

基于方向感知网络的场景偏振三维成像技术(特邀)

张俊杰¹, 刘飞^{1*}, 王鑫², 何飘¹, 樊昭¹, 邵晓鹏^{1,2**}¹西安电子科技大学光电工程学院, 陕西 西安 710071;²西安电子科技大学杭州研究院, 浙江 杭州 311231

摘要 针对场景偏振三维成像中光照不均匀、色彩、材料复杂和大视场下观测方向变化等原因造成的偏振法线梯度不准确和真实三维信息获取困难的问题,提出一种基于方向感知卷积神经网络的场景偏振三维成像新方法。首先,搭建具有方向感知能力的场景深度估计网络结构;其次,利用卷积神经网络所估计的场景深度对偏振法线梯度进行校正;最后,利用校正后的梯度通过基于梯度的积分算法进行三维重建。实验结果表明,所提方法解决了偏振固有的方位角模糊,提高了在光照不均匀、大视场范围场景条件下获取的法线梯度的准确性,最终在恢复场景真实三维形状的同时保留了丰富的纹理细节信息。实验结果证明了所提技术的有效性与优越性。

关键词 偏振三维成像; 深度估计; 梯度场校正; 神经网络

中图分类号 O436.3

文献标志码 A

DOI: 10.3788/LOP232498

Scene Polarization 3D Imaging Technology Based on Direction-Aware Network (Invited)

Zhang Junjie¹, Liu Fei^{1*}, Wang Xin², He Piao¹, Fan Zhao¹, Shao Xiaopeng^{1,2**}¹School of Optoelectronic Engineering, Xidian University, Xi'an 710071, Shaanxi, China;²Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, Zhejiang, China

Abstract To overcome challenges arising from inaccurate polarization normal gradients and difficulties in obtaining real three-dimensional (3D) information in scene polarization 3D imaging—attributed to factors like uneven illumination, complex colors, materials, and changes in observation direction under a large field of view—a new approach utilizing a direction-aware convolution neural network is explored. The method involves constructing a scene depth estimation network with direction perception abilities, correcting polarization normal gradients using the convolutional neural network estimated scene depth, and ultimately reconstructing the 3D image through a gradient-based integration algorithm. Experimental results showcase this approach's effectiveness in resolving azimuth ambiguity inherent in polarization, enhancing normal gradient accuracy in scenes with uneven illumination and wide field of view, and successfully restoring the real 3D shape of the scene while preserving intricate texture details. The findings affirm the efficacy and superiority of the proposed technology.

Key words polarized 3D imaging; depth estimation; gradient field correction; neural network

1 引言

光学三维成像技术随着数字化的快速发展,在越来越多的应用场景中扮演着更为重要的角色^[1-4]。而目前的三维成像方式按照获取图像的方法分成扫描和快照两种方式^[5]:扫描型三维成像技术是通过目标

物体所在的平面进行逐点扫描,获取对应区域内的目标图像,进而进行三维重建的。这种方法可以较为精确地还原目标的三维特性,但是容易受到大视场场景的限制^[6-7],且成本较高、重建过程复杂、效率低。快照三维成像技术直接获取目标区域的所有像素信息,之后根据不同的理论模型进行图像的三维重建,主要包

收稿日期: 2023-11-14; 修回日期: 2023-12-21; 录用日期: 2023-12-22; 网络首发日期: 2023-12-25

基金项目: 国家自然科学基金(62205259, 62075175, 61975254, 62375212, 62005203, 62105254)、西安电子科技大学基本科研业务费资助项目(QTZX22016)

通信作者: *feiliu@xidian.edu.cn; **xpshao@xidian.edu.cn

括飞行时间法 (ToF)、双目立体视觉、偏振三维成像等。ToF 和双目立体视觉由于测量方法所限,在远距离成像时的适用性仍然存在一定问题^[8-9]。

偏振作为光波的固有特性之一,能够建立空间几何形貌与光波物理信息之间的关系,所以在目前三维成像技术中受到了国内外研究者的广泛关注。Koshikawa^[10]提出了偏振三维重建技术,并推导出目标表面法线与入射光方向、出射光的偏振信息之间的物理关系。美国的 Wolff^[11]对偏振三维重建技术进行了改进,在求取物体表面法线方向过程中引入方位角与偏振度。而 Atkinson 等^[12]提出了一种利用偏振和阴影信息进行三维表面重建的新方法,并最终对目标实现三维重建。Li 等^[13]在近红外波段进行偏振三维重建,有效消除了由目标表面色彩不同导致的偏振三维重建畸变。但是上述方法多数对光照条件有较高的要求,应用场景较少,为了解决上述问题,Zou 等^[14]将偏振信息作为输入引入神经网络技术,采用深度学习方法估计人体形状,增强了偏振三维成像技术的可靠性。而 Deschaintre 等^[15]将偏振三维成像与深度学习进行融合,实现了对单目标形状和深度的高质量评估。

综上所述,虽然偏振三维成像技术已经在不同的

领域取得了一定的成果,但是由于算法和环境的限制,目前该技术的应用还局限于物体级,针对大场景的偏振三维成像尚未成熟,还存在以下问题:1)场景不连续,第一类间断点导致目标表面不可积;2)大场景中复杂自然光照下的法线场信息分布更加复杂,难以准确解译;3)大视场下观测方向改变导致法线信息不准确。

针对以上问题,本文提出一种基于方向感知卷积神经网络(CNN)的场景级偏振三维成像新方法。该方法通过建立基于偏振信息表征深度的神经网络,有效估算出场景的粗深度信息,并以此作为先验信息,对偏振法线梯度进行校正,解决大场景偏振三维重建过程中复杂光强分布及场景不连续的问题,并进行实验验证及分析,最终实现三维形状准确且纹理细节丰富度较高的室内场景偏振三维成像。本研究为场景级目标的偏振三维成像应用提供了理论方法和实施方案,有利于拓宽偏振三维成像技术的应用领域,具有良好的应用前景。

