

结合深度线索和几何结构的稀疏光场密集重建^{*}

李 梅 张旭东 孙 锐 范之国

(合肥工业大学计算机与信息学院 合肥 230601)

摘要:光场一次成像可以同时获得空间和角度的四维信息。现有方法进行角度超分辨率重建时多用于小基线场景的光场图像,在处理大基线场景重建时存在模糊等现象,同时在光场重建过程中遮挡区域重建效果差、长距离的空间关系难以捕获。针对上述问题,提出一种结合深度线索和几何结构的稀疏光场密集重建方法。该方法采用空间金字塔池化提取多尺度特征,更好地保留了图像的纹理细节和高频信息;通过在深度估计模块的部分引入空洞卷积并进行密集连接,扩大了感受野,提高了大基线场景深度估计的精度;利用视图细化模块对图像进行优化处理,在保留视差结构的同时重建了遮挡区域。实验结果表明,本文方法较好地解决了大基线场景光场重建问题,在光场大基线场景数据集上超越了其他算法,PSNR提高了2 dB,SSIM提高了0.018,重建图像的质量均优于现有的算法。

关键词:光场;角度分辨率;深度估计;大基线

中图分类号: TP391.41; TN91 文献标识码: A 国家标准学科分类代码: 510.4050

Sparse light fields dense reconstruction combining depth cues and geometric structures

Li Mei Zhang Xudong Sun Rui Fan Zhiguo

(School of Computer and Information, Hefei University of Technology, Hefei 230601, China)

Abstract: The four-dimensional information of space and angle can be obtained by one-time imaging of light field. The existing methods are mostly used for the light field image of small baseline scene in the angular super-resolution reconstruction, and there are some phenomena such as blur when reconstructing the large baseline scenes. At the same time, the reconstruction effect of the occlusion area is poor in the process of light field reconstruction, and the long-distance spatial relationship is difficult to capture. To solve this problem, a sparse light field intensive reconstruction method combining depth clues and geometric structure is proposed. This method uses spatial pyramid pool to extract multi-scale features, which can preserve the texture details and high frequency information of images better. By introducing void convolution and dense connection in the depth estimation module, the receptive field is expanded and the accuracy of depth estimation of large baseline scene is improved. The view refinement module is used to optimize the image and reconstruct the occlusion area while preserving the parallax structure. Experimental results show that the proposed method solves the problem of large-baseline scene optical field reconstruction well, and exceeds other algorithms in the large-baseline scene data set, with PSNR increased by 2 dB and SSIM increased by 0.018. The quality of reconstructed images is superior to the existing algorithms.

Keywords: light field; angular super-resolution; depth estimation; large baseline

0 引言

光场相机不仅记录光线的空间位置信息还记录其方

向信息。光场具有多视角特性,应用于三维显示^[1]、深度估计^[2-3]、虚拟现实^[4-5]等领域。由于传感器的限制,光场图像的空间分辨率和角度分辨率之间相互制约,即在空间上获得高分辨率的同时只能获得稀疏采样的光场,而

获得密集采样的光场则要以牺牲空间分辨率代价。因此光场的角度超分辨率重建已经成为了一个重要的研究热点。

现有工作光场角度超分辨率重建包含传统方法和基于深度学习的重建方法,主要针对基线较窄的真实场景,然而在处理远距离的输入子孔径图像或大基线场景的图像时,现有方法的泛化性能较弱且应用范围较窄,重建得到的效果较差。目前在合成场景以及远距离稀疏采样的真实场景中多为大基线,其角度超分辨率重建的质量对其应用于三维重建^[6]等领域具有重要意义,因此如何对大基线场景进行稀疏光场密集重建成为了本文的研究重点,解决大基线场景的角度重建可将方法泛化到更多的光场图像。大基线场景光场重建的难点在于相比于窄基线场景输入视图之间的重叠较少,需要重建大量遮挡区域。光场一次成像可以获得多视角的图像,场景中物体相互遮挡,一些像素仅在某输入视图中出现,而在其他输入视图中被遮挡,因而无法对其搜索到匹配像素,获得的深度并不准确,由遮挡带来的深度突变在重建的合成视角中会出现空洞。通过提取多视角的特征学习如何补偿遮挡,利用输入视图中物体的遮挡关系来推出合成图像中遮挡关系发生的变化,保证重建过程中遮挡关系的变化平滑,光场的遮挡区域得以重建。当输入视图的基线变大时,因为常用的局部卷积在建模远距离关系时存在困难,重影和模糊现象变得很严重。受 Jin 等^[7]启发,针对上述问题,本文提出一种结合深度线索^[8]和几何结构的稀疏光场密集重建方法,利用深度估计预测的深度图进行扭曲获得新视角的图像,经过视图细化保持光场几何结构最终重建得到高角度分辨率的光场图像。

本文的贡献如下:

1) 采用了特征提取模块获取多尺度特征,将全局上下文信息整合到图像特征中,获得并重建了更多的纹理细节。

2) 针对普通卷积难以建模远距离关系实现大基线场景深度估计的问题,在深度估计模块采用空洞卷积扩大感受野,并且对空洞卷积进行密集连接,加强了特征传递保留了高频信息,提高了预测深度图的精度。

