能量公式 $\varepsilon(u)$ 如下所示:

$$\varepsilon(u) = \sum_p c_p(u_p - f_p)^2 + \lambda\Omega(u). \quad (8)$$

其中:输入图像为初始深度图 f ,输出图像为最终深度图 u , u 的初始值为经彩色图引导初始深度图 f 得

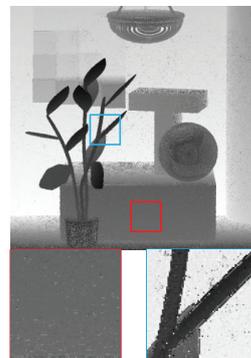
到的深度图; p 为像素点的空间坐标 (x, y) ; c_i 为保真项,用于协调输出图像 u 和输入图像 f ,表示像素点 p 深度信息的可靠性; λ 为加权系数; $\Omega(u)$ 为平滑项,用于约束相邻像素的一致性,当两个相邻点颜色接近时,两点的平滑约束较强,反之约束较弱.与单一静态或动态引导不同,本文采用的联合引导的平滑项 $\Omega(u)$ 由彩色平滑 ϕ_μ 和深度非凸平滑 ψ_ν 构成,如下式所示:

$$\Omega(u) = \sum_{p,q} \phi_\mu(g_p - g_q)\psi_\nu(u_p - u_q), \quad (9)$$

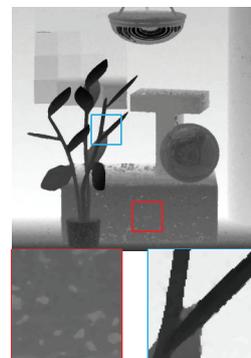
$$\phi_\mu(z) = \exp(-\mu z^2), \quad (10)$$

$$\psi_\nu(z) = \frac{1 - \phi_\nu(z)}{\nu} = \frac{1 - \exp(-\nu z^2)}{\nu}. \quad (11)$$

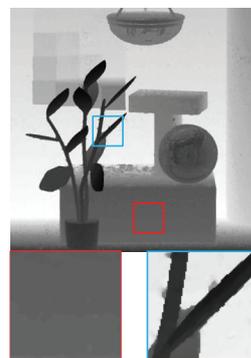
其中: q 表示为像素点 $p(x, y)$ 的邻域像素点坐标, μ 和 ν 表示平滑带宽, ϕ_μ 和 ψ_ν 分别用于约束彩色图和深度图相邻像素的一致性.联合滤波对异常值具有鲁棒性,当 ψ_ν 减小时发生反向扩散,保留在平滑期间深度图中高频结构(边缘和角落)的特征;但是 ψ_ν 在保留深度图边缘特征的同时,也会增强深度图的噪声,而在彩色图中噪声较少,因而 ϕ_μ 对于噪声有着很好的控制效果.在保留深度信息边缘的同时利用彩色图的结构信息控制噪声,进一步提升了深度信息的准确性.



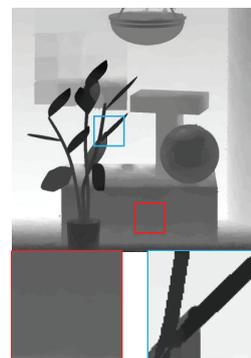
(a) 初始深度图



(b) 单一深度图加权中值滤波



(c) 单一彩色图引导滤波



(d) 联合引导滤波

图9 不同优化方法对比

图9为不同深度图的优化方法对比,本文对比了单一彩色图引导滤波(图9(b))、单一深度图加权中

值滤波(图9(c))以及本文所用联合引导滤波方法(图9(d)). 通过对比可以看出,本文所用方法可以有效去除噪声,保持深度图空间连续性以及边缘信息,提升深度估计准确性.

3 实验结果及分析

为了验证本文算法的有效性,本文分别在合成数据集和真实数据集上进行实验,并与近3年的抗遮挡深度估计方法进行比较. 分别使用文献[24-27]的方法及本文方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

实验使用的PC机配置为Core i7-3770k CPU 3.5 GHz X8cores, RAM 12 GB, Windows 764 位操作系统, 编程软件为Matlab2015b 和 Microsoft Visual Studio2015.

