

文章编号: 1002-1582(2025)04-0385-07

一种新型复振幅全息图编码方法

田仪凡, 魏晨晓, 刘娟*

(北京理工大学 光电学院 北京市混合现实与新型显示工程技术研究中心, 北京 100081)

摘 要: 当前全息图计算存在效率低和编码信息丢失的问题。为解决以上问题, 提出了一种新的复振幅全息图编码方法。方法将角谱法计算得到的复振幅分布拆解成振幅信息和相位信息, 通过多相位编码生成包含全部振幅信息和相位信息的全息图。相比传统GS算法, 在相同计算时间下, 多相位编码方法全息图重建图像PSNR提高了15.49dB, SSIM提高0.65。方法为提高三维显示中全息图重建图像质量提供了新途径。

关 键 词: 计算全息; 三维显示; 全息编码; 复振幅编码

中图分类号: TP394.1; TH691.9

文献标识码: A

A novel encoding method for complex amplitude hologram

TIAN Yifan, WEI Chenxiao, LIU Juan*

(Beijing Engineering Research Center for Mixed Reality and New Display Technology,
School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Current hologram computation suffers from low efficiency and loss of encoded information. To address these issues, a novel encoding method for complex amplitude hologram is proposed. The proposed approach decomposes the complex amplitude distribution calculated by the angular spectrum method into amplitude and phase information, then generates a hologram containing complete amplitude and phase information through multi-phase encoding. Compared with the traditional Gerchberg-Saxton (GS) algorithm, under the same computation time, the proposed multi-phase encoding method improves the PSNR of the reconstructed holographic image by 15.49 dB and the SSIM by 0.65. This method provides a new approach for enhancing the quality of holographic reconstructed images for three-dimensional display.

Keywords: computer-generated hologram; three-dimensional display; holographic encoding; complex amplitude encoding

0 引 言

全息显示是一种三维显示形式, 因其在波前重建方面的优势而展现出广阔的发展前景^[1,2]。空间光调制器和计算技术的进步为电子全息技术开辟了新的途径, 无需光敏介质和复杂的干涉记录过程即可使用空间光调制器呈现计算机生成的全息图^[3-5]。由于传统的空间光调制器无法在单个面板中实现完全

的复振幅调制, 计算机生成的复振幅全息图通常被转换为仅振幅或仅相位全息图。虽然仅振幅全息图^[6]可以通过再现图像的振幅轻松控制, 但这些图像会受到杂散光(如共轭光和直射光)的污染。相比之下, 仅相位全息图不受直射光或共轭光的影响, 具有高衍射效率和易于控制的再现图像相位, 但对再现图像的振幅控制则具有挑战性。

光波本质上由复振幅表征。然而, 当前的空间

收稿日期: 2025-04-09; 收到修改稿日期: 2025-05-17

基金项目: 国家自然科学基金项目(62035003 & U22A2079)

作者简介: 田仪凡(1999—), 女, 硕士研究生, 从事计算全息研究。18811023918@163.com

刘娟(1970—), 女, 教授, 博士, 从事计算全息、微纳光学研究。

* 通讯作者: juanliu@bit.edu.cn

光调制器只能调制入射光的振幅或相位，因此复振幅全息图必须转换为仅振幅或仅相位全息图^[6]。因此，在仅相位全息图中编码复振幅是一个重要问题。目前已经提出了多种编码方案，包括迭代方法、非迭代方法以及近年来的深度学习方法。迭代方法，如 Gerchberg-Saxton (GS) 算法^[7-9]和梯度下降算法^[10-12]，可以生成质量更好的再现图像，但需要大量时间来优化全息图。一种结合迭代和非迭代方法的混合技术^[13]被提出，但其图像质量仍不够理想。基于深度学习的方法^[14-16]可以快速推断出优化的全息图，但需要大量的训练时间。

非迭代方法包括误差扩散^[17]、采样方案^[18]以及双相位全息图^[19,20]等。其中，双相位编码是最主流的编码技术。它可以通过两个单位复向量表示一个复像素，并且能够生成高质量的仅相位全息图，广泛应用于全息显示和光束生成等领域。近年来，提出了更复杂的双相位编码方法^[21-23]，例如利用高速空间光调制器提高双相位全息图的空间带宽，以及通过滤波方法提升双相位编码全息图的再现图像质量；然而，这些方法的缺点是光效率有所降低；或者采用改进的二值化编码方法在相位全息图中保留部分振幅信息，然而，这种方法无法在全息图中表征全部振幅信息^[24]。

