

利用光场合成与弥散圆渲染的单幅图像重聚焦

王奇^{1,2,3}, 傅雨田^{1,2*}

¹中国科学院红外探测与成像技术重点实验室, 上海 200083;

²中国科学院上海技术物理研究所, 上海 200083;

³中国科学院大学, 北京 100049

摘要 提出一种单幅图像动态重聚焦方法, 结合基于深度学习的光场合成与基于几何结构的弥散圆渲染方法, 模拟出光场重聚焦效果。该方法仅输入单幅图像, 首先进行深度估计, 然后将深度图转换为视差, 最后在不同深度确定弥散圆直径以进行像素重采样。设计两种神经网络结构, 分别以光场相机的多视角子图和重聚焦图像为样本进行有监督的深度学习。在多个数据集和实际场景中的实验结果表明, 本文方法在可接受的计算成本下能获得优于其他方法的视觉效果与评价指标, 峰值信噪比达到了 34.55, 结构相似度达到了 0.937。

关键词 成像系统; 计算成像; 光场; 重聚焦; 深度估计; 弥散圆渲染

中图分类号 TP751

文献标志码 A

doi: 10.3788/AOS202040.0111021

Single-Image Refocusing Using Light Field Synthesis and Circle of Confusion Rendering

Wang Qi^{1,2,3}, Fu Yutian^{1,2*}

¹Key Laboratory of Infrared System Detection and Imaging, Chinese Academy of Sciences, Shanghai 200083, China;

²Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;

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

Abstract A method to dynamically refocus a single image is presented; by combining deep learning-based light field synthesis with geometric structure-based circle of confusion rendering, it simulates the light field refocusing effect. In the proposed method, the depth map is estimated and converted into disparity, and then the circle of confusion diameter is measured at different depths to resample the pixels. Two neural network structures are designed, supervised by multi-views and refocused images of the light field camera. Experiments are conducted on multiple datasets and real scenes. Compared with other techniques, the results obtained using the proposed method show superior visual performance and evaluation indicators, along with an acceptable computational cost, with the peak signal-to-noise ratio and structural similarity index reaching 34.55 and 0.937, respectively.

Key words imaging systems; computational imaging; light field; refocusing; depth estimation; circle of confusion rendering

OCIS codes 110.1758; 330.1715; 150.0155; 100.2000

1 引 言

光场相机通过在相机镜头和图像传感器之间放置微透镜阵列来捕捉 4 维(4D)光场, 以记录场景中所有光线的辐射和方向。依据光场信息可以实现重聚焦、视点变换、深度估计等多项应用^[1]。在光场的典型应用中, 需要从原始 4D 光场中计算出聚焦于

不同平面的多个视角全聚焦图像, 称为聚焦堆栈^[2]。相机阵列^[3]获取光场的方式不同, 通过相机在空间的一定排布来同时抓取一系列视角略有差别的图像, 从而重构出光场数据, 获得比全光相机更高的分辨率, 但需要复杂的软硬件设计。

光场图像经过数字重聚焦后的结果类似数字单反相机的浅景深视图效果。然而不管是光场相机还

收稿日期: 2019-08-05; 修回日期: 2019-09-04; 录用日期: 2019-11-06

基金项目: 国家 863 计划(2015AA7015090, 2015AA7015097)、国家自然科学基金(11573049)

* E-mail: yutianfu@mail.sitp.ac.cn

是数字单反相机都需要高质量的镜头与高分辨率的探测器,体积大、成本高,为解决此问题,多种方法已应用于模拟光场相机的重聚焦效果。McGraw^[4]使用2D卷积,将输入图像与它的深度图传输至低秩线性滤波器中,以达到近似散焦的效果,但是没有真实的浅景深图像作为参考,缺乏视觉效果;Zhang等^[5]使用边界模型渲染方法,仅能在前景和背景两个深度中进行重聚焦模拟;Liu等^[6]从相机阵列中获取多视角子图进行立体匹配,渲染了模糊效果,但仍需输入多个子图。

基于图像的渲染(IBR)是一种从给定的视图中渲染不同物体和场景的基本计算机图形学问题^[7]。获得多视角视图(聚焦堆栈)是光场研究中基本的IBR问题,除了全光相机和相机阵列的采样方法,还可以归结为两个方向,即基于几何结构的视图合成和基于深度学习的视图合成。基于几何结构的方法有:Cayon等^[8]通过计算每幅输入图片的几何结构,将相邻图片重新投影混叠出新的视图,从而避免繁杂的网格估计;光流反馈算法^[9]根据不同帧光流估计相机相对位置,以构建稀疏点云。基于深度学习的方法如Kalantari等^[10]输入四个角的图片,以光场真实视角图作为样本,使用深度网络单独预测每个视角的深度,以恢复出整个光场。这些方法都需要多幅图像,并且需精心的场景布置,计算成本很高。

本文提出一种基于深度学习的单幅图像模拟光场重聚焦的方法,将光场合成与弥散圆渲染相结合,

搭建网络,构建虚拟相机结构,实现单幅图像动态变焦。本文方法仅需合成 3×3 的聚焦堆栈,无需恢复整个光场,同时将深度估计转换为视差,避免了视差估计所需的立体视角匹配计算,因此大大降低了计算成本。本文方法将真实光场重聚焦视图作为训练样本,提升了峰值信噪比(PSNR)和结构相似度(SSIM)。