1) 实验数据集: 实验采用光场合成数据集与实际场景数据集.

① HCI 光场合成数据集^[33]. 该数据集采用

Blender 软件对虚拟场景图像进行渲染, 得到多幅场景复杂的合成图像, 图像空间分辨率为 768×768 像素, 角度分辨率为 9×9 , 提供场景的深度真值 d_{GT} 可用于定量分析;

② Lytro Illum 光场相机真实数据集. 该数据集由 Stanford Lytro Illum 数据集^[34] 和本实验室使用 Lytro Illum 相机拍摄的数据组成.

选取的场景包含不同程度的遮挡情况, 图像空间分辨率为 375×541 像素, 采取该数据集的 9×9 个视角作为角度分辨率.

2) 评价指标: 采用均方误差作为量化指标评价算法性能, 即

$$MSE_{100} = \frac{\sum (d_{GT} - d)^2}{H \times W} \times 100. \quad (12)$$

其中: H 和 W 分别表示图像的高度和宽度, d 表示算法预测的深度图, d_{GT} 表示深度图真值.

3.2 定性分析

3.2.1 合成场景

图10给出了光场HCI数据集上文献[24-27]与本文方法深度获取结果以及局部放大对比. 可以看



图10 合成数据集深度图及局部放大图对比

出,在 Monas 场景中,针对单一遮挡且较为细长(如植物的根茎,Monas 中蓝框部分)的场景,本文方法与文献[25]方法、文献[27]方法均取得较好的结果.这是因为文献[25]方法针对单一遮挡场景进行建模,对于细长的场景,具有很好的效果,但针对多遮挡场景(Monas 中红框部分)效果较差;而文献[27]方法对于多遮挡与单遮挡都有较好的鲁棒性,但是整体深度信息噪声较多(如 Monas).针对多遮挡问题(如植物的绿叶,Monas 中红框部分),本文方法与文献[24]方法、文献[27]方法效果较好.这是因为文献[24]利用相位平移表征多视角图像间的视差变化,针对场景不密集的遮挡区域有着较好的处理结果,但是对于细长且背景与前景相似的场景(Monas 中蓝框部分)效果较差.在 Horse 场景中,由于此场景纹理较少,所有方法的效果都不是很好,但是只有本文方法与文献[25]方法能够正确获取到 Horse 中马尾的形状(细长的场景, Horse 中红框部分).

图 11 给出了光场 HCI 数据集上文献[24-27]与本文方法深度获取结果对比.可以看出,在 Buddha 场景

中,文献[24]方法与文献[25]方法在一些孔洞区域出现错误估计;在 Papillon 场景中,文献[24]方法边缘出现伪影,文献[26]方法出现白色孔洞;在 Medieval 场景中,文献[26]方法与文献[27]方法在场景窗口区域出现了错误估计的区域;在 stillLife 场景中,所有方法均取得了较高的估计精度.

综合来看,对于纹理丰富且少遮挡的场景,文献[24-27]与本文方法都可以获取较为精确的深度信息.当场景中出现单遮挡时,文献[25]方法、文献[27]方法与本文方法效果较好;对于多遮挡场景,与其他方法相比,本文方法有着较大的优势;对于纹理较少的场景,所有方法的效果均不是很好.

3.2.2 真实场景

图 12 给出了 Lytro Illum 光场相机拍摄的场景在文献[24-27]与本文方法深度获取的结果.其中场景 a ~ 场景 c 为 Stanford Lytro Illum 数据集中的场景,场景 d ~ 场景 e 为本实验室利用 Lytro Illum 相机拍摄的场景.可以看出,本文方法在深度图像的结构还原