2 偏振三维成像原理

根据场景目标表面出射光偏振特性的变化与其表面三维形状之间的固有关系,可对场景目标表面法线方向进行求解^[16],如图 1 所示。

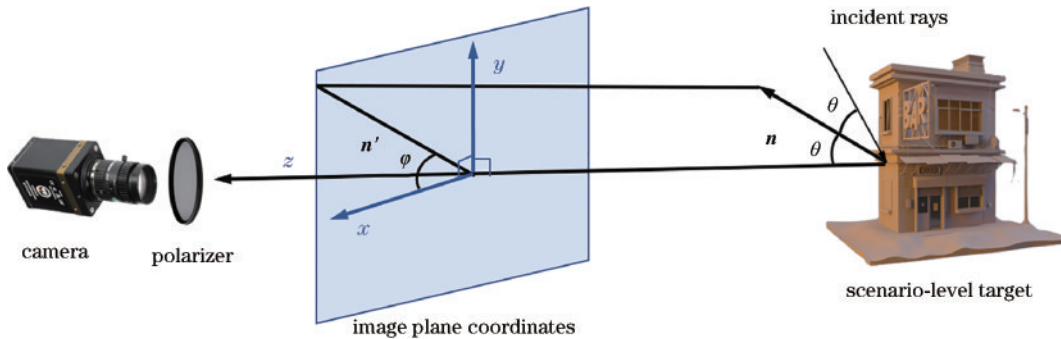


图 1 场景偏振三维成像模型

Fig. 1 Scene-level polarization 3D imaging model

光波入射到场景目标表面后发生反射,且反射角等于入射角,都为 θ 。 xoy 平面为偏振相机所处平面,物体表面法线 \mathbf{n} 在该平面上的投影为 \mathbf{n}' ,其与 x 轴的夹角为 φ ,其中, θ 和 φ 分别为法线的天顶角和方位角。

根据图 1 中所示几何关系可知, θ 和 φ 共同约束物体表面该点的法线方向:

$$\mathbf{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{bmatrix} = \begin{bmatrix} \tan \theta \cos \varphi \\ \tan \theta \sin \varphi \\ 1 \end{bmatrix}, \quad (1)$$

式中: n_x 、 n_y 、 n_z 分别表示法线在 x 轴、 y 轴和 z 轴方向上的分量。

在获取目标表面法线之前需要求解出法线的天顶角与方位角,而根据菲涅耳(Fresnel)公式可知,当发生漫反射时天顶角与反射光偏振度之间存在以下关系:

$$P_{\text{diff}} = \frac{\left(n - \frac{1}{n}\right)^2 \sin^2 \theta}{2 + 2n^2 - \left(n + \frac{1}{n}\right)^2 \sin^2 \theta + 4\cos \theta \sqrt{n^2 - \sin^2 \theta}}, \quad (2)$$

式中: P_{diff} 为漫反射光的偏振度; n 为目标材质的折射率; θ 为漫反射光的天顶角也是入射角。

为了更加直观地体现二者之间的关系,根据式(2)计算出漫反射光的偏振度随着入射角度和目标折射率的变化曲线,具体如图 2 所示。

从图 2 可以看出,当目标的折射率一定的情况下,漫反射光偏振度随着入射角增大而单调增大,故在已知漫反射光偏振度的条件下根据式(2)可以求出与之唯一对应的入射角,即法线的天顶角。

当完全非偏振光入射到目标表面发生反射或折射

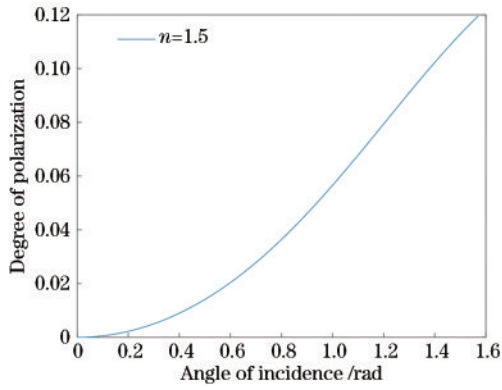


图 2 漫反射光偏振度随入射角变化曲线图

Fig. 2 Graph of the degree of polarization of diffuse light as a function of the angle of incidence

时,光波的偏振状态将发生改变,变成部分偏振光或完全偏振光,而在目标表面漫反射光分量中平行于入射平面的分量总是占优的,因此当偏振器的透光轴与入射平面平行时,通过偏振器的偏振光强度最大。为了得到目标表面法线的方位角,可以通过旋转探测器获取光强随偏振器旋转角度变化的曲线,计算出光强曲线最大值对应的相位角,该角度即为法线方位角,图 3 为偏振器旋转时光强的变化曲线。

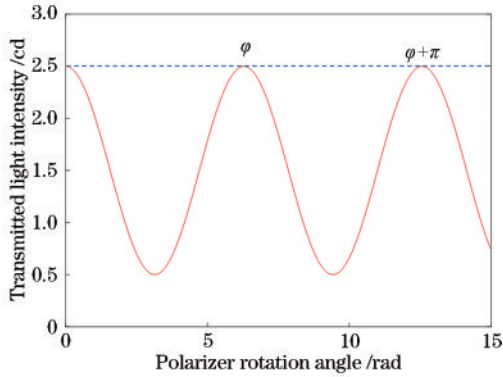


图 3 光强随偏振器旋转角变化曲线

Fig. 3 The change curve of light intensity with polarizer rotation angle

由图 3 可以看出,光强曲线最大值所对应的相位角有两个,分别是 φ 、 $\varphi + \pi$, 因此方位角不能唯一确定,求出的目标法线不准确,导致目标表面形状恢复出现严重的畸变。针对方位角的模糊性问题,需要结合其他方法获取先验信息,对方位角进行校正,从而获取到正确的法线方向。

结合式(1),物体表面高度用函数 $Z(x, y)$ 表示,可得到偏振法线梯度^[17]的表达式:

$$\begin{cases} p = Z_x = n_x = \tan \theta \cos \varphi \\ q = Z_y = n_y = \tan \theta \sin \varphi \end{cases}, \quad (3)$$

式中: p 和 q 分别代表物体表面法线在 x 轴方向和 y 轴方向上的梯度; Z_x 和 Z_y 表示物体表面高度函数在 x 轴

方向和 y 轴方向上的梯度。

通过偏振信息直接求取的法线梯度,其数值准确,只在方向上存在多值性问题。传统的光强校正方法将光强信息降维成梯度,以此作为先验信息对偏振法线梯度进行校正,但在自然场景中由于复杂环境光的影响而失效,而利用 CNN 估计的场景深度信息对场景的形状轮廓信息描述准确^[18],不存在法线梯度方向错误的问题。因此,将通过该方法求取的场景深度作为场景三维重建的先验信息,以此对偏振求取的法线梯度方向进行校正。

将神经网络深度估计结果降维成梯度信息,作为已知先验信息对偏振法线梯度进行校正,如式(4)所示^[19]:

$$\hat{\Lambda} = \underset{\Lambda}{\operatorname{argmin}} \left\| N_{\text{CNN}} - \Lambda(N_{\text{pol}}) \right\|_2^2, \quad \Lambda \in \{-1, 1\}, \quad (4)$$

