

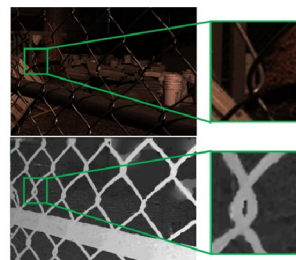


DOI: 10.12086/oe.2021.200422

基于光场内联遮挡处理的噪声场景深度获取

吴迪, 张旭东*, 范之国, 孙锐

合肥工业大学计算机与信息学院, 安徽 合肥 230601



摘要: 光场相机通过单次曝光同时从多个视角采样单个场景, 在深度估计领域具有独特优势。噪声场景下的深度获取是光场图像深度估计的难点之一。传统针对噪声场景的深度获取方法大多仅适用于非遮挡情况, 无法较好处理包含遮挡区域的噪声场景。针对含遮挡的噪声场景深度估计问题, 提出了基于内联遮挡处理的深度估计方法。该方法采用内联遮挡处理框架, 通过将遮挡处理集成进抗噪成本量中, 在保证抗噪性能的同时提升算法的抗遮挡能力。在成本量建立完成后, 为进一步滤除剩余噪声, 采用提出的适应遮挡的多模板滤波策略对成本量进行遮挡感知优化, 该策略通过为不同方向的遮挡分别设计滤波模板, 在滤波的同时能较好保留图像的边缘结构, 有效改善了传统滤波算法无法保留遮挡边界的问题。实验结果表明, 相比其它先进深度估计算法, 该方法在高噪声场景下具有显著优势, 并能更好处理噪声场景深度估计的遮挡问题。

关键词: 光场; 深度估计; 散焦线索; 噪声抑制; 遮挡处理

中图分类号: TP391; TB811

文献标志码: A

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

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

Depth acquisition of noisy scene based on inline occlusion handling of light field

Wu Di, Zhang Xudong*, Fan Zhiguo, Sun Rui

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China

Abstract: A light field camera can simultaneously sample a scene from multiple viewpoints with a single exposure, which has unique advantages in portability and depth accuracy over other depth sensors. Noise is a challenging issue for light field depth estimation. Most of the traditional depth estimation methods for noisy scenes are only suitable for non-occluded scenes, and cannot handle the noisy scenes with occluded regions. To solve this problem, we present a light field depth estimation method based on inline occlusion handling. The proposed method integrates the occlusion handling into the anti-noise cost volume, which can improve the anti-occlusion capability while maintaining the anti-noise performance. After the cost volume is constructed, we propose a multi-template filtering algorithm to smooth the data cost while preserving the edge structure. Experimental results show that the proposed method has better performance over other state-of-the-art depth estimation methods in high noise scenes, and can

收稿日期: 2020-11-16; 收到修改稿日期: 2021-05-12

基金项目: 国家自然科学基金资助项目(61876057, 61971177)

作者简介: 吴迪(1996-), 男, 硕士研究生, 主要从事计算机视觉、光场技术的研究。E-mail: wudi19960512@gmail.com

通信作者: 张旭东(1966-), 男, 博士, 教授, 主要从事智能信息处理、机器视觉的研究。E-mail: xudong@hfut.edu.cn

版权所有©2021 中国科学院光电技术研究所

better handle the occlusion problem of depth estimation in noisy scenes.

Keywords: light field; depth estimation; defocus cue; noise suppression; occlusion handling

1 引言

光场相机^[1]通过对单个场景进行多视角密集采样,使得高精度深度信息的挖掘成为可能,在深度传感器中不论在便携性还是精度方面都具有显著优势,适用于电影游戏特效、增强现实、人机交互等多种应用场景。

目前许多学者利用光场的各种特性挖掘深度信息,例如,基于极平面图像(epipolar plane images, EPI)的方法^[2-3]、基于多视角立体匹配(multi-view stereo, MVS)的方法^[4-5]和基于多视角图(angular patch, AP)的方法^[6-7]。但是大部分方法仅考虑深度估计中的遮挡问题,忽略了算法对图像噪声的适应性要求。近年来,一些结合抗遮挡和抗噪的深度估计方法相继提出^[8],但其大多利用多线索融合(多成本量叠加)策略来提高方法对不同场景的适应能力。当噪声和遮挡较弱时,多线索融合策略可通过优势成本量的强响应覆盖劣势成本量的弱响应。但在高噪场景中由于成像一致性^[9]无法满足,遮挡处理难度将显著提升,原本在遮挡区域的优势成本量将具有较低置信度,因此无法通过多线索融合策略覆盖劣势成本量。

针对含遮挡的噪声场景下的深度估计问题,提出了基于内联遮挡处理的深度估计方法,该方法将遮挡处理集成进抗噪成本量中,既保留了成本量原本的抗噪优势,又提升了抗遮挡性能。对于抗噪成本量的构建,本文方法改进了传统焦堆栈匹配度量,为单次匹配引入更多样本点,进一步提升了成本量的抗噪能力。对于内联遮挡处理,设计了基于视角掩模的遮挡处理算法,可在成本量的构建过程中较好地去除遮挡影响。最后,为进一步提升方法对遮挡和噪声的处理能力,提出了适应遮挡的多模板滤波策略,通过为不同方向的遮挡分别设计滤波模板,在滤除噪声的同时较好保留了图像的边缘结构。

本文的主要贡献如下:

1) 针对传统成本量无法同时处理遮挡和噪声的问题,设计了内联遮挡处理框架,通过将遮挡处理内联进抗噪成本量中,可在保证成本量抗噪能力的同时,提升其抗遮挡性能。

2) 提出基于双向散焦的焦堆栈匹配度量,并将其

作为内联遮挡处理框架的成本量构建度量。该度量采用双向散焦的方式构造焦堆栈,相比传统焦堆栈匹配度量具有更多信息量,具备更优抗噪性能。

3) 提出基于视角掩模的遮挡处理算法,并将其作为内联遮挡处理框架的遮挡处理算法。该算法符合框架的内联遮挡处理思想,可在成本量的构建过程中较好滤除遮挡。

4) 提出了适应遮挡的多模板滤波策略,通过将遮挡信息集成进滤波器中,可在滤波的同时较好保留图像的边缘结构。

2 国内外现状

光场的多视角采样特性使得深度信息的挖掘成为可能。然而,受噪声影响,深度估计的重要前提—成像一致性无法严格满足,进而会导致深度估计精度下降。特别对于存在遮挡的高噪场景,噪声将增大遮挡检测难度,错误的遮挡判断会进一步降低深度估计质量。因此增加方法对遮挡的适应能力对于光场抗噪深度估计是十分必要的。

大部分深度估计算法仅考虑深度估计的遮挡问题。一些先进的光场深度估计算法^[2,10]通过设计基于EPI的成本量,以提高算法的抗遮挡性能。EPI通过限制光场采样方向,可有效应对不同方向遮挡。但基于EPI的方法仅利用光场的小部分视角,造成视角信息的大量丢失,极大限制了算法的抗噪性能。针对算法对噪声和遮挡的适应问题,Williem等人^[7]设计了角度熵(angular entropy, AE)和自适应散焦(angular defocus, AD)成本量,并将两者融合以提高算法对不同场景的适应能力。其中AE属于多视角图成本量,能够较好地处理遮挡问题;AD属于焦堆栈成本量,对噪声有较低敏感度。该多线索融合策略在仅有遮挡或仅有噪声的场景中有较好表现。但当遮挡和噪声同时存在,两种线索的表现均会受限(如善于处理遮挡的成本量会被噪声干扰,而善于处理噪声的成本量会被遮挡干扰),进而导致深度估计的不精确。Park等人^[6]通过向文献^[7]中的两种成本量添加约束,进一步提升其抗遮挡和抗噪性能,但依然没有解决多线索融合策略的主要问题。Chen等人^[4]使用超像素正则化对深度图进行约束,通过在超像素内对深度信息进行传播,可较好

解决场景的遮挡和噪声问题。但该算法的约束能力依赖超像素精度, 若超像素错误混合前景和背景像素, 将加剧深度图的前后景混淆。为使算法能适用于遮挡和噪声同时存在的情况, Guo 等人^[11]首先利用抗遮挡算法计算初始深度图, 之后使用引导滤波对其平滑处理, 使方法同时具备较优的抗遮挡和抗噪能力。但基于引导滤波的平滑处理没有利用光场结构信息, 因此对高噪场景的处理具有一定局限性。