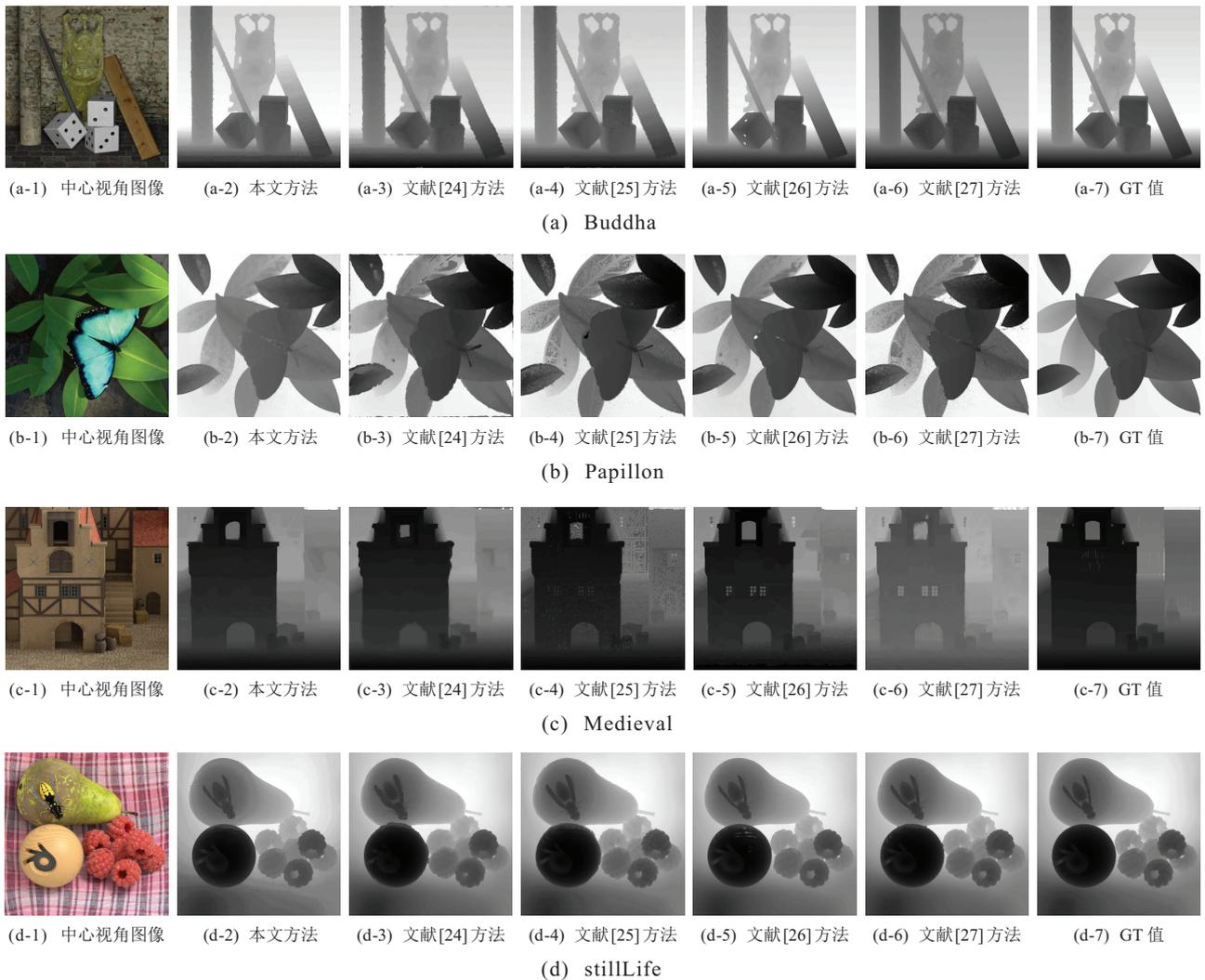


图 11 合成数据集深度图对比

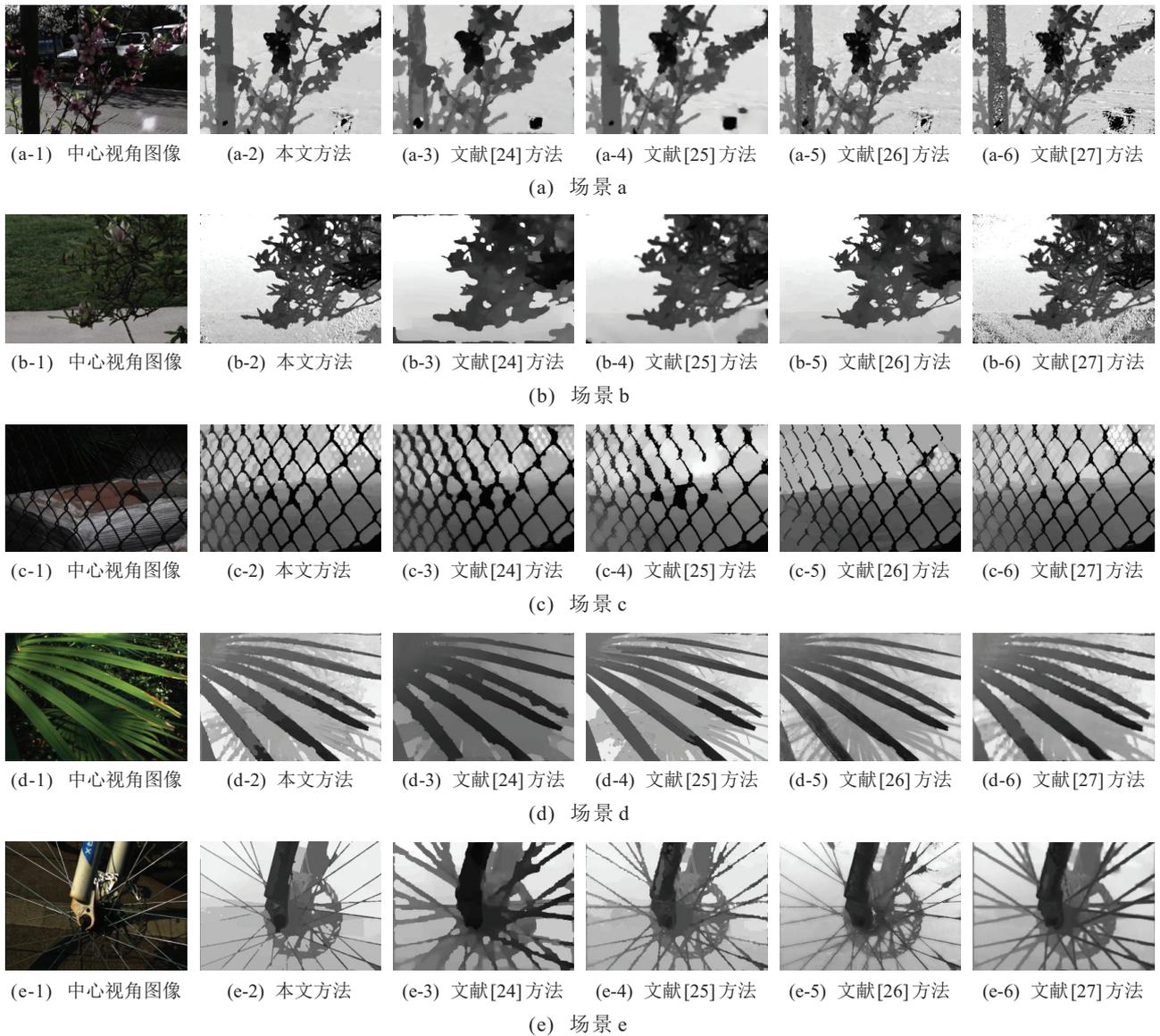


图 12 Lytro Illum相机光场数据集深度图对比

以及边缘保留上优于其他4种方法. 文献[25]和文献[26]方法在场景a和场景b中出现前景膨胀效应, 整体深度估计效果较差. 文献[24]方法在场景b和场景e中出现较为严重的涂抹现象, 文献[27]方法整体效果较好, 但噪声较为明显, 本文方法能够准确估计场景深度信息, 且能较好保留边缘信息. 在场景c中, 本文方法能较完整还原网格细节, 文献[25]和文献[26]方法在网格区域丢失部分信息. 在场景d中, 文献[24]

方法出现轻微涂抹现象, 其余方法效果较好. 在场景e中, 本文方法是唯一能够捕捉自行车手刹细节的方法.