2 算法概述

单幅图像动态重聚焦的整体框架如图1所示,包含3方面内容:

1) 光场图像的预处理。计算子图像堆栈 $L(x, u)$,如图1(f)所示,其中 x 表示重聚焦平面坐标, u 表示主镜平面坐标;计算目标重聚焦图像,如图1(h)所示,即浅景深视图(DoF);

2) 基于聚焦堆栈的弥散圆渲染,即图1(a)~(d),首先用卷积神经网络(CNN)对全聚焦图像(输入中心图像) $I(x)$ 作单目深度估计,获得深度图映射 $D(x; I)$,然后将深度图转换为视差关系,计算亚像素偏移量,用双线性插值合成 3×3 的聚焦堆栈,之后作立体角点匹配并计算弥散圆尺寸,经过弥散圆渲染获得重聚焦图像 $S(x; I)$,再使用目标视角图(target views)对focal stack进行监督学习;

3) 基于视差的弥散圆渲染,即图1(i)~(l),直接基于深度图像 $D(x; I)$,计算弥散圆的大小,使用弥散圆渲染得到重聚焦图像 $S(x; I)$,使用目标重聚焦图像target DoFs对 $S(x; I)$ 进行监督学习。

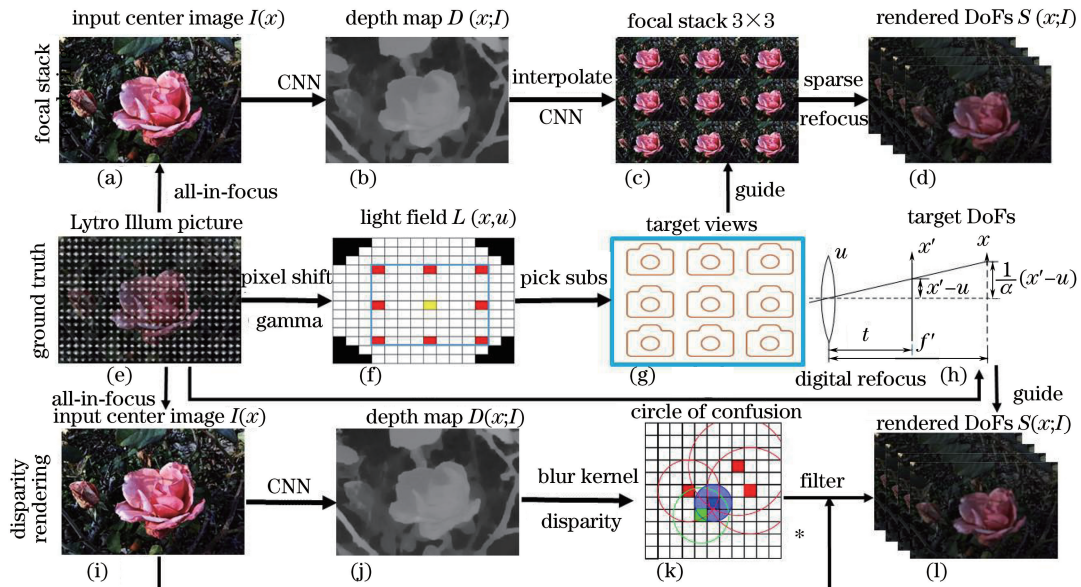


图1 单幅图像动态重聚焦算法框架

Fig. 1 Framework of single image dynamic refocusing algorithm

3 基本原理

3.1 光场数字重聚焦

Ng 等^[2]提出了数字重聚焦方法,定义 4D 光场 $L_F(x', y', u, v)$, 其中 (u, v) 表示主镜平面, (x', y') 表示传感器成像平面, (x, y) 表示合成的重聚焦平面, t 表示像距, f' 表示重聚焦平面与主镜平面的距离, 如图 1(h) 所示。引入缩放因子 $\alpha = t/f'$, 在平面 x 的重聚焦图像为

$$I(x, y) = \iint L_F\left(u, v, u + \frac{x-u}{\alpha}, v + \frac{y-v}{\alpha}\right) du dv. \quad (1)$$

重聚焦结果如图 2 所示。图 2(a) 为 Lytro Illum 拍摄的传感器图像, 该图像具有 $14 \text{ pixel} \times 14 \text{ pixel}$ 的

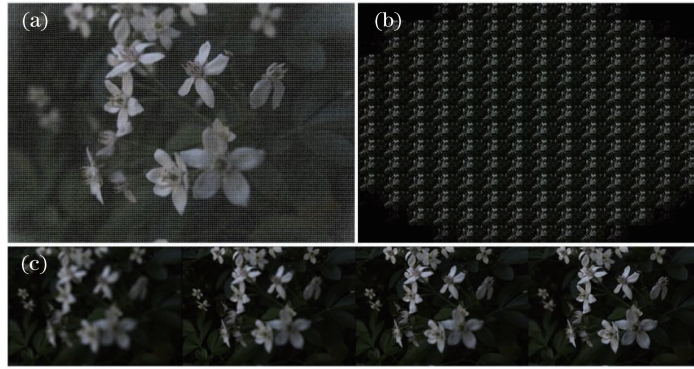


图 2 光场数字重聚焦。(a)传感器图像;(b)子孔径图像;(c)重聚焦图像

Fig. 2 Light field digital refocusing. (a) Sensor image; (b) sub-aperture image; (c) refocused image

$$2r_c = \left| D \frac{f(h-h_q)}{h_q(h-f)} \right|, \quad (2)$$

式中: r_c 为 CoC 半径; h 为物距; f 为系统焦距, 满足薄透镜方程 $1/h + 1/t = 1/f$; D 为有效孔径, 可以用 F 数(光圈数)计算得出, 即 $D = f/F$; h_q 为渲染平面与薄透镜中心距离, 此时的 CoC 为圆盘形。

