

光学学报

实时全息三维显示技术研究进展

刘娟^{1,2,3*}, 皮大普^{1,2,3}, 王涌天^{1,2,3}

¹北京理工大学光电学院, 北京 100081;

²北京市混合现实与新型显示工程技术研究中心, 北京 100081;

³光电成像技术与系统教育部重点实验室, 北京 100081

摘要 全息三维显示技术能有效地重建三维物体的波前,并为人眼提供完整的深度线索,已经成为三维显示领域的研究热点。相比于光学全息,计算全息通过计算机模拟全息图的记录过程,并采用可刷新的空间光调制器替代传统的光学记录材料作为全息图的承载媒介,因而成为理想的实现实时全息三维显示的技术方案。然而,复杂三维场景数据量巨大、空间光调制器调制能力不足以及全息三维显示系统展示度不高等问题仍阻碍了实时全息三维显示的发展。为了克服这些不足,研究者在算法和硬件两方面做出了许多创新工作。本文综述了实时全息三维显示的进展。首先概述了全息术的基本原理和发展简史,接着详细介绍了全息图快速计算方法和针对现有空间光调制器的波前编码方法,然后讨论了深度学习对实时全息三维显示做出的贡献并介绍了一些典型的全息显示系统,最后对实时全息三维显示的未来发展进行了展望。

关键词 全息三维显示; 计算全息; 全息图; 深度学习

中图分类号 O436 **文献标志码** A

DOI: 10.3788/AOS230744

1 引言

人类通过视觉感知获得的信息占其获取信息总量的 80% 以上。作为展示图像信息的载体,显示技术一直都是光学领域研究的重点之一。随着对信息获取需求的提高,人们对显示技术的要求日益提高,显示技术的发展经历了从黑白显示到彩色显示,从二维显示到三维显示的过程。在现有的三维显示技术中,全息三维显示技术由于能够完整地记录和重建三维物体的波前,提供人眼视觉系统所需的全部深度信息,因此被视为具有广阔应用前景与发展潜力的终极三维显示技术,并且已经成为全球研究热点之一^[1-2]。

全息术于 1948 年被 Gabor^[3] 提出,其最初的目的是提高电子显微镜的分辨率。全息术按其物理意义可分为波前记录和波前重建两部分。在波前记录过程中,物光波与参考光波在全息图平面发生干涉,干涉条纹的强度被记录下来。由于已知参考光波的振幅和相位,全息术就是把物光波的复振幅信号转换为强度信号,从而记录了物光波的全部信息。在波前重建过程中,用一束重建光波对全息图进行照射,重建光波经过全息图的衍射光波在特定的位置上可以重建出物光波的振幅与相位信息。Gabor 从理论和实验上证明了全

息术可以记录和重建物光波的波前,并因此获得了 1971 年度的诺贝尔物理学奖。但是由于当时没有好的相干光源,且同轴全息图产生的孪生像不能在空间上分离,因此全息术的应用受到了极大的限制。1962 年,Leith 和 Upatnieks^[4-6] 提出离轴全息图,通过增加参考光的载频来分离孪生像,从而极大地扩展了全息术的适用性。传统的光学全息利用感光材料记录全息图的干涉条纹,并在一定条件下重建物光波。光学全息的记录和重建过程均需在实际光路中进行,且需要显影、定影、烘干等处理,过程繁琐复杂,材料也不可重复擦写。

随着计算机技术和光电显示技术的发展,计算全息受到了广泛的关注。计算全息在计算机中模拟波前记录过程,相比光学全息,其具有三个优势:1) 计算全息避免了复杂的记录光路;2) 计算全息可以借助空间光调制器实现动态刷新,进而实现实时全息三维显示;3) 计算全息可以记录真实世界中不存在的物光波。计算全息主要包括波前计算、波前编码与波前重建这三个步骤。波前计算即根据传播距离、采样间隔等条件计算物光波在全息图平面上的复振幅分布。波前编码是将全息图平面上的复振幅分布编码为与空间光调制器调制方式相匹配的函数,通常为纯振幅分布或纯相

收稿日期: 2023-03-29; 修回日期: 2023-05-08; 录用日期: 2023-05-22; 网络首发日期: 2023-06-28

基金项目: 国家自然科学基金(62035003, U22A2079, 61975014)、北京市科委中关村管理委员会科技计划项目(Z211100004821012)