3.3 定量分析

由于Stanford Lytro Illum数据集未提供深度真值, 本实验仅在HCI合成数据集上进行定量分析. 表1给出了5种方法的深度估计结果评价指标, 最优结果加粗显示, 次优结果加下划线显示.

表 1 深度估计MSE₁₀₀比较

HCI数据集	调焦系数 α	文献[24]方法	文献[25]方法	文献[26]方法	文献[27]方法	本文方法
Buddha	-1.54 ~ 1.54	0.3009	<u>0.2418</u>	0.2724	0.2848	0.2196
Horse	-3.69 ~ 3.69	4.0335	3.6008	2.1549	<u>1.9979</u>	1.7414
Papillon	-1.17 ~ 1.17	2.0210	0.6349	0.4993	0.6156	<u>0.5281</u>
Monas	-0.79 ~ 0.79	0.5580	0.3577	0.3683	<u>0.2917</u>	0.2901
Medieval	-1.93 ~ 1.93	0.5835	0.7601	0.4701	0.6137	<u>0.4955</u>
stillLife	-2.71 ~ 2.71	0.1012	0.1186	<u>0.0523</u>	0.0541	0.0511

从表1可以看出,本文方法的深度估计结果误差更小。相较次优结果,本文方法所得到的 MSE_{100} 平均降低0.0469,降低比例约为7.8%。表2给出了各方法的计算时间复杂度,比较了5种方法在不同数据集上每幅图像的平均计算时间,最优结果加粗显示,次优结果加下划线显示。可以看出,本文方法的计算时间最少,在保证计算精度的同时降低了计算时间。

表2 各方法耗费时间对比 单位: s

方法	HCI数据集	Lytro Illum
文献[24]方法	2033	855
文献[25]方法	1064	674
文献[26]方法	1553	905
文献[27]方法	2011	866
本文方法	925	513

4 结论

针对现有光场深度估计方法尚不能很好解决复杂遮挡场景下的深度估计问题,本文在光场数字重聚焦的框架下设计了一种对复杂遮挡有效的光场深度估计方法。通过分析不同遮挡程度下场景多视角的强度概率分布,首先设计了一种基于基尼指数的抗遮挡成本量;然后根据成本最小原则自适应得到最佳成本量,计算得到初始深度图;再结合彩色图联合引导滤波,获得最终的全局高精度的深度图像。实验结果表明,相比于其他先进方法,本文方法简单有效,并且在保证深度计算精度的同时,降低了时间复杂度。但是,本文方法没有考虑光照环境对估计结果的影响,不能有效处理低纹理场景,限制了本文方法的使用场景,接下来的工作是尝试解决高光对深度估计的影响。

参考文献(References)

