第 44 卷 第 9 期/2024 年 5 月/光学学报 研究论文

光学学报

基于双螺旋相位板的单目三维编码成像

张越，蔡怀宇*，盛婧，汪毅，陈晓冬

天津大学精密仪器与光电子工程学院光电信息技术教育部重点实验室，天津 300072

摘要 提出一种能够同时获得场景深度信息并实现景深拓展的成像方法。通过在相机的光瞳处引入双螺旋相位调制，将深度信息编码在图像中，使用端到端的深度学习技术对成像过程进行反演，最终得到景深拓展的图像和深度图。分析了相位板参数和物距对成像性能的影响，讨论了在给定的深度范围内合理选择相位板参数的方法。在NYU Depth V2数据集的深度范围内进行了仿真，深度估计相对误差最低可达8.3%，景深拓展后图像的峰值信噪比(PSNR)和结构相似度(SSIM)最高分别可达35.254 dB和0.960，所提方法与传统光学系统相比景深可拓展数十倍以上，并且结果证明了缩小探测范围和增大物距可提升平均深度估计精度。针对闸机人脸识别等潜在的应用场景，以1.1～1.32 m为探测范围搭建了实物系统，在真实场景中深度估计的相对误差为2.2%，所提方法相比传统光学系统景深拓展约10倍。本文方法仅需在传统成像系统中加入一块相位板即可同时实现场景的深度估计和景深拓展功能，在低成本三维成像和检测领域具有一定的应用潜力。

关键词 成像系统；计算成像；单目深度估计；景深拓展；双螺旋相位板；点扩散函数

中图分类号 O438 **文献标志码** A **DOI**：10.3788/AOS231957

1 引 言

自然界中的物体均以三维的形式存在，而传统的成像系统只能将三维世界记录到二维平面上，丢失了深度信息，并且分辨率和景深之间存在相互制约的关系，二者不能兼顾。计算成像作为新一代成像技术，可突破传统成像系统的限制[1-2]，获得更高的分辨率、更大的视场、更远的探测距离和更多维的信息等。计算编码成像是计算成像的一种，它遵循光学编码-计算解码的范式，通过在镜头中加入一个特殊设计的光学元件，对入射光进行振幅或相位调制，将感兴趣的物理量以纹理特征的方式编码在图像中，设计算法对编码图像进行解码，即可重建出想要的信息。

光学编码按调制方式可分为振幅调制[3-4]和相位调制[5]两类，都是通过在孔径平面处放置滤波器并将点扩散函数(PSF)调制为想要的形状来实现，但是振幅调制会降低系统光通量，进而导致系统分辨率和信噪比的降低，而相位调制几乎不影响系统光通量。因此，相位调制逐渐成为主要的调制方法，根据不同的目的可将PSF调制为不同的形式。对于景深拓展，通常将PSF调制为不随离焦变化的形式，使不同深度的物点获得一致性模糊，经过解码复原实现景深的拓展[5-6]。相反，对于深度估计，则希望将PSF调制为离焦敏感的形式，通过PSF的形状估计对应物点的三维位置信息[7]，典型的PSF包括双螺旋点扩散函数[8](DH-PSF)和鞍形点扩散函数[9]等。目前该方法主要应用于三维荧光显微领域[10]中，拍摄荧光分子可近似为点光源成像，直接对PSF的形状进行测量即可，无需使用算法解码。

然而，在日常的拍摄场景中，物点是连续分布的，位于不同深度的物点形成的PSF无限密集地重叠起来，形成编码图像，必须设计解码算法从编码图像中提取信息。Quirin等[11]在相机中同时利用三次相位板和双螺旋相位板进行编码，并利用维纳滤波算法进行解码，同时获得大景深图像和深度图，但两块相位板增加了系统复杂度和成本；Berlich等[12]利用双螺旋点扩散函数进行编码并使用基于倒频谱的算法进行解码，获得了清晰图和深度图，尽管解码算法设计得很精巧，但分块处理的操作造成横向分辨率降低。解码算法的本质就是对成像过程进行反演，近年来，卷积神经网络(CNN)在逆向问题中表现出色。研究人员利用CNN进行解码，并且端到端地进行光学-算法联合设计，在训练解码网络的同时实现相位板面型参数的优化，有效地弥补了传统解码算法设计困难、横向及深度分辨率低等不足。该方法已用于实现深度估计[13-16]、景深拓展[17-18]和高动态范围成像[19-20]等任务中，但目前研究中各任务仍相互独立，在成像特性和光学参数的确定上缺乏讨论，并且完全从数据中无约束地学习到的相

收稿日期：2023-12-21；**修回日期**：2024-02-18；**录用日期**：2024-02-23；**网络首发日期**：2024-03-13

通信作者：*hycai@tju.edu.cn

位板面型奇异，不仅计算复杂且难于加工。

为此，本文提出一种基于双螺旋相位板的三维成像方法。利用DH-PSF随离焦灵敏旋转的特点和双螺旋光束具有更长焦深的特性，仅需一块相位板即可同时实现深度估计和景深拓展成像。在此基础上，分析了相位板参数和物距对成像性能的影响，讨论了根据给定深度范围合理选择相位板参数的方法。在公开数据集上进行了仿真，仿真结果验证了本文方法的深度估计和景深拓展能力，以及成像范围对深度估计精度的影响。此外，相位板实物验证结果表明本文方法在设计范围内可有效地实现景深拓展并得到深度图，证明了本文方法的实用性。

2 双螺旋相位编码成像原理

2.1 双螺旋相位调制原理

根据傅里叶光学理论[21]，在光瞳处引入环式双螺旋相位调制后，非相干成像系统的PSF为

$$h_{\mathrm{I}}(x_{\mathrm{i}},y_{\mathrm{i}})=\left|\mathcal{F}\left\{p(\xi,\eta)\exp\left\{\mathrm{j}\left[\phi^{\mathrm{DF}}(\xi,\eta)+\phi^{\mathrm{M}}(\xi,\eta)\right]\right\}\right\}\right|^2_{f_x=\frac{x_{\mathrm{i}}}{\lambda d_{\mathrm{i}}},f_y=\frac{y_{\mathrm{i}}}{\lambda d_{\mathrm{i}}}},\tag{1}$$

式中：$(x_{\mathrm{i}},y_{\mathrm{i}})$为像面坐标；$d_{\mathrm{i}}$为像距；$(\xi,\eta)$为光瞳面的坐标；$p(\xi,\eta)$为光瞳的形状；$\phi^{\mathrm{M}}(\xi,\eta)$为环式双螺旋相位板引入的相位调制；$\phi^{\mathrm{DF}}(\xi,\eta)$为物体离焦引起的相位偏差。

$\phi^{\mathrm{DF}}(\xi,\eta)$[16]可用归一化坐标表示为

$$\phi^{\mathrm{DF}}(u,v)=\frac{\pi D^2}{4\lambda}\left(\frac{1}{d_{\mathrm{o}}}-\frac{1}{d_{\mathrm{of}}}\right)(u^2+v^2)=\varphi(u^2+v^2),\tag{2}$$

式中：D为系统出瞳直径；(u,v)为归一化的光瞳坐标，$u=2\xi/D$，$v=2\eta/D$；d_{o}为物点的深度；d_{of}为系统对焦平面的位置；φ为离焦的程度。利用高斯公式可将φ转换到像空间计算：

$$\varphi=\frac{\pi D^2}{4\lambda}\left(\frac{1}{d_{\mathrm{o}}}-\frac{1}{d_{\mathrm{of}}}\right)=\frac{\pi D^2}{4\lambda}\left(\frac{1}{d_{\mathrm{if}}}-\frac{1}{d_{\mathrm{i}}}\right)\approx\frac{\pi D^2(d_{\mathrm{i}}-d_{\mathrm{if}})}{4\lambda d_{\mathrm{if}}^2},\tag{3}$$

式中：d_{i}和d_{if}分别为d_{o}和d_{of}对应的成像位置，由于相机拍摄时通常物距远大于像距，因此$d_{\mathrm{i}}d_{\mathrm{if}}$可近似为$d_{\mathrm{if}}^2$。式(3)表明$\varphi$的大小与高斯像面和实际像面之间的位置偏差近似为线性关系，描述了二者之间的偏离程度。当物点O位于对焦平面，传感器放置于$d_{\mathrm{if}}+\Delta$（$\Delta\ll d_{\mathrm{if}}$，$\Delta$为相对于$d_{\mathrm{if}}$位置的一个小偏移量）处成像时，对应的$\varphi$为

$$\varphi=\frac{\pi D^2}{4\lambda}\left(\frac{1}{d_{\mathrm{if}}+\Delta}-\frac{1}{d_{\mathrm{if}}}\right)\approx-\frac{\pi D^2\Delta}{4\lambda d_{\mathrm{if}}^2},\tag{4}$$

表明φ从大到小变化对应的PSF形状与成像光束沿轴向的光强分布相同。

$\phi^{\mathrm{M}}(\xi,\eta)$[22]可用极坐标表示为

$$\phi^{\mathrm{M}}(\rho,\theta)=(2n-1)\cdot\theta,\quad\left(\frac{n-1}{N}\right)^{\gamma}R_{\mathrm{mask}}<\rho\leqslant\left(\frac{n}{N}\right)^{\gamma}R_{\mathrm{mask}},\tag{5}$$

式中：(ρ,θ)为极坐标，$\rho=\sqrt{\xi^2+\eta^2}$，$\theta\in[0,2\pi)$为方位角，$\tan\theta=\eta/\xi$；N为总环数；$n=1,2,\cdots,N$；R_{mask}为相位板半径；$\gamma\in[0,1]$为指数因子。经过调制后，出瞳处出射的会聚球面波在高斯像面附近形成双螺旋光束的形式，光强分布为沿轴向近似线性旋转的两个光斑，可通过相位参数N和γ灵活调节双螺旋光束的旋转速率和主瓣间距（详细讨论见3.1节），如图1(a)所示。

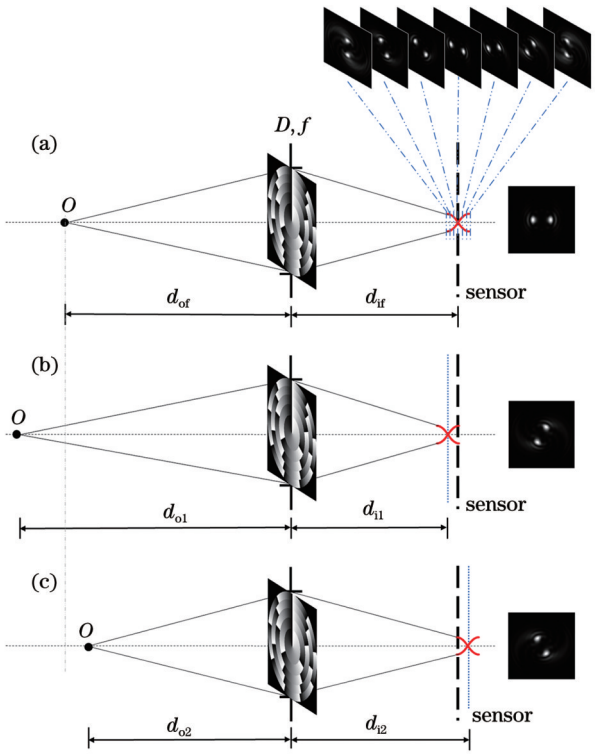

图1 双螺旋相位编码成像系统示意图。(a)无离焦；(b)负离焦；(c)正离焦

Fig. 1 Schematic diagram of double helix phase encoding imaging system. (a) No defocus; (b) negative defocus; (c) positive defocus