在 4D 光场模型 $L_F(x', y', u, v)$ 中, 定义 γ 为深度, 对于 x 平面, $\gamma = f'$, 视差为 d , 参考背景长度为 B , 深度与视差的关系可以描述为^[12] $d = fB/\gamma$ 。(2)式中, 令 $h = \gamma_i, h_q = \gamma_p, \gamma_i, \gamma_p$ 分别为主镜平面和聚焦平面的深度, 代入视差 d , 可以得到

$$r_c = \left| \frac{D(d_p - d_i)}{2(B - d_i)} \right| = K |d_p - d_i|, \quad (3)$$

式中: d_i, d_p 分别为主镜平面和聚焦平面的视差, B, d_i, D 都为常数, $K = D/2(B - d_i)$ 为可调参数。

最后, 使用聚集法进行 CoC 渲染, 如图 3 所示, 方块 a 表示需要渲染的像素 P , 周围有四个像素 Q_i ($i=1, 2, 3, 4$) (方块 b、c、d、e), 圆圈表示围绕各个

角度分辨率以及 $520 \text{ pixel} \times 375 \text{ pixel}$ 的空间分辨率。图 2(b) 为子孔径图像, 具有 14×14 个视角。本课题组计算了 20 个深度的重聚焦图像, 获得了丰富的浅景深效果, 四个重聚焦图像如图 2(c) 所示, 缩放因子 α 分别为 $1/20, 5/20, 10/20, 15/20$, 聚焦位置依次从背景叶子变换到前景花朵上。

3.2 弥散圆渲染

将模糊渲染的方法引入到光场的重聚焦模拟中, 与以往的滤波核估计方法相比, 模糊渲染方法有很大优势, 可以对单幅图像在深度维度上进行模糊渲染。弥散圆 (CoC) 渲染的过程为: 根据深度图或者光场聚焦堆栈表示的物面相对关系, 在不同的深度处计算 CoC 直径, 渲染模糊效果, 模拟出景深视图。在薄透镜系统中, CoC 直径表示为^[11]

像素点的 CoC, 左图吸收所有四个像素的值, 而右图只吸收方块 e 的像素值。具体做法是, 按照深度排序, 选取符合条件的 k 个点 q_i ($i=1, 2, \dots, k$) 对 P 点进行聚集重采样, 计算得到 P 点渲染重采样后的像素值为

$$I(p) = \frac{1}{\omega_a(p)} \times \sum_{i=1}^k [I(q_i) \times \alpha_c(q_i)], \quad (4)$$

式中: $I(p), I(q_i)$ 为像素值; $\alpha_c(q_i) = 4/r_{q_i}^2, r_{q_i}$ 为 q_i 的 CoC 半径; $\omega_a(p) = \sum_{i=1}^k \alpha_c(q_i)$ 。

需要说明的是, 不同相机系统在相同离焦量下的弥散圆尺寸是不同的, 本文采用的深度是一个相对量, 而实际弥散圆尺寸取决于具体的系统和成像距离(深度的绝对量), 不同的绝对量模糊效果是完全不同的。在具体的、复杂的成像系统与拍摄场景中, 需要对本文的建模方法进行改进, 完善渲染方法, 从而符合实际成像模型或者观察者的直观理解。

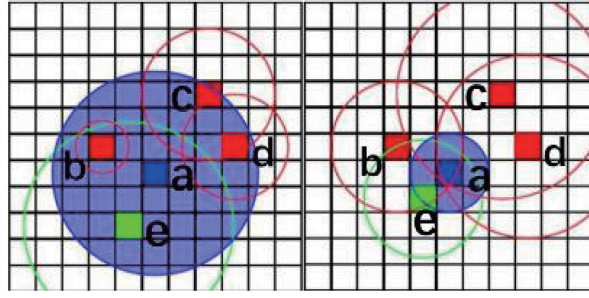


图 3 聚集法 CoC 渲染原理

Fig. 3 Principle of CoC rendering by gathering method

3.3 单目深度估计和 3×3 聚焦堆栈合成

针对单幅图像的动态重聚焦, 本文从两个角度进行研究。采用先进的对抗生成网络的方法, 设计两种如图 4 所示的神经网络架构, 该网络与图 1 描述一致。首先使用编码器(encoder)将输入中心图像 $I(x)$ 映射到高维分布 z , 再经过生成网络(GN)上采样为深度图(depth), 然后分别经光场重建与弥散圆渲染获得 3×3 的聚焦堆栈和重聚焦图像, 最后分别将合成的图像(聚焦堆栈和重聚焦图像)与目标图像(原始光场聚焦堆栈和原始重聚焦图像)送入判别网络(DN)进行对抗式训练。图 4(a)为基于聚焦