3) 针对图像深度不连续和扭曲后图像发生变形等问题设计视图细化模块,充分考虑不同视角像素间的相关性以及光场的视差结构,在遮挡和边缘区域取得了较好的重建效果。

1 光场角度重建的研究进展

光场角度超分辨率重建方法根据是否使用深度信息可以分为非基于深度估计的角度超分辨率重建和基于深度估计的角度超分辨率重建。

非基于深度估计的方法无需利用深度信息来进行角度分辨率重建。LFCNN^[9]首次将卷积神经网络(convolutional neural networks, CNN)引入光场的超分辨率重建,可以同时实现光场图像的空间和角度超分辨率重建。该方法在重建中间视角图像时将水平或垂直方向的两个子孔径图像输入到网络中生成新视角图像,仅利用单个方向的视角信息,而忽略了光场其他方向视角的信息。Wu 等^[10]为利用光场多视角特性利用极平面(epipolar plane image, EPI)图像的结构,将稀疏视图重建建模为基于 EPI 的角度细节恢复,平衡了空间域和角度域的信息。该方法存在的问题在于仅针对二维的 EPI 图像进行处理,没有充分利用光场的四维空间和角度信息。Yeung 等^[11]为有效利用光场数据的高维空间角度信息和多视角特性,首先使用空间角度交替卷积^[12]进行特征转换以从粗到精的方式合成新视图,然后采用 4D 卷积及引导残差学习^[13]细化中间新视图。该方法为了合成得到高质量密集光场需要多张输入子孔径图像,当输入视角减少时,重建质量明显变差。非基于深度估计的方法在没有建模场景深度的情况下合成新视图,此类方法的重建效果往往随着输入视图基线的增加而变差,而基于深度估计的方法可以处理基线较大的输入光场。

通常基于深度估计的视图合成方法分为 3 个步骤,这些方法首先利用稀疏光场视图估计输入视图的深度,然后对深度图进行扭曲变形合成目标视图,最后进行视图细化操作充分利用不同视角间的互补性,进一步优化获得高质量的角度重建图像。视图重建的质量取决于深度图估计的精度以及细化模块对视图的优化操作。Kalantari 等^[14]首先提出基于深度学习的利用深度估计实现光场角度重建的方法,在视图细化部分采用颜色估计模块对扭曲图像进行优化处理获得高角度分辨率的光场图像。但由于该方法深度估计模块预测精度较低,因此重建的光场图像质量较差。Navarro 等^[15]为进一步提高深度图预测精度,在深度估计步骤中首先采用特征提取模块提取重要特征来获得更多高频信息。此外,该方法对应于每个输入视图的 4 个视差映射,然后利用选择 CNN 从扭曲图像中选择像素进行细化。Srinivasan 等^[16]提出在只提供中心视图的情况下重建整个光场,在深度估计部分通过使用深度一致性正则化来预测 4D 光场的深度,经过扭曲操作后在视图优化部分通过残差网络进一步细化。由于输入视图少,该方法忽略了视角之间的相关性且处理复杂场景及遮挡区域时效果较差。上述方法在处理过程中使用的深度估计模块均无法捕捉长距离的空间对应关系,仅适用于处理真实场景即小基线的光场图像,对于大基线场景进行重建时效果较差。Jin 等^[7]提出一种针对大基线场景的光场重建方法,在深度估计模块利用空洞卷积实现对远距离空间关系的建模,在细

化部分利用光场混合模块融合了空间角度信息并保持了光场的几何结构即视差一致性获得最终多视角的光场图像,但缺乏对光场图像多尺度特征的提取和上下文信息利用,在纹理区域表现效果较差。因此本文针对大基线场景稀疏光场重建问题,提出了一种结合深度线索和几何结构的光场角度超分辨率重建方法,通过预测深度图以及进行扭曲重建,引入光场细化模块保持几何结构,实现大基线场景下重建出高角度分辨率光场图像。

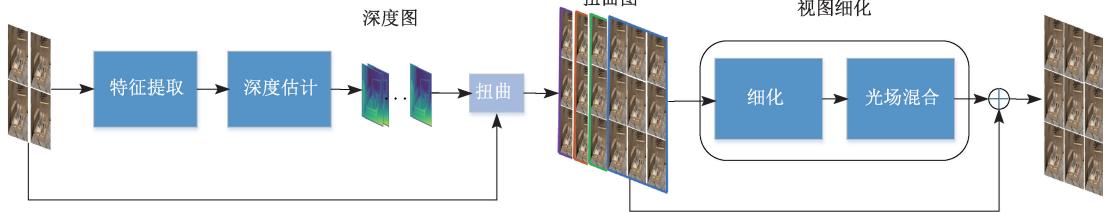


图 1 算法总体框图

Fig. 1 The algorithm framework

网络主要由 4 个模块组成:特征提取模块 (feature extraction),深度估计模块 (depth estimation),扭曲模块 (warp),视图细化模块 (view refinement)。本文算法首先从输入光场图像中获得深度图,然后对深度图进行扭曲获得新视角图像,最后以特定方式组合扭曲图像获得高角度分辨率的光场图像。具体可以分为以下步骤:首先通过提取模块提取出输入图像的多尺度特征,深度估计模块估计每个子孔径图像的深度图,然后这些深度图在扭曲模块中用于扭曲输入视图,以合成新视角。扭曲过程中不可避免带来了图像失真,因此在最后一个部分视图细化模块中,融合扭曲的新视角图像并进行进一步处理保持视差结构,获得高角度分辨率的光场图像。

4D 光场可以表示为函数 $L(x, y, u, v)$,其中 x, y 表示空间坐标, u, v 表示角坐标。在本文中将光场函数缩写为 $L(x, y)$ 。实验的目标是从稀疏采样的输入中重建出一个高角度分辨率的 4D 光场,这个问题可以表述为:

$$\hat{L}(x, u) = f(L'(x, u')) \quad (1)$$

其中, $\hat{L}(x, u)$ 是重建的高角度分辨率的光场图像,

2 结合深度线索和几何结构的稀疏光场密集重建

为了实现大基线场景下光场的角度分辨率重建,本文提出了基于深度线索和几何结构的光场重建算法,算法整体框图如图 1 所示。

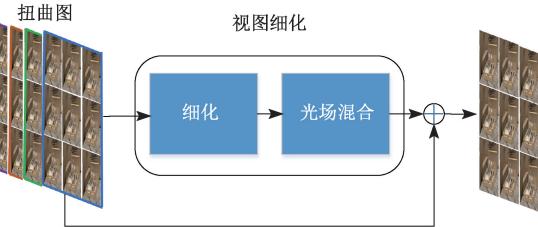


图 1 算法总体框图

Fig. 1 The algorithm framework

f 是稀疏采样的输入视图 $L'(x, u')$ 和重建得到的光场 $\hat{L}(x, u)$ 之间的对应关系。

2.1 特征提取模块

单从像素强度确定上下文关系是不准确的,因此丰富的上下文信息有利于获得图像的整体特征及结构。纹理信息是图像超分辨率重建中的重要部分,在光场重建过程中,对高频信息的提取和利用决定了光场超分辨率的结果能否还原其细节纹理特征。受 Chang 等^[17]在立体匹配网络中使用空洞空间卷积进行特征提取的启发,针对以往工作在纹理区域重建效果较差等问题,本文采用空间金字塔池化模块对输入子孔径图像进行特征提取,大小不同的池化层将特征压缩为多个尺度并进行结合获得多尺度特征^[18],加入特征提取模块后的网络能够更加准确地获得高频信息和合并层次上下文信息,对于输入图像提取纹理细节信息以及重建光场图像不同区域间结构关系具有重要意义。特征提取模块的结构如图 2 所示。

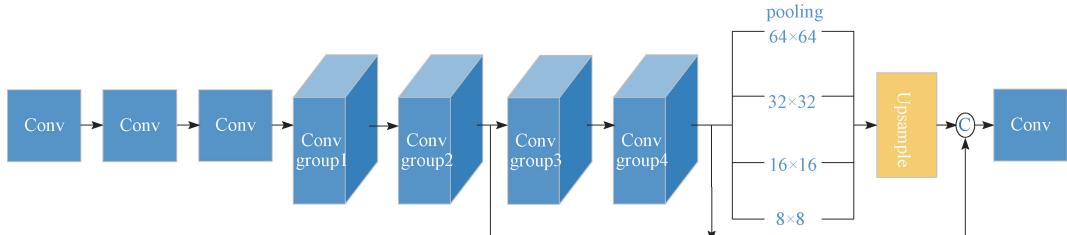


图 2 特征提取模块框图

Fig. 2 The feature extraction module framework

网络首先采用 3 个连续的 3×3 卷积级联来构建一个更深层次的网络, 在获得相同感受野的情况下, 减少训练参数, 增加网络拟合能力。Conv group1, Conv group2, Conv group3 和 Conv group4 这 4 个卷积组分别由多个卷积组成, 学习图像的深层特征, 在 Conv group3 和 Conv group4 中采用空洞卷积进一步扩大感受野。为了获得深层特征, 采用空间金字塔池化部分, 从其结构中可以看到空间金字塔池化部分使用平均池化层将特征压缩为 4 个尺度, 然后经过 1×1 卷积降低特征维数, 通过双线性插值将图像采样到与原始特征图相同的大小, 合并不同层次上下文信息, 将不同级别的特征图连接为最终的空间金字塔池化(spatial pyramid pooling, SPP)特征图, 在此过程中用到了上下文信息有助于借助环境信息丰富图像全局特征的提取。本文中将 SPP 设置为 4 个固定大小的平均池化层, 分别是 64×64 、 32×32 、 16×16 和 8×8 。特征提取模块提取了不同感受野下的特征, 并将其进行了融合得到了多尺度特征, 其中包含了重建过程中需要的纹理细节等信息。特征提取模块可以表示为:

$$F(x, u') = f_f(L'(x, u')) \quad (2)$$

其中, f_f 是输入图像与深度提取后的特征图 $F(x, u')$ 之间的对应关系。

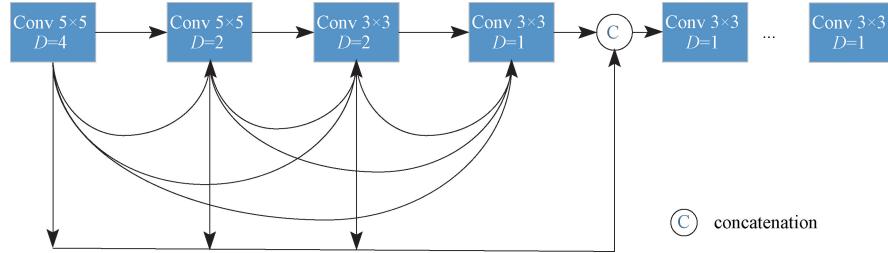


图 3 深度估计模块框图

Fig. 3 The depth estimation module framework

该模块由 9 层卷积组成, 除最后一层外, 每层卷积后都有一个 ReLU 激活层。普通卷积组成的深度估计网络无法获得长距离的对应关系, 因此为了找到视差较大的输入视图之间的对应关系进行深度估计, 网络需要有足够的感受野。深度估计模块首先采用 4 个空洞卷积, 膨胀率分别为 4、2、2、1, 卷积核大小为 5×5 、 5×5 、 3×3 和 3×3 , 后面采用 5 个 3×3 大小的普通卷积进行连接, 这样的设置为网络提供了大小为 44 的感受野, 对应视差范围为 $[-22, 22]$, 满足大基线场景的视差范围。同时为了保证深度估计过程中的细节信息不被丢失和有效利用, 采用了 4 个空洞卷积密集连接的方式。该模块通过将多个空洞卷积的结果进行密集连接, 更好地利用了各层的特征, 获得了更为准确的深度估计结果。光场图像具有视差结构一致性, 基于该特性深度估计网络可以获得深度图, 可以表示为:

$$L(x, u) = L(x + d\Delta x, u + \Delta u) \quad (4)$$

$u')$ 之间的对应关系。

2.2 深度估计模块

现有工作中光场角度超分辨率重建多针对小基线场景, 利用输入视图间的冗余信息以及光场结构一致性作为先验进行重建, 这些方法并不适用于大基线场景, 在基于深度估计的光场重建方法中光场重建的质量受到深度图预测精度的制约。本文研究针对大基线场景的角度超分辨率重建, 所以需要设计感受野足够大的深度估计模块预测深度图。本文中采用空洞卷积成功扩大了网络感受野, 能够获取长距离范围之间的空间对应关系, 适用于大基线场景的深度估计。对空洞卷积进行密集连接^[19], 加强了特征传递, 有效利用了整体的局部和全局特征, 能够更加准确预测出深度图。深度估计模块采用空洞卷积和普通卷积串联的方式, 卷积按照大的卷积核在前小的卷积核在后的方式排列, 模块的功能为估计每个子孔径图像的视差图 $D(x, u)$, 可以表示为:

$$D(x, u) = f_d(F(x, u')) \quad (3)$$

其中, f_d 是输入图像与 4D 光场深度 $D(x, u)$ 之间的对应关系。针对合成场景输入图像大基线的特点, 本文设计了如图 3 所示的深度估计模块。

其中, d 表示 $L(x, u)$ 的深度。与 Jin 等^[7]方法类似, 本文通过最小化扭曲视图与真值之间的误差来训练深度估计网络。深度估计模块的训练损失函数如下:

$$l_d = \sum_{x, u} \left(\sum_u |L(x, u) - W(x, u, u')| + \nabla_x D(x, u) \right) \quad (5)$$

2.3 扭曲模块

扭曲操作是通过输入视图中相邻像素进行插值来合成新视图的像素。由深度估计获得的深度图, 可以通过扭曲输入视图来合成新视图。扭曲可以表示为:

$$W(x, u, u') = f_w(L(x, u'), D(x, u)) = L(x + D(x, u)(u - u'), u') \quad (6)$$

其中, $W(x, u, u')$ 表示在 u' 处扭曲输入视图而产生的角度位置 u 处的新视图。

2.4 视图细化模块

从输入的光场图像扭曲得到的新视图不可避免地包

含扭曲,这包含两个原因,首先,深度估计模块难以准确预测光线深度,尤其在物体边缘或复杂区域,在无真值(ground truth, GT)监督的条件下难以解决。另外因为在输入视图中找不到原像素,扭曲模块会在遮挡区域引入误差。因此,扭曲光场的 EPI 中的线性结构可能会变形,

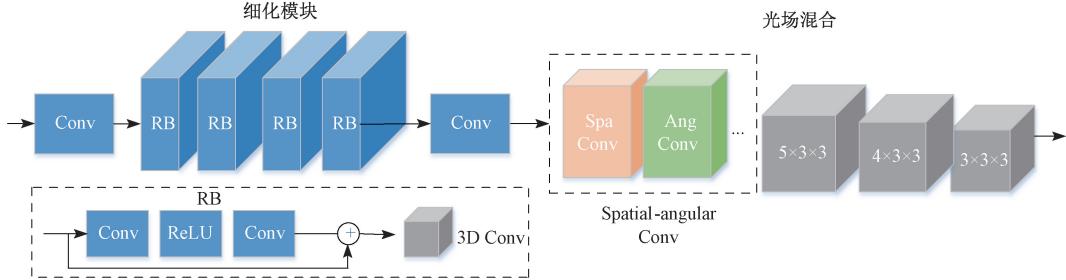


图 4 视图细化模块框图

Fig. 4 The view refinement module framework

由于扭曲视图中的像素是独立合成的,因此没有考虑它们之间的相关性即光场的几何结构,为了进一步提高光场图像的质量,提出了细化模块,通过使用 CNN 从输入子孔径图像恢复其空间相关性。细化模块的作用是将产生的新视角的视图进行进一步处理,针对深度估计不准确、边缘模糊等问题设计了该模块,包含 4 个残差块和一个卷积层,更有效地融合了各部分特征,使得到的图像更加清晰。细化模块可以有效地利用合成的子孔径图像之间的互补信息,并改善光场图像质量。光场混合模块的作用是探索合成的扭曲图像间的角度关系,以恢复光场的几何结构。为了探索光场各视图之间的关系,采用空间角度交错卷积,即在空间和角度维度上交替使用 2D 卷积层,这能够获得输入视图中空间和角度之间的关系,同时与直接使用 4D 卷积对四维光场同时提取空间角度特征相比,节约了资源减少了计算量。为了增加感受野,在空间卷积中采用空洞卷积,在空间角度交替卷积后,使用 3 层 3D 卷积来重建残差图,最后将光场图像重建为:

$$\hat{L}(x, u) = W(x, u, u'_1) + f_b(f_r(W(x, u, u'))) \quad (7)$$

其中, $W(x, u, u'_1)$ 是从第 1 个输入视图扭曲得到的光场, f_r 为细化模块对深度图的操作, f_b 为光场混合操作。视图细化模块的损失函数是 l_v , 其可以表示为:

$$l_v = \sum_{x, u} |L(x, u) - \hat{L}(x, u)| \quad (8)$$

为了保持光场的几何视差结构,采用基于 EPI 的损失函数^[12]。水平 EPI 表示为 $E_{y^*, v^*}(x, u) = L(x, y^*, u, v^*)$, 垂直 EPI 表示为 $E_{x^*, u^*}(y, v) = L(x^*, y, u^*, v)$ 。通过分析光场的结构特征,将四维光场通过固定一个角度维度和空间维度的坐标投影得到二维 EPI 图像, EPI 线条的斜率反映了与场景深度之间的关系,因此采用 EPI

并在被遮挡的区域中出现伪影等现象。视图细化模块(view refinement)主要由细化网络和光场混合^[7]部分组成,其作用在于对扭曲图像做进一步优化处理并保持其光场结构的视差一致性,生成最终的高角度分辨率的光场图像。其具体结构如图 4 所示。