综上所述, 现有深度估计方法无法较好处理遮挡和噪声同时存在的场景。本文采用内联遮挡处理框架, 将遮挡处理直接集成进抗噪成本量。相比多线索融合策略, 该框架能在保证成本量抗噪能力的同时提升算法抗遮挡性能, 因此能够有效应对遮挡和噪声同时存在的情况。对于抗噪成本量的设计, 相比于传统基于焦堆栈的成本量, 本文提出的基于双向散焦的焦堆栈匹配度量包含更多信息量, 更充分利用了光场结构信息, 因此具备更优秀的抗噪能力。算法最后采用的多模板滤波策略通过将遮挡信息集成进滤波器中, 可在保持遮挡边缘的情况下进一步滤除剩余噪声。

3 基于内联遮挡处理的噪声场景深度获取

针对噪声场景下的深度估计问题, 尤其是对于包

含遮挡的高噪场景, 本文设计了基于内联遮挡处理的深度估计算法, 通过将遮挡处理内联于抗噪成本量, 同时保证了成本量对于遮挡和噪声的适应能力。算法框图如图 1 所示, 光场图像首先进行成本量构建, 之后通过成本量融合和成本量优化和求解得到视差图, 最后经过映射得到最终深度图。具体分为以下步骤:

- 1) 成本量构建。该步骤利用提出的基于双向散焦的焦堆栈匹配度量, 构建双向散焦成本量, 并在构建过程中使用一系列视角掩模进行内联遮挡处理;
- 2) 成本量融合。该步骤在成本量集合上采用最小池化操作, 筛选出被遮挡影响最小的成本量;
- 3) 成本量优化及求解。该步骤使用所提出的多模板滤波策略, 在保证图像边缘结构的情况下对成本量进行逐标签平滑处理, 之后使用 TG2^[12]求解最终视差图。
- 4) 视差-深度映射。该步骤根据成像设备参数完成视差向深度的映射。

本文使用多目成像的表述方式描述光场捕获的光线, 记作 $L(s, p)$, 其中 $s = (u, v)$ 表示采样到该光线的视角在视角平面 Π 内的坐标, $p = (x, y)$ 为该光线在视角 s 所成的像中的坐标。特别地, 使用 $s_c = (u_c, v_c)$ 表示中心视角对应的视角平面 Π 内的坐标, 并将其作为生成视差图的参考视角。

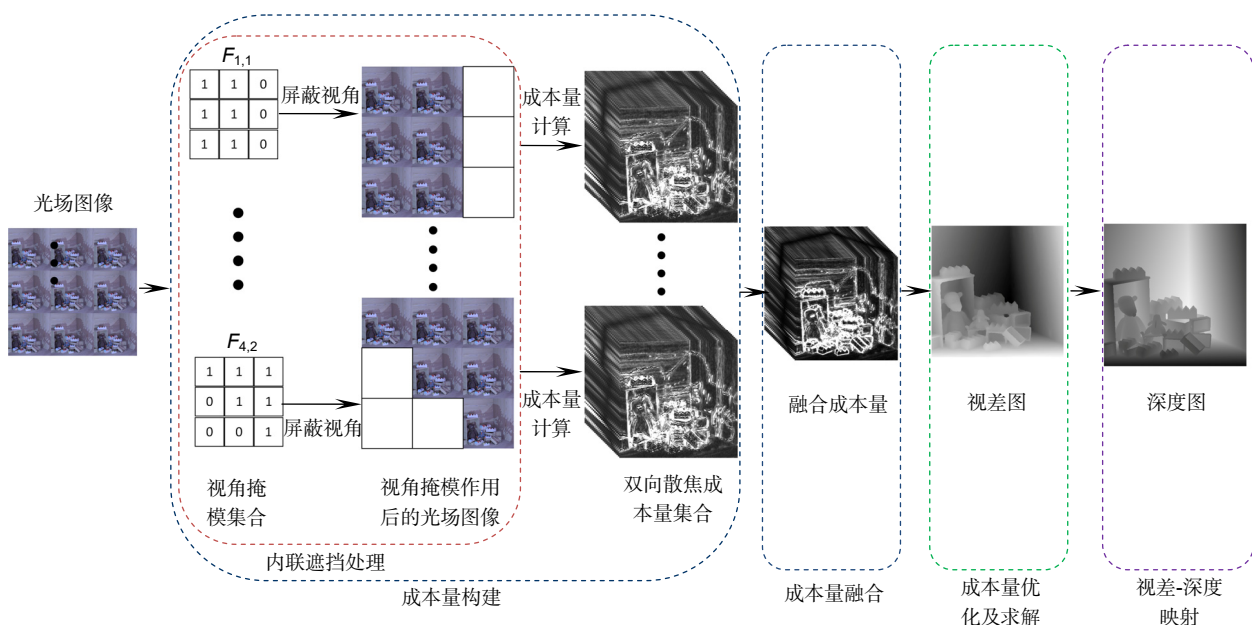


图 1 算法总体框图
Fig. 1 The algorithm framework

3.1 基于双向散焦的焦堆栈匹配度量

光场焦堆栈匹配度量(focal stack matching measure)^[13]利用中心图像(central image)构建合成焦堆栈, 并与光场焦堆栈进行栈与栈之间的匹配, 其度量函数如下所示:

$$s_p(f) = \int_0^{\delta_{\max}} \rho(\psi_p(\delta) - \varphi_p(f + \delta)) d\delta, \quad (1)$$

式中: $\varphi_p(f + \delta)$ 表示标签为 $f + \delta$ 的光场焦堆栈切片(focal slice)中点 p 处的像素值, 其中光场焦堆栈切片由光场多视角图进行逐视角累加得到; $\psi_p(\delta)$ 表示标签为 δ 的合成焦堆栈切片(rendered focal slice)中点 p 处的像素值, 其中合成焦堆栈切片由合成多视角图进行逐视角累加得到;

$$\rho(x) = 1 - \exp\left(-\frac{|x|}{2\sigma^2}\right)$$

为距离函数(distance function), 其作用是增加匹配度量对噪声的鲁棒性。该度量通过考虑散焦范围 $\delta \in [0, \delta_{\max}]$ 内的匹配成本, 相比传统切片匹配更加充分利用了光场数据, 提升了算法对噪声的鲁棒性。但其依然没有充分发掘光场结构信息, 匹配过程仅涉及近向散焦匹配, 将散焦范围限制在 $\delta \in [0, \delta_{\max}]$, 忽视了远向散焦(即 $\delta < 0$ 方向)的信息, 极大地限制了抗噪能力。

本文对该度量公式进行扩充改进, 提出了基于双

向散焦的焦堆栈匹配度量, 可同时从远向散焦和近向散焦方向捕获信息, 并选取最优捕获方向, 其公式如下所示:

$$s_p^{\text{double}}(f) = \min(s_p^{\text{near}}(f), s_p^{\text{far}}(f)), \quad (2)$$

其中: $s_p^{\text{near}}(f) = s_p(f)$ 表示近向散焦匹配成本量, $s_p^{\text{far}}(f)$ 表示远向散焦匹配成本量, 其拥有与 $s_p^{\text{near}}(f)$ 相反的散焦方向, 表达式为

$$s_p^{\text{far}}(f) = \int_{-\delta_{\max}}^0 \rho(\psi_p(\delta) - \varphi_p(f + \delta)) d\delta. \quad (3)$$

基于双向散焦的焦堆栈匹配度量在构建光场焦堆栈过程中, 采用远向散焦和近向散焦两个散焦方向。如图 2 所示, 近向和远向散焦分别以不同的散焦方向对同一区域进行采样, 因此均可与采样该区域的合成焦堆栈进行匹配。当场景存在噪声, 对同一区域的多次采样允许算法从中选择匹配代价最低的样本, 从而提升抗噪性能。在最小池化(式(2))操作后, 筛选出的成本量将包含较少被噪声影响的像素, 因此与原本的单向散焦(式(1))相比, 基于双向散焦的焦堆栈匹配度量能更有效应对噪声场景。