本研究专注于不需要迭代计算和训练的非迭代方法。为将全部复振幅场信息编码于全息图中以提高全息再现像质量，我们提出了复振幅全息图的多相位编码方法。我们通过角谱法生成复振幅全息图，并构建巨像素生成具备全部振幅信息和相位信息的全息图，所提出的方法能够实现具有更高图像质量的全息再现像。

1 原理与方法

1.1 模型构建

本文主要通过角谱传输的方式来实现计算全息图的生成和模拟再现过程。整个模型的建立如图 1 所示。

首先物图像 $I(x,y)$ 通过角谱传输距离 z 获取复振幅全息图 $U_H(x,y)$ ，之后通过多相位编码获取生成具备全部振幅信息和相位信息的全息图 $U_\phi(x,y)$ ，最后通过角谱传输距离 z 获得模拟再现图像 $I(x,y)$ 。

$$H(f_x, f_y, z) = \begin{cases} \exp\left[j\frac{2\pi}{\lambda} z \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right] & \text{if } \sqrt{(f_x)^2 + (f_y)^2} < \frac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

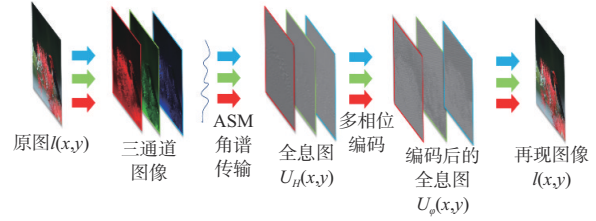


图 1 算法模型图

$U_H(x,y)$ 是根据角谱传输的原理计算得到的。

用平面波角谱理论解决衍射问题的关键是：根据 $z=0$ 平面的空间频谱 $A(f_\xi, f_\eta)$ ，求出观察面 Π 上的空间频谱 $G(f_\xi, f_\eta)$ ，再通过傅里叶变换得到观察面 Π 上的光场分布 $E(x,y)$ 。式 (1) 即平面波角谱理论求解衍射问题的基本公式。平面波角谱理论的实质是傅里叶分解和综合的过程，即首先将输入函数 $U(\xi, \eta)$ 分解为一系列简谐平面波，然后各个简谐平面波独立传播到输出平面，最后再将传播过程中经历了不同相位延迟的平面波成分相加，综合出输出面上衍射的复振幅 $E(x,y)$ 。输出面上的光场 $E(x,y)$ 可以写作以下形式

$$E(x,y) = \iint_{-\infty}^{\infty} G(f_\xi, f_\eta) e^{j2\pi(f_\xi x + f_\eta y)} df_\xi df_\eta \\ = \iint_{-\infty}^{\infty} A(f_\xi, f_\eta) \exp\left[jkz \sqrt{1 - \lambda^2 f_\xi^2 - \lambda^2 f_\eta^2}\right] \exp\left[j2\pi(f_\xi x + f_\eta y)\right] df_\xi df_\eta \quad (1)$$

$$f_\xi^2 + f_\eta^2 \leq \frac{1}{\lambda^2} \quad (2)$$

式中， λ 是光波长； f_ξ, f_η 是表示 ξ 方向和 η 方向的空间频率。

为简化上述公式，可以用 $H(f_\xi, f_\eta)$ 表示 $G(f_\xi, f_\eta)$ 和 $A(f_\xi, f_\eta)$ 的比值，则有

$$H(f_\xi, f_\eta) = \frac{G(f_\xi, f_\eta)}{A(f_\xi, f_\eta)} = \left[jkz \sqrt{1 - \lambda^2 f_\xi^2 - \lambda^2 f_\eta^2} \right] \quad (3)$$

将 $A(f_\xi, f_\eta)$ 和 $G(f_\xi, f_\eta)$ 看作系统的输入和输出，则 $H(f_\xi, f_\eta)$ 就是该系统的传递函数。因此，可以将角谱传输的过程简化为

$$E(x,y) = f(I,z) \quad (4)$$

$$f(I,z) = \mathcal{F}^{-1}\left\{\mathcal{F}(U(x,y)) \cdot H(f_x, f_y, z)\right\} \quad (5)$$

