基于深度学习的光场成像三维测量方法研究

伍俊龙[1,2,3]，郭正华[1,2,3]，陈先锋[1,2,3]，马帅[1,2,3]，晏旭[1,2,3]，朱里程[1,2,3]，王帅[1,3]，杨平[1,3*]

[1]中国科学院自适应光学重点实验室，四川 成都 610209；

[2]中国科学院光电技术研究所，四川 成都 610209；

[3]中国科学院大学，北京 100049

摘要 为了解决光场相机应用于三维测量时，在弱纹理区域和精细结构区域难以获得准确视差估计结果问题，提出了基于深度学习技术对光场深度估计进行建模，并建立了光场视差与真实深度之间的转换关系。将所提方法应用于多种复杂场景中，实验结果均表明：该方法可以准确获取弱纹理区域和精细结构区域的视差信息，较好地复原场景的三维结构，视差估计处理时间压缩到 1 s 量级，相比传统的基于代价优化的方法，降低了 1～2 个数量级。

关键词 测量；三维测量；光场成像；深度估计；深度学习

中图分类号 TP2 **文献标志码** A **doi**：10.3788/CJL202047.1204005

Three-Dimensional Measurement Method of Light Field Imaging Based on Deep Learning

Wu Junlong[1,2,3], Guo Zhenghua[1,2,3], Chen Xianfeng[1,2,3], Ma Shuai[1,2,3], Yan Xu[1,2,3], Zhu Licheng[1,2,3], Wang Shuai[1,3], Yang Ping[1,3*]

[1]*Key Laboratory on Adaptive Optics, Chinese Academy of Sciences, Chengdu, Sichuan* 610209, *China*;

[2]*Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan* 610209, *China*;

[3]*University of Chinese Academy of Sciences, Beijing* 100049, *China*

Abstract To estimate the accurate disparity in weak texture region and fine structure region when the light field camera is used for three-dimensional measurement, a model of the light field depth estimation based on deep learning technology is proposed. Moreover, the relationship between the disparity and corresponding depth is also established. The proposed method is applied to a variety of complex scenes, and the experimental results show that the proposed method can accurately estimate the disparity information in the weak texture region and fine structure region, and leading to a good reconstruction of three-dimensional structure. The processing time of the proposed method is compressed to the order of 1 s, which is 1 to 2 orders of magnitude lower than the traditional methods based on cost optimization.

Key words measurement; three-dimensional measurement; light field imaging; depth estimation; deep learning

OCIS codes 120.4820; 040.1880; 100.6890

1 引 言

传统的视觉三维测量方案主要指立体视觉技术，这类技术首先通过多相机捕获场景的多视角信息，然后基于立体匹配算法估计出场景的视差[1-4]。在标定相机参数之后，立体视觉技术能方便地实现视差与深度值之间的转换。与之相比，光场三维测量采用光场相机对场景进行成像。

在光场相机参数标定方面，文献[5]基于光场相机的成像特点，提出采用线特征建立图像特征与

收稿日期：2020-06-09；**修回日期**：2020-07-13；**录用日期**：2020-07-27

基金项目：国家自然科学基金(61805251，61875203，11704382)、中国科学院青促会(2017429)

*E-mail：pingyang2516@163.com；

三维目标之间的线性方程组，并通过求解线性方程组获得相机的几何参数。文献[6]借鉴普通相机的参数标定方法，并从解码后的光场子孔径图中提取特征点。在光场视差估计方面，现有的主流方法基于对场景几何特性的分析，通过手动设计特征将视差估计转化为代价函数优化问题。这类方法通常采用光场子孔径图(SAI)或极平面图(EPI)作为算法输入[7]。例如，文献[8]首先基于子孔径图像的逐像素特征构建匹配代价和代价聚合，然后采用胜者为王(WTA)策略决策出初始视差值；为了消除初始值中的坏点，该方法进一步构建了包含数据项和平滑项的优化函数。文献[9]依据极平面图中的线段斜率与物点深度成正比的关系，提出采用自旋平行四边形算子(SPO)划分极平面图区域，通过最大化不同区域之间的分布间距估计线段斜率。总体上，这类方法[9-14]需要进行大量的计算，并且手动设计的特征难以充分表达场景的复杂特性，使其在弱纹理和精细结构区域难以获得准确结果。近年来，采用深度学习技术对视差估计进行建模的方法受到了广泛关注。文献[15]基于卷积神经网络构建光场与视差之间的映射关系，该模型以端到端方式训练，采用多幅子孔径图作为输入，输出端直接输出场景的视差。文献[16]采用神经网络模型估计极平面图中线段的斜率，该模型以极平面图作为输入，直接输出线段斜率。相比而言，基于深度学习的方法[15, 17-21]能够自动地从训练数据中学习场景的特性，且训练数据的质量和数量是这类方法的关键。由于光场成像本身的复杂特性，在实际应用中往往只能获取有限数量的训练数据[17]，极大地限制了这类模型的泛化性能，且模型难以准确处理复杂的几何结构。