堆栈的渲染方法, 该方法使用原始聚焦堆栈监督学习, 并使用 Srinivasan 等^[13]方法中的光场重建方法, 即像素偏移与双线性插值, 在频域进行 shearlet transform, 计算视差一致性, 经对抗 GN 训练得到 3×3 的聚焦堆栈, 由聚焦堆栈计算物面相对关系, 以优化视差, 计算弥散圆大小, 最后对像素重采样获取模糊渲染图像, 以此模拟出重聚焦图像。图 4(b)为基于视差的方法, 该方法使用光场重聚焦图像进行监督学习, 直接将深度信息转换为视差, 计算弥散圆大小并进行弥散圆渲染。

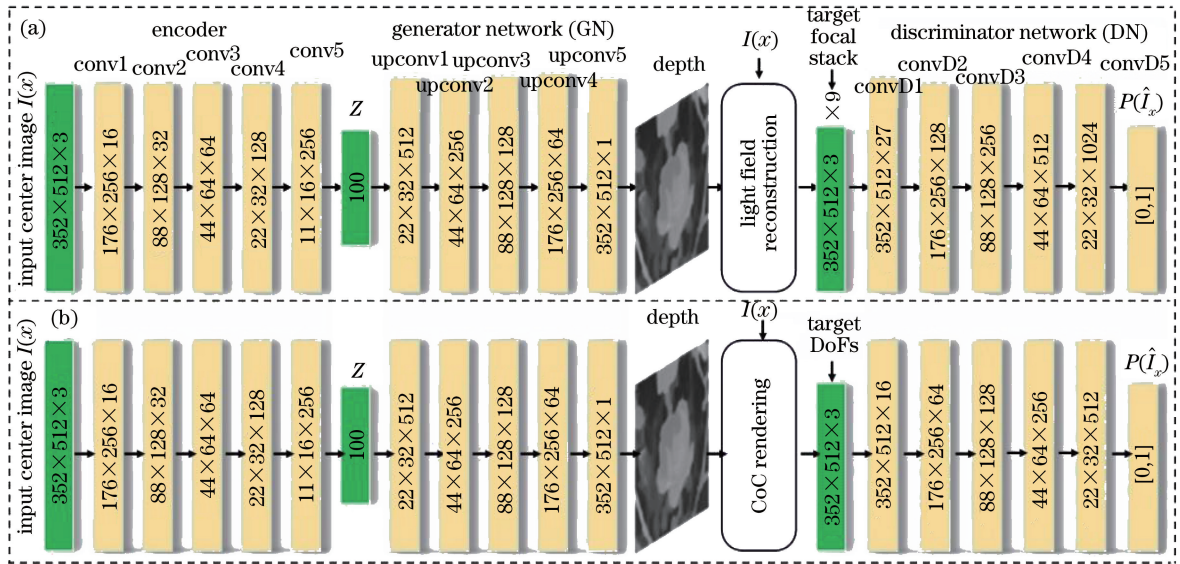


图 4 深度网络架构。(a)基于聚焦堆栈的方法;(b)基于视差的方法

Fig. 4 Deep network architecture. (a) Focused stack based method; (b) parallax based method

本文方法将总损失函数 L_{total} 结合对抗损失 L_{GAN} 与内容损失 L_{cont} , 最终的优化目标为

$$L_{total} = L_{GAN} + L_{cont} \quad (5)$$

DN 的作用是在生成图像(采样于分布 P_z)与真实图像(采样于数据分布 p_{data})对抗之间判别, 两个网络架构中的 p_{data} 分别为真实聚焦堆栈和真实重聚焦图像的分布, 对抗损失 L_{GAN} 为

$$L_{GAN} = \min_{L_G} \max_{L_D \sim R} V(L_G, L_D) = E_{I \sim P_{data}} \{ \log [L_D(I)] \} + E_{z \sim P_z} \{ \log (L_D [L_G(z)]) \}, \quad (6)$$

式中: $L_D(I)$ 为 DN 输出; $L_G(z)$ 为 GN 输出; $V(L_G, L_D)$ 为 GN 与 DN 的距离; DN 使 $V(L_G, L_D)$ 最大化, 从而增大对抗损失; GN 使 $V(L_G, L_D)$ 最小化,

从而减小对抗损失。

将生成的图像内容 S_s 与原始数据内容 S_{data} 的 L1 范数作为内容损失,即

$$L_{cont} = \min \sum \| S_{data} - S_s \|_1, \quad (7)$$

式中:对于基于聚焦堆栈的方法, S_{data} 为原始聚焦堆栈, S_s 为合成的聚焦堆栈;对于基于视差的方法, S_{data} 为原始重聚焦图像, S_s 为重聚焦渲染图像。

4 实 验

4.1 实验平台和数据集

实验平台硬件配置为 3.41 GHz CPU (Intel Core i7-6700), 24 GB RAM, GPU 采用 NVIDIA GeForce GTX-1070ti, 环境配置为 Python3.6.8, TensorFlow1.9, CUDA9.2。

选取 4 个公开光场数据集,即 Stanford^[14]、UCSD^[10]、Flower^[13]、EPFL^[15],如表 1 所示。四个数据集都由 Lytro 光场相机拍摄,Stanford 包含 cars、vegetable、buildings 等 12 个场景,EPFL 拍摄 people、studio、light 等 10 个场景,UCSD 获取 bicycle、cars、trees、foliage 四个场景,Flower 拍摄种类丰富的花草,具有不同的姿态与动态范围。

表 1 光场数据集参数

Table 1 Parameters for light field datasets

Dataset	Quantity	Resolution	Format
Stanford	720	14×14×540×375	LFR/mat/npz
UCSD	100	14×14×540×372	PNG
Flower	3343	14×14×540×372	PNG
EPFL	118	14×14×552×383	LFR/mat/npz