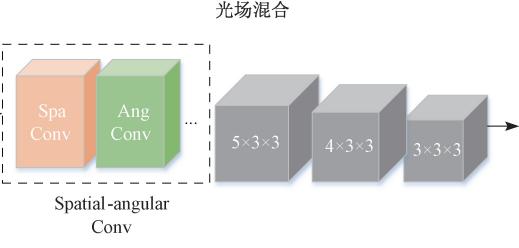


图 4 视图细化模块框图

Fig. 4 The view refinement module framework

损失函数有利于保持光场的几何结构即视差一致性。

$$\begin{aligned} l_e = & \sum_{y, v} (| \nabla_x E_{y, v}(x, u) - \nabla_x \hat{E}_{y, v}(x, u) | + \\ & | \nabla_u E_{y, v}(x, u) - \nabla_u \hat{E}_{y, v}(x, u) | + \\ & \sum_{x, u} (| \nabla_y E_{x, u}(y, v) - \nabla_y \hat{E}_{x, u}(y, v) | + \\ & | \nabla_v E_{x, u}(y, v) - \nabla_v \hat{E}_{x, u}(y, v) |) \end{aligned} \quad (9)$$

最终网络的损失函数为 $\min l_d + l_v + \lambda l_e$, 其中 λ 是 EPI 损失函数的权重。

3 实验结果与分析

本文搭建的神经网络在 Nvidia TITAN GPU 的 PC 上基于 PyTorch 框架训练,模型使用 Adam^[20]优化方法。模型的初始学习率设置为 1×10^{-4} , 每 500 个周期衰减 0.5 倍, 经过 3 000 个周期后停止训练。实验在 MATLAB 2018b 上进行数据集的预处理。

实验中采用了来自 HCI^[21]、HCI old^[22]以及 Inria DLFD^[23]3 个合成场景数据集的图像,训练集和测试集分别包含 20 和 48 张光场图像。训练集中的光场图像均由 9×9 个子孔径图像组成,空间分辨率为 512×512 ,首先对这些图像进行预处理,即从 9×9 的子孔径图像中取出 7×7 的子孔径图像,取其 4 个角的子孔径图像作为输入的 2×2 的低角度分辨率光场图像,最后经过重建得到 7×7 的高角度分辨率的光场图像,重建过程的视角图如图 5 所示。商业光场相机所拍摄的视差范围通常小于一个像素,测试集的视差范围远大于商业相机拍摄的光场图像的视差范围,实验中所用到的合成场景数据集的视差范围为 $[-4, 4]$, 2×2 光场的输入角视图之间的基线范围是视差范围的 6 倍,因此输入图像的基线范围为 $[-24,$

24]。经过深度估计模块的网络的感受野为 $[-22, 22]$ ，满足大基线光场重建的要求。在训练过程中，将光场图像随机裁剪成 96×96 的图像块作为输入，并在所有卷积层中使用零填充，以保持空间分辨率不变。从角度分辨率 2×2 的输入子孔径图像经角度超分辨率重建得到 7×7 的光场图像。

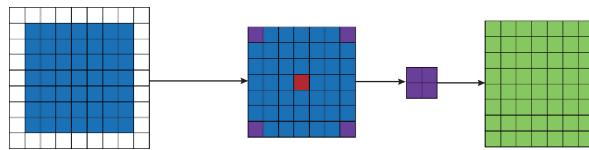


图 5 角度超分辨率重建过程

Fig. 5 The process of angular super-resolution reconstruction

本文采用峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似性 (structural similarity index metric, SSIM) 作为评价指标对实验结果进行评估^[24]，PSNR 值越大说明重建效果越好，SSIM 值越接近 1 说明重建结果越接近真值。

3.1 消融实验

本文基于光场的多视角特性和视差一致性原理，研究在大基线场景下光场的角度超分辨率效果。其中特征提取模块通过空间金字塔池化提取输入图像的多尺度特征，有利于重建出细节纹理信息；深度估计模块通过采用

空洞卷积的密接连接和普通卷积的级联扩大网络感受野，适用于大基线场景，预测得到精确的深度图；在网络的后层采用细化模块探索光场视差结构的一致性。本文以 Jin 等^[7]的网络为 baseline，探索上述 3 个模块对网络性能的影响。

为了探索本文的深度估计模块对网络的影响，添加了特征提取模块，与无特征提取模块直接进行深度估计的基线方法相比，实验结果得到了一定的提升，验证了多尺度特征提取模块有利于重建性能的提升。

深度估计模块在采用多个连续空洞卷积连接的基础上使用了密集连接，获取了更加丰富的信息。并且本文方法在 baseline 基础上改进了空洞卷积的卷积核和膨胀率的大小，使得感受野加大，能够探索出更大范围的上下文信息，对于大基线场景的光场图像更加适用。由实验结果可见，采用本文中提出的深度估计模块的重建效果优于 baseline 方法，更加准确地预测了深度图。

通过在 baseline 方法上加上特征提取模块以及改用密集连接的深度估计模块，研究同时采用以上两个模块对实验产生的影响，从表 1 中可以看出相比于单独加上特征提取模块或改变深度估计模块实验结果提升较少，同时采用特征提取和密集连接的深度估计模块共同作用有更好的实验结果，对上下文信息以及高频信息的提取利用大幅提高了重建图像的质量。

表 1 特征提取、深度估计和细化网络对实验影响 (PSNR/SSIM)

Table 1 The influence of the feature extraction module, depth estimation module and refinement (PSNR/SSIM)

特征提取模块	深度估计模块	细化模块	数据集		
			HCI	HCI old	DLFD
√			34. 60/0. 937	40. 84/0. 960	34. 83/0. 933
	√		34. 94/0. 943	41. 71/0. 967	35. 62/0. 940
√	√		34. 71/0. 938	40. 94/0. 961	35. 17/0. 935
		√	35. 83/0. 946	41. 98/0. 970	35. 84/0. 941
	√	√	35. 10/0. 946	41. 39/0. 964	35. 94/0. 939
		√	36. 35/0. 959	42. 95/0. 978	36. 21/0. 949