为了实现准确的光场三维测量，本文首先基于光场相机的几何成像模型推导了光场视差与深度之间的转换关系。针对现有视差估计方法在复杂结构区域(弱纹理区域、精细结构)存在的困难，提出一种基于多尺度特征的光场深度估计方法。针对训练数据匮乏的问题，提出了一种针对光场数据集的数据增强方案。实验结果表明，本文方法能够准确地测量出场景的三维信息，且多尺度特征的引入显著改善了模型在复杂结构区域的处理效果。

2　基本原理

2.1　相机成像模型

如图1所示，光场相机通过安装在传感器与主镜头之间的微透镜阵列(MLA)实现光场信息编码。为了实现光场视差与场景深度之间的转换，本文首先借助光场相机的几何成像模型[5]，建立光场视差与三维物点深度之间的转换关系。

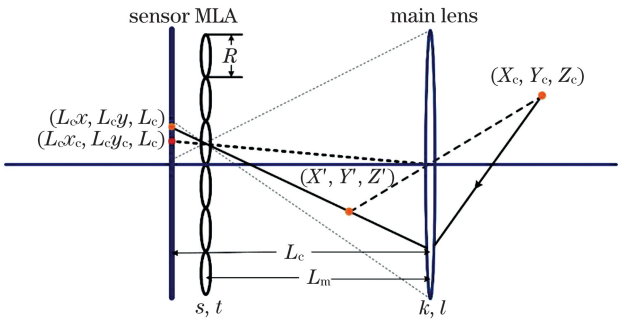

图1　光场相机成像模型示意图

Fig. 1　Projection model of light field camera

记光场相机主镜头的焦距为 F，任意三维物点 (X_c, Y_c, Z_c) 关于主镜头的像为 (X', Y', Z')。以主镜头的主点为坐标原点，记微透镜阵列与主镜头间距为 $-L_m$，传感器与主镜头间距为 $-L_c$，则该物点发出的光线在传感器平面上的投影点坐标可记为 $(L_c x, L_c y, L_c)$，与该投影点对应的微透镜投影中心可记为 $(L_c x_c, L_c y_c, L_c)$。根据光场相机的几何成像模型可得：

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}=\begin{bmatrix}x-x_c\\ y-y_c\end{bmatrix}=\frac{L_m-L_c}{(L_m-Z')L_c}\begin{bmatrix}X'-Z'x_c\\ Y'-Z'y_c\end{bmatrix}。\tag{1}$$

另记传感器单位长度内的像元数为 m，传感器的像素坐标为 (u, v)，传感器中心的像素偏移为 (c_x, c_y)。根据主镜头的物像关系，$(\Delta x, \Delta y)$ 对应的像素坐标 $(\Delta u, \Delta v)$ 将满足：

$$\begin{bmatrix}\Delta u\\ \Delta v\end{bmatrix}=\frac{1}{K_1 Z_c+K_2}\begin{bmatrix}-mL_cX_c-Z_c(u_c+c_x)\\ -mL_cY_c-Z_c(v_c+c_y)\end{bmatrix}。\tag{2}$$

其中，

$$\begin{bmatrix}u_c\\ v_c\end{bmatrix}=\begin{bmatrix}-mL_cx_c-c_x\\ -mL_cy_c-c_y\end{bmatrix},\tag{3}$$

$$K_1=-\frac{(L_m+F)L_c}{(L_m-L_c)F},\tag{4}$$

$$K_2=\frac{L_mL_c}{L_m-L_c},\tag{5}$$

式中：(u_c, v_c) 表示微透镜投影中心的像素坐标；$(\Delta u, \Delta v)$ 对应光场的角度采样坐标，若提取像平面上具有相同 $(\Delta u, \Delta v)$ 的像素，并将这些像素按照其在原图中的相对位置重新排列，得到的新图像称为光场子孔径图。光场子孔径图等效于周期排布的虚拟相机阵列所成的像。

2.2　光场视差与物点深度转换

为了说明视差与物点深度之间的关系，图 2 为光场视差与物点深度之间的关系示意图。图中 D 表示物点深度，b 表示相邻虚拟相机的光心间距，称为相邻子孔径图之间的等效基线。若记 s 为物点在子孔径图中的像点坐标，则物点 P 对应的视差可表示为