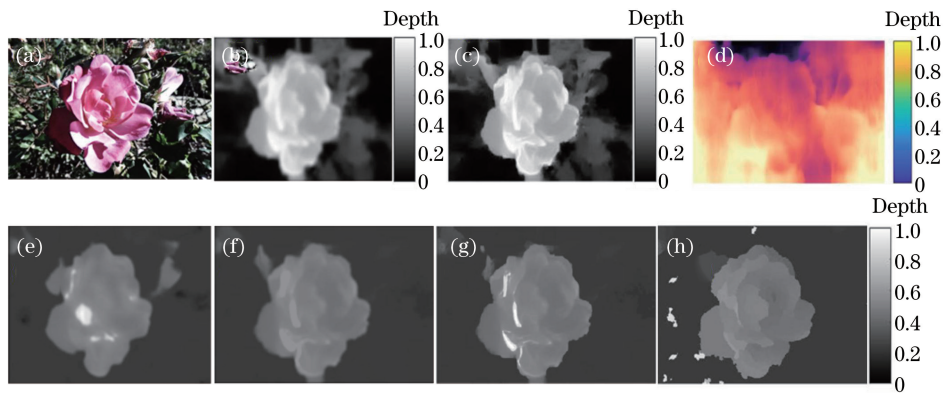


图 5 单目深度估计方法结果比较。(a)中心图像;(b)视差方法;(c)聚焦堆栈方法;(d) Godard 等^[16]的方法;(e) Cheng 等^[17]的归一化方法(SSIM:0.756);(f)归一化的视差方法(SSIM:0.823);(g)归一化的聚焦堆栈方法(SSIM:0.895);(h) Jeon 等^[18]的方法

Fig. 5 Comparison of monocular depth estimation results. (a) Center image; (b) disparity method; (c) focal stack method; (d) method proposed by Godard et al.^[16]; (e) normalized method proposed by Cheng et al.^[17] (SSIM:0.756); (f) normalized disparity method (SSIM: 0.823); (g) normalized focal stack method (SSIM: 0.895); (h) method proposed by Jeon et al.^[18]

4.2 深度估计

采用最先进的生成对抗网络无监督单目深度估计方法,该方法将真实光场子孔径图像或者真实重聚焦的景深图作为样本,进行监督学习,能获取丰富的立体角度信息,可以学习到基于光场的深度估计函数表达。将本文方法与多种单目深度估计方法进行对比,如图 5 所示,色度条的数值表示归一化的深度值,0 为背景,色度条数值越大深度值越小。如图 5(d)所示,Godard 等^[16]的方法使用了无监督学习,在 KITTI Benchmark 数据集上表现很好,但是在本文 Flower 数据集上很难辨认深度变化。如图 5(e)所示,Cheng 等^[17]的方法采用了雷达测距原理,使用线性传播卷积网络学习一个深度的密集采样矩阵,并进行单目深度估计,可以较好地表现物体深度关系。本文的两种深度网络(图 5(b)基于视差的方法与图 5(c)基于聚焦堆栈的方法)中作为中间变量的深度估计结果都较好地表现了不同物体的深度关系,聚焦堆栈的方法纹理较为清晰,深度变化更为明显。为了便于定量分析,以图 5(h)中 Jeon 等^[18]的光场深度估计方法作为 ground truth 参考图像,经归一化后,计算各种算法的深度估计结果与 Jeon 等^[18]的结果之间的 SSIM。Jeon 等^[18]的方法是通过原始 4D 光场中 81 个子图,经复杂的视角匹配计算得出深度估计结果,属于多目深度估计,对 81 张子图迭代 30 次耗费 30 min 以上可获得真实的光场深度信息,具有较高的参考价值。从图 5 的结果可以看出,基于归一化的聚焦堆栈的方法的 SSIM 达到了 0.895,有效地估计了单幅图像的深度。

4.3 聚焦堆栈合成

本文聚焦堆栈合成方法结合了生成对抗网络的细节优化特点和在频域处理高维运动信息的优点,可以很好地学习到像素偏移信息,进行细节修复。从图 2 可以看出,光场原始子孔径图像有缺失,需选取有效图像,对于 14×14 的图像,选取 $[4, 4]$ 、 $[4, 8]$ 、 $[4, 12]$ 、 $[8, 4]$ 、 $[8, 8]$ 、 $[8, 12]$ 、 $[12, 4]$ 、 $[12, 8]$ 、 $[12, 12]$ 这 9 个子图进行监督学习。将 Kalantari 等^[10] 和 Srinivasan 等^[13] 的方法进行对比,采用 PSNR 和 SSIM 两个评价方法进行分析,视觉效果与指标参数如图 6 所示。Kalantari 等^[10] 的方法输入具有物理结构关系的四个角的子图像,子图像包含大量的光场信息,属于多视图光场重建,而本文方法与 Srinivasan 等^[13] 的方法都属于