参考式 (1) ~ (6), 在算法模型中, 复振幅全息图 $U_H(x, y)$ 可以通过角谱传输得到。其公式表示为

$$H(f_x, f_y, z) = \begin{cases} \exp\left[j\frac{2\pi}{\lambda}z\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right] & \text{if } \sqrt{(f_x)^2+(f_y)^2} < \frac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

式中, $\mathcal{F}$ 表示傅里叶变换, $\mathcal{F}^{-1}$ 表示傅里叶逆变换; f_x, f_y 表示 x 方向和 y 方向的空间频率; λ 表示光波长; $H(f_x, f_y, z)$ 是传递函数。

1.2 多相位编码原理

在获取原始复振幅全息图 $U_H(x, y)$ 后, 对其进行多相位编码操作。多相位编码的原理如图2所示, 其编码对象是通过1.1原理得到的物图像的复振幅全息图。将空间光调制器上的 2×2 个像素看作是一个巨像素, 即一个复振幅调制单元 U_H , 设该调制单元内的元素分别为 $U_{H_i} = A_i \exp(j\varphi_i)$ ($i = 1, 2, 3, 4$)。对巨像素单元中的每一个元素进行双相位分解得到双相位分量 θ_{i+}, θ_{i-}

$$U_{H_i} = \exp(j\theta_{i+}) + \exp(j\theta_{i-}) \quad (10)$$

$$\theta_{i+} = \varphi_i + \arccos[A_i/A_{i\max}] \quad (11)$$

$$\theta_{i-} = \varphi_i - \arccos[A_i/A_{i\max}] \quad (12)$$

式中, $A_{i\max}$ 为所有复振幅调制单元相应位置处 A_i 的最大值。

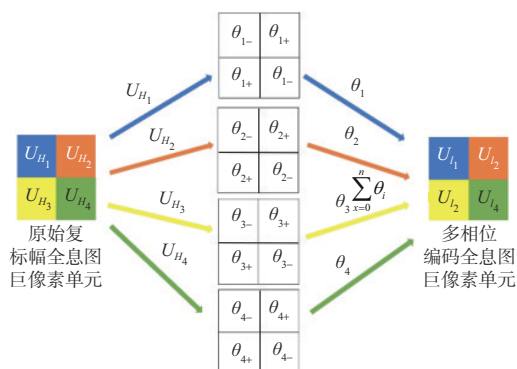


图2 多相位编码原理

然后将双相位分量在一个复振幅单元内沿对角线交错排列, 得到新的巨像素单元 θ ($i = 1, 2, 3, 4$), 如图3所示。

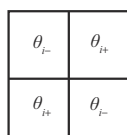


图3 巨像素单元

对巨像素单元 θ_i 进行累加操作得到巨像素单元 U_L

$$U_L = \sum_{i=0}^4 \theta_i \quad (13)$$

$$U_H(x, y) = f(I, z) \quad (7)$$

$$f(I, z) = \mathcal{F}^{-1}\{\mathcal{F}(U_H(x, y)) \cdot H(f_x, f_y, z)\} \quad (8)$$

设由巨像素单元 U_L 构成的全息图为 $U_\varphi(x, y)$, 则其重建图像 I 可表示为

$$I = f(\exp(jU_\varphi(x, y)), -z) \quad (14)$$

2 仿真模拟

2.1 仿真实验参数

为了说明所提方法的有效性, 对上述方法进行数值仿真和光学实验测试。实际进行数值模拟时的计算机配置: Intel®Core™i7-13700 @ 3400MHz, 32GB RAM 以及 NVIDIA GeForce RTX3060Ti, 具体的光学仿真模拟和实验参数见表1。

表1 仿真模拟和光学实验参数

软件参数	编译环境	MatlabR2020b
硬件参数	CPU	i7-13700
	GPU	RTX3060Ti
	内存/GB	32
物图参数	分辨率	1024×1024
	像元尺寸	8um
全息图参数	分辨率	1024×1024
	像元尺寸	8um
光波波长	R	638nm
	G	520nm
	B	450nm
传播距离	30mm	

2.2 评价函数

为评价生成全息图质量, 使用结构相似度 (Structural Similarity, SSIM) 和峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 来作为重建图像的评价函数。SSIM通过亮度、对比度和结构三个方面来比较两幅图像相似性, 其计算公式表示为

$$SSIM = \frac{(2\mu_I\mu_U + c_1)(\sigma_{IU} + c_2)}{(\mu_I + \mu_U + c_1)(\sigma_I^2 + \sigma_U^2 + c_2)} \quad (15)$$