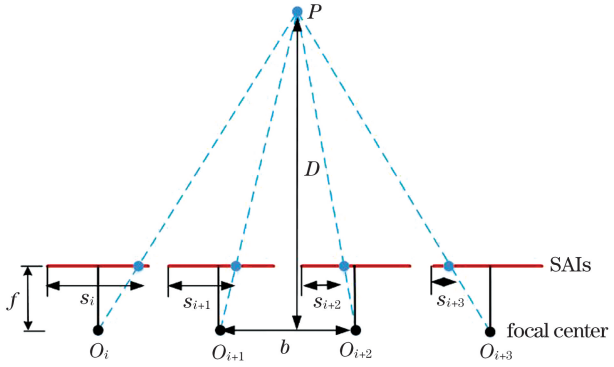

图 2　视差与深度转换关系

Fig. 2　Relationship between disparity and depth

$$d = s_i - s_{i+1}。\tag{6}$$

对于图 2 中的子孔径图，$i=1,2,3$。根据几何关系，视差 d 与深度 D 将满足

$$D = \frac{b \cdot f}{d},\tag{7}$$

式中：视差 d 由视差估计算法从光场中估计。为了确定物点的深度值，需要进一步确定等效基线长度 b 和变量 f 的取值。

如图 3 所示，对于光场相机而言，传感器与微透镜阵列之间的间距等于微透镜焦距 f_s，因而子孔径图的成像光线会聚于主透镜的主面，不同子孔径图对应的会聚点等效为虚拟相机的光心。记任意相邻子孔径图在传感器上的像素坐标分别为 u_i 和 u_{i+1}，并取主透镜主面与光轴的交点为坐标原点，则

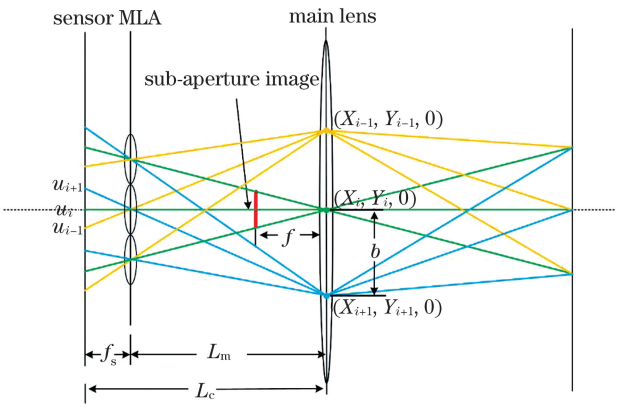

图 3　等效基线示意图

Fig. 3　Illustration of equivalent baselines

相邻子孔径图对应的等效光心坐标分别为 $(X_i, Y_i, 0)$ 和 $(X_{i+1}, Y_{i+1}, 0)$，可得：

$$b = \sqrt{(X_i - X_{i+1})^2 + (Y_i - Y_{i+1})^2}。\tag{8}$$

另外，由(2)式可知，相邻子孔径图满足

$$\begin{bmatrix}\Delta u_{i+1}\\ \Delta v_{i+1}\end{bmatrix} = \begin{bmatrix}\Delta u_i\\ \Delta v_i\end{bmatrix} + \begin{bmatrix}1\\ 0\end{bmatrix},\tag{9}$$

或

$$\begin{bmatrix}\Delta u_{i+1}\\ \Delta v_{i+1}\end{bmatrix} = \begin{bmatrix}\Delta u_i\\ \Delta v_i\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}。\tag{10}$$

联立(2)式、(8)式、(9)式或(10)式可得

$$b = \frac{-L_m}{m(L_m - L_c)},\tag{11}$$

式中：L_c 表示传感器与主镜头间距。

为了求取变量 f 的值(图 2 和图 3)，记传感器的尺寸为 L_s，微透镜的半径为 R，那么

$$\frac{L_s}{L_s/2Rm} = \frac{-L_m}{f}。\tag{12}$$

因而

$$f = \frac{-L_m}{2Rm}。\tag{13}$$

结合(7)式、(11)式和(13)式，即可实现光场视差与物点深度值的转换。

3　模型设计

3.1　多尺度特征提取与融合

本文采用的网络结构如图 4 所示，网络整体分三部分组成：初始特征提取、多尺度特征提取和特征融合网络。在输入端，初始特征提取部分旨在提取场景中丰富的细节信息，该部分采用四条支路分别处理不同方向的输入。卷积操作采用 3×3 的卷积核，步长和 padding 均设置为 1。因此，初始特征提取网络的每个模块保持输入与输出分辨率一致。

在初始特征的基础上，多尺度特征提取部分由两条支路构成。首先，为了确保模型能有效处理场景中的细节，第一条支路采用多个连续卷积层提取逐像素特征。为了改善模型在复杂结构区域的处理效果，第二条支路基于 U-net[22] 构建多尺度网络，并在 U-net 结构的基础上进行如下改进：1)移除 U-net 第一级，仅使用 4 级解码与编码结构；2)每一级由两个卷积层组成；3)与 U-net 逐层降低特征图分辨率不同，每一个卷积层的输入与输出保持分辨率一致。在编码部分，相邻两级之间通过最大池化降低特征图分辨率，以增大后续网络层的感受野，实现更大尺度范围的特征提取。编码网络的各级

特征图输出逐级增加，后一级是前一级的 2 倍。解码部分在结构上与编码部分相互对称。不同的是，解码网络通过上卷积操作逐级增大输出特征图的分辨率，将其前一级上卷积输出与同级编码网络输出相叠加，作为后一级的输入。网络的最后一部分为特征融合网络，且模型最后一层采用 1×1 卷积操作处理融合后的特征，以回归的方式获得最终的视差。在模型训练过程中，以网络模型的输出与真值（GT）视差之间的误差作为模型参数更新的依据，因此模型参数表征了从输入视图到场景视差的转换过程。因而，训练好的模型能够学习到输入视图与场景视差之间的映射关系。在测试阶段，网络模型基于训练得到的模型参数，实现由输入直接估计出场景视差的过程。