通过对扭曲后的图像添加细化模块进行处理，融合了各视角间的空间和角度信息，对遮挡区域以及边缘区域进行优化处理，更加准确地预测出新视角像。相比于未添加细化模块的网络，添加了细化模块的网络性能有很大的提升，验证了其有效性，细化模块对网络边缘及遮挡区域的处理以及对空间角度关系的探索有助于网络整体性能的提升。

3.2 定性实验

本文中的方法与其他先进方法进行对比，包括 Wu 等^[25]，Wang 等^[26]，Kalantari 等^[14]，Jin 等^[7]提出的方法。在 HCI, HCI old 和 India DLFD 数据集上评估了各方法的重建效果。其中除了 Wang 等^[26] 和 Wu 等^[25] 提出的方

法，其他方法均为基于深度估计的光场重建算法。在实验中本文采用相同的训练测试数据集，对实验结果进行比较。图 6 展示了本文方法与其他方法的定性比较，在 Still Life 图像中以往方法均出现了不同程度的伪影现象，在 Coffee beans and vase 图像中以往方法在边缘区域出现了变形扭曲现象，而本文方法在这些边缘以及纹理复杂的区域均表现良好。如图 6 所示，在非基于深度估计的 Wang 等^[26] 和 Wu 等^[25] 提出的方法中，重建图像存在明显伪影。在光场图像 Still Life 图像中边缘区域出现了明显的扭曲变形，在纹理信息丰富的区域，存在模糊的问题。在 Camera and brush 图像中 Kalantari 等^[14] 提出的方法由于深度估计精度较低图像出现模糊以及伪影等现

象,Jin 等^[7]提出的方法由于高频信息较少且缺少针对模糊区域的特定处理,在边缘区域出现扭曲变形等现象。本文所提出的方法在 HCI、HCI old 以及 Inria DLFID 这 3 种大基线的合成场景光场图像中均证明了有较好的性

能,在边缘区域、遮挡区域以及纹理信息丰富的区域,均有较好的视觉效果。针对边缘模糊,扭曲变形以及伪影等问题,得到了较好的解决。

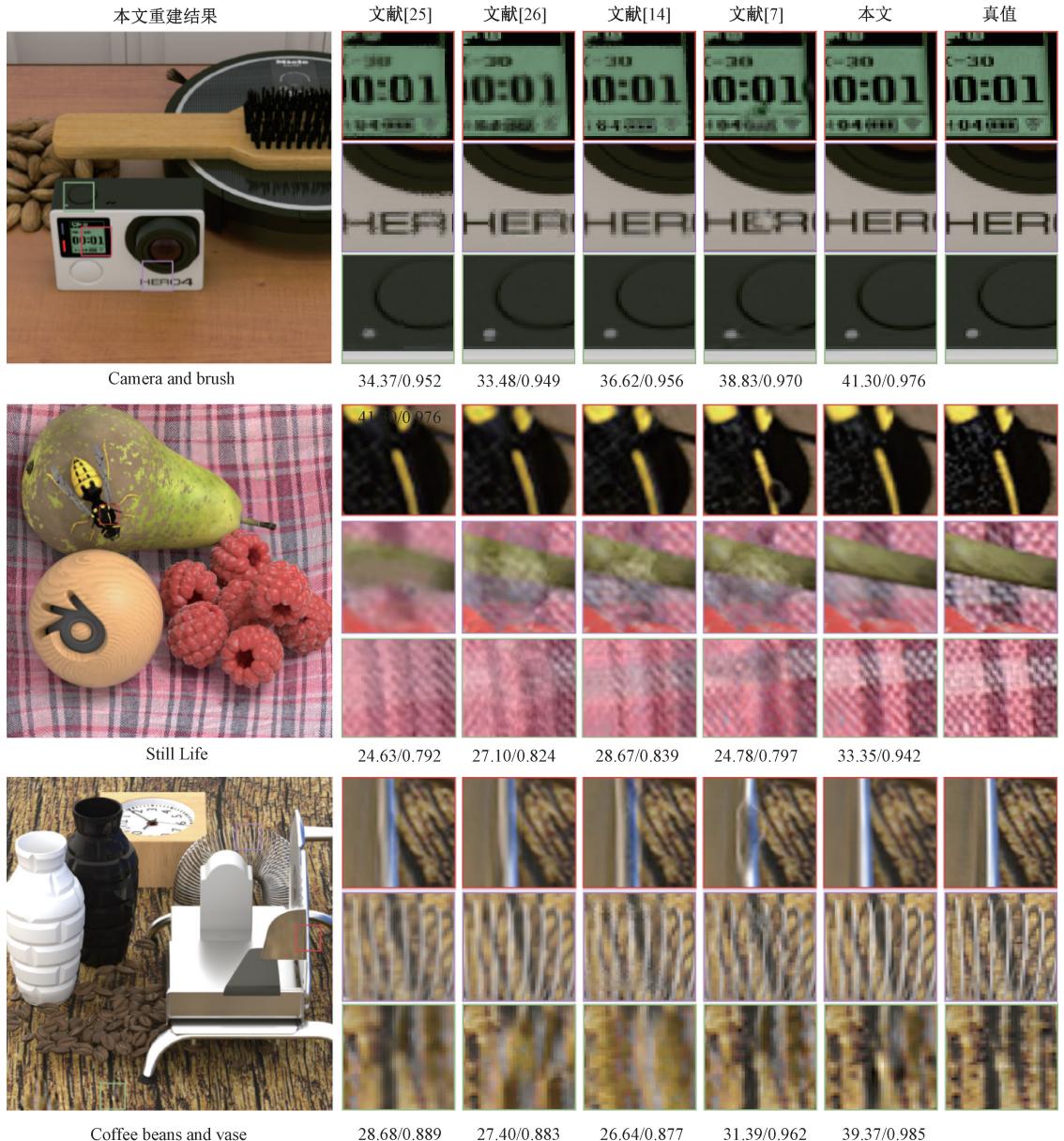


图 6 各方法在数据集上的定性比较 (PSNR/SSIM)