式中: N_{CNN} 和 N_{pol} 分别为利用神经网络和偏振信息获取的场景表面法线梯度; $\hat{\Lambda}$ 代表二元操作数集合,用以判断通过偏振信息求取的法线梯度方向是否与真实法线梯度方向一致; $\operatorname{argmin}(\cdot)$ 函数是对函数求集合的函数,它的结果是使括号中的函数取极小值的自变量的集合。式(4)中 $\operatorname{argmin}(\cdot)$ 函数的自变量为 Λ , Λ 是一个

二元操作符, Λ 的值为 1 或 -1, 对应的两个值分别表示方位角是否需要通过加上 π 来进行校正。当 $\Lambda = 1$ 时, N_{CNN} 与 N_{pol} 的值相近, 此时认为通过偏振信息求取的法线梯度方向是正确的, 不需要进行校正; 而当 $\Lambda = -1$ 时, N_{CNN} 与 N_{pol} 的值相差较远, 此时认为法线梯度方向不正确, 需要进行校正。

通过式(4)优化后的二元操作数集合 $\hat{\Lambda}$ 对利用偏振解算出来的自然场景表面法线梯度方向进行校正:

$$N_{\text{cor}} = \hat{\Lambda} \cdot N_{\text{pol}}, \quad (5)$$

式中: N_{cor} 表示经卷积神经网络校正之后的正确法线梯度。

经过偏振法线梯度的校正后,设法线梯度场是处处可积的,此时可根据局部^[20-21]或全局^[22]积分法重建物体表面高度。但是在实际应用中,通过偏振信息直接求取的法线梯度不准确,在局部区域上具有离散性,因此不能假设物体表面的法线梯度是处处可积的。

把表面高度函数 $Z(x, y)$ 关于梯度场的表达式^[23]定义为

$$Z = F^{-1} \left\{ -j \frac{uF(p) + vF(q)}{u^2 + v^2} \right\}, \quad (6)$$

式中: F 和 F^{-1} 分别表示离散傅里叶变换和傅里叶逆变换操作; u 和 v 分别为 x 轴方向和 y 轴方向的角频率。不可积的表面法线梯度场在进行傅里叶变换后变为了可积的函数集合,把积分问题转至频域中^[24],从而实现积分重建。

3 卷积神经网络设计

3.1 网络结构

图 4 为卷积神经网络结构模型。所使用的卷积神经网络结构包含 3 个部分:进行图像特征提取的编码器、进行特征信息复原的解码器^[25],以及进行提升神经网络方向感知能力的 CoordConv 模块。该网络结构的流程如下:首先,将原图裁剪为 4 张偏振子图像,并通过斯托克斯公式求解出偏振度、偏振角图像,与 CoordConv 模块一起输入网络,经过编码器网络进行

特征提取,将图像的特征信息转换为高阶信息;随后,将提取到的高阶特征信息传递给解码器将特征信息融合为一维信息,再通过监督学习的方式对网络进行训练,最终获取估计出的场景深度信息。

在卷积神经网络中输入光强信息,经过训练之后也能够得到深度信息,但是在复杂光照条件下,建立的光强信息与图像深度信息之间的映射关系并不稳定。相对于光强信息,由 4 幅偏振子图所求解出的偏振信息与场景深度信息之间存在着稳定的映射关系。因此,在该网络结构中输入偏振子图进行网络的训练。

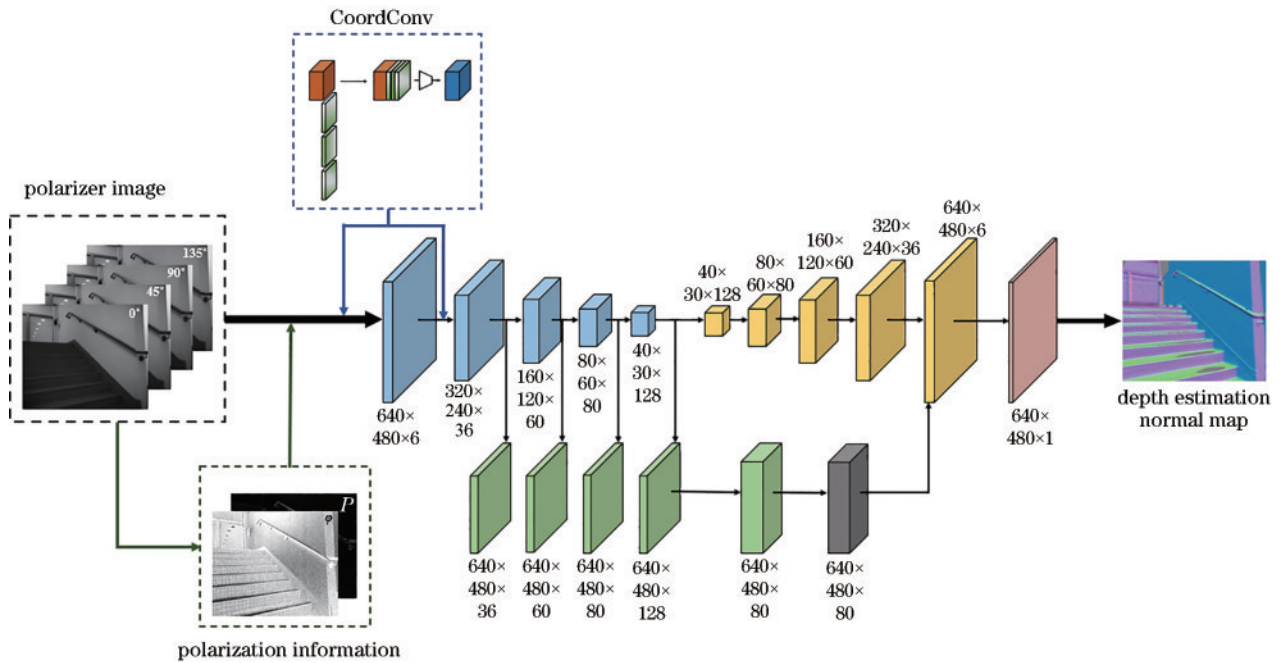


图 4 场景深度估计网络结构图

Fig. 4 Network structure of scene depth estimation

相对于目标级深度估计对象,场景中存在复杂的环境和物体,这对编码器的特征提取能力提出了更高的要求。常见的编码器网络采取大小固定的卷积核^[26-28],限制了编码器网络特征的提取能力。为了弥补现有常见编码器网络的缺陷,采用 Res2Net^[29]编码器网络来提升其特征提取能力。相对于 ResNet, Res2Net 在单独残差分块之间进行划分,并将其卷积相连,增加了层级之间的特征提取通道,提高了编码器的特征提取能力,从而提高了最终深度估计图像的可靠性。