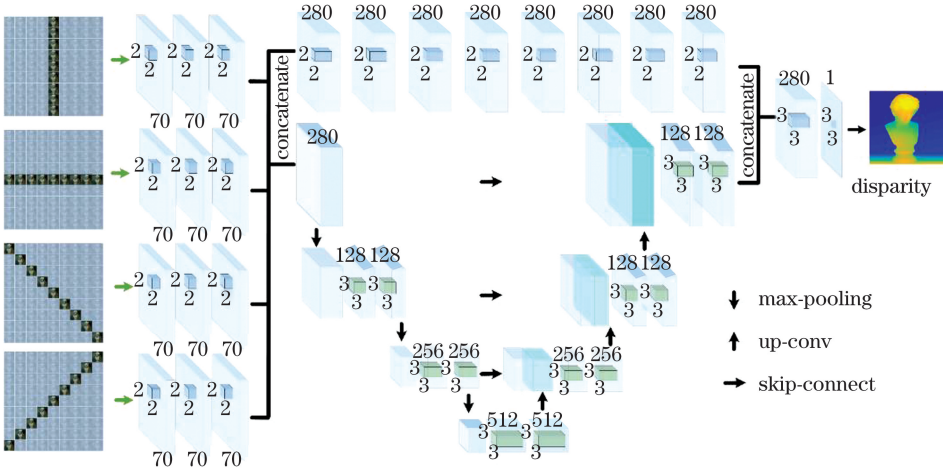

图 4　网络结构

Fig. 4　Structure of network

3.2　模型输入

上述网络模型采用多幅光场子孔径图作为输入，通过光场解码[6]可从光场相机采集的原始图像中获得如图 5 所示的子孔径图。图中给出了角度采样数为 9×9 的光场所对应的全部子孔径图。理论上，模型输入采用的子孔径图数量越多，越可能获得更高的准确率。然而，如果直接将光场所有的子孔径图作为输入，那么网络的整体规模和参数量势必会很大，这将加重模型训练和推理的负担。出于降低模型复杂度并确保模型准确率的考虑，本文选取一部分子孔径图代替整个光场作为模型输入。具体而言，本文选取沿对角方向、横向和纵向等 4 个方向的子孔径图的组合作为输入，如图 5 所示。

为说明上述被选取的子孔径图与视差之间的关系，将光场表示为 $F(s,t,k,l)$，其中 (s,t) 表示光场的空间坐标，(k,l) 表示光场的角度坐标（图 5）。另外，记光场的视差图为 $d(s,t)$，中心位置的子孔径图为 $F(s,t,k_0,l_0)$，那么中心子孔径图与其余的子孔径图之间满足

$$F(s,t,k_0,l_0)=$$
$$F[s+d(s,t)\times k,t+d(s,t)\times l,k,l],\quad(14)$$

式中：$k\neq k_0$；$l\neq l_0$。需要说明的是，上述输入视图的选取策略，一方面保证了输入能覆盖整个光场的范围，减少信息丢失；另一方面，被选取的子孔径图沿角度坐标相互对称，这有利于卷积神经网络从输入中学习到场景的几何特性。

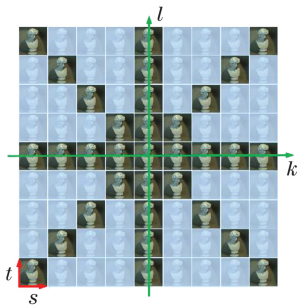

图 5　网络输入示意图

Fig. 5　Illustration of network input

为了避免模型在有限的训练数据上过拟合，本文对训练集进行数据增强处理。对每一幅光场随机进行如下处理：

1）旋转变换。将光场随机旋转 90°、180°、−90°。这些特定的旋转角度不改变光场角度采样所对应的视角，且模拟了光场相机绕光轴旋转。

2）图像放缩与颜色变换。为了模拟不同视差范围变化，随机地将所有的光场子孔径图下采样为

原始分辨率的 1/2 或 1/4。同时，为了增强模型对颜色和光照变化的鲁棒性，将训练数据每个颜色通道的初始灰度值乘以区间(0.6，1)之间的随机数。

3)添加噪声与 Gamma 变换。为了增强模型的抗噪性能，在训练数据中随机添加不同方差的零均值高斯噪声。同时，为了模拟光场成像过程 Gamma 值变化，在(0.6，1.1)之间随机调整训练数据的 Gamma 值。