Fig. 6 Qualitative comparison of methods on a dataset (PSNR/SSIM)

从图 6 中可以看出,使用本文方法重建的图像在遮挡区域和边缘区域清晰,而其他方法在这些区域均出现严重的模糊或扭曲变形等现象。此外,与其他方法相比,本文的方法可以在纹理区域产生更好的高频细节^[27],这表明在大基线的条件下,本文方法可以实现高质量的光场重建。

3.3 定量实验

本文方法与其他方法在 3 个数据集上角度超分辨率重建的评价指标 PSNR 和 SSIM 对比如表 2、3、4 所示。表中给出了各图像的视差范围,相对于真实场景视差范围而言,实验中图像的视差范围较大,因此可作为大基线的光场图像进行。表 2、3、4 展示了本文方法与其他方法在 HCI、HCI old 和 India DLFID 数据集上的定量比较结

果。其中最优结果和次优结果分别用加粗和下划线表示。3 种数据集上本文方法重建结果均为最优, 指标明显高于其他先进的光场角度重建方法。在 3 种数据集上与 baseline 方法对比本文方法的重建结果有明显提升, PSNR 平均增加了 2 dB, SSIM 平均增加了 0.018, 上述结果表明本文方法在大基线场景的数据集上有优越的表现。相比于现有工作中多针对小基线场景的光场重建方

法, 本文提出的方法更加适用于大基线场景, 在定量结果上有明显提高, 相比于针对大基线光场重建的 baseline 方法, 本文在深度估计阶段针对深度图精度较低和扭曲图像的优化阶段针对变形模糊等问题进行了进一步处理, 因此本文提出的方法在大基线合成场景下取得了更好的重建结果。

表 2 各方法在 HCI 数据集上的定量比较 (PSNR/SSIM)

Table 2 Quantitative comparison of methods on HCI datasets (PSNR/SSIM)

光场图像	视差范围	文献[25]	文献[26]	文献[11]	文献[14]	文献[7]	本文
Bedroom	[-1.7, 2.2]	39.15/0.961	39.38/0.959	38.22/0.957	38.77/0.959	<u>41.98/0.975</u>	42.76/0.978
Bicycle	[-1.7, 1.7]	30.84/0.924	31.73/0.923	32.92/0.945	32.37/0.935	<u>34.03/0.954</u>	35.01/0.962
Herb	[-3.1, 1.8]	30.80/0.831	31.42/0.839	31.05/0.836	31.70/0.847	<u>32.76/0.882</u>	35.53/0.936
dishes	[-3.1, 3.5]	26.59/0.876	26.35/0.875	27.00/0.863	28.56/0.893	<u>29.63/0.938</u>	32.08/0.959
Avg. over 4 light fields		31.84/0.898	32.22/0.899	32.30/0.900	32.85/0.909	<u>34.60/0.937</u>	36.35/0.959

表 3 各方法在 HCI old 数据集上的定量比较 (PSNR/SSIM)

Table 3 Quantitative comparison of methods on HCI old datasets (PSNR/SSIM)

光场图像	视差范围	文献[25]	文献[26]	文献[11]	文献[14]	文献[7]	本文
buddha	[-0.85, 1.54]	42.91/0.986	42.67/0.985	44.03/0.988	42.47/0.985	<u>45.65/0.991</u>	46.82/0.993
buddha2	[-0.70, 1.20]	38.03/0.966	40.26/0.970	40.61/0.973	39.51/0.969	<u>41.48/0.975</u>	41.98/0.977
stilllife	[-2.71, 2.56]	24.63/0.792	25.10/0.824	24.14/0.771	24.78/0.797	<u>25.67/0.854</u>	33.35/0.942
papillon	[-1.17, 0.89]	41.42/0.981	39.89/0.978	44.73/0.986	43.04/0.983	<u>45.51/0.987</u>	46.07/0.988
monasroom	[-0.79, 0.72]	41.06/0.983	38.43/0.968	44.92/0.989	43.09/0.985	<u>45.88/0.990</u>	46.55/0.991
Avg. over 5 light fields		37.61/0.942	37.27/0.945	39.69/0.941	38.58/0.944	<u>40.84/0.960</u>	42.95/0.978

表 4 各方法在 DLFM 数据集上的定量比较 (PSNR/SSIM)

Table 4 Quantitative comparison of methods on DLFM datasets (PSNR/SSIM)

光场图像	视差范围	文献[25]	文献[26]	文献[11]	文献[14]	文献[7]	本文
Black & White	[-1.62, 0.10]	33.73/0.969	29.46/0.933	29.31/0.923	30.62/0.925	<u>34.69/0.974</u>	39.43/0.986
Furniture	[-2.06, 1.92]	36.62/0.949	36.79/0.936	38.36/0.948	36.73/0.935	<u>40.62/0.962</u>	42.20/0.970
White Roses	[-1.52, 3.38]	35.60/0.962	36.23/0.956	36.27/0.960	36.28/0.960	<u>40.59/0.981</u>	41.93/0.984
Avg. over 39 light fields		32.35/0.911	31.77/0.896	31.65/0.892	32.53/0.899	<u>34.83/0.933</u>	36.21/0.949

4 结 论

针对大基线场景光场重建边缘模糊等问题设计了结合深度线索和几何结构的稀疏光场密集重建算法。本文提出的网络框架中, 通过空间金字塔池化来提取多尺度特征, 然后经过密集连接的空洞卷积进行深度估计获得深度图, 扩大感受野探索更深层次的信息, 加强了特征传递, 经过扭曲得到新视图后使用细化网络和光场混合对边缘和遮挡等区域进一步优化处理并保持光场的视差结构, 获得最终的高角度超分辨率的光场图像。实验结果表明, 该方法在大基线的光场合成场景中表现出了优秀的性能, 在边缘、纹理和遮挡区域也有较好的表现。本文的方法仅针对特定视差范围的大基线场