3.2 内联遮挡处理

虽然上述基于双向散焦的焦堆栈匹配度量对噪声场景的深度估计具有独特优势, 但当遮挡存在, 光场焦堆栈与合成焦堆栈之间将无法完成匹配, 导致其无

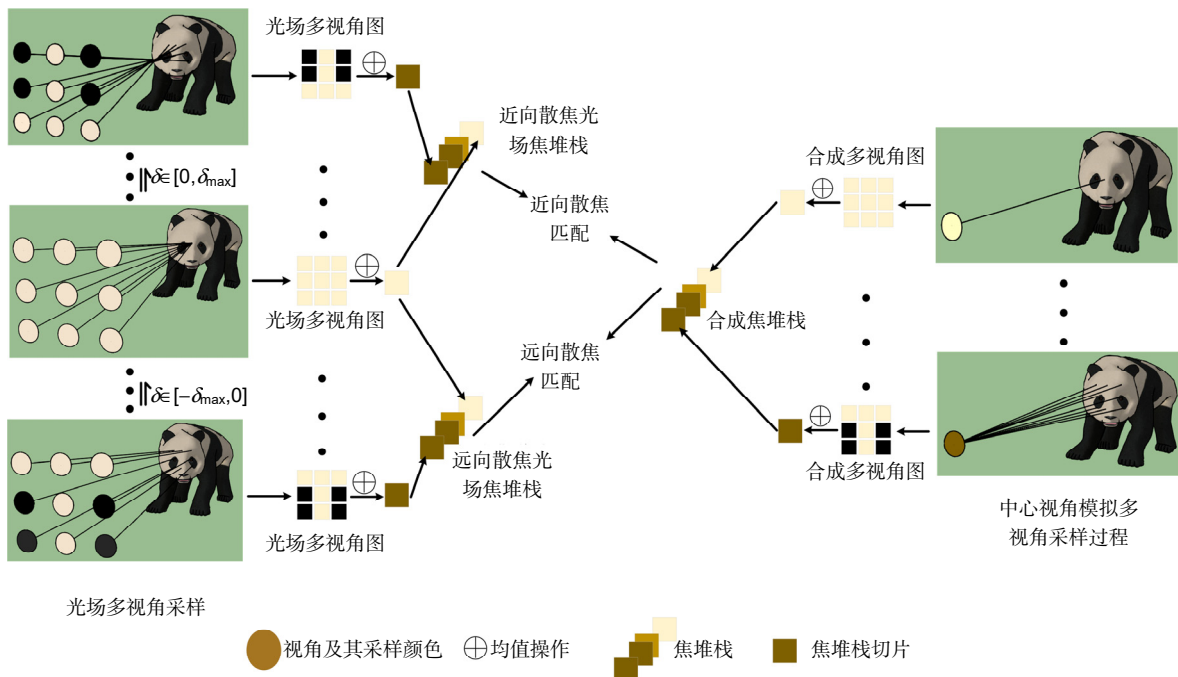


图 2 基于双向散焦的焦堆栈匹配度量(角度分辨率 3×3)

Fig. 2 Focal stack matching measurement based on double-directions defocusing (angel resolution 3×3)

法处理包含遮挡的场景。如图 3 所示, 为方便描述, 此处将 3D 场景简化为 2D 模型。图 3(a)表示光场焦堆栈切片中某一点 $\varphi_p(\delta)$ 的构建过程, 图中视角在散焦状态下对蓝色物体进行采样, 但由于采样区域附近存在深度更小的绿色物体, 视角 $C_0 \sim C_1$ 无法正确采样蓝色物体, 导致 $\varphi_p(\delta)$ 包含不同深度物体的色彩信息。图 3(b)表示与之对应的合成焦堆栈切片的构建过程, 即使 $\psi_p(\delta')$ 已处于正确标签, 由于遮挡存在, 光线 $l_0 \sim l_1$ 所采样位置无法与视角 $C_0 \sim C_1$ 所采样位置相匹配, 导致 $\varphi_p(\delta)$ 与 $\psi_p(\delta')$ 无法匹配, 最终将导致整个焦堆栈匹配失败。使用多线索融合策略虽可提高算法对遮挡和噪声的适应能力, 但其遮挡处理能力受噪声等

级影响较大, 因此不适用于高噪场景。并且由于散焦状态下的不同视角均采样不同空间点, 因此无法利用成像一致性检测遮挡, 导致遮挡处理难度进一步加大。

为使方法能有效处理包含遮挡的高噪场景, 设计了内联遮挡处理框架, 该框架采用内联思想, 将遮挡处理直接集成于成本量的构建过程, 相比多线索融合策略能较好地处理含遮挡、高噪场景。对于遮挡处理方法的选取, 本文设计了基于视角掩模的遮挡处理算法, 该算法符合框架的内联遮挡处理思想, 可在成本量的构建过程中较好滤除遮挡。如图 4 所示, 在构建焦堆栈时, 若把(a)受遮挡影响的视角 $C_0 \sim C_1$ 与所对应(b)中的光线 $l_0 \sim l_1$ 去除, 剩余视角和光线所构建的

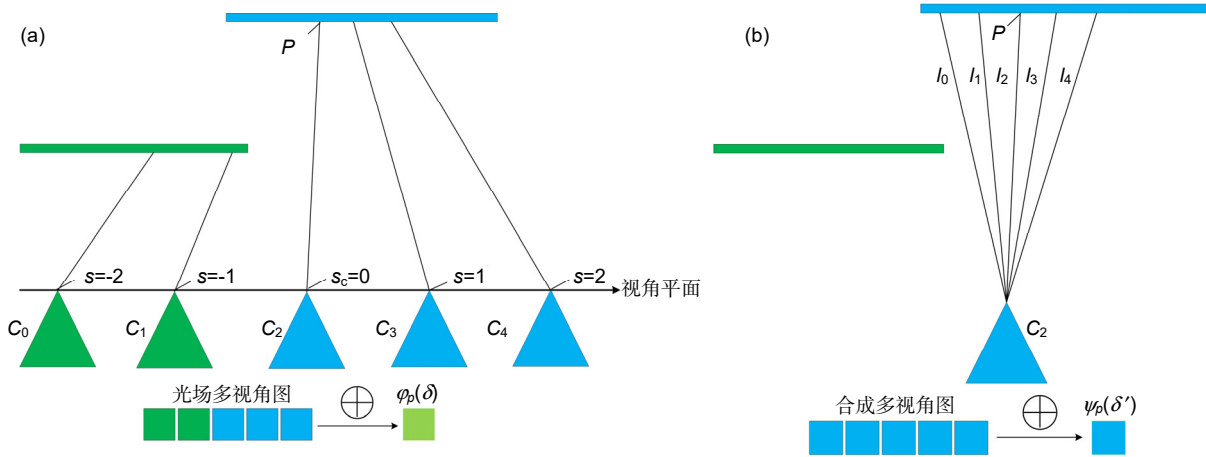


图 3 遮挡对焦堆栈匹配的影响(角度分辨率 5×1), 顶部蓝、绿线条表示待采样物体, 底部三角形为视角, 填充色为采样颜色。(a) 所有视角构建光场焦堆栈切片; (b) 所有合成视角构建合成焦堆栈切片

Fig. 3 The influence of occlusion on focal stack matching (angel resolution 5×1), the bottom triangles are the view points. (a) Focal slices based on all viewpoints; (b) Rendered focal slices based on all viewpoints

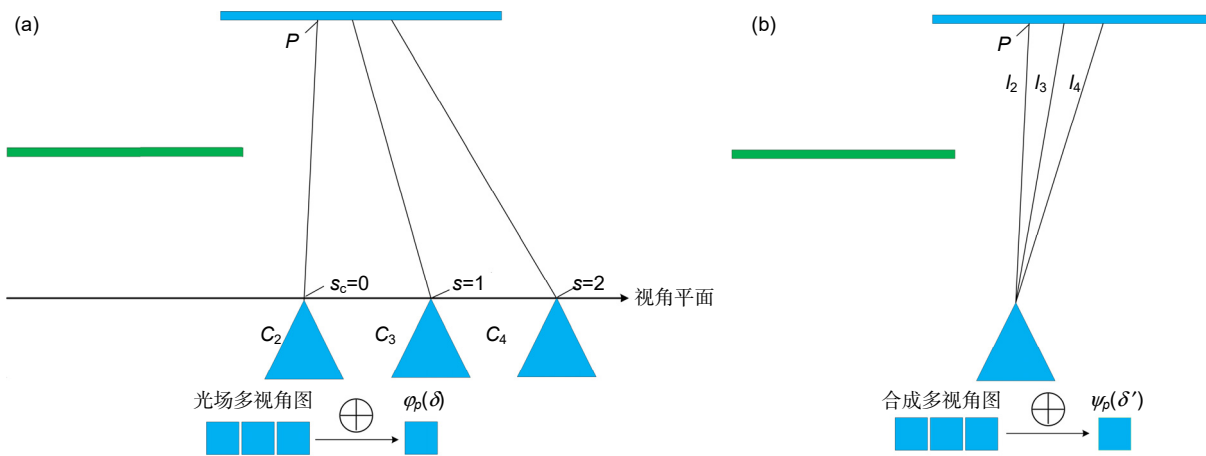


图 4 焦堆栈的局部匹配(角度分辨率 5×1)。

(a) 部分视角构建光场焦堆栈切片; (b) 部分合成视角构建合成焦堆栈切片

Fig. 4 Partial matching of focal stack (angle resolution: 5×1).