加入双螺旋相位调制后，形成的DH-PSF随φ的变化关系与双螺旋光束沿轴向的光强分布相同，如图2所示。在光学系统成像过程中，位于不同深度的物点对应的高斯像面与传感器平面有不同程度的偏离，对应不同大小的φ，所形成的双螺旋光束的不同截面投射到图像传感器上，形成具有不同旋转角度的DH-PSF，如图1(b)、(c)所示，通过测量DH-PSF的旋转角度，推断其离焦大小φ，便可由式(3)得到对应物点的深度d_{o}。

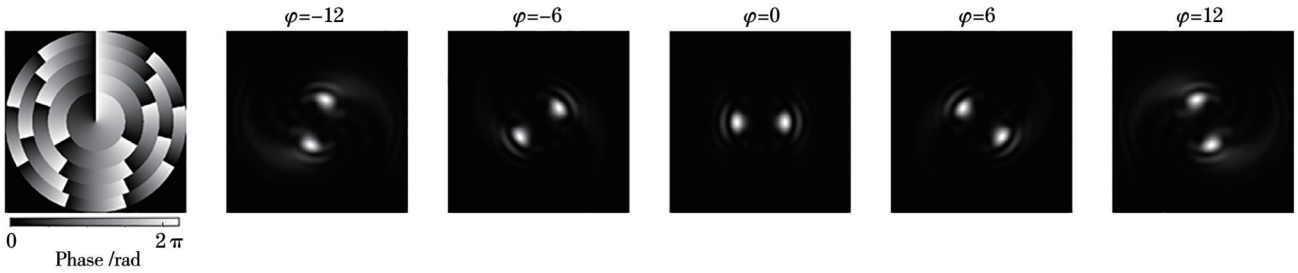

图 2　环式双螺旋相位编码孔径及对应的 PSF

Fig. 2　Annular double helix phase encoded aperture and corresponding PSF

由式(3)也可看出，若经过相位调制后可以在更大的 φ 下保持有效的 PSF 形式，说明调制后的光束具有更大的焦深，可以覆盖传感器所在的位置。根据 Hopkins 准则，传统成像系统的离焦极限 $\varphi \approx 1$[23]，而 DH-PSF 可以在更大的 φ 下保持有效形式。例如，图 2 中的 DH-PSF 在 $\varphi \approx 12$ 时仍保持清晰的双螺旋形式，表明可将焦深拓展 10 倍以上。

2.2　解码与复原算法

实际场景中的物点是连续分布的，对具有不同深度分布的场景进行光学成像的过程可以建模为

$$i = \int_{-\infty}^{+\infty} o(z) * h(z)\,\mathrm{d}z + n, \tag{6}$$

式中：i 为采集的空间离散分布的图像；$o(z)$ 为离散后的物体表面的亮度；n 为一个加性噪声；$h(z)$ 为编码点扩散函数；$*$ 为卷积运算。采用双螺旋相位调制时，PSF 分裂为两个光斑，在编码图像上形成一组孪生像，视觉效果为重影，重影的方向和间隔分别对应编码 DH-PSF 的旋转角度和主瓣间距。各光斑存在一定的横向扩散，使孪生像变得模糊，如图 3 所示。

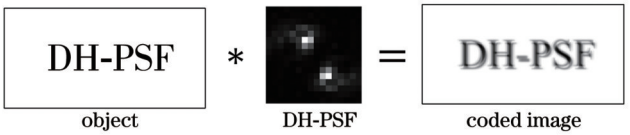

图 3　DH-PSF 编码成像过程示意图

Fig. 3　Schematic diagram of DH-PSF encoded imaging process

真实成像过程中位于不同深度的物点形成的重影特征密集地重叠起来，因此，需要利用解码算法根据图像局部的重影方向和间隔推断离焦大小 φ，进而由式(3)计算场景物点的深度，同时对编码图像进行复原以实现在更大景深范围内的清晰成像。

解码与复原算法的本质是根据图像特征对光学成像过程进行反演，因此首先需要通过计算模拟编码成像的过程，对应图 4 中的光学层(光学层中对应数字的单位为 m)。为便于计算，对式(6)进行离散化处

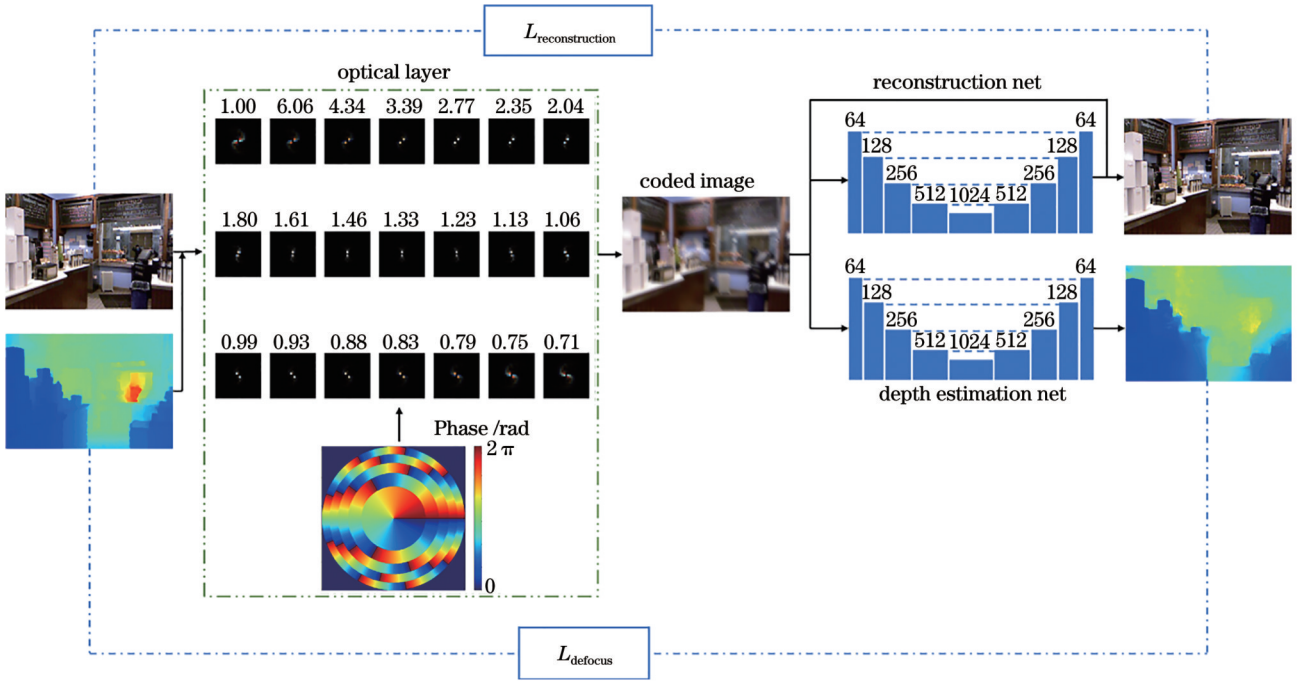

图 4　DH-PSF 编码成像系统联合优化算法框架

Fig. 4　Joint optimization algorithm framework for DH-PSF encoded imaging system

理，得到分层的景深模型[16]：

$$I_\lambda=\sum_{j=1}^{K}\left(L_\lambda * h_{\lambda,j}\right)\cdot M_j+n, \tag{7}$$

式中：L_λ是波长为λ的清晰图像；K为分层的数目；$h_{\lambda,j}$为对应第j个φ值，波长为λ的编码PSF；M_j为第j个φ值对应深度图的二值掩模，由每张清晰图像的深度标签形成，对于每个像素，有

$$\sum_j M_j=1。 \tag{8}$$

解码与复原算法采用经典的逐像素预测网络U-Net实现[24]，如图4所示。解码算法的输入为编码图像，输出为深度图，在U-Net输出层使用1×1卷积和sigmoid函数输出范围为(0,1)的单通道图像。复原算法的输入同样为编码图像，输出为景深拓展的清晰图像。不同于解码算法，复原算法采用残差的方式使网络学习清晰图像和编码图像之间的差异：一方面，清晰图像和编码图像之间的差异较小，网络更容易训练且能够更快收敛；另一方面，残差结构具有恒等映射特性，有利于高频信息的恢复。因此复原算法的U-Net输出层使用1×1卷积和Tanh函数输出范围在(−1,1)区间的三通道彩色图像。

3　光学参数对成像性能的影响

不同的应用需求对成像范围有不同的要求，本节详细分析了相位板的环数N和指数γ对成像性能的影响，并在给定的成像范围内给出对应的选择方法。另外，分析了成像范围对深度估计性能的影响。

3.1　相位板参数对成像性能的影响

1）环数N

本文利用连通域分割提取出DH-PSF的主瓣区域，通过计算质心之间的角度来表征DH-PSF的旋转角度。使用Python计算出$\gamma=0.7$下，5～13环DH-PSF的旋转角度与离焦φ之间的关系，如图5(a)所示，可以看出不同环数的DH-PSF随φ均表现为范围略小于180°近似线性的旋转，环数越多则旋转速率越低。在图示范围之外，两主瓣的相对位置趋于不变，旁瓣逐渐变大，最终双螺旋形式破裂，无法利用。图5(b)为不同γ值下，可保持有效双螺旋形式的最大离焦和环数N之间的关系，图5表明环数N越大，DH-PSF可在越大的离焦范围内保持有效形式，结合式(3)，也意味着对焦深的拓展能力越强。环数不改变总的旋转范围，所以对深度估计精度无显著影响。

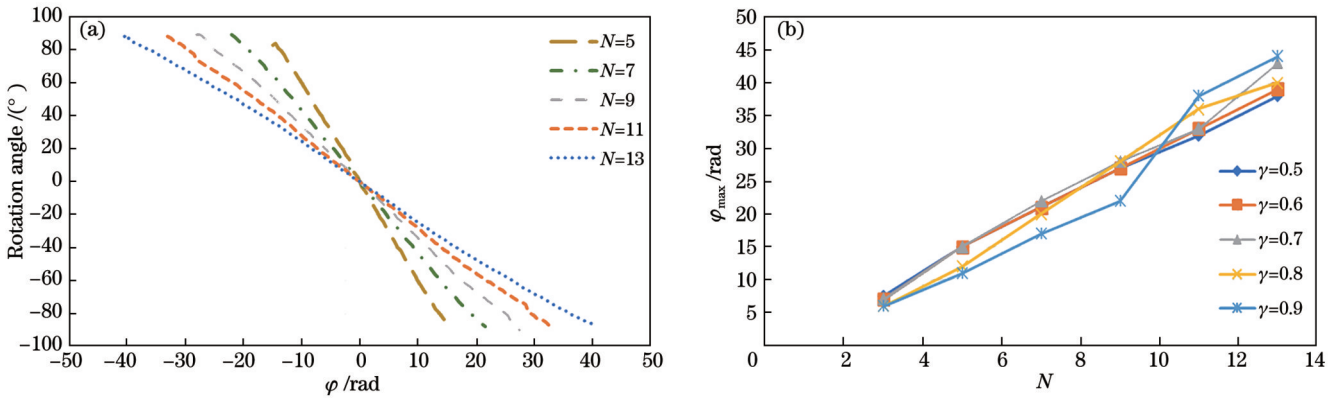

图5　环数对成像特性的影响。(a) $\gamma=0.7$下不同环数的DH-PSF旋转角度和离焦的对应关系；(b)不同γ下环数和保持有效DH-PSF的最大离焦的关系