式中, I 和 U 是输入的两幅图像, μ_I 和 σ_U 分别表示 I 和 U 的平均值; σ_I 和 σ_U 分别表示 I 和 U 的标准差; σ_{IU} 表示 I 和 U 的协方差; c_1, c_2 分别为常数; SSIM 越大表示两幅图像越相似, 质量越高。

PSNR通过计算两幅图像像素值差值来衡量两幅图像相似性, 其计算公式表示为

$$\text{PSNR} = 10 \lg \left[\frac{\text{MAX}^2}{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i,j) - U(i,j)]^2} \right] \quad (16)$$

式中, m 和 n 为图像大小; I 为参考图像; U 为待评价图像; MAX 值与图像位宽有关, 本文图像为 8bit 位宽, MAX 取值为 255。PSNR 越大表示图像失真越小, 质量越高。在数值仿真中, 以原图像作为参考图像 I , 以再现图像作为待评价图像 U 。

2.3 数值仿真

为进行数值模拟, 选取以下两张图像作为所提出方法的测试物图像, 如图 4 (a)、(g) 所示。根据表 1 中的参数, 计算 G 通道的多相位编码全息图。两张测试图像在表 1 软硬件环境下生成的多相位编码全息图及其再现结果如图 4 所示, 具体数据由表 2 所示。根据数值模拟结果可以看出, 对于测试物图像, 通过新方法编码的全息图其再现像 PSNR 均大于 25dB, 图像质量较高。根据图 4 (d)、(e)、(g)、(k) 可以看出, 多相位编码方法的再现像较大程度地保留了原始物图像的高频信息。

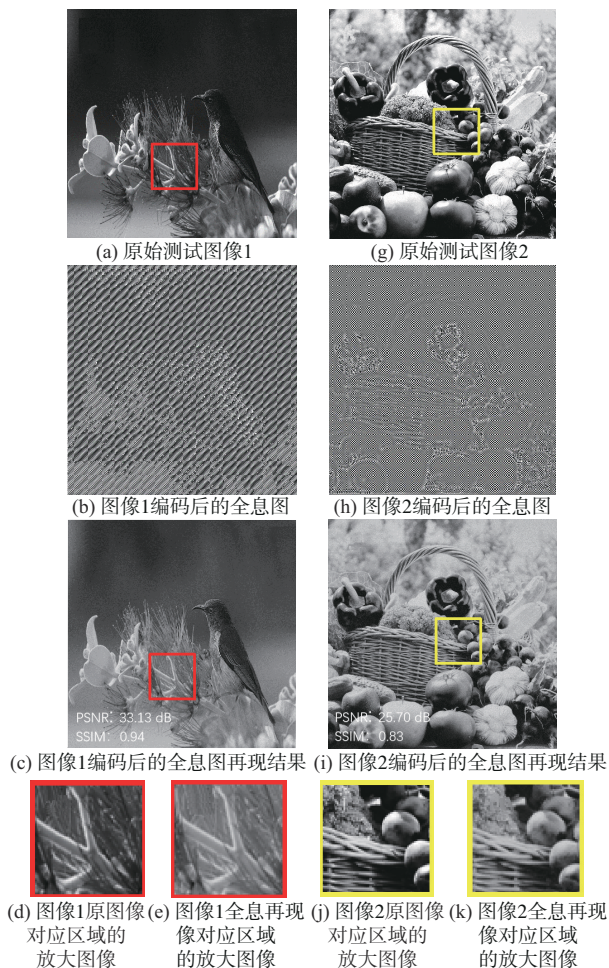


图 4 绿色通道图像多相位编码全息图与再现结果

表 2 G 通道编码全息图再现质量

测试图像	PSNR	SSIM
1	33.13dB	0.94
2	25.70dB	0.83

为比较多相位编码对不同波长全息图的编码效果, 图 5 比较了针对同一彩色图像不同通道生成的多相位全息图的再现像质量, 具体数据由表 3 所示。由图 5 (a)、(b)、(c) 及表 3 看出, 图像在 R, G, B 三通道的再现像 PSNR 均达 33dB 以上, SSIM 均达 0.93 以上, 说明多相位编码对于不同波长均具有普适性。