单视图光场合成。Kalantari 等^[10] 的方法通过密集采样重新渲染子视图,可学习到轮廓与视差信息,光场恢复效果很好,但是渲染时间很长,渲染一幅子图需要 70 s 以上,而本文方法只需 10 s。Srinivasan 等^[13] 的方法使用中心视图,经深度估计、遮挡物渲染、插值、像素偏移和卷积操作,最终重建出光场。而本文方法是重聚焦渲染,仅需 9 个视图,不需要恢复整个光场,大大降低了运算成本。从图 6 中可以看出:本文方法评价指标高于 Srinivasan 等^[13] 的方法,并接近 Kalantari 等^[10] 的方法;相比于 Kalantari 等^[10] 方法,本文方法仅需输入单幅图片,并且不需要仔细调整拍摄位置与拍摄角度,数据获取方式简单、计算成本小、合成速度快。

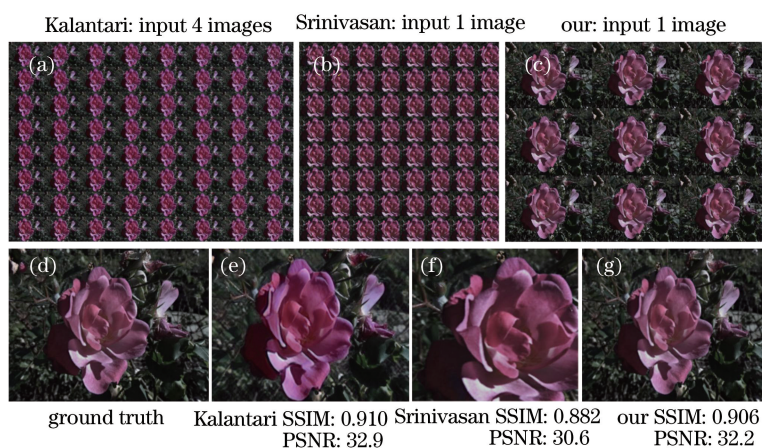


图 6 光场合成方法结果定性与定量对比。(a) Kalantari 等^[10] 的方法;(b) Srinivasan 等^[13] 的方法;(c) 本文方法;(d) 真实中心图像;(e)(f)(g) Kalantari 等^[10] 的方法、Srinivasan 等^[13] 的方法和本文方法渲染的中心图像

Fig. 6 Quantitative and qualitative comparison of light field synthesis methods. (a) Method proposed by Kalantari et al.^[10]; (b) method proposed by Srinivasan et al.^[13]; (c) our method; (d) ground truth center image; (e)(f)(g) rendered center images obtained by methods proposed by Kalantari et al.^[10] and Srinivasan et al.^[13], and our method

4.4 弥散圆渲染

首先在 Flower 数据集上进行实验。使用重聚焦方法计算 20 个深度的重聚焦图像作为 ground truth,用于计算渲染结果的 SSIM 和 PSNR。基于聚焦堆栈的方法深度估计精度更高,可进一步提高视差精度,其渲染结果优于直接的视差方法,但是需要大量的插值计算。基于视差渲染的方法实现简单,效果接近聚焦堆栈方法。为得到更好的结果,在网络架构上进行改进,便于开发端到端的应用。

将本文两个方法与 Zhang 等^[5] 的边界模型渲染方法和 Wang 等^[19] 的相机阵列方法进行对比,如图 7 所示。Zhang 等^[5] 的方法可以检测到前景和背景,可以变换聚焦和离焦对象,但由于缺乏深度信

息,不能扩展聚焦和离焦范围。Wang 等^[19] 的方法使用真实的相机阵列子孔径图像,再经立体匹配方法与视差校正优化后,可取得较好的评价指标,但该方法需要获取精心布置的相机阵列图像,比 Kalantari 等^[10] 的方法复杂。图 7 表明:聚焦堆栈的方法在渲染效果上最优,取得了最高的 SSIM 与 PSNR 指标,分别为 0.937 和 34.55;基于视差的方法是聚焦堆栈方法的简化与拓展,可进行优化,通过改进网络架构可快速、端到端地实现接近聚焦堆栈方法的效果。由图 5~7 的结果可以看出,聚焦堆栈合成与弥散圆渲染的结果取决于深度估计的结果,而本文将弥散圆渲染结果作为监督信号,通过神经网络促进了深度估计结果的优化。