Fig. 5　Influence of the number of rings on imaging characteristics. (a) Corresponding relationship between the rotation angle and defocus of DH-PSF for different numbers of rings with $\gamma=0.7$; (b) relationship between maximum defocus for maintaining effective DH-PSF and number of rings under different γ

为了将DH-PSF接近180°的旋转范围充分地映射到待测深度区间$[d_{\text{omin}},d_{\text{omax}}]$内，可列出如下方程组：

$$\begin{cases}\varphi_{\max}=\dfrac{\pi D^2}{4\lambda}\left(\dfrac{1}{d_{\text{omin}}}-\dfrac{1}{d_{\text{of}}}\right)\\ \varphi_{\min}=\dfrac{\pi D^2}{4\lambda}\left(\dfrac{1}{d_{\text{omax}}}-\dfrac{1}{d_{\text{of}}}\right)\end{cases}, \tag{9}$$

式中：$\varphi_{\max}$和$\varphi_{\min}$是DH-PSF的离焦极限，互为相反数。由式(9)可得

$$\varphi_{\max}-\varphi_{\min}=\frac{\pi D^2}{4\lambda}\left(\frac{1}{d_{\text{omin}}}-\frac{1}{d_{\text{omax}}}\right), \tag{10}$$

在实际应用中，D的大小由分辨率的要求决定，相位板半径R_{mask}也随之确定，为使景深范围拓展为$[d_{\text{omin}},d_{\text{omax}}]$，应根据图5(b)选择合适的环数$N$，即合适的焦深拓展倍率，使式(10)成立。

2）指数γ

指数γ决定了双螺旋相位板环半径的分布，γ越接近1，各环半径分布越均匀。γ主要影响了DH-PSF两主瓣间距以及在不同离焦下形状的一致性。计算了不同环数的DH-PSF两主瓣间距在无离焦下随γ的变化，如图6(a)所示，图6(a)表明γ越大，两主瓣间距越

大。由于数字图像是对光强分布的离散采样，越大的间距在数字图像上的角度分辨率越高，因此较大的γ值有助于深度分辨能力的提高。然而，较大γ值对应的旋转一致性较差。分别计算了5环双螺旋相位板在0.5、0.7和0.9三个不同γ值下的DH-PSF，如图6(b)所示。从图中可以看出，$\gamma=0.5$时在对焦和有较大离焦下的DH-PSF除有一定角度的旋转外，形状上无明显区别，而当γ增大时，各相环产生的光学涡旋在轴上的旋转速率差异增大[22]，随离焦的增大光斑更加分散，如图6(b)中$\varphi=12$对应的图像所示，这不利于深度估计。因此γ不宜太大也不宜太小，实际应用中，可以将环式双螺旋相位板近似为可微的形式[14]：

$$P(\rho,\theta)=\sum_{n=0}^{n=N}P_n(\theta)\times\frac{1}{2}\left\{\tanh\left\{100\left[\rho-R(n-1)\right]\right\}-\tanh\left\{100\left[\rho-R(n)\right]\right\}\right\},\tag{11}$$

式中：$P_n(\theta)=(2n-1)\theta$为第n环的相位分布；$R(n)=(n/N)^{\gamma}R_{\mathrm{mask}}$为第$n$环的半径大小。如式(11)所示，将$\gamma$作为图4光学层中一个可优化的权重，在训练解码算法的同时，端到端地学习γ的取值。

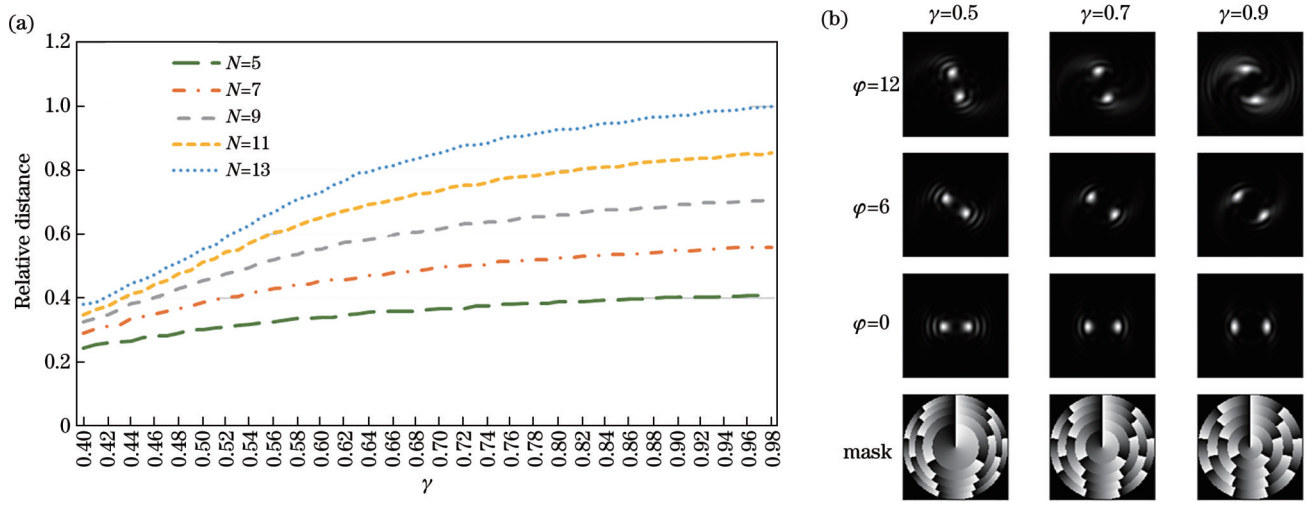

图6　γ对成像特性的影响。(a)不同环数的双螺旋相位板在对焦时两主瓣间的相对距离与γ之间的关系；(b)不同γ值的五环双螺旋相位板在不同离焦下的DH-PSF

Fig. 6　Influence of γ on imaging characteristics. (a) Relationship between the relative distance of the two main lobes of double helix phase mask with different numbers of rings and γ under focus; (b) DH-PSF of five-ring double helix phase mask with different γ under different defocus

3.2　成像范围对深度估计性能的影响

由式(9)可得对焦位置d_{of}为

$$d_{\mathrm{of}}=d_{\mathrm{omin}}+\left(\frac{1}{2}-\frac{1}{4\times\dfrac{d_{\mathrm{omin}}}{d_{\mathrm{omax}}-d_{\mathrm{omin}}}+2}\right)(d_{\mathrm{omax}}-d_{\mathrm{omin}})。\tag{12}$$

受成像特性限制，双螺旋光束在像方沿轴向近似线性旋转，但映射到物空间后，DH-PSF的旋转角度与物距d_{o}呈反比例函数关系，造成灵敏度在深度上分布不均匀，远处相当大的深度区间可能占有的旋转量很小(图7)，降低了平均深度估计精度。可定义一个比例因子α，表征旋转角度与深度之间的非线性程度：

$$\alpha=\frac{1}{2}-\frac{1}{4\times\dfrac{d_{\mathrm{omin}}}{d_{\mathrm{omax}}-d_{\mathrm{omin}}}+2},\tag{13}$$

α为d_{of}在$[d_{\mathrm{omin}},d_{\mathrm{omax}}]$中的相对位置。$\alpha$越接近0.5，线性度越好；$\alpha$越接近0，线性度越差。由式(13)可知，$\alpha$的理论上限为0.5，即完全线性，对应$d_{\mathrm{omin}}$为无穷大或$d_{\mathrm{omax}}-d_{\mathrm{omin}}$为无穷小，这无实际意义。在实际应用中，增大$d_{\mathrm{omin}}$或减小估计范围可实现平均精度的提升。合理设置待探测的深度范围可以提高α，提高DH-PSF旋转角度与物点深度之间的线性度，进而提高平均估计精度。

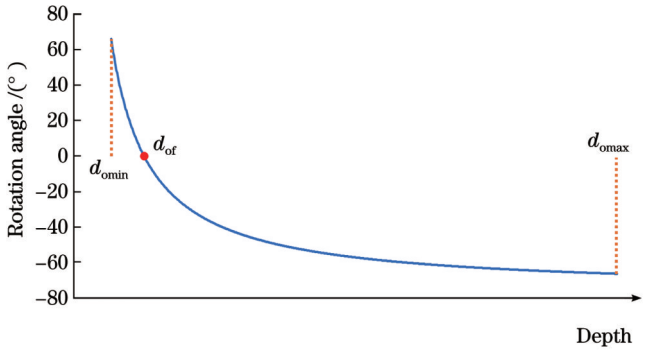

图7　DH-PSF旋转角度与深度之间的关系

Fig. 7　Relationship between DH-PSF rotation angle and depth

4　结果分析与讨论

4.1　仿真

4.1.1　仿真设置

1）数据集

由2.2节可知，本文方法的本质是识别由物体离焦和引入的双螺旋相位形成的重影特征推断深度，与具体的场景无关，因此可选用非真实场景但数据量充足的合成数据集FlyingThings3D[25]进行训练和测试。为验证本文方法在真实场景中的有效性和泛化性，使用真实场景拍摄的NYU Depth V2[26]数据集对训练好的模型进行测试。NYU Depth V2由微软Kinect的RGB-D相机采集的真实室内场景图像及对应深度图组成。

2）评价指标

本文使用单目深度估计常用的性能指标评价训练结果，包括均方根误差（RMS）、绝对相对误差（REL）、对数均方根误差（Log10）和阈值范围内的深度估计精度（常用的阈值为1.25、1.25^2和1.25^3）[27]，其中RMS为与尺度相关的指标，其余均为与尺度无关的指标，可用于评价不同深度范围下的性能。图像复原采用峰值信噪比（PSNR）[28]和结构相似性（SSIM）[29]评价。

3）损失函数

深度估计的损失函数L_{depth}由均方根损失L_{RMS}和梯度损失L_{grad}组合而成：

$$L_{\text{depth}}=L_{\text{RMS}}+L_{\text{grad}}。\tag{14}$$

均方根损失使估计值逼近真实值的大小，即

$$L_{\text{RMS}}=\frac{1}{\sqrt{N}}\left\|\hat{I}-I\right\|_2,\tag{15}$$

式中：$\hat{I}$为预测值；I为真值。

梯度损失使网络学习到清晰的边界，可表示为

$$L_{\text{grad}}=\frac{1}{\sqrt{N}}\left(\left\|\frac{\partial\hat{I}}{\partial x}-\frac{\partial I}{\partial x}\right\|_2+\left\|\frac{\partial\hat{I}}{\partial y}-\frac{\partial I}{\partial y}\right\|_2\right)。\tag{16}$$