(a) Focal slices based on partial points; (b) Rendered focal slices based on partial view points

$\varphi_p(\delta)$ 与 $\psi_p(\delta')$ 将完全匹配。但去除被遮挡视角需计算场景遮挡模式，而噪声场景下的遮挡模式拥有较大计算难度。为消除遮挡对焦堆栈匹配的影响且不被噪声干扰，算法为不同遮挡模式预先分别设计对应的视角掩模，并在此基础上分别构成本量，之后根据匹配成本进行自适应选择。如图 5 所示，由于 p 点附近存在遮挡物，光场多视图图包含部分遮挡物像素，该部分像素将影响后续匹配。在视角掩模 $F_{i,j}$ 的作用下，部分视角被掩模去除，其中 $F_{1,1}$ 去除了所有被遮挡视角，在其作用下的焦堆栈切片将拥有最小匹配成本，因此可使用最小池化选择受遮挡影响最小的匹配度量。权衡精确率与计算复杂度，算法设计的视角掩模包含 4

个角度的两个侧面的遮挡模式，足够应对大部分遮挡情况，如图 6 所示。

内联视角掩模的匹配度量：

$$s_p^{occ}(f) = \min_{\substack{i=1,2,3,4 \\ j=1,2}} \{ \min(s_{p,i,j}^{near}(f), s_{p,i,j}^{far}(f)) \} \quad (4)$$

式中： $s_{p,i,j}^{near}(f)$ 和 $s_{p,i,j}^{far}(f)$ 分别表示经过视角掩模 $F_{i,j}$ 处理后的近向和远向散焦成本量，其构建过程：

$$s_{p,i,j}^{near}(f) = \int_0^{\delta_{max}} \rho(\psi_{p,i,j}(\delta) - \varphi_{p,i,j}(f + \delta)) d\delta \quad (5)$$

$$s_{p,i,j}^{far}(f) = \int_{-\delta_{max}}^0 \rho(\psi_{p,i,j}(\delta) - \varphi_{p,i,j}(f + \delta)) d\delta \quad (6)$$

式(5)、式(6)采用与式(1)、式(3)相同的计算形式，但据图 5 可知， $\psi_{p,i,j}$ 和 $\varphi_{p,i,j}$ 内联了视角掩模 $F_{i,j}$ 。当 $F_{i,j}$

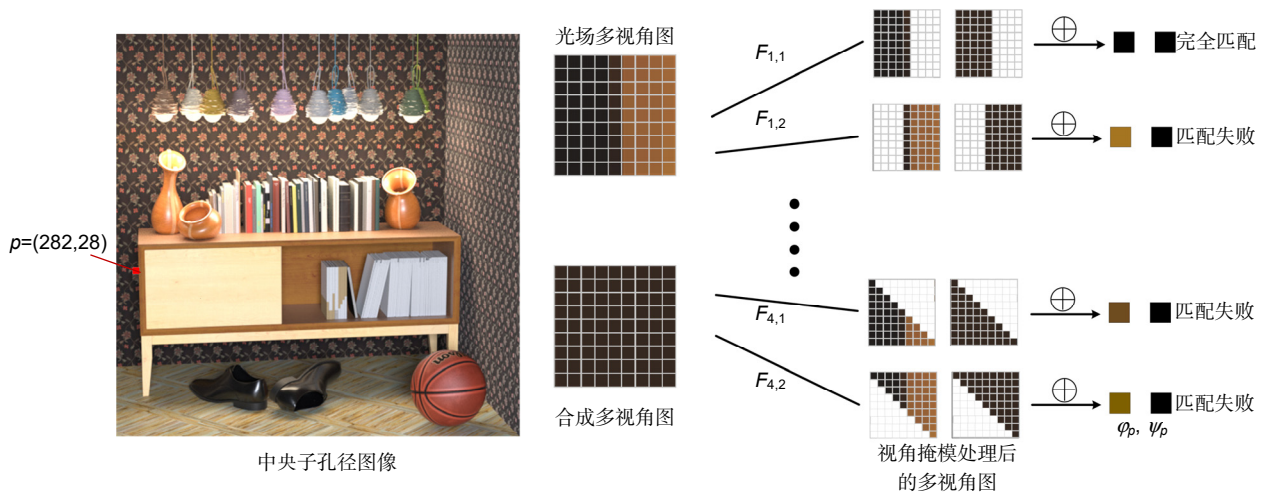


图 5 集成视角掩模的焦堆栈切片匹配(角度分辨率 9×9)
Fig. 5 Focal stack matching with an integrated view mask (angle resolution: 9×9)

$F_{1,1}$			$F_{2,1}$			$F_{3,1}$			$F_{4,1}$		
1	1	0	1	1	1	0	0	1	1	0	0
1	1	0	1	1	1	0	1	1	1	1	0
1	1	0	0	0	0	1	1	1	1	1	1
$F_{1,2}$			$F_{2,2}$			$F_{3,2}$			$F_{4,2}$		
0	1	1	0	0	0	1	1	1	0	0	1
0	1	1	1	1	1	1	1	0	0	1	1
0	1	1	1	1	1	0	0	0	1	1	1

图 6 应对不同方向和不同侧的遮挡的视角掩模(以角度分辨率 3×3 为例)
Fig. 6 View masks for occlusion in different directions and sides (angle resolution 3×3)

与遮挡模式相匹配的情况下, 式(5)、式(6)在正确标签处将拥有最低的匹配成本, 因此式(4)筛选出的成本量将包含最少的遮挡物像素。

对于 $\varphi_{p,i,j}$ 的构建, 将光场多视角图与视角掩模进行与操作并取均值获得:

$$\varphi_{p,i,j}(f+\delta) = \frac{1}{\|F_{i,j}\|_{s \in \Pi}} \sum_{s \in \Pi} L(s, p+(s-s_c) \times (f+\delta)) \times F_{i,j}(s) \quad (7)$$

其中: $\psi_{p,i,j}$ 采用相同的方式构建, 若 $F_{i,j}$ 屏蔽了被遮挡的合成视角, 可认为中心视角与剩余合成视角的采样位置均处于同一深度平面。据文献[13]可知, 在等深度情况下, 合成多视角图将对应应在中心图像以目标点为中心的均匀采样结果, 其构建过程:

$$\psi_{p,i,j}(\delta) = \int_q k_{p,i,j}^\delta(q) L(s_c, q) dq \quad (8)$$

其中: $k_{p,i,j}^\delta$ 为经过视角掩模处理的以 p 为中心的均匀采样模板, 其表达式:

$$k_{p,i,j}^\delta(q) = \begin{cases} 1 & q = p + (s - s_c) \times \delta, F_{i,j}(s) = 1 \\ 0 & \text{else} \end{cases} \quad (9)$$

3.3 成本量融合

式(7)、式(8)默认多视角图始终存在部分视角被遮挡, 但若目标点周围不存在遮挡物, 屏蔽一部分视角

将造成有用信息的丢失, 并影响算法抗噪性能。针对这一问题, 在成本量构建完成后, 利用其能量分布对遮挡状态进行推断, 并依据该状态决定是否遮挡处理。遮挡判断原理如图 7 所示, 在不存在遮挡的情况下, 若成本量采用相反的散焦方向, 且散焦间隔相等并使用不同侧的视角掩模, 多视角图将由相同像素点构成, 如图 7(a)、7(b)中两视角均采样 P_1 和 P_2 点, 因此得到的成本量将是相似的。而对于遮挡场景, 将至少存在一对不同侧的视角掩模, 使光场在两种散焦方向下的被遮挡视角是不一致的, 如图 7(c)、7(d)中部分视角被遮挡, 导致采样点不一致, 因此得到的成本量将会有较大差异。