经过编码器特征提取的图像需要进行上采样操作还原偏振子图像分辨率。由于在上采样过程中并没有增多图像本身的信息,图像的质量会随之降低,为了减小上采样对图像质量的影响,同时控制计算成本,采取较简便的双线性插值法进行上采样^[30]。解码器使用的网络架构虽然具有强大的学习能力,但是网络的误差控制能力会随着网络复杂度的增加而下降。为了解决该问题,解码器结构通过跳跃连接与上采样之间特征

的融合实现,跳跃连接是指跳过神经网络中的某些层,通过该方式减小网络的构架复杂度,增强网络的误差控制能力。

在偏振三维成像中存在对目标所有像素而言观测方向都在目标的正前方的假设,但对图 5 所示的场景级目标而言,场景边缘处的像素点与探测器接收处所成夹角 β_1 和 β_2 陡然增大,此时观测方向在像素点正前方的假设不再成立。而神经网络的卷积层具有平移等

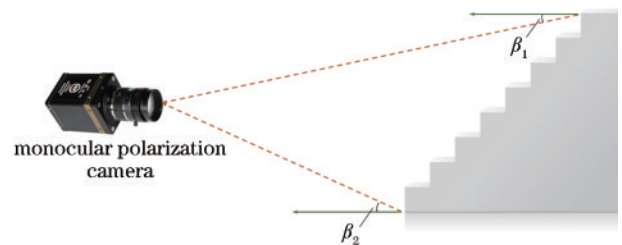


图 5 大场景条件下观测方向对偏振信息获取的影响

Fig. 5 The influence of observation direction on polarization information acquisition under large-scene conditions

变性的特点,该性质的存在使得当每个滤波器被应用于输入以产生输出时,网络无法知晓每个滤波器的位置,如果不将观测方向显式地提供给 CNN,它将无法知晓逐个像素点的观测方向,从而使最终恢复的深度图产生明显误差,导致偏振表示也受到观测方向的影响,且当视场增大时误差会急剧增大,从而进一步增大场景重建结果的畸变。通过在输入中添加 x 坐标、 y 坐

标和 z 坐标,产生 CoordConv 模块,对卷积层的等变程度进行学习,从而让神经网络知晓像素点的观测方向,提升其感知能力。

CoordConv 模块为标准卷积的简单扩展,其实例化了额外的通道并填充了未经训练的坐标信息,然后将它们按通道加入输入端,图 6 描述了将 3 个坐标 x 、 y 和 z 相加的操作。

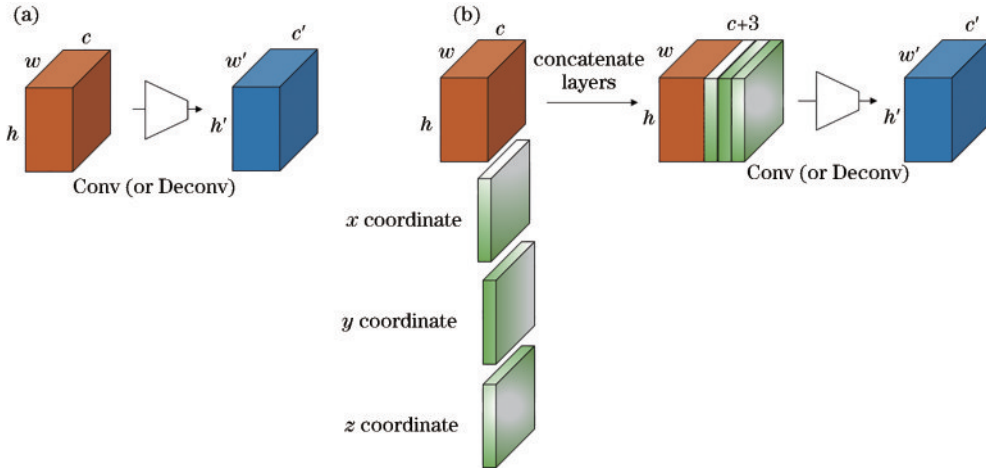


图 6 卷积层与 CoordConv 层。(a) 卷积层; (b) CoordConv 层

Fig. 6 Convolutional layer and CoordConv layer. (a) Convolutional layer; (b) CoordConv layer

具体而言, x 坐标是一个 $h \times w$ 的矩阵,其第 1 行填充 0,第 2 行填充 1,第 3 行填充 2,以此类推。同理, y 坐标的第 1 列填充 0,第 2 列填充 1,第 3 列填充 2。对于 z 坐标,中间点填充 0,逐圈向外扩散为 $1, 2, 3, \dots, n$,其中, $z = \sqrt{(x - h/2)^2 + (y - w/2)^2}$ 。最终对 x 、 y 和 z 坐标都进行线性缩放,并使其归一化为 $[-1, 1]$ 。对于三维图像的卷积, (x, y, z) 坐标便足以指定一个输入像素。

CoordConv 模块的引入有效减轻了观测方向对场景偏振信息的影响,优化了利用偏振图像集进行网络训练的深度估计结果,因此在深度估计和细节还原上有较好的表现。

3.2 网络模型的训练

对场景深度估计网络的训练过程如下:将场景级偏振图像及其真实深度图输入网络,通过多个阶段的特征采集,将浅层细节特征与深层全局特征融合,经过网络学习,得到目标尺寸的特征图,由融合和精细化模块输出最终深度图。

网络训练中所使用的硬件设施配置如下:CPU 为 Intel (R) Core (TM) i9-11900K @4.8 GHz, GPU 为 NVIDIA Geforce RTX,内存为 32 GB,网络结构是通过 Windows 10 系统下的 PyTorch 1.8.0 和 CUDA 11.3 学习框架实现的,利用该图像处理工作站进行深度学习及场景偏振三维重建,且根据设备配置性能,设置 batch size 为 8,初始学习率为 0.0001。训练过程所使用的数据集 SPW Dataset 由香港科技大学团队创建,