图像复原的损失函数$L_{\text{reconstruction}}$采用均方根损失函数：

$$L_{\text{reconstruction}}=L_{\text{RMS}}。\tag{17}$$

4）光学参数的设置

以NYU Depth V2的深度范围[0.71, 9.99] m作为$[d_{\text{omin}}, d_{\text{omax}}]$，首先以$N=5$、$\gamma=0.5$作为初值，使用端到端的网络学习到的$\gamma$值收敛在0.7附近，如图8所示，0.7是一个折中的数值，兼顾了形状不变性和深度分辨率，这符合理论的分析，因此将γ设为0.7。为验证不同景深拓展倍率下的成像性能，在13环相位板的基础上按照环半径缩减光圈，在[0.71 m, 9.99 m]内分别验证了5、7、9、11、13环相位板的成像性能，对应的景深拓展能力逐渐增强，由式(9)计算出的各项光学参数如表1所示。很多照相物镜的结构具有较强的对称性，可校正垂轴像差，入瞳直径和出瞳直径数值相近，因此以出瞳直径代替入瞳直径计算F数，焦距$f'=50$ mm。不同环数相位板在不同深度下的PSF如图9所示，可以看出对应位置DH-PSF旋转角度基本一致。

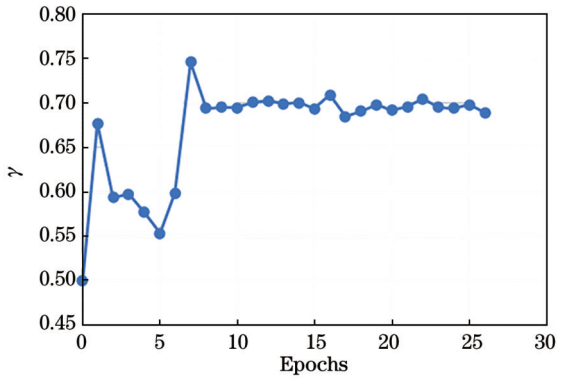

图8　端到端联合优化过程中γ随epochs的变化曲线

Fig. 8　Variation of γ with epochs during end-to-end joint optimization process

表1　不同环数下的光学参数

Table 1　Optical parameters for different numbers of rings

Number of rings	$\varphi_{\max}$	Pupil diameter /mm	Focusing position /m	F number
5	11.2	3.4	1.3084	14.7
7	17.6	4.2	1.3084	11.9
9	25.5	5.1	1.3084	9.8
11	33.0	5.8	1.3084	8.6
13	43.0	6.6	1.3084	7.6

5）训练过程

采用随机裁剪和随机翻转等几何变换以及亮度、对比度、饱和度等随机颜色变换对数据进行增强，采用

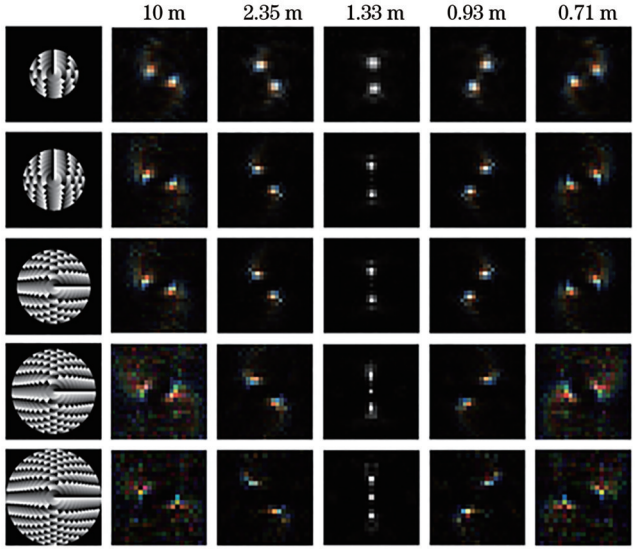

图9　不同光圈大小的双螺旋相位板在不同深度处的PSF

Fig. 9　PSF of double helix phase masks with different aperture sizes at different depths

Adam优化器(学习率为0.001,其他设置保持默认)对模型进行优化,光学层分为21层,batchsize为16,深度估计网络在NVIDIA GeForce GTX 3090(24 GB)上训练约60轮后收敛,图像复原网络训练约20轮后收敛。

4.1.2　深度估计性能验证

使用NYU Depth V2对在FlyingThings3D上训练好的模型进行测试,并以对焦位置进行划分,将小于和大于对焦位置的区间分别标记为near和far,并分段进行测试,测试结果如表2所示,表2中V_{near}和V_{far}分别代表near和far区间内深度估计的指标。可以看出,小于对焦位置的区间内估计误差更小,这与图7中的变化率一致。整体指标主要由大于对焦位置的区间决定,因为数据集中绝大多数像素的深度都在大于对焦位置的区间内。环式DH-PSF编码成像与目前优秀的同类方法性能相近,特别是小环数下表现出略优的性能,随着环数的增多,深度估计精度略有下降。从图9可以看出,虽然环数的增加会增大两主瓣间距、提高深度分辨能力,但同时会使DH-PSF在离焦较大时主瓣明显分散,导致编码图像变模糊,解码效果变差。因此,在达到分辨率要求的前提下,若景深拓展后能覆盖待测的深度范围,应使用小环数相位板,以获得更高的深度估计精度。模型在FlyingThings3D和NYU Depth V2的测试结果分别如图10和图11所示,虽然并未在

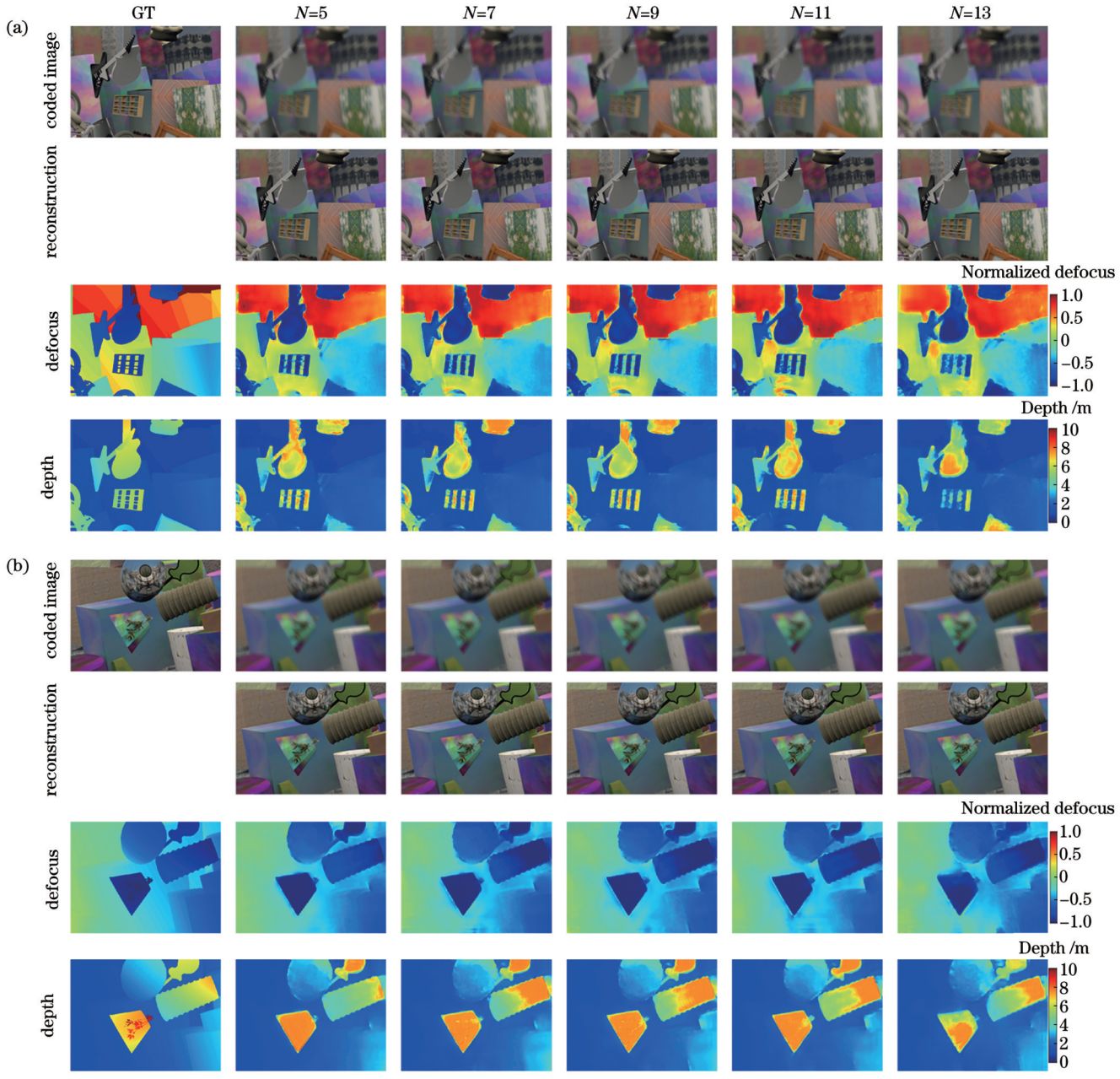

图10　不同环数双螺旋相位板在FlyingThings3D上的深度估计与景深拓展图像复原效果图。(a)场景1;(b)场景2

Fig. 10　Depth estimation and recovery images with extended depth for double helix phase masks with different numbers of rings on FlyingThings3D dataset. (a) Scene 1; (b) scene 2

表 2　不同环数下的深度估计结果评价

Table 2　Evaluation of depth estimation results for different numbers of rings

Parameter	Error			Accuracy		
	RMS	REL (V_{near}, V_{far})	Log10 (V_{near}, V_{far})	Threshold is 1.25	Threshold is 1.25^2	Threshold is 1.25^3
PhaseCam3D[16]	0.382	0.093	0.050	0.932	0.989	0.997
DeepOptics[13]	0.433	0.087	0.052	0.930	0.990	0.999
N=5	0.392	0.083 (0.075, 0.084)	0.047 (0.035, 0.047)	0.947	0.995	0.999
N=7	0.398	0.091 (0.078, 0.091)	0.050 (0.037, 0.050)	0.944	0.994	0.998
N=9	0.403	0.093 (0.084, 0.092)	0.051 (0.039, 0.050)	0.943	0.993	0.998
N=11	0.428	0.098 (0.084, 0.097)	0.054 (0.040, 0.053)	0.933	0.992	0.997
N=13	0.421	0.100 (0.083, 0.099)	0.056 (0.041, 0.054)	0.937	0.991	0.997