根据这一原理, 利用近向和远向散焦成本量的差异判断遮挡状态:

$$E_{occ}(p) = \prod_{i=1}^4 T\left(\int_{f_{min}}^{f_{max}} |s_{p,i,1}^{near}(f) - s_{p,i,2}^{far}(f)| df > T\right) \quad (10)$$

其中: $T(\text{cond}) = \begin{cases} 1 & \text{cond} = \text{true} \\ 0 & \text{else} \end{cases}$,

T 为条件判断语句, 是判断成本量相似度的阈值。若在任意一对相反视角掩模 $F_{i,1}$ 和 $F_{i,2}$ 的作用下, 点 p 在两个散焦方向的任意深度标签处均拥有相近的成本量, 则判定为不被遮挡。若点 p 不被遮挡, 将改用成

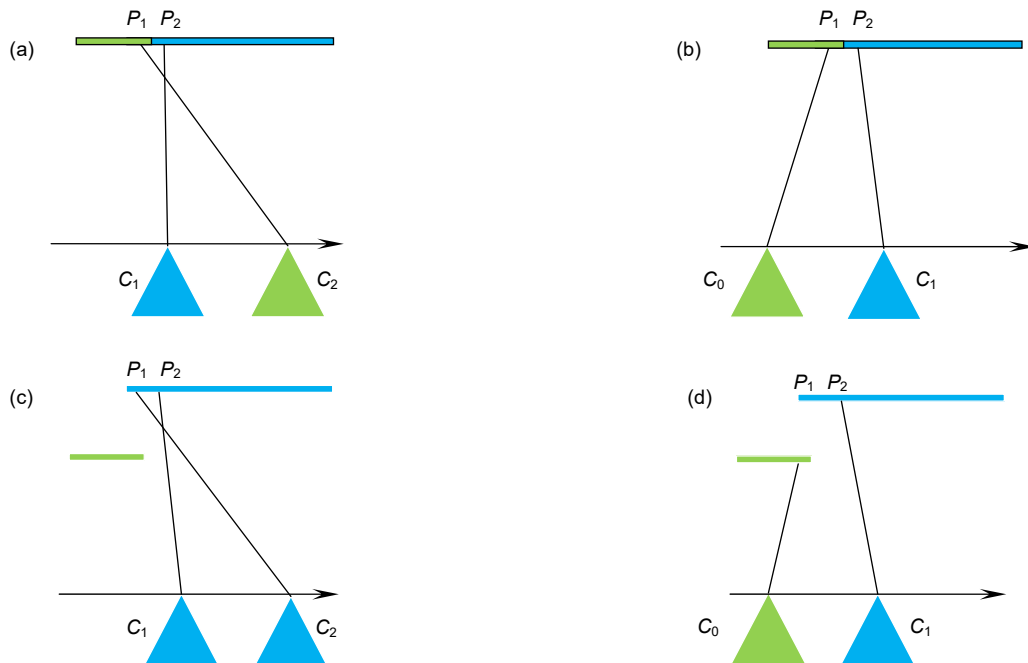


图 7 采用相反的散焦方向及不同侧视角掩模的采样情况对比。

(a), (c) 近向散焦且左侧视角被屏蔽; (b), (d) 远向散焦且右侧视角被屏蔽

Fig. 7 Comparison of sampling using opposite defocusing directions and different side view masks.

(a), (c) Near defocus at the left view is blocked; (b), (d) Far defocus at the right view is blocked

本量集合(即池化操作(式(4))前的结果)的均值作为该点的最终成本量。即:

$$s_p^{\text{local}}(f) = \begin{cases} s_p^{\text{occ}}(f) & E_{\text{occ}}(p) = 1 \\ \sum_{i,j,C} \frac{1}{16} s_{p,i,j}^C(f) & E_{\text{occ}}(p) = 0 \end{cases}, \quad (11)$$

其中 C 表示近向散焦和远向散焦。

3.4 多模板滤波

内联遮挡处理在遮挡区域屏蔽了部分视角, 将造成抗噪能力的下降。为弥补抗噪性能的缺失, 同时进一步滤除剩余噪声, 在成本量融合完成后, 本文使用滤波算法对成本量进行平滑处理。传统滤波算法如均值滤波、高斯滤波和中值滤波都无法保持图像边缘结构, 极易造成遮挡边界附近前后景的混淆。为使方法能在滤除噪声的同时保持遮挡边缘, 提出了适应遮挡的多模板滤波策略。

据 3.2 节分析可知, 若内联的视角掩模与遮挡模式相匹配, 在正确标签处将拥有最小匹配成本, 基于这一性质, 对于遮挡边界附近的一点 p , 其遮挡方向和遮挡侧可由式(12)计算的 $i(p)$, $j(p)$ 查找视角掩模 $F_{i(p),j(p)}$ (图 6) 得出。

$$i(p), j(p) = \arg \min_{i,j} \left(\min_{\substack{f \in \{f_{\min}, f_{\max}\} \\ C = \text{near, far}}} (s_{p,i,j}^C(f)) \right). \quad (12)$$

在滤波过程中, 为避免不同深度像素点的错误混淆, 首先使用式(12)得出的视角掩模对卷积核进行加权, 之后再对成本量进行滤波操作, 如下所示:

$$s_p^{\text{filtered}}(f) = \sum_{p' \in N(p)} s_{p'}^{\text{local}}(f) * F_{i(p),j(p)}(p' - p) * G(p' - p), \quad (13)$$

其中: G 表示高斯卷积核, $N(p)$ 表示像素 p 的邻域。向卷积核引入遮挡信息后, 若 p 的邻域内存在被遮挡像素, 该部分像素将不再参与卷积操作, 有效避免了信息的错误融合。对成本量进行滤波操作后, 采用 TGV2^[12] 对成本量进行求解得到视差图, 经映射^[3] 后得到最终深度图。

4 实验结果与分析

4.1 实验配置

实验使用的 PC 机配置为 Intel Core i5-8250U @ 1.60 GHz, RAM 8 GB, Windows10 64 位操作系统。基于 Matlab 2016a 和 TDM-GCC 5.1.0 进行编译。实验分别在合成数据集和真实数据集上对提出的算法进行测试。其中合成数据集由 4D light Field Benchmark 数据集^[15] 添加加性高斯噪声来生成。该数据集中的每个光场图像由 9×9 个子孔径图像组成, 实验添加噪声的标

准差分 $\sigma_n = 1, 5, 10$ 三个等级。真实场景选用 Stanford Lytro Illum 数据集, 该数据集中的每个光场图像由 14×14 个子孔径图像组成, 实验选择中央 9×9 个子孔径图像作为输入。定量实验使用均方误差(MSE $\times 100$, 用 $E_{\text{MSE}} \times 100$ 表示)和错误像素比例(BadPix(t), 用 $R_{\text{BadPix}}(t)$ 表示)对实验结果进行评估, 其定义:

$$E_{\text{MSE}} \times 100 = \frac{1}{|M|} \sum_{p \in M} (d(p) - d_{\text{gt}}(p))^2 \times 100, \quad (14)$$

$$R_{\text{BadPix}}(t) = \frac{|\{p \in M : |d(p) - d_{\text{gt}}(p)| > t\}|}{|M|} \times 100, \quad (15)$$

其中: d 和 d_{gt} 分别表示估计视差图和真实视差图, t 表示错误像素的评价阈值(本文设为 0.07), M 表示视差图的像素集合。MSE $\times 100$ 和 BadPix(t) 为深度估计领域最常用的两个评价标准, 其中 MSE $\times 100$ 计算真实值和估计值之间的均方误差, 该指标越小则说明真实值和估计值有越高逼近程度; BadPix(t) 统计错误像素比例, 该指标越小则说明所提方法在一定误差范围内有越高的达标率。