是包含场景偏振图像与对应表面法线图像的数据集^[31]。

为了提高神经网络对场景的深度估计能力,采取损失函数来衡量深度估计图像与真实深度图像在像素之间的差异程度。模型会根据损失函数衡量的差异反向更新网络中的各个参数,进一步降低真实值与估计值之间的损失值。采取的损失函数为

$$L(y, y') = L_M(y, y') + \lambda L_p(y, y'), \quad (7)$$

式中: L_M 是多尺度结构相似性损失; L_p 是尺度不变性损失; y 为真实深度; y' 为最终估计的深度; λ 为调节两损失项的超参数。

多尺度结构相似性损失能够最大程度地保留生成图像的局部纹理等细节信息,其表达式为

$$L_M(y, y') = \frac{1 - \text{MSSSIM}(y, y')}{2}, \quad (8)$$

式中: $\text{MSSSIM}(\cdot)$ 为多尺度结构相似性损失函数。

多尺度结构相似性损失函数以真实深度图和预测深度图作为输入,对滤波后的图像进行下采样,对图像的结构相似性在不同尺度上进行融合^[32],其表达式为

$$\text{MSSSIM}(y, y') =$$

$$\left[L_M(y, y') \right]^{\alpha_M} \prod_{j=1}^M \left[S_j(y, y') \right]^{\beta_j} \left[C_j(y, y') \right]^{\gamma_j}, \quad (9)$$

式中: $L(y, y') = \frac{2\mu_y\mu_{y'} + C_1}{\mu_y^2 + \mu_{y'}^2 + C_1}$, $S(y, y') = \frac{\sigma_{yy'} + C_3}{\sigma_y\sigma_{y'} + C_3}$, $C(y, y') = \frac{2\sigma_y\sigma_{y'} + C_2}{\sigma_y^2 + \sigma_{y'}^2 + C_2}$; σ_y^2 、 $\sigma_{y'}^2$ 、 μ_y 、 $\mu_{y'}$ 和 $\sigma_{yy'}$ 分别表示

y, y' 的方差、均值, 以及 y 与 y' 之间的协方差; C_1, C_2, C_3 为正值常数; $\alpha_M, \beta_j, \gamma_j$ 为调整各分量权重的参数; M 为图像下采样的次数减 1。

尺度不变性损失能够保证估计深度图像与真实图像的差异不受图像尺寸与比例的影响, 其表达式为

$$L_p(y, y') = \alpha \sqrt{\frac{1}{T} \sum_i g_i^2 - \frac{\beta}{T^2} \left(\sum_i g_i \right)^2}, \quad (10)$$

式中: $g_i = \ln y'_i - \ln y_i$; T 代表具有深度值的像素点个数; $\beta = 0.8$; $\alpha = 10$ 。

3.3 网络模型的训练结果分析

为了验证 CoordConv 模块的有效性, 在使用相同的编解码器结构和损失函数的前提下, 分别去除和加入 CoordConv 模块, 在 SPW 数据集上进行测试, 图 7 为神经网络在加入与去除 CoordConv 模块下, 深度估计相对误差随像素点与图像中心之间的像素数目变化曲线。可以看出, 随着像素点与图像中心之间的像素数目的增大, CoordConv 模块的引入对误差的改善能力也随之提高, 这是因为在图像边缘和角落区域的像素点与探测器形成的夹角更大, 从而使得观测方向对偏振信息的准确获取造成的干扰更加严重。因此, CoordConv 模块的加入提高了所用网络对特征的提取能力, 所得场景深度估计图误差更小。

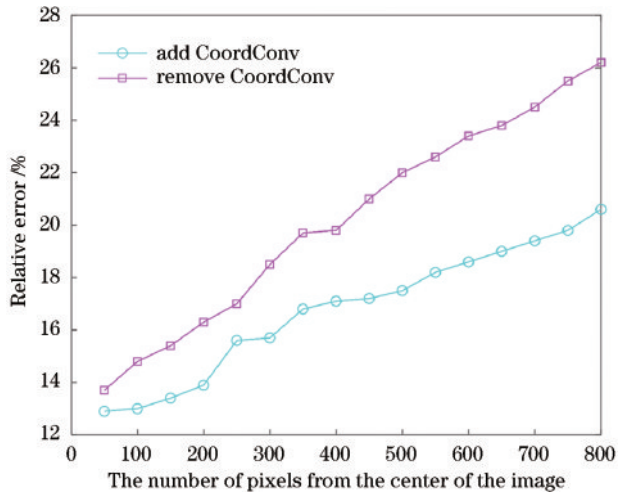


图 7 CoordConv 模块对场景深度估计的影响

Fig. 7 The effect of the CoordConv module on scene depth estimation

为了表征网络结构的有效性, 对部分深度估计的重建结果进行讨论。图 8 为两个场景以及基于方向感知网络的三维重建结果, 从场景的三维重建结果可以看出: 场景一重建的三维模型中墙壁与门较为平滑、门上的窗户边界清晰; 场景二重建的三维模型中阶梯与扶手轮廓清晰, 符合真实目标的情况。因此, 所使用的三维重建算法以及网络结构能够达到比较好的效果, 该网络结构重建的场景三维模型能较好地还原真实场景。

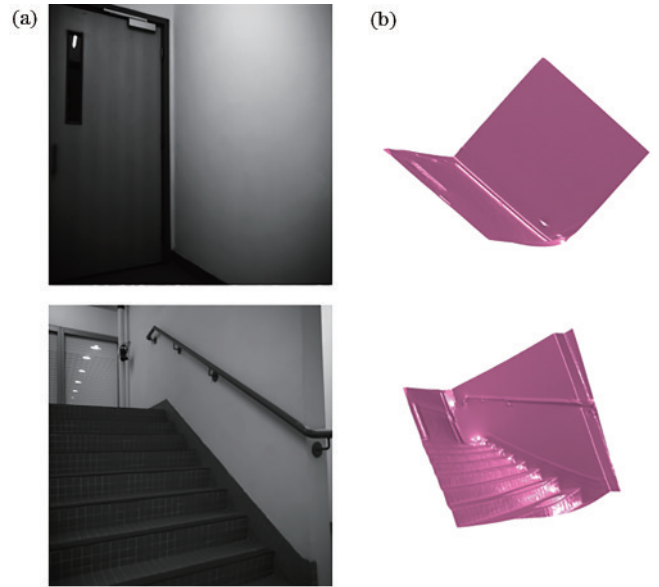


图 8 基于方向感知网络的场景重建结果。(a) 场景一与场景二; (b) 基于方向感知网络的重建结果

Fig. 8 Scene results reconstructed by direction-aware network. (a) Scene 1 and scene 2; (b) results reconstructed by direction-aware network

4 实验与讨论

为了验证基于方向感知网络场景偏振三维成像方法的有效性, 通过实验对复杂光照条件下的场景进行三维重建, 并将基于方向感知网络对偏振校正后的重建结果、基于光强对偏振进行校正后的重建结果和基于方向感知网络深度估计的重建结果进行对比。