3.3　训练损失函数

由于本文模型采用端到端工作方式，模型输出为场景的逐像素视差图。因此，为了准确衡量训练过程中的模型输出与真值之间的误差，本文基于平均绝对误差(MAE)构建模型训练的损失函数。具体地，记模型输出的视差图为 $d(s,t)$，场景真实的视差图为 $g(s,t)$，则模型训练过程中的损失函数可以表示为

$$L_{\text{loss}}=\sum_{s,t}\left\|d(s,t)-g(s,t)\right\|_1, \tag{15}$$

式中：L_{loss} 代表训练过程中的损失；(s,t) 代表子孔径图的像素坐标。

4　实验验证和分析

为了验证上述方法的有效性，采用 pytorch 对提出的模型进行训练，训练数据采用文献[23]提供的 16 幅合成场景。该合成数据集提供的光场维度为 512×512×9×9，其中 512×512 表示光场子孔径图的分辨率，9×9 表示光场的角度分辨率。在模型训练过程中，输入 batch 的尺寸被设置为 128×128×33×16，其中 128×128 表示从子孔径图中随机选取 128×128 的子区域，33 表示从每幅光场中选取的子孔径图数量(图 5)，16 表示 batch 数量。测试集采用文献[23]提供的另外 8 幅合成场景和文献[24]提供的 4 幅真实场景。作为比较，本文选取多种具有代表性的主流算法，即 EPI2[15]、LF[25]、EPInet[15]、FocalstackNet[23]、LF_OCC[4]。

4.1　标准测试集测试

本文首先在合成标准测试数据集上对算法进行了测试，图 6 给出了测试场景的中心视图和视差真值图。这些场景内容的设置考虑了现有光场深度估计算法存在的困难和挑战，采用这些数据能够客观地评估算法的性能。

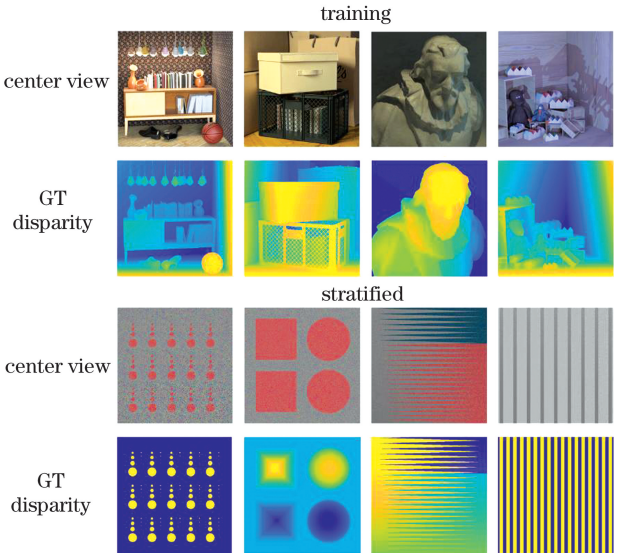

图 6　合成数据集的中心视图与视差真值图

Fig. 6　Center view and GT disparity of synthetic datasets

为定性评价，图 7 给出了标准测试集上的运行结果。以 Boxes 场景为例，说明本文方法对场景中精细结构的处理性能。从图 6 可以看出，该场景包含大量的网状精细结构。现有的方法在处理这些重复的精细结构时，存在不同程度的困难。相比于其他方法在网状区域不能正确区分“网孔”和“网状线”的情况，本文方法通过整合多尺度特征，在对应区域获得了锐利的边缘。通过将实验结果与该场景的真实视差图对比可知，本文方法在 Boxes 场景的网状区域获得了最好的结果。以 Dino 场景为例，说明本文方法对弱纹理区域的处理性能。对于 Dino 场景，该场景中心处的三角形阴影区域(图 7 中方框所示)缺少明显的纹理信息。从图 6 给出的真实视差图可知，该三角形区域的视差存在缓慢的变化。对比 Dino 场景的真实视差图和实验结果可知，在该弱纹理区域本文方法给出了与真实视差十分相近的结果，而其他几种方法处理该区域时存在不同程度的困难。

为定量评估，本文进一步统计了测试结果的 Badpixel(0.07)(误差大于 0.07 pixel 的像素比例)和均方误差(MSE)值。记算法输出的视差图为 $d(s,t)$，测试场景的真实视差图为 $g(s,t)$，待评价的视差图区域为 M，则

$$P_{\text{BadPixel}}(t)=\frac{\left|\{(s,t)\in M:\left|d(s,t)-g(s,t)\right|>0.07\}\right|}{|M|}, \tag{16}$$