4.2 定性分析

4.2.1 合成场景

图 8 给出了 4D light Field Benchmark^[14] 上 SPO^[2]、POBR^[4]、CAE^[6]、OADE^[9]、LF_PAC^[11] 同本文方法的结果对比, 其中错误像素用红色标出。场景 cotton 的背景部分为无纹理区域, 噪声的引入会使此区域的深度估计难度急剧增加。SPO 在 $\sigma_n = 1$ 的 cotton 场景中取得最好成绩, 但随着噪声等级的提高其在背景区域生成大量错误像素, 这是由于基于 EPI 的成本量无法捕获足够信息, 导致深度估计精度受噪声影响较大。而本文焦堆栈匹配度量充分发掘光场信息, 具有较低噪声敏感度, 因此即使在 $\sigma_n = 10$ 的高噪场景下也拥有较少错误像素。场景 dino 和 sideboard 包含较多遮挡边缘, 虽然 SPO 和 CAE 在低噪情况下有着较优表现, 但其遮挡处理水平受限于噪声等级, 因此无法适用于高噪情况。LF_PAC 在中、低噪场景表现较好, 但其使用的引导滤波方法没有利用光场结构信息, 因此极大限制了抗噪能力, 导致在高噪场景拥有较多错误像素。本文内联遮挡处理算法无需对被遮挡视角进行识别, 因此其遮挡处理能力受噪声等级影响较小, 能够有效处理高噪情况下的遮挡。

4.2.2 真实场景

图 9 给出了 Stanford Lytro Illum 上 SPO、POBR、CAE 与本文方法的视差估计结果对比。实验采用的场

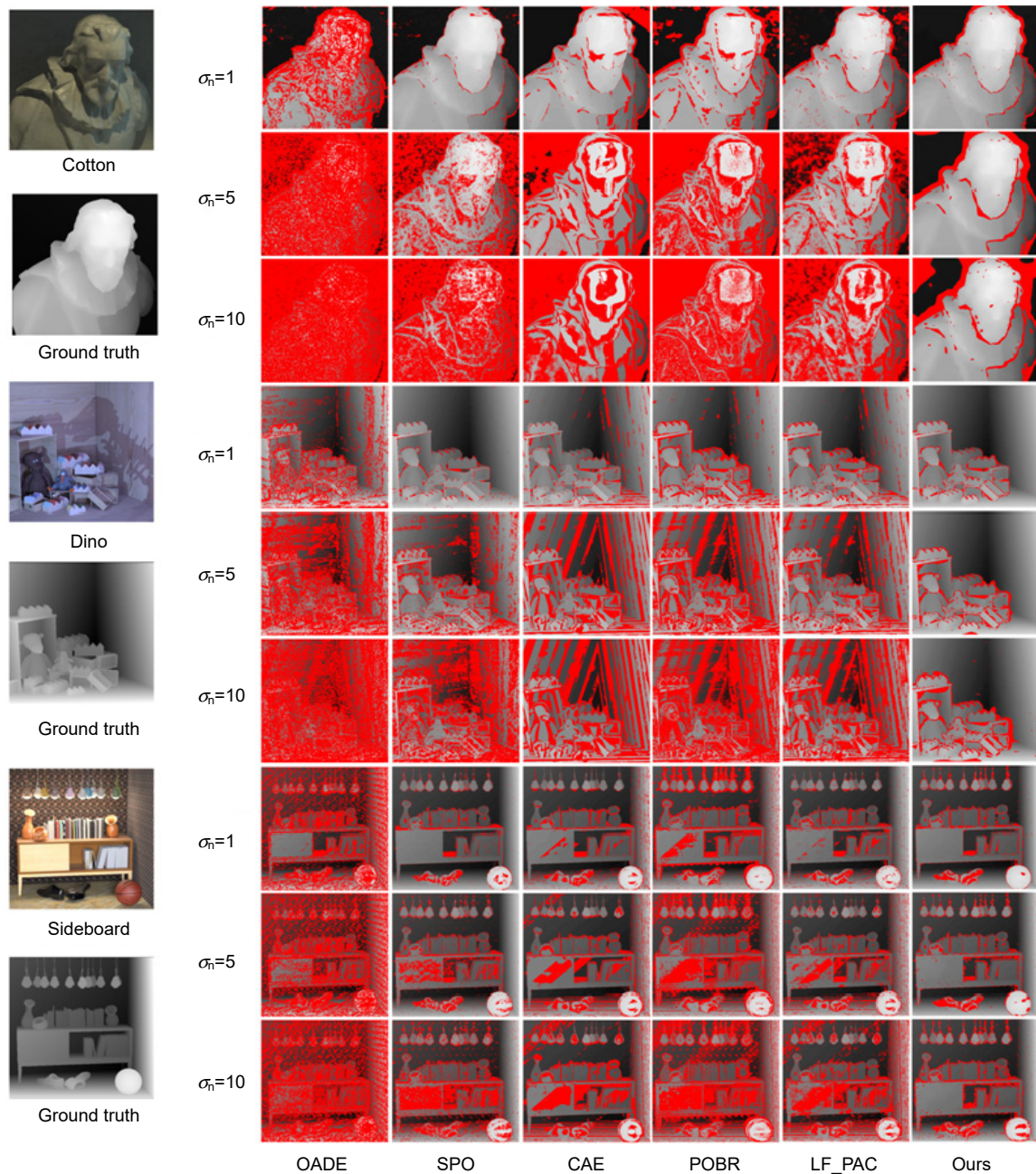


图 8 不同噪声等级的合成场景视差估计结果对比

Fig. 8 Comparison of disparity estimation results of synthetic scenes with different noise levels

景均在光照较暗的条件下拍摄，且部分场景包含大量网格结构，具有较大的噪声和遮挡处理难度。由于内联遮挡处理可从噪声场景精确滤除被遮挡视角，本文算法在场景 I、III 估计出清晰的网格结构。CAE 采用的多线索融合策略由于无法处理噪声和遮挡同时存在的情况，估计结果出现了前景向背景的平滑过度和网格结构的断裂。虽然 POBR 的结果没有出现网格断裂，但由于采用的聚类方法无法较好适应低对比度和高噪