图 9 为场景一原图、分别基于光强、方向感知网络对偏振进行校正后的重建结果, 以及两种重建方式在虚线处的深度分布图。由图 9(b) 可以看出, 基于光强校正偏振的三维重建结果的右侧墙壁正上方出现了明显的凸出畸变, 而在右侧墙壁右方又出现了凹陷畸变。这是由于场景一处于复杂的光照条件下, 由光强所求解出的梯度信息不准确, 这些错误的梯度信息在对偏振梯度进行校正的过程中会传递给目标的法线方向, 从而导致重建结果出现畸变。而在训练神经网络时, 以偏振信息为输入对象, 故基于方向感知网络对偏振梯度进行校正时不受复杂光照的影响, 重建结果也更加符合实际情况。

在图 9(c) 中, 第 537、850、1222 列分别处于墙壁拐角、光强极值点、边缘位置, 在这 3 个像素点处, 基于光强校正的深度值分别为 266、392、266, 同一像素位置基于方向感知网络校正的深度值分别为 19、194、413。在图 9(a) 中, 上述 3 个像素点位于墙壁的同一水平线上, 故在图 9(c) 中 3 点也应在同一直线上。因此, 利用 3 个像素点将图 9(c) 中右侧墙壁的截线分割成左右两段, 通过对比两侧截线斜率值之间的差异来表征两种校正方法的重建效果。经过计算可知: 基于方向感知网络校正的两侧截线的斜率分别是 0.56、0.59, 差值

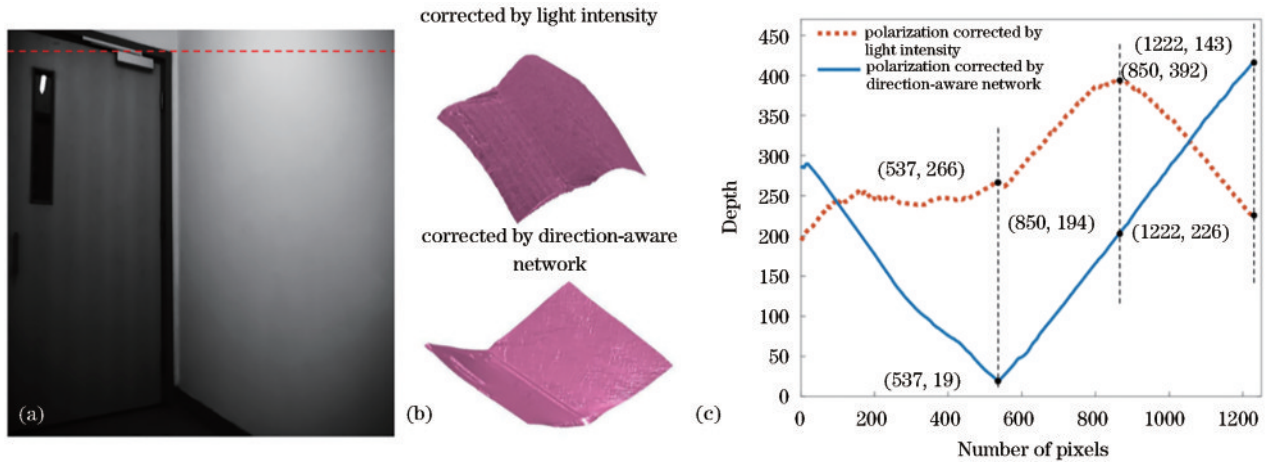


图 9 场景一及其重建结果。(a)原图;(b)重建结果;(c)场景一第 51 行的像素位置

Fig. 9 Scene 1 and 3D reconstruction results of scene 1. (a) Original figure; (b) 3D reconstruction results; (c) location of pixels in line 51 of scene 1

较小;而基于光强校正的两侧截线的斜率分别是 0.40、-0.34,两斜率之间的差值较大。即基于方向感知网络校正后的三维表面形状变化趋势更加符合真实情况。

为了进一步验证基于方向感知网络的场景偏振三维成像方法在不同材料和颜色场景目标上的鲁棒性,选择了细节纹理更加丰富和材料种类更复杂的楼梯场景(场景二)作为目标场景,具体如图 10 所示。

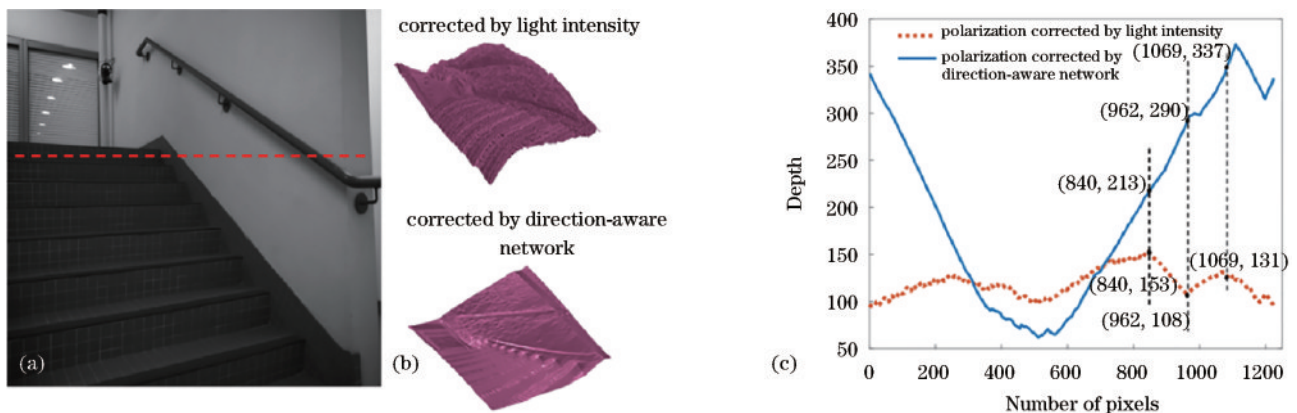


图 10 场景二的三维重建结果。(a)原图;(b)重建结果;(c)场景二第 382 行的像素位置

Fig. 10 3D reconstruction results of scene 2. (a) Original figure; (b) results reconstructed; (c) location of pixels in line 382 of scene 2

由图 10 场景二基于光强校正偏振、基于方向感知网络校正偏振的三维重建结果对比可知,在图 10(a)中右侧楼梯扶手存在局部的弱光,导致基于光强校正偏振的三维重建结果表面出现了明显的凹陷畸变,而在扶手两侧区域由于光强增大又出现了凸出畸变。由图 10(b)中基于方向感知网络校正偏振的三维重建结果可以看出,通过所提方法对法线梯度进行校正后楼梯的三维表面在高光和阴影处的凹凸畸变显著减小。