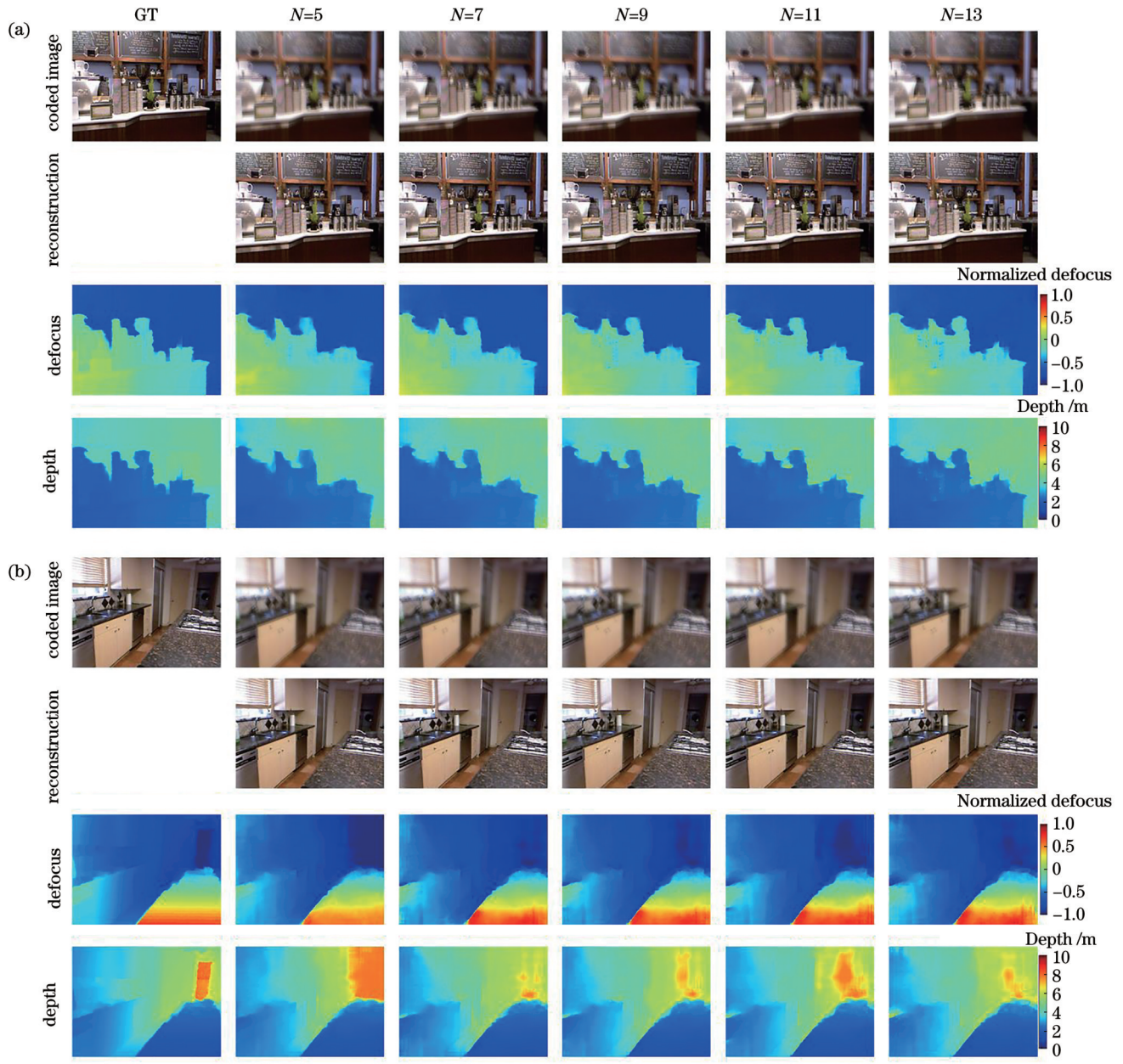

图 11　不同环数双螺旋相位板在 NYU Depth V2 的测试结果。(a)场景 1;(b)场景 2

Fig. 11　Test results of double helix phase masks with different ring numbers on NYU Depth V2 dataset. (a) Scene 1; (b) scene 2

NYU Depth V2上进行训练，但仍有出色表现，证明本文方法不依赖场景的高级语义信息，具有优秀的泛化性，并且在不同的景深拓展倍率下均有良好的深度估计能力。

4.1.3 景深拓展性能验证

不同的环数N对应了不同的景深拓展倍率，对于标准孔径，按照离焦极限$\varphi \approx 1$进行计算，可得在表1所示的对焦位置和出瞳直径下，对应的景深范围分别为0.201、0.1313、0.089、0.0688、0.0531 m，对应的景深拓展倍率分别为46、70、104、134、174，可见环数越多则景深拓展能力越强。使用NYU Depth V2对不同环数下的图像复原性能进行测试，并与$\alpha=60$和$\alpha=80$的三次相位板进行了比较，指标如表3所示。$\alpha=60$和$\alpha=80$下三次相位板的拓展能力分别与$N=9$和$N=11$的双螺旋相位板相近，因此使用表1中对应的光学参数进行训练。可以看出双螺旋相位板的复原效果略差于拓展能力相近的三次相位板，因为三次相位板的PSF几乎不随离焦变化，解码难度更低，但丢失了深度信息，无法实现深度估计。环式双螺旋相位板的复原质量表现出随环数的增大而降低的趋势，与深度解码类似，图像复原也是根据重影的方向和间隔对成像过程进行反演的过程，因此，大环数相位板在离焦较大时主瓣分散造成的图像模糊同样会影响复原的效果。如图10和图11所示，即使是13环，也可实现高质量的复原，这表明在不同的景深拓展倍率下均可实现清晰成像。

表3 不同环数下图像复原结果

Table 3 Image reconstruction results for different numbers of rings

Method	PSNR /dB	SSIM
$N=5$	35.254	0.960
$N=7$	34.508	0.956
$N=9$	33.239	0.943
$N=11$	33.560	0.949
$N=13$	33.346	0.947
$\alpha=60$	34.497	0.954
$\alpha=80$	35.565	0.955

4.1.4 不同成像范围下深度估计性能验证

为验证成像范围对深度估计精度的影响，调整了数据集的深度范围，以$N=5$、$\gamma=0.7$的相位板为例进行测试，测试结果与原始深度范围的估计精度的对比如表4所示。表4中的5个范围对应的α分别为6.66%、15.07%、24.48%、27.58%和34.88%，可以看出，随着线性度的提升，对焦位置两侧的误差逐渐接近，整体的平均估计精度也随之提升。在此项对比中，RMS是尺度相关的评价指标，无实际意义。绘制了0.71～9.99 m、0.71～2.20 m和10.71～19.99 m三个范围内DH-PSF的旋转角度和深度之间的关系曲线，如图12所示，可以看出成像范围的调整使DH-PSF旋转角度在对应的成像范围内的线性度得到较大提升。对应地，从图13中可以看出线性度提升后，原来远处PSF变化不灵敏导致的较大估计误差得到了明显的改善，这证明缩小深度范围或增大物距可提升深度估计的平均精度。

表4 不同范围对深度估计性能的影响

Table 4 Influence of different ranges on depth estimation performance

α /%	Range /m	Error			Accuracy		
		RMS	REL (near, far)	Log10 (near, far)	Threshold is 1.25	Threshold is 1.25^2	Threshold is 1.25^3
6.66	0.71–9.99	0.392	0.083 (0.075, 0.084)	0.047 (0.035, 0.047)	0.947	0.995	0.999
15.07	0.71–4.00	0.138	0.049 (0.049, 0.052)	0.031 (0.027, 0.032)	0.986	0.998	0.999
24.48	0.71–2.20	0.045	0.028 (0.027, 0.031)	0.018 (0.016, 0.018)	0.997	0.999	1.000
27.58	5.71–14.99	0.298	0.025 (0.024, 0.028)	0.016 (0.015, 0.016)	0.997	0.999	1.000
34.88	10.71–19.99	0.330	0.018 (0.017, 0.018)	0.011 (0.010, 0.010)	0.998	1.000	1.000

4.1.5 局限性分析

本文方法由于需要依赖图像局部的特征推理深度，因此理论上并不适用于没有纹理的场景。如图10和图11所示，对于弱纹理的场景，卷积神经网络也表现出强大的拟合能力，不会失效。然而，如图14所示，过曝等因素造成的纹理完全丢失会导致本文方法无法工作。

4.2 真实场景下的可行性验证

4.2.1 系统搭建

为验证仿真中光学层对实际物理成像过程近似的有效性及本文方法在真实场景下的可行性，针对闸机人脸识别和工业零件检测等潜在应用，以[1.1, 1.32] m为待测范围搭建了系统。镜头采用永诺50 mm F1.8 AF定焦镜头，最近可对焦于0.45 m处，镜头结构为5

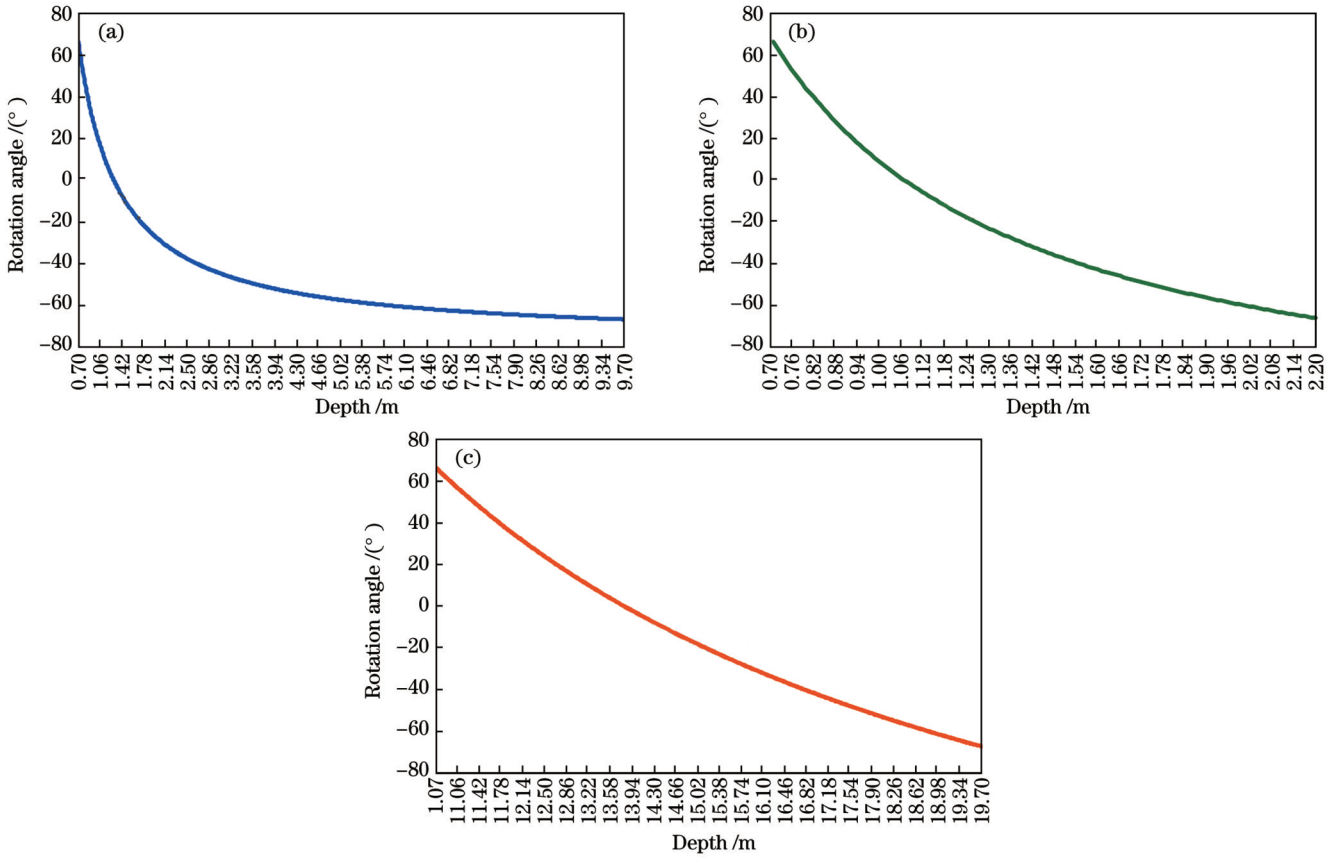

图 12　不同范围内 DH-PSF 的旋转角度和深度之间的关系。(a) 0.71～9.99 m;(b) 0.71～2.20 m;(c) 10.71～19.99 m