表 3 RGB 通道编码全息图再现质量

编码方法	图像通道	PSNR	SSIM
GS 算法	R	23.63dB	0.71
	G	25.41dB	0.79
	B	23.70dB	0.75
	彩色	24.36dB	0.75
多相位编码	R	32.59dB	0.93
	G	33.13dB	0.94
	B	33.43dB	0.94
	彩色	33.03dB	0.94

为说明多相位编码全息图在再现像质量上的优势, 图 5 比较了同一物图像通过 GS 编码、多相位编码和复振幅全息图的再现结果。

仿真实验结果表明, GS 算法在经过 500 次迭代后能够获得高质量的重建图像, 但由于初始相位的随机性和不确定性, GS 算法的再现图像具有明显的散斑噪声。因此, 它无法满足实时全息显示的要求。复振幅编码方法展现出了最高的再现像质量, 但由于目前单个空间光调制器无法同时调制光场的振幅和相位信息, 其编码的全息图不能够直接用于光学再现。使用所提出的多相位编码方法时, 重建图像中没有明显的散斑噪声, 且其编码的相位全息图可以直接用于光学再现。此外, 对于上述测试图像, 所提出的方法在 PSNR 和 SSIM 方面优于 GS 算法, 且图像质量与复振幅编码结果更为接近; 这表明所提出的方法具有更出色的成像质量。

图 6 展示了不同相位全息图生成算法的运行时间与成像质量之间的定量关系。研究采用数据集中的 100 幅大小为 1024×1024 的彩色 BMP 图像作为测试样本, 计算平均运行时间。其中双相位法采用棋盘格编码形式^[19]。所有对比实验均在相同条件下进行, 包括数据图像集、训练周期、衍射模型和设备配置等参数。量化数据表明, 在