- [1] Fujimura K, Zhu Y. Target orientation estimation using depth sensing[P]. US: US7620202B2, 2009.
- [2] Xu Y, Maeno K, Nagahara H, et al. Light field distortion feature for transparent object classification[J]. Computer Vision and Image Understanding, 2015, 139(C): 122-135.
- [3] Waller L, Tian L. 3D intensity and phase imaging from light field measurements in an LED array microscope[J]. Optica, 2015, 2(2): 104-111.
- [4] Qi X, Liao R, Jia J, et al. 3D Graph neural networks for RGBD semantic segmentation[C]. IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 5209-5218.
- [5] Zhang J, Wang M, Gao J, et al. Saliency detection with a deeper investigation of light field[C]. International Conference on Artificial Intelligence. Austin: AAAI Press, 2015: 2212-2218.
- [6] Adelson E H, Wang J Y A. Single lens stereo with a plenoptic camera[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 14(2): 99-106.
- [7] Ng R. Digital light field photography[J]. Stanford: The Department of Computer Science, Stanford University, 2006.
- [8] 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.
- [9] 高隽, 王丽娟, 张旭东, 等. 光场深度估计方法的对比研究[J]. 模式识别与人工智能, 2016, 29(9): 769-779. (Gao J, Wang L J, Zhang X D, et al. Comparative study of light field depth estimation[J]. Pattern Recognition and Artificial Intelligence, 2016, 29(9): 769-779.)
- [10] 张旭东, 李成云, 汪义志, 等. 遮挡场景的光场图像深度估计方法[J]. 控制与决策, 2018, 33(12): 2122-2130. (Zhang X D, Li C Y, Wang Y Z, et al. Light field depth estimation for scene with occlusion[J]. Control and Decision, 2018, 33(12): 2122-2130.)
- [11] Li J, Lu M, Li Z N. Continuous depth map reconstruction from light fields[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3257-3265.
- [12] Kim C, Zimmer H, Pritch Y, et al. Scene reconstruction from high spatio-angular resolution light fields[J]. Acm Transactions on Graphics, 2017, 32(4): 1-12.
- [13] Diebold M, Jahne B, Gatto A. Heterogeneous light fields[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2016: 1745-1753.
- [14] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]. IEEE International Conference on Computer Vision (ICCV). Sydney: IEEE, 2013: 673-680.
- [15] Tao M W, Srinivasan P P, Malik J, et al. Depth from shading, defocus, and correspondence using light-field angular coherence[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 1940-1948.
- [16] Heber S, Ranftl R, Pock T. Variational shape from light field[C]. International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition. Heidelberg: Springer, 2013: 66-79.
- [17] Sabater N, Seifi M, Drazic V, et al. Accurate disparity estimation for plenoptic images[C]. IEEE European Conference on Computer Vision (ECCV). Zurich: IEEE, 2014: 548-560.
- [18] Heber S, Pock T. Shape from light field meets robust PCA[C]. IEEE European Conference on Computer Vision (ECCV). Zurich: IEEE, 2014: 751-767.
- [19] Chen C, Lin H, Yu Z, et al. Light field stereo matching using bilateral statistics of surface cameras[C].

- IEEE Computer Vision and Pattern Recognition (CVPR). Washington D C: IEEE, 2014: 1518-1525.
- [20] Heber S, Pock T. Convolutional networks for shape from light field[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2016: 3746-3754.
- [21] Heber S, Yu W, Pock T. Neural EPI-volume networks for shape from light field[C]. IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2271-2279.
- [22] Jeon H G, Park J, Choe G, et al. Depth from a light field image with learning-based matching costs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2): 297-310.
- [23] Bowen A, Mullins A, Rajpoot N. Photo-consistency and multiresolution methods for light field disparity estimation[C]. IEE International Conference on Visual Information Engineering. Glasgow: IET, 2005: 1-8.
- [24] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 1547-1555.
- [25] Wang T C, Efros A A, Ramamoorthi R. Occlusion-aware depth estimation using light-field cameras[C]. IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 3487-3495.
- [26] 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(C): 148-159.
- [27] Williem, Park I K, Lee K M. Robust light field depth estimation using occlusion-noise aware data costs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(10): 2484-2497.
- [28] 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.
- (Xiong W, Zhang J, Gao X J, et al. Anti-occlusion light-field depth estimation from adaptive cost volume[J]. Journal of Image and Graphics, 2017, 22(12): 1709-1722.)
- [29] 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.
- [30] Zhang Q, Xu L, Jia J. 100+ times faster weighted median filter (WMF)[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Washington DC: IEEE, 2014: 2830-2837.
- [31] He K, Sun J, Tang X. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397-1409.
- [32] Ham B, Cho M, Ponce J. Robust image filtering using joint static and dynamic guidance[C]. IEEE Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 4823-4831.
- [33] Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields[C]. Annual Workshop on Vision, Modeling and Visualization: VMV. Lugano: IEEE, 2013: 225-226.
- [34] Raj A S, Lowney M, Shah R. Light-field database creation and depth estimation[R]. Palo Alto: Stanford University, 2016.

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, E-mail: xudong@hfut.edu.cn;

董运流 (1994—), 男, 硕士生, 从事光场成像技术、计算机视觉的研究, E-mail: dy12877@mail.hfut.edu.cn;

石梦迪 (1994—), 女, 硕士生, 从事光场成像技术、计算机视觉的研究, E-mail: shismdshi@163.com.

(责任编辑: 齐 霖)