Fig. 12　Relationship between rotational angle of DH-PSF and depth in different ranges. (a) 0.71–9.99 m; (b) 0.71–2.20 m; (c) 10.71–19.99 m

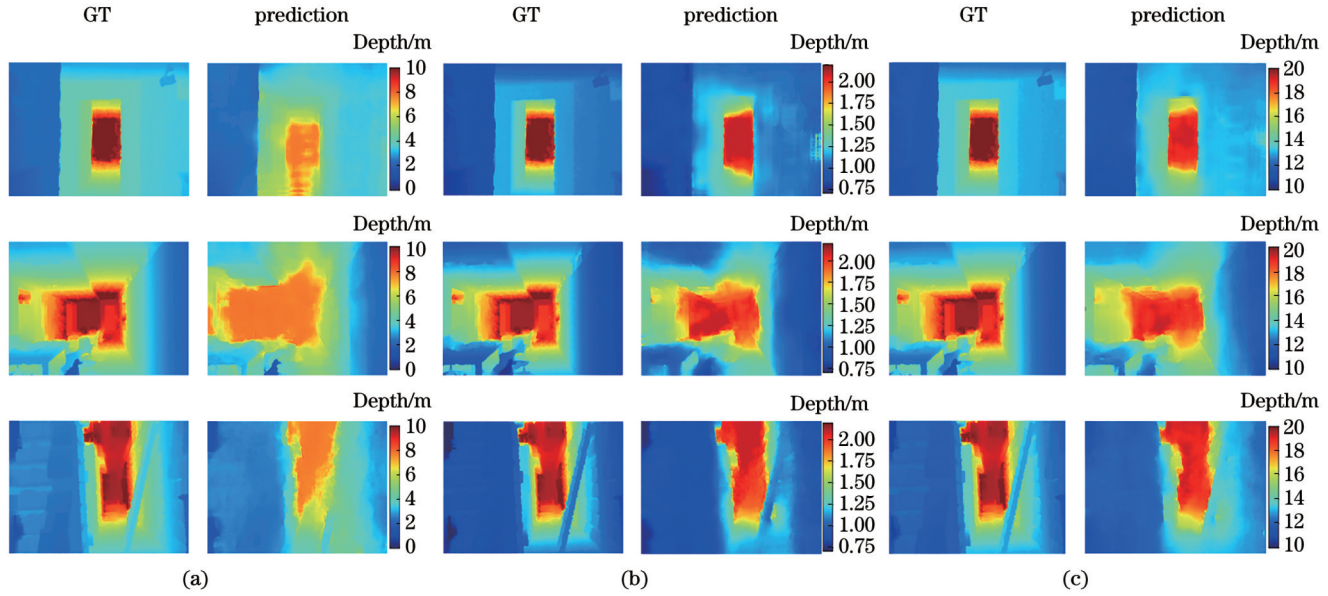

图 13　不同范围对深度估计性能的影响。(a) 0.71～9.99 m;(b) 0.71～2.20 m;(c) 10.71～19.99 m

Fig. 13　Influence of different ranges on depth estimation performance. (a) 0.71–9.99 m; (b) 0.71–2.20 m; (c) 10.71–19.99 m

组 6 片式双高斯照相物镜，如图 15(a)所示，孔径光阑位于两半部分中间，便于改装。采用工业相机 MER-2000-19U3M/C(-L)作为传感器，分辨率为 5496×3672，像元尺寸为 2.4 μm。使用 5 环进行调制，对应的出瞳直径为 10 mm，相位板的实际直径为 8 mm，对焦位置为 1.2 m。由于深度解码网络学习的是 DH-PSF 形状与离焦大小 φ 之间的映射关系，该关系不随深度范围的变化而变化，通过改变网络超参数 D 和 d_{of} 的值

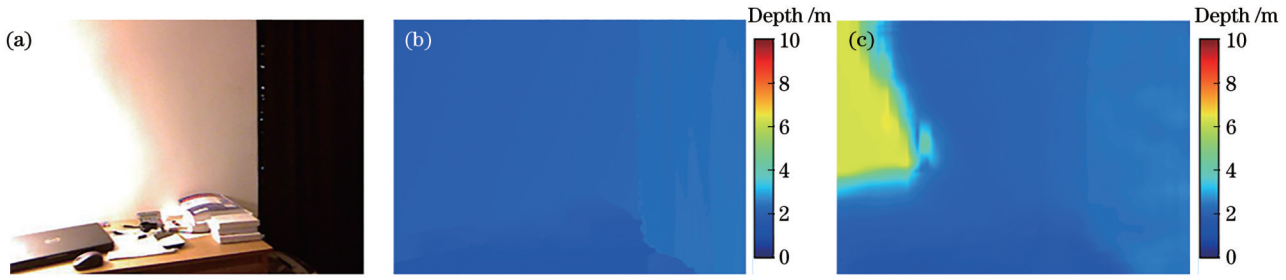

图 14　过曝对深度估计性能的影响。(a)清晰图像;(b)深度真值图;(c)预测的深度图

Fig. 14　Influence of overexposure on depth estimation performance. (a) Clean image; (b) depth ground truth; (c) predicted depth map

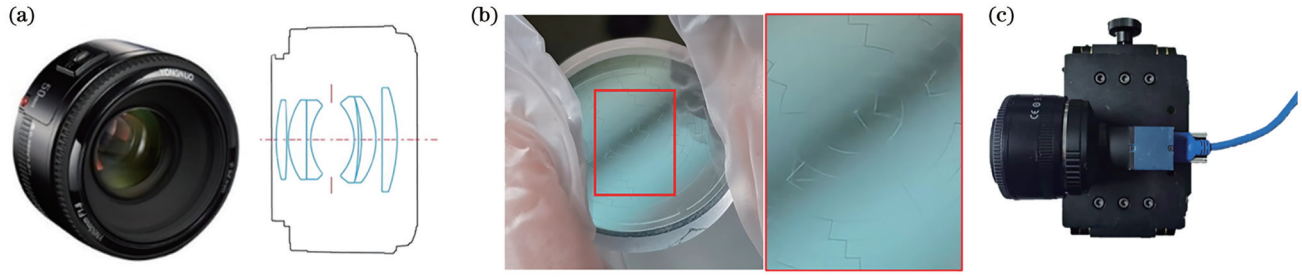

图 15　相位板与成像系统图。(a)永诺 YN50 mm F1.8 标准定焦镜头;(b)加工好的环式双螺旋相位板;(c)双螺旋编码成像实物系统

Fig. 15　Phase mask and imaging system diagram. (a) Yongnuo YN50 mm F1.8 standard prime lens; (b) processed ring-type double helix phase mask; (c) double helix encoded imaging system

即可将离焦大小映射到对应的深度区间,因此可直接利用仿真中的 5 环解码网络,图像复原网络同理。环式双螺旋相位板是一种多拓扑荷数的涡旋相位板,出于成本和精度的考虑,利用文献[30]所提出的“一次曝光制备连续表面涡旋相位板”的方法进行加工,得到的相位板如图 15(b)所示。将双螺旋相位板固定在镜头孔径光阑处,得到双螺旋相位编码成像系统,如图 15(c)所示。

4.2.2　PSF 标定

图 16(a)为不同深度下的仿真 DH-PSF。由于训练好的模型是基于仿真的 PSF 得到,而仿真仅考虑了离焦的影响,没有考虑其他像差以及相位板加工和装配误差等对 PSF 形状的影响。为消除仿真和真实物理系统之间的差异,需要采集真实的 DH-PSF 取代仿真 DH-PSF 对训练好的网络进行微调。

用白光光源照射光学微孔的方法模拟点光源,使用搭建的编码成像系统在光学暗室中按照仿真的各个 DH-PSF 的离焦值对应的深度采集图像,得到标定的 PSF,如图 16(b)所示。可以看出,虽然深度范围不同,但由于各个 DH-PSF 的离焦大小 φ 是一样的,因此旋转角度相同。图 16(b)中实际 PSF 的色差更加明显,除镜头本身会引入色差外,相位板的石英材质对不同波长有不同的折射率,对不同波长入射光的调制效果不同,也会引入色差。由于深度估计的本质需求是 PSF 随深度尽可能地变化明显,而色差的存在引入了颜色通道的差异,提高了变化的自由度,因此适量的色差无需消除。

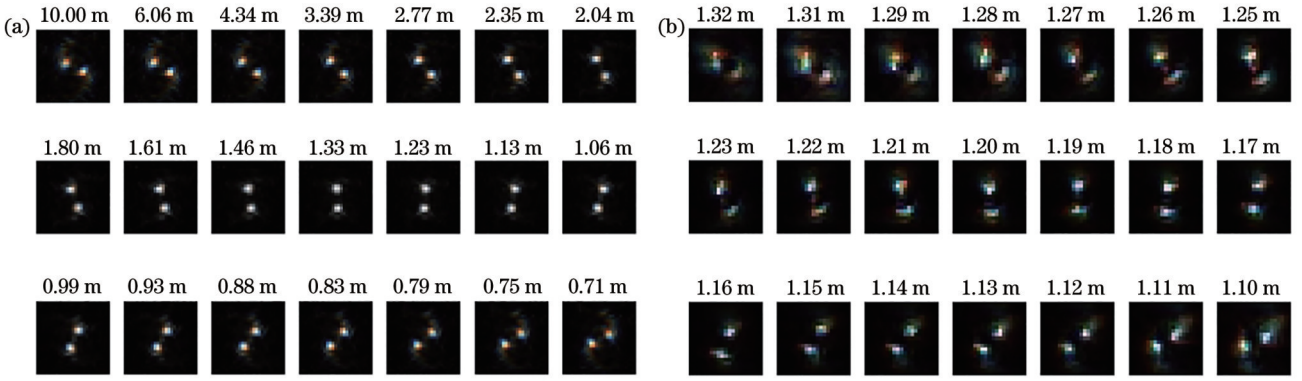

图 16　仿真 DH-PSF 和实际采集的 DH-PSF 对比图。(a)不同深度下的仿真 DH-PSF;(b)不同深度下实际采集的 DH-PSF

Fig. 16　Comparison of simulated DH-PSF and actually acquired DH-PSF. (a) DH-PSF simulated at different depths; (b) actually captured DH-PSF at different depths