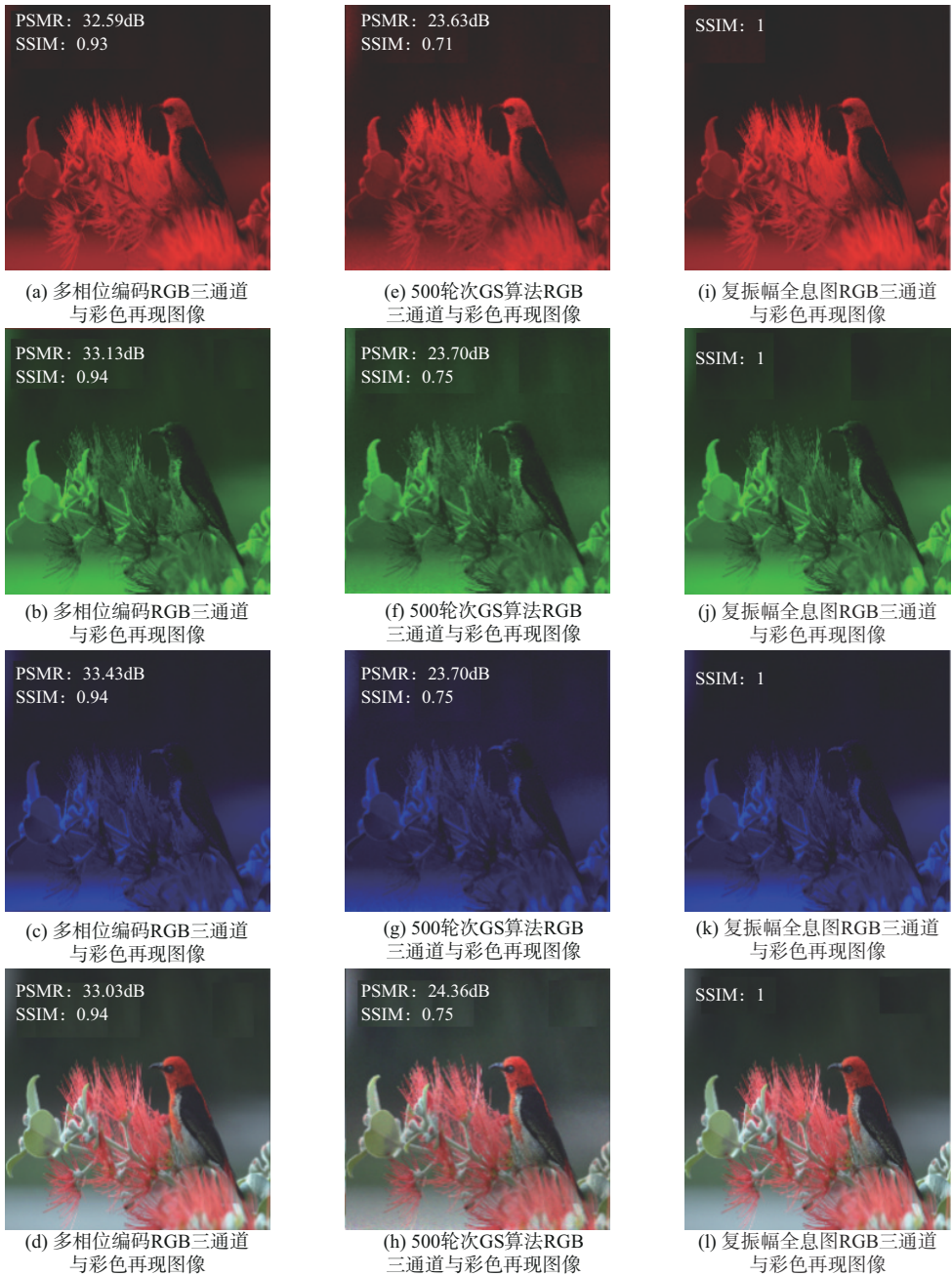


图 5 RGB通道多相位编码全息图再现结果

当前硬件条件下，多相位编码针对 1024×1024 图像的编码速度可达每秒 10 帧，同时保持 29.24dB 的高 PSNR，有效平衡了计算效率与成像质量。

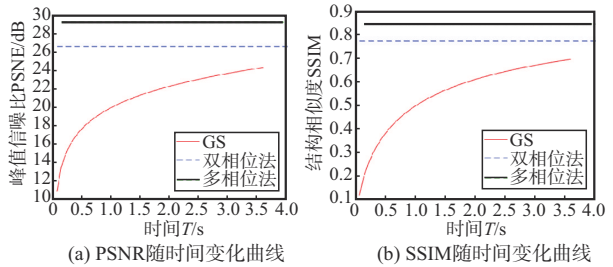


图 6 不同相位全息图生成算法运行时间与成像质量之间的定量关系

结果显示，对于传统 GS 迭代算法，重建质量随着迭代次数增加而提升，但这一过程耗时较长；在经过约 4s 的计算和 50 轮次迭代，GS 算法的重建质量仍无法达到多相位编码的水平。在相同运行时间条件下，多相位编码的重 PSNR 比 GS 算法高出约 15.49dB，SSIM 高出约 0.65。

为说明本方法相较于其它相位编码方法的优势，图 6 也展示了双相位编码方法和多相位编码方法的比较。经理论分析可知，由于双相位编码时将分解的两幅纯相位全息图按照棋盘格取样后叠加，损失了部分未取样的复振幅信息；而多相位编码没有采样过程，其复振幅信息更为完整，其

再现像质量更高。由图 6 中量化数据, 相比双相位编码方法, 多相位编码方法的 PSNR 高出 2.64dB, SSIM 高出 0.07, 说明多相位编码图像的再现像质量更高, 与理论一致。

3 光学实验

本节通过光学实验进一步验证所提方法的可行性。全息显示系统如图 7 所示。实验采用波长为 638nm、520nm、450nm 的激光光束, 经扩束准直后照射反射型纯相位空间光调制器。该空间光调制器选用 Holoeye Pluto2.1 NIR-145 型号, 具有 1920×1080 分辨率及 8 微米像素间距。所有算法生成的 1024×1024 分辨率计算全息图均经零填充处理至 1920×1080 后加载至空间光调制器。经空间光调制器调制反射的光波传输至 $4f$ 系统——该系统由两个傅里叶透镜 L_2 , L_3 一个滤波器组成, 用于消除全息重建产生的高阶衍射信息。最终重建的全息图像由 Thorlabs cs126cu 型号 CCD 相机捕获。