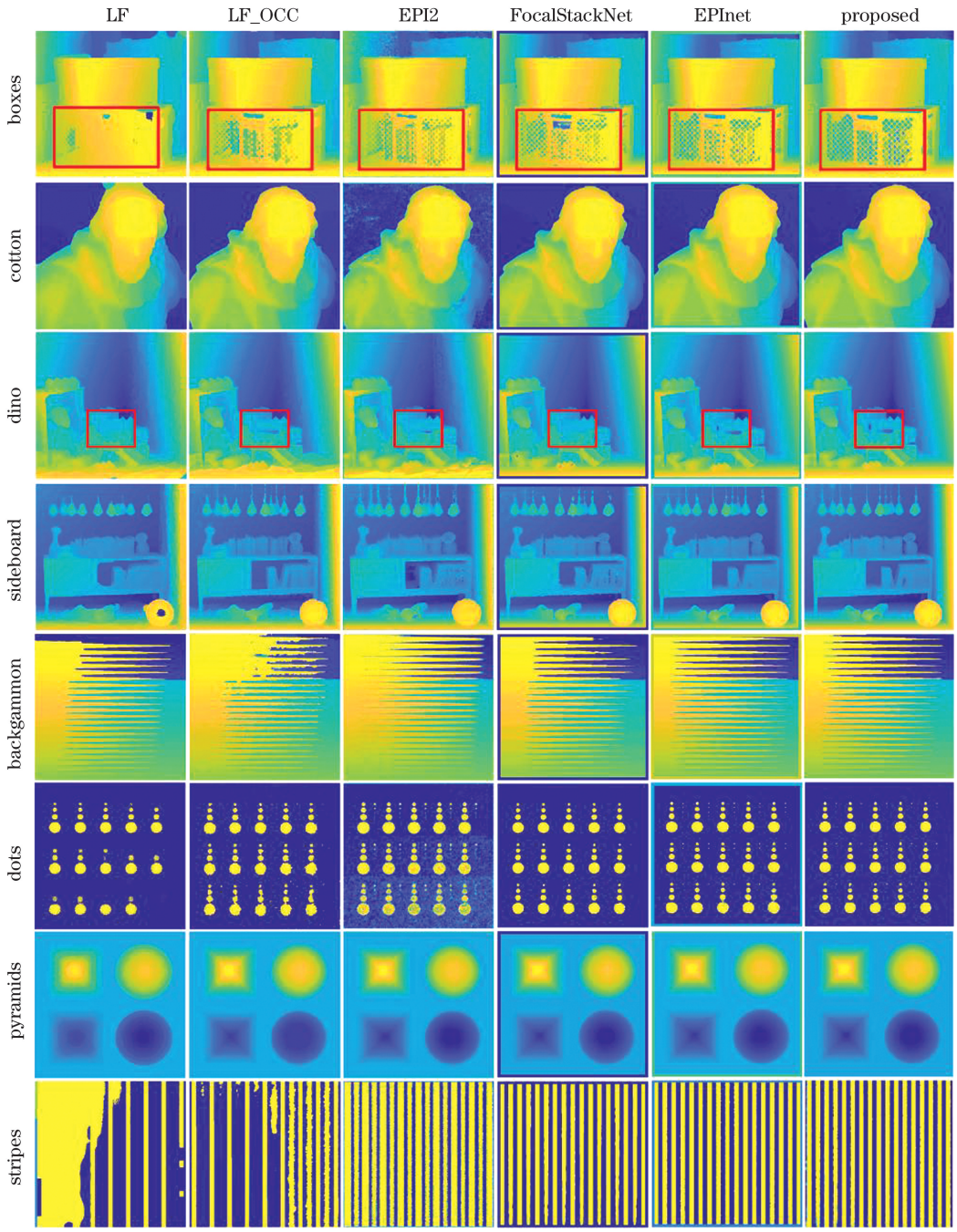

图 7　合成数据集测试结果

Fig. 7　Test results of synthetic dataset

$$E_{\mathrm{MSE}}=\frac{\sum_{(s,t)\in M}[d(s,t)-g(s,t)]^2}{|M|}\times 100。\tag{17}$$

表 1 给出了不同算法结果的 Badpixel(0.07)，该值反映了算法结果和真实值之差大于 0.07 的像素比例。从表 1 可以看出，本文方法在 Badpixel(0.07)值上，取得了比其他方法更好的效果。表 2 统计了不同算法结果的 MSE 值，可以看出，本文方法在多数合成场景中取得了较低的 MSE。这主要得益于网络模型提取到的多尺度特征能有效处理弱纹理区域和精细结构。表 3 统计了不同算法的运行时间，可以看出，本文方法由于采用深度学习实现，处理速度较传统方法提升了 1～2 个数量级，这对大多数应用来说显得至关重要。

表 1　不同算法的 Badpixel(0.07)比较

Table 1　Badpixel(0.07) comparison of different algorithms

Scene	LF_OCC	EPI2	LF	FocalStackNet	EPInet	Proposed
Boxes	26.52	29.80	23.02	14.33	12.24	11.46
Cotton	6.22	16.69	7.83	0.58	0.46	0.51
Dino	14.91	15.67	19.03	2.53	1.26	1.14
Sideboard	18.50	18.95	21.99	5.40	4.78	4.56
Backgammon	19.01	22.08	5.52	4.34	3.28	1.87
Pyramids	3.17	1.08	12.35	0.29	0.15	0.28
Dots	5.82	46.53	2.90	1.02	1.98	3.49
Stripes	18.41	23.81	35.74	3.72	0.91	0.85