在图 10(c)中,第 962 列为场景二对应的楼梯扶手位置,第 840 列与 1069 列分别为两侧墙壁的位置,基于光强校正的结果在这 3 处的深度值分别为 108、153、131,同一像素位置基于方向感知网络校正的深度值分别为 290、213、337。将扶手处基于光强校正深度结果 108 与两侧墙壁处的深度结果 153、131 进行比较发现:

在弱光区域内基于光强校正的结果深度变化迅速,先减小后增大;而基于方向感知网络校正后扶手处的深度 290 则处于两侧墙壁深度 213、337 之间,表明基于三维表面形状变化平滑,更加符合实际情况。

因此,综合对场景一与场景二的重建结果在三维形状上的分析可知,基于方向感知网络场景级偏振三维成像方法得到的场景一、场景二的三维重建结果相较于基于光强对偏振校正的结果在由光照不均匀产生的畸变问题上有了很大的改善,凸出与凹陷位置的三维形状起伏更符合场景真实情况。为了进一步突出基于方向感知网络场景级偏振三维成像方法在重建细节上的优势,将基于方向感知网络的三维重建结果与基于方向感知网络对偏振进行校正后的三维重建结果进行比对。

图 11 为场景二的三维重建结果细节放大图,对

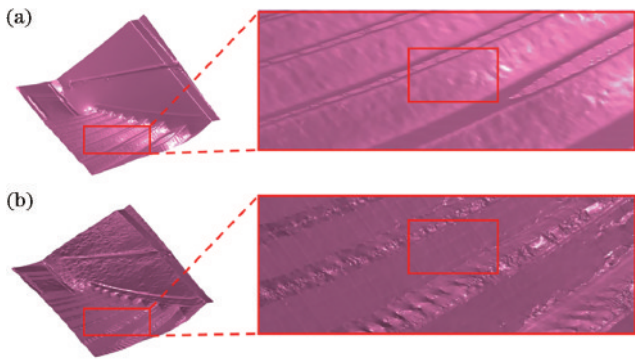


图 11 场景二的三维重建结果细节。(a)方向感知网络重建结果；(b)方向感知网络校正后重建结果
 Fig. 11 Details of the 3D reconstruction results of scene 2. (a) Results reconstructed by direction-aware network; (b) direction-aware network-corrected reconstruction results

比图 10(a)中原图相应部分可以看出,基于方向感知网络对偏振进行校正后的三维重建结果在楼梯台阶上的细节纹理清晰,而基于方向感知网络的三维重建结果则不存在上述细节。因此,基于方向感知网络场景级偏振三维成像方法在保留场景真实三维形貌的同时,引入精度更高的偏振信息,更大程度地还原了场景的细节纹理信息。为了表征该技术的普适性,还对具有更多目标的复杂场景进行了实验,如图 12 所示。

由图 12 可见,在针对具有不同空间位置的多目标场景进行重建时,利用方向感知网络估计深度对偏振进行校正的结果可以保留场景中各目标之间的相对空间位置以及真实三维形貌,同时更大程度上还原目标的细节纹理信息,表明基于方向感知网络的场景法线校正方法具有一定的鲁棒性及普适性。

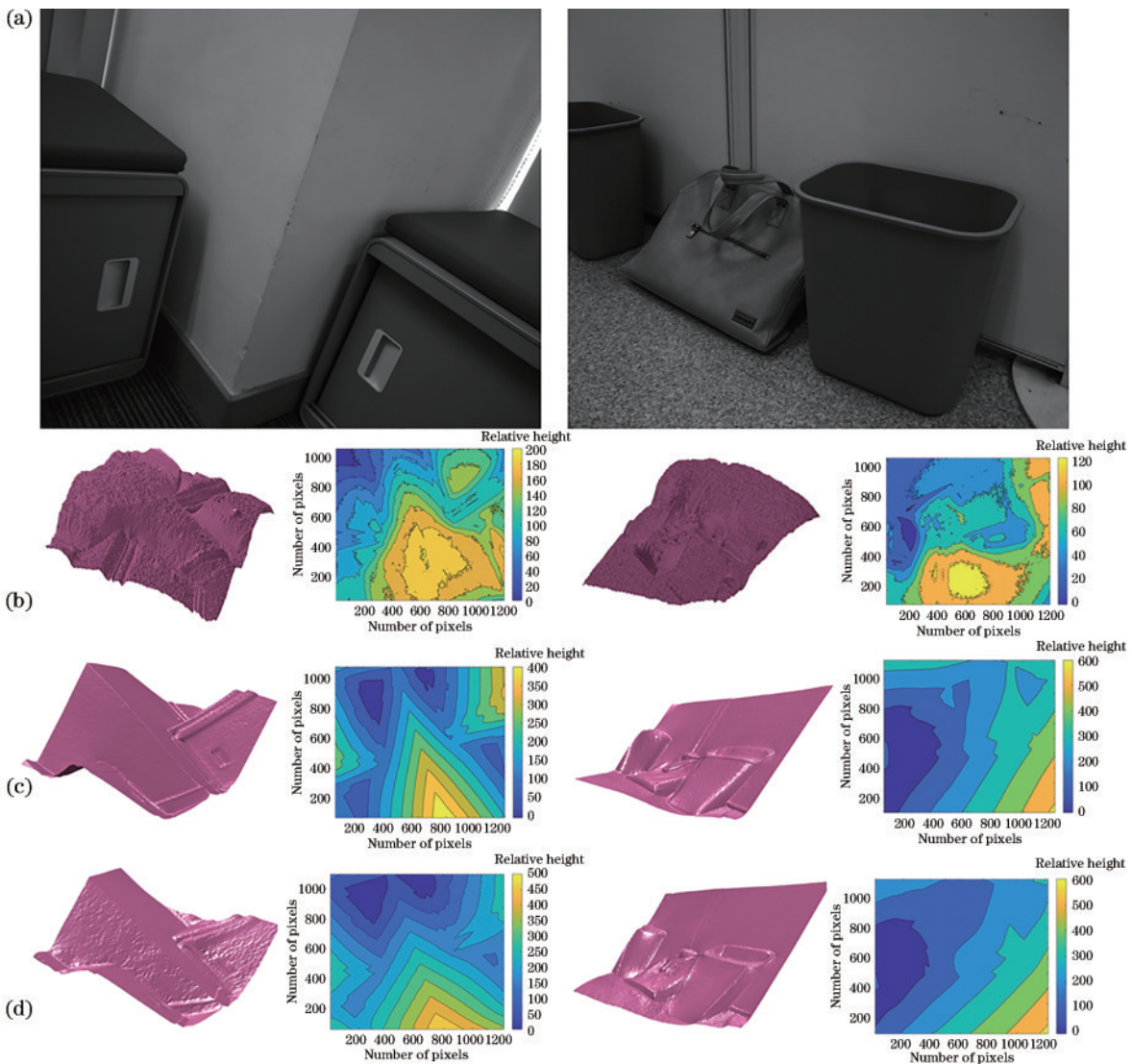


图 12 场景三和场景四的三维重建结果。(a)原图；(b)光强校正后重建结果；(c)方向感知网络重建结果；(d)方向感知网络校正后重建结果
 Fig. 12 3D reconstruction results of scene 3 and scene 4. (a) Original images; (b) reconstruction results corrected by light intensity; (c) reconstruction results of the direction-aware network; (d) reconstruction results corrected by direction-aware network