4.2.3　真实场景成像实验

利用实际采集的DH-PSF对深度估计网络和图像复原网络在FlyingThings3D上进行微调后，使用双螺旋编码成像系统拍摄实际场景，得到编码图像，将编码图像输入到两个网络中，得到景深拓展后的清晰图像和深度图，如图17所示，实际采集的图像与训练数据集完全不同，再次证明了本文方法优秀的泛化性。从图17中可以看出复原后的清晰图像中物体边缘明显变得锐利，纹理也变清晰了。图17(c)中编码图像的橙子边缘表现出明显的重影特征，复原后重影被消除，边缘恢复清晰。实验结果表明，经过图像复原，在设计的范围内均可实现清晰成像。计算可得，相同大小的标准孔径对应的景深范围约为2 cm，景深的拓展倍率约为10倍。

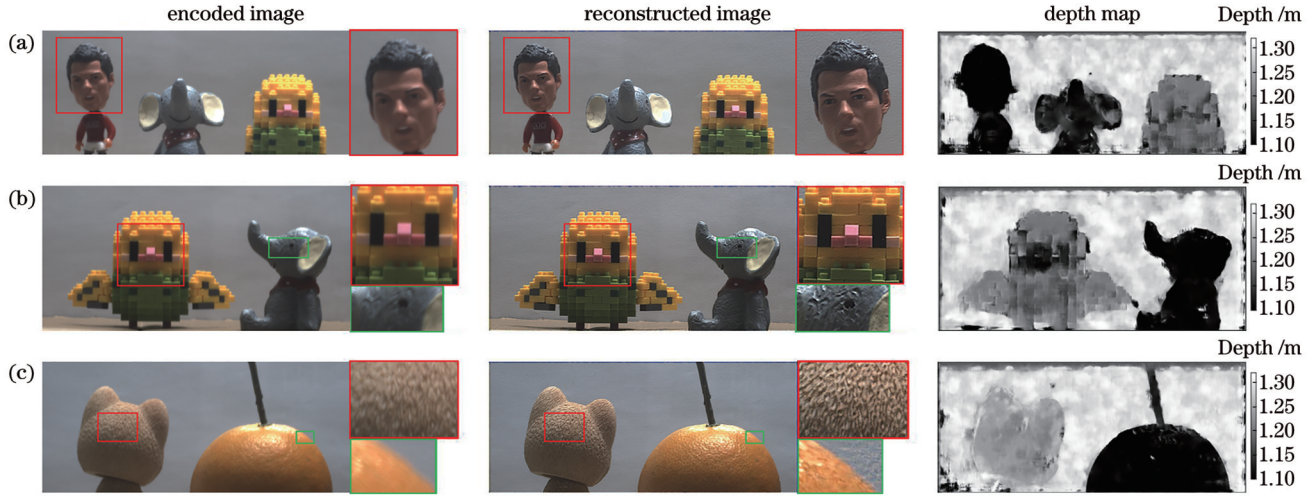

图17　双螺旋编码成像系统在真实场景下的表现。(a)场景1；(b)场景2；(c)场景3

Fig. 17　Performance of the double helix encoded imaging system in real scenes. (a) Scene 1; (b) scene 2; (c) scene 3

从图17的深度图中可明显看出各物体的前后关系，利用平面物体估计系统输出深度的精度，以卷尺测量值作为真值，真值与输出值的差异如图18所示，计算出的相对误差为2.2%，与数据集表现接近，证明了本文深度估计方法在真实场景下的可行性和有效性。但其中存在个别区域深度错误、平面区域深度值波动的问题，对系统进行多次调试后发现出现该问题的原因在于噪声的影响。使用$N=5$、$\gamma=0.7$的相位板分别测试了无噪声、加入$\sigma=0.005$和$\sigma=0.01$的高斯噪声后的各项指标，如表5所示。可以看出算法性能随噪声水平的增大而下降，深度图中会出现明显波动，错误点增多，复原图像中出现细微的伪影，图像质量下降，但算法仍可以有效工作，如图19所示。真实场景中的噪声情况更加复杂，因此对编码图像进行了高斯滤波以抑制噪声。通过拍摄两个不同间距的平面物体，测试了系统的深度分辨率。图20为两个平面物体

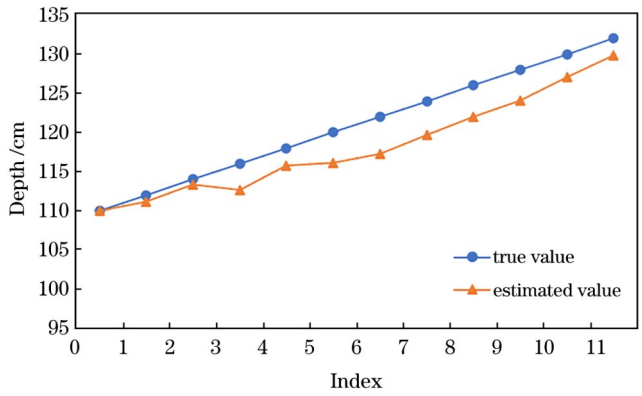

图18　系统实测的不同深度下的估计值与真值差异

Fig. 18　Difference between estimated values measured by the system and true values at different depths

间隔1 cm时的深度估计情况，图20(b)中虚线框内各列均值的曲线如图20(c)所示，从中可清晰看出两平

表5　噪声对深度估计和图像复原结果的影响

Table 5　Influence of noise on depth estimation and image reconstruction results

Noise	Depth estimation						Reconstruction	
	RMS	REL	Log10	Threshold is 1.25	Threshold is 1.25^2	Threshold is 1.25^3	PSNR /dB	SSIM
$\sigma=0$	0.392	0.083	0.047	0.947	0.995	0.999	35.254	0.960
$\sigma=0.005$	0.433	0.092	0.053	0.927	0.992	0.999	32.632	0.860
$\sigma=0.01$	0.498	0.132	0.071	0.854	0.980	0.995	29.597	0.758

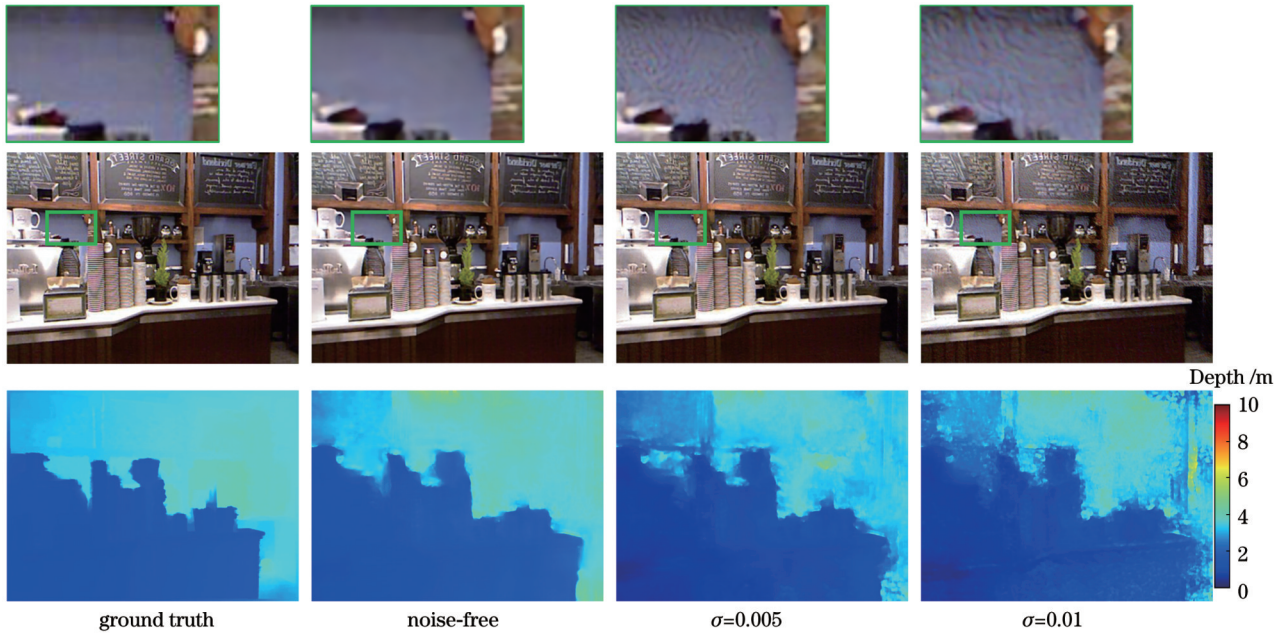

图 19　不同噪声水平下深度估计和图像复原效果图

Fig. 19　Visual results of depth estimation and image reconstruction under different noise levels

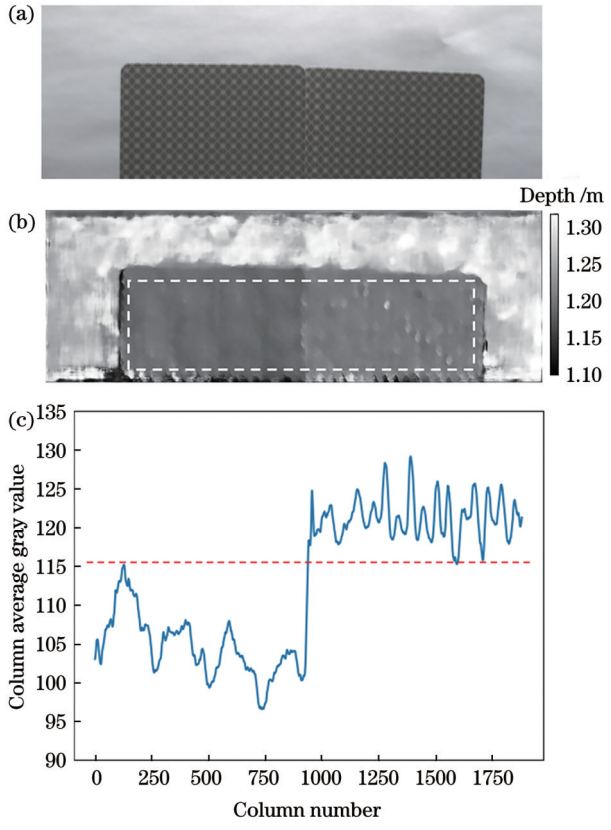

图 20　系统深度分辨率测试图。(a)拍摄间隔 1 cm 的两平面物体得到的编码图像；(b)图 20(a)对应的深度图；(c)图 20(b)中虚线框内的列均值曲线图

Fig. 20　System depth resolution test images. (a) Encoded image obtained from two flat objects taken at 1 cm interval; (b) depth map corresponding to Fig. 20(a); (c) column average value within the dashed box in Fig. 20(b)

面的前后关系，表明系统可正确分辨 1 cm 以上的深度差，若深度差小于 1 cm，深度估计值的波动可能导致部分像素的相对前后关系判断错误。整体而言，本文方法实现了从单目捕获的单帧图像中估计场景深度和拓展系统景深的目标，证明了模拟的编码成像过程对实际物理成像过程近似有效以及本文方法的可行性与实用性。

5　结　　论

提出一种双螺旋相位编码的三维成像方法，仅需一块相位板即可同时实现从单镜头捕获的单帧图像中估计场景深度并实现景深拓展成像的功能。分析了相位板参数和物距对成像能力的影响，讨论了根据给定深度范围确定相位板参数的方法。在 FlyingThings3D 和 NYU Depth V2 数据集上进行了仿真，结果表明本文方法在不同的景深拓展倍率下均可出色地估计出场景深度并获得清晰图像。此外，缩小深度估计范围和增大物距有助于平均深度估计精度的提升。加工相位板并搭建了样机，在真实场景下验证了本文方法的可行性和实用性。本文方法通过识别双螺旋相位编码形成的重影特征推断深度，结果与具体场景无关，因此表现出优秀的泛化性。

本文方法仍存在一些不足之处。首先，虽然解码过程依赖图像纹理，但本文方法对弱纹理场景仍然有效，仅对于由过曝等因素造成的纹理严重缺失的场景会失效。另外，在真实场景下受到噪声的影响会导致部分深度值错误、系统平均深度估计精度降低以及复原图像存在轻微伪影等问题，后续可考虑通过将噪声抑制加入解码算法解决上述问题。

参 考 文 献

[1] 邵晓鹏, 刘飞, 李伟, 等. 计算成像技术及应用最新进展[J]. 激光与光电子学进展, 2020, 57(2): 020001.
Shao X P, Liu F, Li W, et al. Latest progress in computational imaging technology and application[J]. Laser & Optoelectronics Progress, 2020, 57(2): 020001.