表 2　不同算法的均方误差比较

Table 2　MSE comparison of different algorithms

Scene	LF_OCC	EPI2	LF	FocalStackNet	EPInet	Proposed
Boxes	9.85	10.93	17.43	11.82	6.01	4.81
Cotton	1.07	4.32	9.17	0.88	0.22	0.23
Dino	1.14	2.07	1.16	0.89	0.15	0.15
Sideboard	2.30	4.65	5.07	1.96	0.81	0.59
Backgammon	21.59	20.78	13.01	6.58	3.91	2.39
Pyramids	0.10	0.02	0.27	0.02	0.007	0.01
Dots	3.30	6.66	5.68	1.87	1.98	4.29
Stripes	8.13	6.10	17.45	1.79	0.91	0.85

表 3　不同算法的运行时间比较

Table 3　Runtime comparison of different algorithms　　unit: s

Scene	LF_OCC	EPI2	LF	FocalStackNet	EPInet	Proposed
Boxes	10408.26	8.91	962.18	85.04	2.03	1.03
Cotton	6325.51	9.07	984.53	84.90	2.03	1.04
Dino	10099.05	8.09	1130.56	85.62	2.03	1.07
Sideboard	13531.30	8.74	987.47	84.77	2.02	1.11
Backgammon	5116.25	6.93	979.87	84.24	2.02	1.04
Pyramids	11688.42	6.88	929.72	92.09	2.02	1.04
Dots	10820.85	7.66	979.87	81.56	2.03	1.08
Stripes	19331.42	8.53	1093.98	91.92	2.03	1.07

4.2　真实场景测试

为了进一步验证模型的泛化性能，本文在真实场景上开展了测试。图 8 为上述方法在真实场景上的测试结果。从图中可以看出，模型在真实场景上具有良好的泛化性能，能够准确地处理真实场景的视差。一方面，测试场景中包含细小的线状目标(方框所示)，测试结果清晰地给出了这些线状物的视差。以 Danger 场景为例，模型很好地估计出了图中的网状区域的视差信息，并且对于网状物后面的区域能给出合理的结果。另一方面，对于真实场景中的弱纹理区域(如地面)，模型同样能够获得连续平滑的结果。从结果中还可以看出，上述方法在真实场景中同样能够准确处理遮挡边缘区域的视差，这也是该方法的另一个优点。根据上述方法估计出的视差，以及视差与深度之间的转换关系，图 9 对部分测试场景的三维形貌进行了重建，可以看出，

上述方法能够准确地复原场景的三维结构，对于弱纹理区域和精细结构能够获得清晰的重建结果。通过对比三维重建结果和相应的视差图可以发现，准确的视差估计是良好重建的关键。

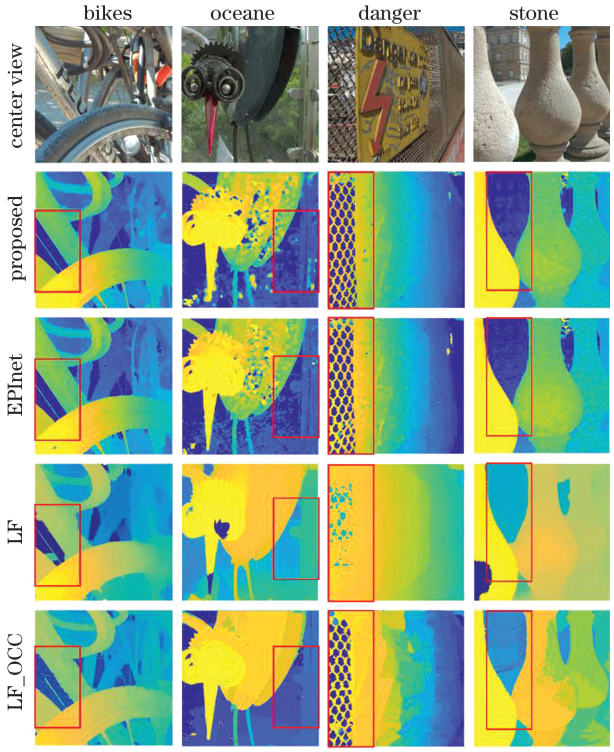

图 8　真实数据集测试结果

Fig. 8　Test results of real dataset

图 9　测试场景三维重建结果

Fig. 9　Three-dimensional reconstruction results of test scene

为定量评估重建精度，本文构建了图 10(a)所示的测试场景。依据上述方法，得到如图 10(b)所示的视差和如图 10(c)所示的三维重建结构。从图 10(a)中的刻度能读取被测量间距的大小。图 10(c)显示了采用三维编辑软件(MeshLab)内置的工具对重建结构进行测量的结果。图中的白色数字对软件显示的测量结果进行了字体放大，与图 10(a)中的刻度进行比较，可以看出，上述方法能够准确地恢复场景的尺寸和结构。

图 10　真实场景尺度测量结果。(a)测试场景；(b)视差图；(c)三维重建结构

Fig. 10　Measurement of real scene scale. (a) Test scene; (b) disparity map; (c) three-dimensional reconstruction structure