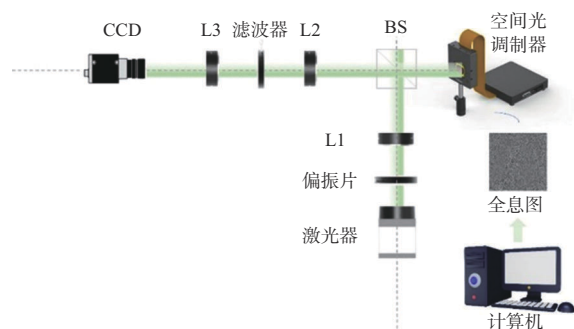


图 7 全息显示系统

全息显示的重建图像如图 8 所示。对多相位编码全息图的重建效果进行了分析。由于实际条件的限制, CCD 的噪声、环境杂散光、空间光调制器填充因子未达 100% 以及实际操作中的人为因素等, 都可能对实验结果产生负面影响。尽管存在以上干扰因素, 实验结果与仿真结果仍保持基本一致。

分析图 8 可以看出, 采用本论文提出的多相位编码方法编码全息图获得的的光学重建图像质量较高; 该方法能够有效抑制散斑噪声, 且重建图像强度分布较为均匀。从再现图像的放大视图可见, 物体轮廓较为清晰, 且细节信息较为丰富, 模糊程度低。这些实验结果证实了本方法在抑制散斑噪声和提升全息图重建质量方面的有效性。

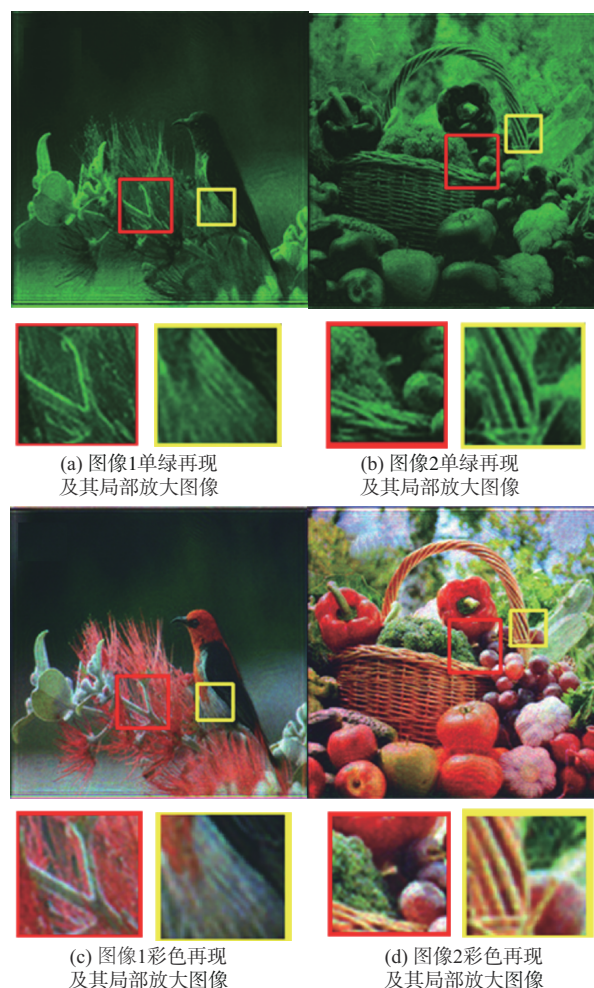


图 8 光学实验结果

4 结 论

本文提出了一种新的多相位全息图编码方法, 其创新型在于编码得到的相位全息图具有复振幅场的全部振幅信息和相位信息, 而非传统全息图中仅包含复振幅场的一部分相位信息和振幅信息。该全息图编码方法通过非迭代计算的方式重新构建了全息图的巨像素单元, 而非通过迭代方式确定全息图的相位分布。详细阐述了多相位编码方法的原理架构和数学模型, 并通过仿真实验验证; 该方法可以在 0.11s 内生成 1K 分辨率的计算全息图, 平均 PSNR 达到 29.24dB, SSIM 达到 0.84。以上实验数据表明本方法可以有效提升重建成像质量。

在光学实验中, 所提多相位编码方法显著抑制了散斑噪声并改善了亮度均匀性, 并保留了更丰富的细节特征。实验结果充分验证了多相位编码方法的合理性与高效性, 为三维显示中理想全息图的设计提供了新思路。

参考文献:

- [1] Yavuz O, Kesim D K, Turnali A, et al. Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors[J]. *Nature Photonics*, 2019, 13(4): 251—256.
- [2] Li X, Ren H, Chen X, et al. A thermally photoreduced graphene oxides for three-dimensional holographic images[J]. *Nature Communications*, 2015, 6(1):6984.
- [3] Yu H, Lee K, Park J, et al. Ultrahigh-definition dynamic 3D holographic display by active control of volume speckle fields[J]. *Nature Photonics*, 2017, 11(3):186—192.
- [4] Li J, Smithwick Q, Chu D P. Scalable coarse integral holographic video display with integrated spatial image tiling[J]. *Optics Express*, 2020, 28(7):9899—9912.
- [5] Liu J S, Collings N, Crossland W A, et al. Simulation and experiment on generation of an arbitrary array of intense spots by a tiled hologram[J]. *Journal of Optics*, 2010, 12(8):085402.
- [6] Tsang P W M. Computer-generated phase-only holograms for 3D displays: a matlab approach[M]. Cambridge: Cambridge University Press, 2021.
- [7] Chang C, Xia J, Yang L, et al. Speckle-suppressed phase-only holographic three-dimensional display based on double-constraint Gerchberg-Saxton algorithm[J]. *Applied Optics*, 2015, 54(23):6994—7001.
- [8] Wu Y, Wang J, Chen C, et al. Adaptive weighted Gerchberg-Saxton algorithm for generation of phase-only hologram with artifacts suppression[J]. *Optics Express*, 2021, 29(2):1412—1427.
- [9] Liu K, He Z, Cao L. Double amplitude freedom Gerchberg-Saxton algorithm for generation of phase-only hologram with speckle suppression[J]. *Applied Physics Letters*, 2022, 120(6):061103.
- [10] Chen C, Lee B, Li N N, et al. Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function[J]. *Optics Express*, 2021, 29(10):15089—15103.
- [11] Peng Y, Choi S, Kim J, et al. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration[J]. *Science Advances*, 2021, 7:eabg5040.
- [12] Liu S, Takaki Y. Gradient descent based algorithm of generating phase-only holograms of 3D images[J]. *Optics Express*, 2022, 30(10):17416—17436.
- [13] Zea A V, Ramirez J F B, Torroba R. Optimized random phase only holograms[J]. *Optics Letters*, 2018, 43(4):731—734.
- [14] Shi L, Li B, Kim C, et al. Towards real-time photorealistic 3D holography with deep neural networks[J]. *Nature*, 2021, 591: 234—239.
- [15] Shi L, Li B, Matusik W. End-to-end learning of 3D phase-only holograms for holographic display[J]. *Light: Science & Applications*, 2022, 11:247.
- [16] Hu C C, Yang G L, Xie H Y. 3D information transmission of a computer-generated hologram using a quantum compensation hybrid neural network[J]. *Optics Express*, 2024, 32(13):23736—23749.
- [17] Tsang P W M, Poon T C. Novel method for converting digital Fresnel hologram to phase-only hologram based on bidirectional error diffusion[J]. *Optics Express*, 2013, 21(20):23680—23686.
- [18] Tsang P, Chow Y T, Poon T C. Generation of phase-only Fresnel hologram based on down-sampling[J]. *Optics Express*, 2014, 22(21):25208—25214.
- [19] Hsueh C K, Sawchuk A A. Computer-generated double-phase holograms[J]. *Applied Optics*, 1978, 17(24):3874—3883.
- [20] Mendoza-Yero O, Mnguez-Vega G, Lancis J. Encoding complex fields by using a phase-only optical element[J]. *Optics Letters*, 2014, 39(7):1740—1743.
- [21] Sui X, He Z, Zhang H, et al. Spatiotemporal double-phase hologram for complex-amplitude holographic displays[J]. *Chinese Optics Letters*, 2020, 18(10):100901.
- [22] Sui X, He Z, Jin G, et al. Band-limited double-phase method for enhancing image sharpness in complex modulated computer-generated holograms[J]. *Optics Express*, 2021, 29(2):2597—2612.
- [23] Sui X, He Z, Jin G, et al. Spectral-envelope modulated double-phase method for computer-generated holography[J]. *Optics Express*, 2022, 30(17):30552—30563.
- [24] Shimobaba T, Wang F, Starobrat J, et al. Comparison of double-phase hologram and binary amplitude encoding: holographic projection and vortex beam generation[J]. *Applied Optics*, 2023, 62(28):1—10.