场景，其在遮挡边界附近生成大量深度伪影。而基于 EPI 的 SPO 由于缺乏足够的信息量，在三个场景背景的黑暗区域均生成大量错误像素。

4.3 定量分析

4.3.1 非遮挡区域表现

由于 Stanford Lytro Illum 未提供深度真值，所以实验仅在 HCI 合成数据集上进行定量分析。图 10 展

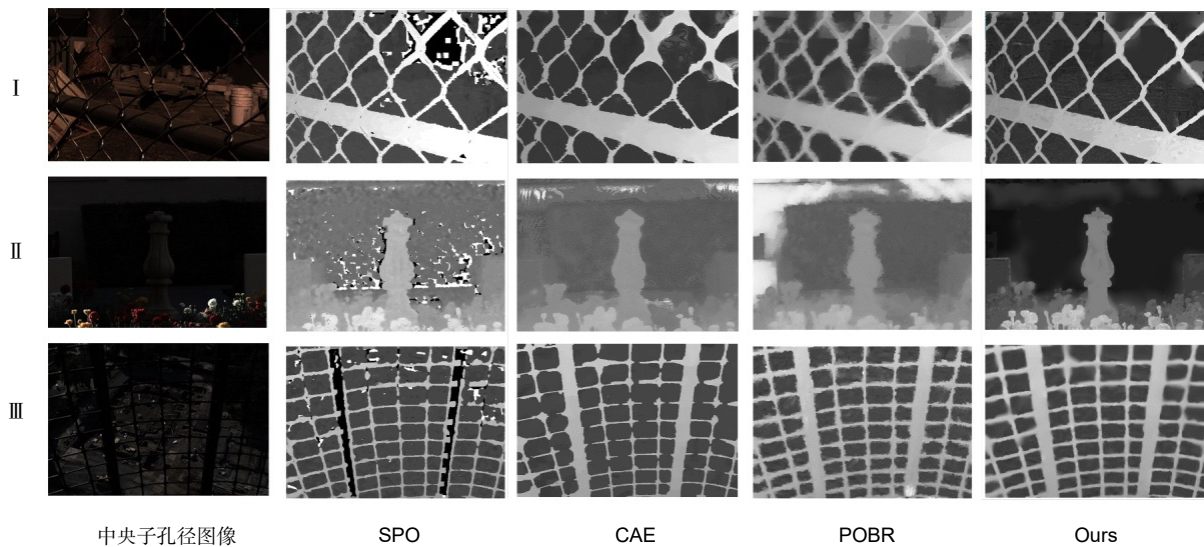


图 9 真实场景视差估计结果对比
Fig. 9 Comparison of real scene disparity estimation results

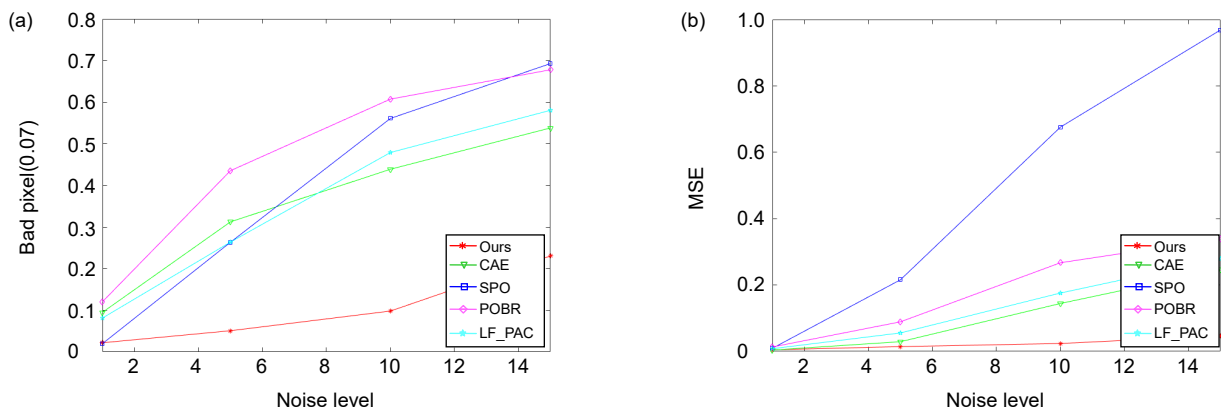


图 10 非遮挡区域各算法性能对比
Fig. 10 Performance comparison of algorithms in non-occluded regions

示了 SPO、POBR、CAE、LF_PAC 和本文算法对含噪场景的非遮挡区域的估计精度对比。本文基于双向散焦的焦堆栈匹配度量充分发掘光场结构信息，拥有极高的抗噪性能，因此在各个噪声等级的表现均领先于其他算法，且优势随着噪声等级的增加逐渐扩大，在 $\sigma_n \geq 10$ 的高噪场景下大幅领先次优算法，充分验证其在非遮挡区域的抗噪能力。

4.3.2 遮挡区域表现

图 11 展示了 SPO、POBR、CAE、LF_PAC 和本文算法对含噪场景的遮挡区域的估计精度对比。遮挡区域拥有较高深度估计难度，所有方法的精度均有所下

降。虽然在低噪情况下本文方法的深度估计精度不占优势，但随着噪声等级的增大，其余算法的精度均受到较大影响，而内联遮挡处理算法在不同的噪声等级都具有较高精度且受噪声等级的影响较小。当噪声等级达到 $\sigma_n = 5$ 时，本文方法在遮挡区域的深度估计精度达到最优水平，并随着噪声等级的升高，优势逐渐扩大。因此该方法能更有效处理包含遮挡的高噪场景。

4.3.3 整体表现

表 1 给出各方法在不同噪声等级场景的错误像素指标，表 2 给出了对应均方误差指标，其中最优和次优分别用加粗和下划线突出显示。结果表明，本文方

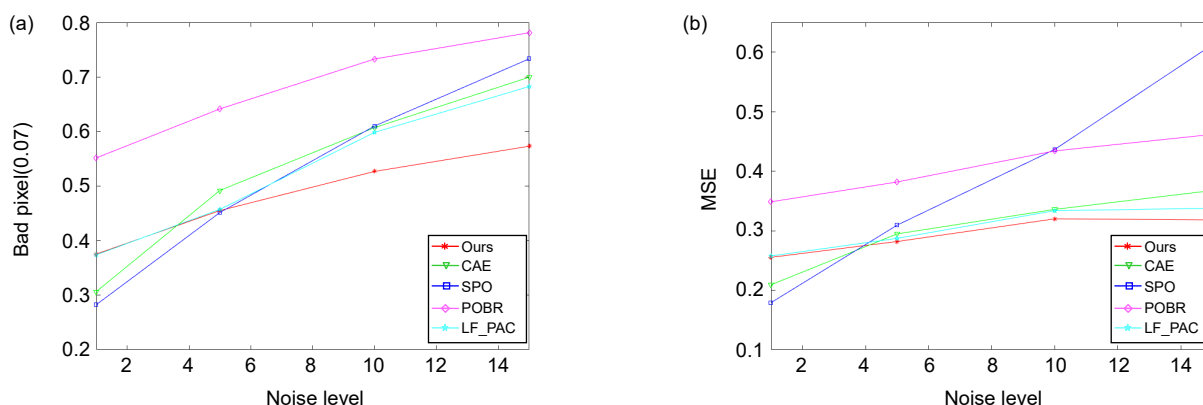


图 11 遮挡区域各算法性能对比
Fig. 11 Performance comparison of algorithms in occluded regions

表 1 坏像素 0.07 评估
Table 1 BadPix(0.07)

名称	噪声等级	OADE	SPO	CAE	POBR	LF_PAC	本文方法
Cotton	$\sigma_n=1$	22.39	3.46	11.02	16.92	9.13	<u>5.92</u>
	$\sigma_n=5$	78.69	46.37	48.75	63.32	<u>37.58</u>	12.19
	$\sigma_n=10$	88.37	78.02	65.75	76.54	<u>64.34</u>	20.54
Dino	$\sigma_n=1$	13.11	3.06	9.23	13.24	9.78	<u>4.62</u>
	$\sigma_n=5$	44.86	<u>23.18</u>	32.24	41.52	24.59	7.83
	$\sigma_n=10$	69.27	59.44	<u>44.38</u>	64.20	46.58	11.07
Sideboard	$\sigma_n=1$	27.41	7.48	11.40	22.82	10.34	<u>8.67</u>
	$\sigma_n=5$	39.90	<u>15.78</u>	19.32	34.98	17.81	9.98
	$\sigma_n=10$	50.98	29.31	<u>28.85</u>	46.35	33.07	14.97
Average		48.33	29.57	30.1	42.21	<u>28.14</u>	10.64

表 2 均方误差评估
Table 2 MSE×100

名称	噪声等级	OADE	SPO	CAE	POBR	LF_PAC	本文方法
Cotton	$\sigma_n=1$	4.55	2.81	<u>4.48</u>	7.81	6.61	5.68
	$\sigma_n=5$	41.25	55.76	<u>14.14</u>	34.37	20.07	9.64
	$\sigma_n=10$	67.47	110.58	54.13	80.13	<u>48.77</u>	12.66
Dino	$\sigma_n=1$	1.69	<u>0.52</u>	0.51	1.95	1.00	0.85
	$\sigma_n=5$	5.92	8.89	<u>1.41</u>	2.81	1.52	1.29
	$\sigma_n=10$	19.58	54.56	<u>2.57</u>	10.16	4.13	1.82
Sideboard	$\sigma_n=1$	4.52	<u>1.10</u>	0.85	5.05	1.17	1.91
	$\sigma_n=5$	5.58	6.38	1.18	6.11	<u>1.50</u>	1.98
	$\sigma_n=10$	7.61	18.43	2.32	6.88	3.77	<u>3.18</u>
Average		17.57	28.78	<u>9.07</u>	17.25	9.84	4.33

法在低噪声等级中与 SPO 均有着优秀的表现。从 4.3.1 节和 4.3.2 节实验可知, 本文算法在遮挡区域和非遮挡区域均拥有较低噪声敏感度, 而 SPO 的噪声敏感度偏高。因此随着噪声等级的提高, SPO 的误差大幅增加, 本文方法凭借在高噪声等级下的抗遮挡和抗噪优势, 评测结果大幅领先次优算法, 充分验证了该方法对含遮挡、高噪场景的处理能力。

4.3.4 消融实验

表 3 给出本文方法在各区域的消融实验结果, 实验基准设为单向散焦(式(1)), 表中 I 表示仅使用单向散焦, II 表示双向散焦, III 表示双向散焦+多模板滤波, IV 表示本文完整方法, 加粗、下划线分别表示各区域的最优和次优值。本文基于双向散焦的焦堆栈度量, 充分利用光场结构信息, 拥有优秀的抗噪性能,