5　结　　论

本文依据光场相机的成像模型推导了光场视差与场景深度之间的转换关系，实现了由场景视差到场景真实深度信息的恢复。为了解决现有光场视差估计方法在弱纹理区域和精细结构区域难以获得准确视差的问题，提出了一种基于深度学习的光场视差估计方法。通过将多尺度特征引入光场深度估计，本文方法实现了对弱纹理区域和精细结构的准确处理。该方法基于卷积神经网络实现，相比于传统的基于代价优化的方法，处理速度提升了 1～2 个数量级。在合成数据集和真实数据集上的实验结果表明，本文方法能够准确地实现场景的视差估计与三维结构复原。

参　考　文　献

[1] Furukawa Y, Hernández C. Multi-view stereo: a tutorial[J]. Foundations and Trends in Computer Graphics and Vision, 2015, 9(1/2): 1-148.

[2] Fu L, Hong H B, Wang X, et al. Non-Lambertian photometric stereo vision based on inverse reflectance model[J]. Acta Optica Sinica, 2020, 40(5): 0520001.
付琳, 洪海波, 王晰, 等. 基于逆向反射模型的非朗伯光度立体视觉[J]. 光学学报, 2020, 40(5): 0520001.

[3] Xiao J S, Tian H, Zou W T, et al. Stereo matching based on convolutional neural network[J]. Acta Optica Sinica, 2018, 38(8): 0815017.
肖进胜, 田红, 邹文涛, 等. 基于深度卷积神经网络的双目立体视觉匹配算法[J]. 光学学报, 2018, 38(8): 0815017.

[4] Shan B H, Huo X Y, Liu Y. A stereovision measurement method using epipolar constraint to correct digital image correlation matching[J]. Chinese Journal of Lasers, 2017, 44(8): 0804003.

单宝华, 霍晓洋, 刘洋. 一种极线约束修正数字图像相关匹配的立体视觉测量方法[J]. 中国激光, 2017, 44(8): 0804003.

[5] Bok Y, Jeon H G, Kweon I S. Geometric calibration of micro-lens-based light field cameras using line features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 287-300.

[6] Dansereau D G, Pizarro O, Williams S B. Decoding, calibration and rectification for lenselet-based plenoptic cameras[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 1027-1034.

[7] Johannsen O, Honauer K, Goldluecke B, et al. A taxonomy and evaluation of dense light field depth estimation algorithms[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1795-1812.

[8] Mishiba K. Fast depth estimation for light field cameras[J]. IEEE Transactions on Image Processing, 2020, 29: 4232-4242.

[9] 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: 148-159.

[10] Anisimov Y, O W, Stricker D. Rapid light field depth estimation with semi-global matching[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 16-20, 2019, Long Beach, CA. New York: IEEE, 2019, 78-89.

[11] Anisimov Y, Wasenmüller O, Stricker D. A compact light field camera for real-time depth estimation[J]. Computer Analysis of Images and Patterns, 2019: 52-63.

[12] Cai Z W, Liu X L, Pedrini G, et al. Accurate depth estimation in structured light fields[J]. Optics Express, 2019, 27(9): 13532-13546.

[13] Cai Z W, Liu X L, Pedrini G, et al. Light-field depth estimation considering plenoptic imaging distortion[J]. Optics Express, 2020, 28(3): 4156-4168.

[14] Li J K, Jin X. Frequency descriptor based light field depth estimation[C]//2019 IEEE Visual Communications and Image Processing (VCIP), December 1-4, 2019, Sydney, Australia. New York: IEEE Press, 2019: 1-4.

[15] 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]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4748-4757.

[16] Heber S, Yu W, Pock T. Neural EPI-volume networks for shape from light field[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2271-2279.

[17] Leistner T, Schilling H, Mackowiak R, et al. Learning to think outside the box: wide-baseline light field depth estimation with EPI-shift[C]//2019 International Conference on 3D Vision (3DV), September 16-19, 2019, Québec City, QC, Canada. New York: IEEE Press, 2019: 249-257.

[18] Schiopu I, Munteanu A. Deep-learning-based depth estimation from light field images[J]. Electronics Letters, 2019, 55(20): 1086-1088.

[19] 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, 2019, 41(2): 297-310.

[20] Kalantari N K, Wang T C, Ramamoorthi R. Learning-based view synthesis for light field cameras[J]. ACM Transactions on Graphics, 2016, 35(6): 1-10.

[21] Sun X, Xu Z M, Meng N, et al. Data-driven light field depth estimation using deep convolutional neural networks[C]//2016 International Joint Conference on Neural Networks (IJCNN), July 24-29, 2016, Vancouver, BC, Canada. New York: IEEE Press, 2016: 367-374.

[22] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.

[23] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[J]. Computer Vision - ACCV 2016, 2017: 19-34. DOI:10.1007/978-3-319-54187-7_2.

[24] Hanhart P, Řeřábek M, Ebrahimi T. Subjective and objective evaluation of HDR video coding technologies[C]//2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), June 6-8, 2016, Lisbon, Portugal. New York: IEEE Press, 2016: 1-6.

[25] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1547-1555.