景光场, 今后工作的方向可以放在设计一款更加通用的框架, 对更大基线范围的真实场景光场图像进行角度超分辨率重建。

参考文献

- [1] WANG Y, WANG L, YANG J, et al. Spatial-angular interaction for light field image super-resolution [C]. European Conference on Computer Vision. Springer, Cham, 2020: 290-308.
- [2] SHIN C, JEON H G, YOON Y, et al. Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4748-4757.
- [3] GUO C, JIN J, HOU J, et al. Accurate light field depth estimation via an occlusion-aware network [C]. IEEE

- International Conference on Multimedia and Expo (ICME), 2020: 1-6.
- [4] HUANG F C, CHEN K, WETZSTEIN G. The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues [J]. ACM Transactions on Graphics, 2015, 34(4CD): 60.
- [5] YU J. A light-field journey to virtual reality [J]. IEEE Multimedia, 2017, 24(2): 104-112.
- [6] CHANG J R, CHEN Y S. Pyramid stereo matching network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5410-5418.
- [7] JIN J, HOU J, YUAN H, et al. Learning light field angular super-resolution via a geometry-aware network [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11141-11148.
- [8] 廖畅, 马秀丽. 基于 TOF 相机的俯视行人检测 [J]. 电子测量技术, 2019, 42(6): 104-108.
- LIAO CH, MA X L, et al. People detection under an overhead time of flight camera [J]. Electronic Measurement Technology, 2019, 42(6): 104-108.
- [9] YOON Y, JEON H G, YOO D, et al. Learning a deep convolutional network for light-field image super-resolution [C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015: 24-32.
- [10] WU G, ZHAO M, WANG L, et al. Light field reconstruction using deep convolutional network on EPI [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6319-6327.
- [11] YEUNG H W F, HOU J, CHEN J, et al. Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 137-152.
- [12] JIN J, HOU J, CHEN J, et al. Deep coarse-to-fine dense light field reconstruction with flexible sampling and geometry-aware fusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 1819-1836.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] KALANTARI N K, WANG T C, RAMAMOORTHI R. Learning-based view synthesis for light field cameras [J]. ACM Transactions on Graphics (TOG), 2016, 35(6): 1-10.
- [15] NAVARRO J, SABATER N. Learning occlusion-aware view synthesis for light fields [J]. Pattern Analysis and Applications, 2021, 24(3): 1319-1334.
- [16] SRINIVASAN P P, WANG T, SREELAL A, et al. Learning to synthesize a 4D RGBD light field from a single image [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2243-2251.
- [17] CHANG J R, CHEN Y S. Pyramid stereo matching network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5410-5418.
- [18] 何群, 邵丹丹, 王煜文, 等. 基于多特征卷积神经网路的运动想象脑电信号分析及意图识别 [J]. 仪器仪表学报, 2020, 41(1): 138-146.
- HE Q, SHAO D D, WANG Y W, et al. Analysis and intention recognition of motor imagery EEG signals based on multi feature convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2020, 41 (1): 138-146.
- [19] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [20] KINGMA D P, BA J. Adam: A method for stochastic optimization [C]. International Conference on Learning Representations, 2015: 1-13.
- [21] 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, Springer, Cham, 2016: 19-34.
- [22] WANNER S, MEISTER S, GOLDLUECKE B. Datasets and benchmarks for densely sampled 4D light fields [C]. VMV, 2013, 13: 225-226.
- [23] SHI J, JIANG X, GUILLEMOT C. A framework for learning depth from a flexible subset of dense and sparse light field views [J]. IEEE Transactions on Image Processing, 2019, 28(12): 5867-5880.
- [24] 王聪, 薛晓军, 李恒, 等. 基于颜色校正和改进二维伽马函数的水下图像增强 [J]. 电子测量与仪器学报, 2021, 35(2): 171-178.
- WANG C, XUE X J, LI H, et al. Underwater image enhancement based on color correction and improved 2D gamma function [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(2): 171-178.
- [25] WU G, LIU Y, DAI Q, et al. Learning sheared EPI structure for light field reconstruction [J]. IEEE Transactions on Image Processing, 2019, 28 (7): 3261-3273.

- [26] WANG Y, LIU F, ZHANG K, et al. High-fidelity view synthesis for light field imaging with extended pseudo 4DCNN [J]. IEEE Transactions on Computational Imaging, 2020, 6: 830-842.
- [27] 田文利. 基于双重滤波与锐化的遥感图像增强算法[J]. 国外电子测量技术, 2017, 36(4): 13-16.
- TIAN W L. Remote sense image enhancement algorithm based on filtering and sharpening [J]. Foreign Electronic Measurement Technology, 2017, 36(4): 13-16.

作者简介



李梅, 2020 年于合肥学院获得学士学位, 现为合肥工业大学硕士研究生, 主要研究方向为计算机视觉。

E-mail: lm@mail.hfut.edu.cn

Li Mei received her B. Sc. degree from Hefei University in 2016. Now she is a M. Sc.

candidate in Hefei University of Technology. Her main research interest includes computer vision.



张旭东(通信作者), 1989 年于合肥工业大学获得学士学位, 1992 年于合肥工业大学获得硕士学位, 2005 年于中国科学技术大学获得博士学位。现为合肥工业大学教授, 主要研究方向为智能信息处理、机器视觉。

E-mail: xudong@hfut.edu.cn

Zhang Xudong (Corresponding author) received his B. Sc. degree from Hefei University of Technology in 1989, M. Sc. degree from Hefei University of Technology in 1992 and Ph. D. degree from University of Science and Technology of China in 2005, respectively. Now he is a professor in Hefei University of Technology. His main research interests include intelligent information processing and machine vision.