5 结 论

针对场景级偏振三维成像中由于光照不均匀、材料复杂性、大视场范围造成的法线梯度不准确和三维重建结果存在畸变的问题,研究了一种基于方向感知网络的场景偏振三维成像方法。首先设计场景深度估计网络结构,然后利用该网络对不同场景进行深度估计,再利用所估计深度对偏振法线梯度的方向进行校正,最后通过校正后的梯度进行场景表面的三维重建,实现场景级的偏振三维成像。

所研究的基于方向感知网络的场景偏振三维成像方法能够解决光照不均匀、大视场范围的自然场景中偏振固有的方位角二值性问题和三维重建表面畸变问题,在获取场景真实三维信息的同时保留丰富的纹理细节信息。该研究成果拓宽了偏振三维成像技术的应用领域,为实现高精度的场景级偏振三维成像技术提供了理论方法和实施方案。

参 考 文 献

- [1] 董鹏辉,柯良军. 基于图像的三维重建技术综述[J]. 无线电通信技术, 2019, 45(2): 115-119.
Dong P H, Ke L J. Overview of 3D reconstruction techniques based on images[J]. Radio Communications Technology, 2019, 45(2): 115-119.
- [2] Qin H L, Meng Z H, Meng W, et al. Autonomous exploration and mapping system using heterogeneous UAVs and UGVs in GPS-denied environments[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2): 1339-1350.
- [3] Wong A J N, Wong M, Kutschera P, et al. Dual-energy CT in musculoskeletal trauma[J]. Clinical Radiology, 2021, 76(1): 38-49.
- [4] Chen Y, Zhang B H, Zhou J, et al. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots[J]. Computers and Electronics in Agriculture, 2020, 175: 105579.
- [5] Várady T, Martin R R, Cox J. Reverse engineering of geometric models: an introduction[J]. Computer-Aided Design, 1997, 29(4): 255-268.
- [6] Towers C E, Towers D P, Jones J D C. Absolute fringe order calculation using optimised multi-frequency selection in full-field profilometry[J]. Optics and Lasers in Engineering, 2005, 43(7): 788-800.
- [7] Schwarz B. Mapping the world in 3D[J]. Nature Photonics, 2010, 4(7): 429-430.
- [8] Zhu J J, Wang L, Yang R G, et al. Fusion of time-of-flight depth and stereo for high accuracy depth maps[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK. New York: IEEE Press, 2008.
- [9] Tippetts B, Lee D J, Lillywhite K, et al. Review of stereo vision algorithms and their suitability for resource-limited systems[J]. Journal of Real-Time Image Processing, 2016, 11(1): 5-25.
- [10] Koshikawa K. A polarimetric approach to shape understanding of glossy objects[J]. Advances in Robotics, 1979, 2(2): 190.
- [11] Wolff L B. Surface orientation from polarization images[J]. Proceedings of SPIE, 1988, 0850: 110-121.
- [12] Atkinson G A, Hancock E R. Recovery of surface orientation from diffuse polarization[J]. IEEE Transactions on Image Processing, 2006, 15(6): 1653-1664.
- [13] Li X, Liu F, Han P L, et al. Near-infrared monocular 3D computational polarization imaging of surfaces exhibiting nonuniform reflectance[J]. Optics Express, 2021, 29(10): 15616-15630.
- [14] Zou S H, Zuo X X, Qian Y M, et al. 3D human shape reconstruction from a polarization image[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12359: 351-368.
- [15] Deschaintre V, Lin Y M, Ghosh A. Deep polarization imaging for 3D shape and SVBRDF acquisition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 15562-15571.
- [16] 高鹰. 漫反射偏振三维成像关键技术研究[D]. 西安: 西安电子科技大学, 2017.
Gao Y. Research on key technologies of diffuse reflectance polarization 3D imaging[D]. Xi'an: Xidian University, 2017.
- [17] 孙壮,姜肇国,乔杨,等. 应用于偏振重构方法中的高精度梯度场积分法[J]. 长春理工大学学报(自然科学版), 2020, 43(5): 35-43.
Sun Z, Jiang Z G, Qiao Y, et al. A high-precision gradient field integration method applied to shape from polarization method[J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2020, 43(5): 35-43.
- [18] Cao Y, Wu Z F, Shen C H. Estimating depth from monocular images as classification using deep fully convolutional residual networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(11): 3174-3182.
- [19] Kadambi A, Taamazyan V, Shi B X, et al. Depth sensing using geometrically constrained polarization normals[J]. International Journal of Computer Vision, 2017, 125(1): 34-51.
- [20] Coleman E N, Jain R. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry[J]. Computer Graphics and Image Processing, 1982, 18(4): 309-328.
- [21] Marr D. Vision: a computational investigation into the human representation and processing of visual information[J]. The Modern Schoolman, 1985, 62(2): 141-142.
- [22] Agrawal A, Chellappa R, Raskar R. An algebraic approach to surface reconstruction from gradient fields[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, October 17-21, 2005, Beijing, China. New York: IEEE Press, 2005: 174-181.

- [23] 杨毅明. 数字信号处理[M]. 北京: 机械工业出版社, 2012: 18-30.
Yang Y M. Digital signal processing[M]. Beijing: China Machine Press, 2012: 18-30.
- [24] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer International Publishing, 2015, 9351: 234-241.
- [25] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [26] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [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: 7132-7141.
- [27] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [28] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43 (2): 652-662.
- [29] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence, May 24, 2016, New York: IEEE Press, 2016: 640-651.
- [30] Liu R, Lehman J, Molino P, et al. An intriguing failing of convolutional neural networks and the CoordConv solution[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems, December 3-8, 2018, Montréal, Canada. New York: ACM Press, 2018: 9628-9639.
- [31] Lei C Y, Qi C Y, Xie J X, et al. Shape from polarization for complex scenes in the wild[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12622-12631.
- [32] Ba Y H, Gilbert A, Wang F, et al. Deep shape from polarization[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer International Publishing, 2020, 12369: 554-571.