在非遮挡区域相比单向散焦有较大提升。但由于两者均没有遮挡处理, 因此在遮挡区域表现一般。多模板滤波可在保留图像边界的同时对图像平滑处理, 在遮挡区域可利用周围非遮挡区域的正确信息修正双向散焦的误差, 但方法同样会将遮挡区域的误差错误扩散到非遮挡区域, 因此造成该区域错误像素增多。在内联遮挡处理后, 遮挡区域错误像素大幅减少, 继而改善了多模板滤波的误差扩散, 使方法在非遮挡区域表现也有所提升。从全局表现来看, 本文各个模块均能有效改善深度估计结果, 在 $\sigma_n=10$ 的高噪等级下, 所有模块同时作用的精度较基准值提升约 15.5%。结果表明, 本文基于内联遮挡处理的深度获取方法能够有效提升含遮挡、高噪场景的深度估计精度。

表 3 坏像素 0.07 评估
Table 3 BadPix(0.07)

名称	噪声等级	非遮挡区域				遮挡区域			
		I	II	III	IV	I	II	III	IV
Cotton	$\sigma_n=1$	4.0	<u>3.8</u>	4.1	2.8	52.6	53.1	50.8	44.6
	$\sigma_n=5$	7.8	7.4	7.7	<u>7.6</u>	<u>55</u>	56.4	56.1	55
	$\sigma_n=10$	25.7	<u>23.2</u>	25.3	17.4	59.2	<u>58.7</u>	60.8	58.4
Dino	$\sigma_n=1$	0.8	0.6	<u>0.6</u>	0.5	42.1	48.6	<u>41.3</u>	29.2
	$\sigma_n=5$	2.2	<u>1.9</u>	1.8	2.0	47	46.3	<u>46</u>	43
	$\sigma_n=10$	6.4	5.0	<u>4.7</u>	4.6	51.8	54.2	<u>51.0</u>	50.3
Sideboard	$\sigma_n=1$	4.9	4.2	<u>4.2</u>	4.1	44.8	43.7	<u>42.1</u>	24.5
	$\sigma_n=5$	5.2	<u>4.9</u>	5.0	4.3	46.8	44.6	<u>44.6</u>	29.7
	$\sigma_n=10$	8.2	6.3	<u>7.3</u>	7.6	49.3	48.2	<u>47.7</u>	40.7
均值		7.3	<u>6.4</u>	6.7	5.6	49.9	50.4	<u>48.9</u>	41.7

名称	噪声等级	全局			
		I	II	III	IV
Cotton	$\sigma_n=1$	<u>7.49</u>	7.83	7.57	5.9
	$\sigma_n=5$	12.1	<u>11.9</u>	11.3	12.1
	$\sigma_n=10$	25.9	<u>25.8</u>	27.4	20.5
Dino	$\sigma_n=1$	8.59	8.2	<u>6.35</u>	4.6
	$\sigma_n=5$	8.59	8.18	<u>8.11</u>	7.8
	$\sigma_n=10$	12.1	11.9	<u>11.2</u>	11.0
Sideboard	$\sigma_n=1$	14.2	13	<u>12.6</u>	8.6
	$\sigma_n=5$	14.8	<u>13.7</u>	13.8	9.9
	$\sigma_n=10$	16.9	15.6	<u>15.2</u>	14.9
均值		13.4	12.9	<u>12.6</u>	10.3

5 结论

针对现有深度估计方法不能很好解决含遮挡、高噪场景下的深度估计问题, 设计了基于内联遮挡处理的深度估计方法。相比传统方法采用的多线索融合策略, 该方法通过将遮挡处理内联于抗噪成本量中, 在保证成本量抗噪能力的同时充分提升了其抗遮挡能力。在后续滤波优化步骤中, 通过将遮挡信息引入滤波模板, 在滤除噪声的同时较好保留了场景的边缘结构。实验结果表明, 相较于其它先进方法, 本文方法简单有效, 对场景噪声有较低的敏感度, 并能较好处理解含噪场景的遮挡问题, 更加适用于含遮挡、高噪场景的深度获取。但本文内联遮挡处理需多次计算成本量, 具有较高时间复杂度, 接下来的工作是研究更高效的内联遮挡处理方案。

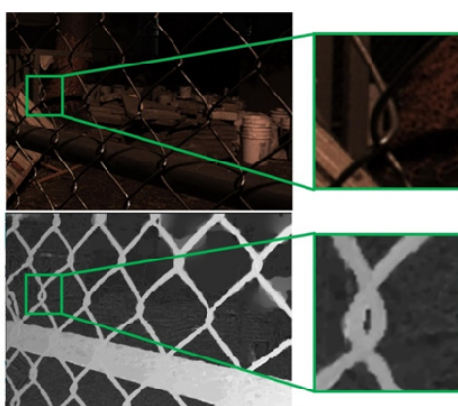
参考文献

- [1] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera[R]. Computer Science Technical Report CSTR2, 2005: 1–11.
- [2] 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.
- [3] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields[C]//2012 *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- [4] Chen J, Hou J H, Ni Y, et al. Accurate light field depth estimation with superpixel regularization over partially occluded regions[J]. *IEEE Trans Image Process*, 2018, **27**(10): 4889–4900.
- [5] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [6] 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*, 2018, **40**(10): 2484–2497.
- [7] Williem W, Park I K. Robust light field depth estimation for noisy scene with occlusion[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [8] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2013.
- [9] Wang T C, Efros A, Ramamoorthi R. Occlusion-aware depth estimation using light-field cameras[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015.
- [10] Sheng H, Zhao P, Zhang S, et al. Occlusion-aware depth estimation for light field using multi-orientation EPIs[J]. *Pattern Recognit*, 2018, **74**: 587–599.
- [11] Guo Z H, Wu J L, Chen X F, et al. Accurate light field depth estimation using multi-orientation partial angular coherence[J]. *IEEE Access*, 2019, **7**: 169123–169132.
- [12] Strecke M, Goldluecke B. Sublabel-accurate convex relaxation with total generalized variation regularization[C]//*German Conference on Pattern Recognition*, 2018.
- [13] Lin H T, Chen C, Kang S B, et al. Depth recovery from light field using focal stack symmetry[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015.
- [14] Sheng H, Zhang S, Cao X C, et al. Geometric occlusion analysis in depth estimation using integral guided filter for light-field image[J]. *IEEE Trans Image Process*, 2017, **26**(12): 5758–5771.
- [15] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4d light fields[C]//*Asian Conference on Computer Vision*, 2016.

Depth acquisition of noisy scene based on inline occlusion handling of light field

Wu Di, Zhang Xudong*, Fan Zhiguo, Sun Rui

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China



Experimental results of the real light field image

Overview: Depth estimation from multiple images is a central task in computer vision. Reliable depth information provides an effective source for visual tasks, such as target detection, image segmentation, and special effects for movies. As one of the new multi-view image acquisition devices, the light field camera makes it more convenient to acquire multiple images data. A light field camera can simultaneously sample a scene from multiple viewpoints with a single exposure, which has unique advantages in portability and depth accuracy over other depth sensors. Noise is a challenging issue for light field depth estimation. Especially for high-noise scenes containing occlusion, the simultaneous presence of occlusion and noise makes depth acquisition more difficult. For this problem, we present a light field depth estimation algorithm that is robust to occlusion and noise. The proposed method uses an inline occlusion handling framework. By integrating the occlusion handling into the anti-noise cost volume, the anti-occlusion ability of the proposed method is improved while maintaining the anti-noise performance. For the construction of the anti-noise cost volume, a focal stack matching measure based on the double-directions defocusing is proposed, which increases the defocus direction of the traditional focal stack and introduces more samples for a single match. More samples allow the algorithm to select the sample with the lowest matching cost, thereby improving the anti-noise performance. For occlusion handling, the occlusion mode in noisy scenes has greater computational difficulty. To eliminate the influence of occlusion on the focal stack and not be interfered by noise, the proposed algorithm designs view masks for different occlusion modes and constructs the cost volume respectively, and then adaptively selects the best volume according to the matching cost. After the cost volume is constructed, we use the filter-based algorithm to further smooth the cost volume. Because of the problem that traditional filtering methods cannot preserve the occlusion boundary, we design a multi-template filtering strategy. This strategy designs filters for occlusion in different directions and can better preserve the edge structure of the scene. Experiments are conducted on the HCI synthetic dataset and Stanford Lytro Illum dataset for real scenes. For quantitative evaluation, we use the percentage of bad pixels and the mean square error to measure the pros and cons of every algorithm. Experimental results show that the proposed method achieves better performance than other state-of-the-art methods for scenes where occlusion and noise exist at the same time.

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; DOI: 10.12086/oe.2021.200422

Foundation item: National Natural Science Foundation of China (61876057, 61971177)

* E-mail: xudong@hfut.edu.cn