[2] 顿雄, 付强, 李浩天, 等. 计算成像前沿进展[J]. 中国图象图形学报, 2022, 27(6): 1840-1876.
Dun X, Fu Q, Li H T, et al. Recent progress in computational imaging[J]. Journal of Image and Graphics, 2022, 27(6): 1840-1876.

[3] Levin A, Fergus R, Durand F, et al. Image and depth from a conventional camera with a coded aperture[J]. ACM Transactions on Graphics, 2007, 26(3): 70-es.

[4] Zhou C Y, Lin S, Nayar S. Coded aperture pairs for depth from defocus[C]//2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE Press, 2009: 325-332.

[5] Dowski E R, Jr, Cathey W T. Extended depth of field through wave-front coding[J]. Applied Optics, 1995, 34(11): 1859-1866.

[6] 王伟, 张露鹤, 傅天文. 基于波前编码的扩展景深短波红外成像系统[J]. 激光与光电子学进展, 2023, 60(10): 1011005.
Wang W, Zhang L H, Fu T W. Wavefront coding-based short-wave infrared imaging system for extended depth of field[J]. Laser & Optoelectronics Progress, 2023, 60(10): 1011005.

[7] Shechtman Y. Recent advances in point spread function engineering and related computational microscopy approaches: from one viewpoint[J]. Biophysical Reviews, 2020, 12(6): 1303-1309.

[8] Pavani S R P, Piestun R. High-efficiency rotating point spread functions[J]. Optics Express, 2008, 16(5): 3484-3489.

[9] Shechtman Y, Sahl S J, Backer A S, et al. Optimal point spread function design for 3D imaging[J]. Physical Review Letters, 2014, 113(13): 133902.

[10] Nehme E, Freedman D, Gordon R, et al. DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning[J]. Nature Methods, 2020, 17(7): 734-740.

[11] Quirin S, Piestun R. Depth estimation and image recovery using broadband, incoherent illumination with engineered point spread functions[J]. Applied Optics, 2013, 52(1): A367-A376.

[12] Berlich R, Bräuer A, Stallinga S. Single shot three-dimensional imaging using an engineered point spread function[J]. Optics Express, 2016, 24(6): 5946-5960.

[13] Chang J L, Wetzstein G. Deep optics for monocular depth estimation and 3D object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2019: 10192-10201.

[14] Haim H, Elmalem S, Giryes R, et al. Depth estimation from a single image using deep learned phase coded mask[J]. IEEE Transactions on Computational Imaging, 2018, 4(3): 298-310.

[15] Ikoma H, Nguyen C M, Metzler C A, et al. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation[C]//2021 IEEE International Conference on Computational Photography (ICCP), May 23-25, 2021, Haifa, Israel. New York: IEEE Press, 2021.

[16] Wu Y C, Boominathan V, Chen H J, et al. PhaseCam3D: learning phase masks for passive single view depth estimation [C]//2019 IEEE International Conference on Computational Photography (ICCP), May 15-17, 2019, Tokyo, Japan. New York: IEEE Press, 2019: 1-12.

[17] Tan S Y, Wu Y C, Yu S I, et al. CodedStereo: learned phase masks for large depth-of-field stereo[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 7166-7175.

[18] Sitzmann V, Diamond S, Peng Y F, et al. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging[J]. ACM Transactions on Graphics, 2018, 37(4): 114.

[19] Metzler C A, Ikoma H, Peng Y F, et al. Deep optics for single-shot high-dynamic-range imaging[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1372-1382.

[20] Sun Q L, Tseng E, Fu Q, et al. Learning rank-1 diffractive optics for single-shot high dynamic range imaging[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1383-1393.

[21] 约瑟夫·W. 古德曼. 傅里叶光学导论[M]. 陈家璧, 秦克诚, 曹其智, 译. 4版. 北京: 科学出版社, 2020.
Goodman J W. Introduction to Fourier optics[M]. Chen J B, Qin K C, Cao Q Z, et al., Transl. 4th ed. Beijing: Science Press, 2020.

[22] Roider C, Jesacher A, Bernet S, et al. Axial super-localisation using rotating point spread functions shaped by polarisation-dependent phase modulation[J]. Optics Express, 2014, 22(4): 4029-4037.

[23] Bartelt H, Ojeda-Castaneda J, Sicre E E. Misfocus tolerance seen by simple inspection of the ambiguity function[J]. Applied optics, 1984, 23 (16): 2693-2696.

[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, 2015, 9351: 234-241.

[25] Mayer N, Ilg E, Häusser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4040-4048.

[26] Silberman N, Hoiem D, Kohli P, et al. Indoor segmentation and support inference from RGBD images[M]//Fitzgibbon A, Lazebnik S, Perona P, et al. Computer vision-ECCV 2012. Lecture notes in computer science. Berlin: Springer, 2012, 7576: 746-760.

[27] Eigen D, Puhrsch C, Fergus R. Depth map prediction from a single image using a multi-scale deep network[EB/OL]. (2014-06-09)[2023-11-12]. https://arxiv.org/abs/1406.2283.

[28] Dosselmann R, Yang X D. Existing and emerging image quality metrics[C]//Canadian Conference on Electrical and Computer Engineering, May 1-4, 2005, Saskatoon, SK, Canada. New York: IEEE Press, 2006: 1906-1913.

[29] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[30] Shi L F, Zhang Z Y, Cao A X, et al. One exposure processing to fabricate spiral phase plate with continuous surface[J]. Optics Express, 2015, 23(7): 8620-8629.

Monocular Three-Dimensional Coding Imaging Based on Double Helix Phase Masks

Zhang Yue, Cai Huaiyu*, Sheng Jing, Wang Yi, Chen Xiaodong

Key Laboratory of Opto-Electronics Information Technology of Ministry of Education, College of Precision Instrument and Opto-Electronic Engineering, Tianjin University, Tianjin 300072, *China*

Abstract

Objective As information technology develops rapidly, cameras are used not only as photography tools to meet users' artistic creation needs but also as hardware devices for visual sensing, serving as the "eyes" of machines. They are now widely applied in 2D computer vision tasks such as image classification, semantic segmentation, and object recognition. However, traditional cameras have two inherent limitations. Firstly, to meet the resolution requirements, the depth of field range needs to be sacrificed. Beyond the depth of field range, image blurring caused by defocus can affect the normal operation of subsequent algorithms. Secondly, as traditional cameras map the 3D world onto a 2D plane, they lose the depth information of the scene, making it difficult to apply to rapidly developing 3D computer vision tasks. Existing methods for depth acquisition, such as structured light, time-of-flight, and multi-view geometry, are inferior to single-lens cameras in terms of power consumption, cost, and size. Therefore, we propose a single-camera 3D imaging method based on a double helix phase mask, which can achieve depth estimation and depth-of-field extension imaging simultaneously with simple hardware modifications.

Methods We propose an imaging method based on a double helix phase mask that can simultaneously acquire scene depth information and achieve depth-of-field extension. By inserting a designed double helix phase mask at the aperture stop of the camera, the imaging beam is modulated into a double helix shape. On the one hand, the depth information is encoded in the image using the sensitive rotation characteristics of the double helix point spread function with defocus. On the other hand, utilizing the longer depth of focus characteristic of the double helix beam, the object points are encoded in the form of a double helix point spread function in a larger depth of field range. The depth information of object points is encoded in the image in the form of local ghosting. We combine convolutional neural networks to decode and reconstruct the encoded image end-to-end, thereby obtaining depth maps and depth of field extended images of the scene and jointly optimizing individual phase mask parameters. We analyze the influence of phase mask parameters and object distance on imaging performance and discuss the method of selecting phase mask parameters reasonably within a given depth range.

Results and Discussions To validate our method, we train it on the FlyingThings3D dataset, and the trained model is tested on the NYU Depth V2 dataset. The relative error of depth estimation on the NYU Depth V2 dataset can reach as low as 0.083 (Table 2). The depth of field extended images can achieve the highest PSNR of 35.254 dB and SSIM of 0.960 (Table 3). Compared to traditional optical systems, the depth of field can be extended by several tens of times. Using a phase mask with more rings can result in a higher depth of field extension imaging, but it may cause a slight decrease in depth estimation accuracy and quality of the depth of field extended images due to increased side lobes of the double-helix point spread function. Nevertheless, the overall performance remains within an acceptable range. The depth estimation accuracy of our method is related to the depth range to be measured. Reducing the detection range or increasing the object distance can improve the average depth estimation accuracy (Fig. 13). For potential application scenes such as gate face recognition, a physical system is built within the test range of 1.1–1.32 m. The relative depth estimation error in real scenes is 2.2%, and the depth of field is extended by about 10 times (Fig. 17), proving the effectiveness and practicality of the proposed method in real scenes.

Conclusions We introduce a three-dimensional imaging method based on a double helix phase, which only requires the addition of a phase mask to the existing lens to simultaneously estimate the depth of the scene from captured single frame images and achieve depth of field extension imaging. This method does not rely on built-in light sources and additional lenses, allowing for further reduction in size and power consumption. Compared to depth estimation algorithms solely based on deep learning, our method has excellent generalization because it identifies optically introduced features to estimate depth without relying on high-level semantic information about the scene. Overall, the

method shows potential applications in low-cost 3D imaging and detection fields. However, there are limitations to the proposed method. It relies on texture and can effectively work in scenes with weak texture, but it may fail in cases where texture is severely missing due to overexposure and other factors (Fig. 14). In addition, being affected by noise in real scenes can lead to errors in some depth values, decreased accuracy of system average depth estimation, and slight artifacts in reconstructed images. Subsequent research could consider incorporating noise suppression into the algorithm to solve this problem.

Key words imaging systems; computational imaging; monocular depth estimation; depth of field extension; double helix phase mask; point spread function