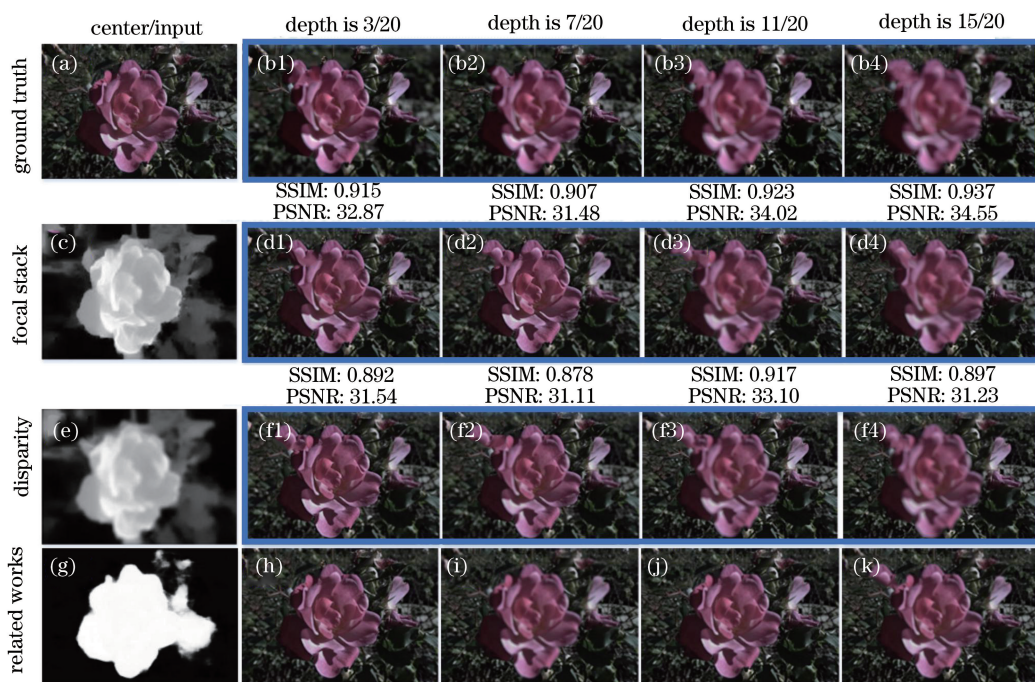


图 7 渲染方法定性与定量对比。(a)中心图像;(b1)~(b4)重聚焦图像;(c)聚焦堆栈方法深度估计;(d1)~(d4)聚焦堆栈方法的渲染结果;(e)视差方法深度估计;(f1)~(f4)视差方法的渲染结果;(g)(h)(i) Zhang 等^[5]的方法,分别为遮挡物检测、前景聚焦、背景聚焦;(j)(k) Wang 等^[19]的方法,分别为深度为 3、SSIM 为 0.902、PSNR 为 32.80 和深度为 3、SSIM 为 916、PSNR 为 33.28

Fig. 7 Quantitative and qualitative comparison of rendering methods. (a) Center image; (b1)–(b4) refocused images; (c) depth estimation using focal stack method; (d1)–(d4) rendering results using focal stack method; (e) depth estimation using disparity method; (f1)–(f4) rendering results using disparity method; (g) (h) (i) results of occlusion detection, focus on foreground, and focus on background obtained by method proposed by Zhang et al. ^[5]; (j)(k) results obtained by method proposed by Wang et al. ^[19]. Depth is 3, SSIM is 0.902, and PSNR is 32.80; depth is 3, SSIM is 916, and PSNR is 33.28

然后在其他数据集上进行实验,使用基于视差的渲染方法,渲染效果如图 8 所示,方框区为部分模糊渲染的位置。同 Flower 数据集一样,以真实重聚焦图片作为 ground truth 参考,计算 20 个

重聚焦模拟图像与真实重聚焦图像的平均 SSIM 与平均 PSNR,如表 2 所示,结果表明,使用弥散圆渲染可以方便地对不同的场景进行圆盘状的模糊渲染。

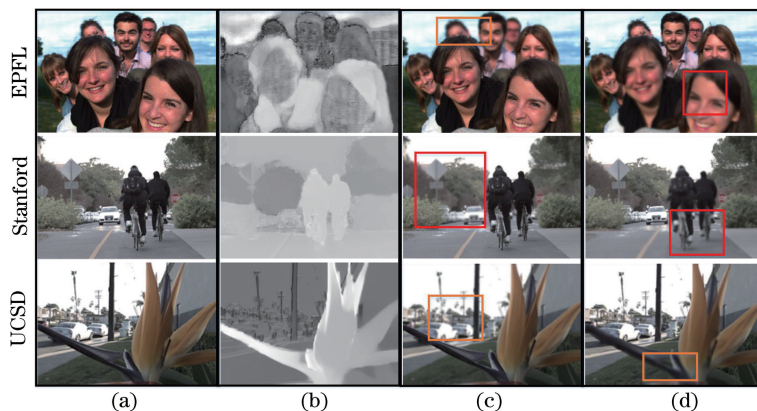


图 8 不同数据集渲染结果的对比。(a)真实中心图像;(b)深度估计;(c)重聚焦于较近位置;(d)重聚焦于较远位置
Fig. 8 Comparison of rendering results on different datasets. (a) Ground truth center image; (b) depth estimation; (c) refocusing on close positions; (d) refocusing on far positions

表 2 光场数据集渲染效果定量分析
Table 2 Quantitative analysis on rendering effects on light field datasets

Dataset	SSIM	PSNR
Stanford	0.897	30.72
UCSD	0.912	32.11
Flower	0.923	32.89
EPFL	0.901	31.03

最后,在真实拍摄的场景中进行实验,对比了 Canon 6D f/1.2 大光圈数字单反相机和 iPhone 双摄像头相机的效果,如图 9 所示,方框处为部分模糊

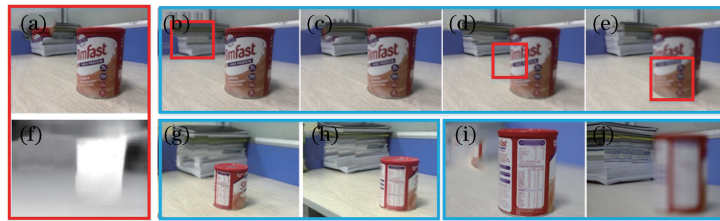


图 9 真实场景下的渲染效果与不同相机获取的图像对比。(a)原图;(b)~(e)本文方法渲染的四个不同深度的重聚焦图像;(f)深度图;(g)(h)双摄像头聚焦于两个位置的成像;(i)(j) Cannon 相机聚焦于两个位置的成像
Fig. 9 Comparison of rendering effects of real scenes with images captured by different cameras. (a) Original image; (b)~(e) refocused images at four depths rendered by our method; (f) depth map; (g)~(h) images shot by dual cameras focused on two positions; (i)~(j) images shot by Cannon focused on two positions

5 结 论

提出一种基于光场合成与弥散圆渲染的单幅图像重聚焦的方法,该方法可以从传统相机的单幅图片中模拟光场重聚焦效果,不需要深度传感器或立体视觉设备的 4D RGBD 数据。使用光场相机数据集进行监督训练,两种神经网络结构在可接受的计算成本上可实现较好的视觉效果,综合评价指标优于其他景深渲染方法。基于聚焦堆栈的方法更注重高性能与视觉效果,基于视差的方法在计算成本与网络设计上具有优势,易于开发端到端的应用。本方法需要在计算成本与网络架构上进一步优化,使基于薄透镜的渲染泛化到其他精密光学元件中。

参 考 文 献

- [1] Zhang X, Li C. Calibration and imaging model of light field camera with microlens array [J]. Acta Optica Sinica, 2014, 34(12): 1211005.
张旭, 李晨. 微透镜阵列式光场成像模型及其标定方法[J]. 光学学报, 2014, 34(12): 1211005.
- [2] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera[R]. USA: Stanford Tech Report CTSR, 2005.
- [3] Wilburn B, Joshi N, Vaish V, et al. High performance imaging using large camera arrays [J]. ACM Transactions on Graphics (TOG), 2005, 24(3): 765-776.
- [4] McGraw T. Fast bokeh effects using low-rank linear filters [J]. The Visual Computer, 2015, 31(5): 601-611.
- [5] Zhang W, Cham W K. Single image focus editing [C] // 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, September 27-October 4, 2009, Kyoto, Japan. New York: IEEE, 2009: 1947-1954.
- [6] Liu D W, Nicolescu R, Klette R. Stereo-based bokeh effects for photography [J]. Machine Vision and Applications, 2016, 27(8): 1325-1337.
- [7] Mildenhall B, Srinivasan P P, Ortiz-Cayon R, et al. Local light field fusion: practical view synthesis with prescriptive sampling guidelines [J/OL]. (2019-05-02) [2019-09-01]. <https://arxiv.org/abs/1905.00889>.
- [8] Cayon R O, Djelouah A, Drettakis G. A Bayesian approach for selective image-based rendering using superpixels [C] // 2015 International Conference on 3D Vision, October 19-22, 2015, Lyon, France. New York: IEEE, 2015: 469-477.
- [9] Li X Z, Yang A L, Qin B L, et al. Monocular camera three dimensional reconstruction based on

渲染的位置。将普通相机拍摄的无浅景深效果的图片作为 input image,使用基于视差的模型,获取深度图,扩展了 4 个深度,如图 9(b)~(e)所示。iPhone 双摄图像如图 9(g)、(h)所示,由图可知,该方法只能在一定距离范围突出前景,虚化背景,而不能进行深度的拓展,且不能使前景模糊。单反相机光圈足够大,获得的图像模糊效果非常强,需要在拍摄时手动调节焦距,如图 9(i)、(j)所示。本文方法可以拓展多个深度,无需调焦,有效地实现了单幅图像重聚焦。

- optical flow feedback[J]. *Acta Optica Sinica*, 2015, 35(5): 0515001.
李秀智, 杨爱林, 秦宝岭, 等. 基于光流反馈的单目视觉三维重建[J]. *光学学报*, 2015, 35(5): 0515001.
- [10] Kalantari N K, Wang T C, Ramamoorthi R. Learning-based view synthesis for light field cameras[J]. *ACM Transactions on Graphics*, 2016, 35(6): 193.
- [11] Zhu S P, Qiang X F. Study on monocular vision method used for camera positioning[J]. *Acta Optica Sinica*, 2001, 21(3): 339-343.
祝世平, 强锡富. 用于摄像机定位的单目视觉方法研究[J]. *光学学报*, 2001, 21(3): 339-343.
- [12] Wang Y Q, Yang J G, Mo Y, et al. Disparity estimation for camera arrays using reliability guided disparity propagation[J]. *IEEE Access*, 2018, 6: 21840-21849.
- [13] Srinivasan P P, Wang T Z, Sreelal A, et al. Learning to synthesize a 4D RGBD light field from a single image [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2262-2270.
- [14] Dansereau D G, Girod B, Wetzstein G. LiFF: light field features in scale and depth[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 8042-8051.
- [15] Rerabek M, Ebrahimi T. New light field image dataset[C] // 8th International Conference on Quality of Multimedia Experience (QoMEX), June 6-8, 2016, Lisbon, Portugal. Switzerland: EPFL, 2016.
- [16] Godard C, Mac Aodha O, Firman M, et al. Digging into self-supervised monocular depth estimation [J/OL]. (2019-08-17) [2019-09-01]. <https://arxiv.xilesou.top/abs/1806.01260>.
- [17] Cheng X J, Wang P, Yang R G. Depth estimation via affinity learned with convolutional spatial propagation network[M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11220: 108-125.
- [18] 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 (CVPR), June 7-12, 2015, Boston, USA. New York: IEEE, 2015: 1547-1555.
- [19] Wang Y Q, Yang J G, Guo Y L, et al. Selective light field refocusing for camera arrays using bokeh rendering and superresolution [J]. *IEEE Signal Processing Letters*, 2019, 26(1): 204-208.