通信作者: *juanliu@bit.edu.cn

位分布。波前重建是将编码后的全息图加载到空间光调制器上,并通过相干光照射,实现物光波的光学重建。当前,实时全息三维显示面临的挑战主要集中在以下方面:1)复杂三维场景数据量巨大,需要开发全息图高速计算方法;2)当前空间光调制器调制能力不足,需要针对空间光调制器开发兼具高运算效率和高重建精度的编码方法;3)全息三维显示系统展示度不高,需要搭建大尺寸、大视角的全息三维显示系统。本文针对当前为实现实时全息三维显示所开展的一系列工作进行了综述:第 1 节回顾了全息术的发展简史,介绍了计算全息的基本概念,并指出实时全息三维显示面临的挑战;第 2 节和第 3 节分别从算法和硬件两方面介绍了全息图高速计算方法;第 4 节介绍了常见的波前编码方法;第 5 节介绍了基于深度学习的计算全息算法;第 6 节介绍了一些典型的全息显示系统;第 7 节对实时全息三维显示未来的发展进行了展望。

2 基于物理模型的算法

目前,三维计算全息算法主要是基于物理模型的算法。这种算法通常将三维物体离散为多个基元,逐个计算每个基元在全息图平面上的复振幅分布,并将其叠加以获得三维物体的全息图。因此,物理模型算法的计算速度主要取决于两个因素:基元的数量和计算每个基元在全息图平面上的复振幅分布所需的时间。根据离散基元的种类,物理模型算法可以分为三类:点源法、面元法和层析法。

2.1 点源法

点源法是目前应用最为广泛的一种计算三维物体全息图的方法。点源法的基本思想是将三维物体离散为一系列的物点,其中每个离散物点被视为一个点光源,发出球面波照射在全息图平面上,通过计算每个点光源发出的球面波在全息图平面上的复振幅分布并进行叠加便可计算出三维物体的全息图,如图 1^[7-8]所示。点源法的计算过程可以表示为

$$H(x, y) = \sum_{i=1}^N A_i \exp \left[jk \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} \right], \quad (1)$$

式中: $H(x, y)$ 为三维物体在全息图平面上的复振幅分布, (x, y) 为全息图平面的坐标; N 为离散物点的个数; A_i 和 (x_i, y_i, z_i) 分别为第*i*个物点的振幅和坐标信息; $k=2\pi/\lambda$ 表示波数, λ 为记录光波的波长; z 为物空间坐标原点全息图平面坐标原点的距离。

在点源法中,三维物体的离散物点数通常会达到百万量级。从式(1)中可以看出,点源法中涉及到的计算主要是每个点光源在全息图平面上的复振幅分布,其包括加法、乘法、平方、开方、指数运算。因此,如果可以提前计算出所有可能的点光源在全息图平面上的

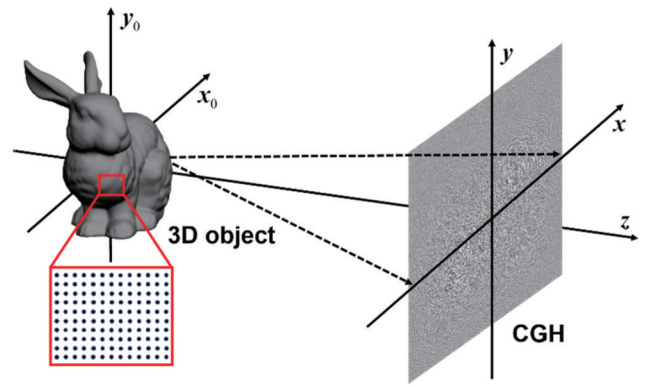


图 1 点源法示意图^[7-8]

Fig. 1 Schematic diagram of point-based method^[7-8]

复振幅分布并将其存储在计算机中,便可以大大减轻计算负担。受到这个想法的启发,Lucente^[9]提出了查找表(LUT)法。LUT法由离线计算和线上计算两部分组成。离线计算主要计算所有可能的点光源在全息图平面上的复振幅分布并将其存储在LUT中。线上计算则通过读取预先存储的点光源在全息图平面上的复振幅分布,然后乘以对应物点的振幅并进行叠加以生成全息图。和点源法相比,LUT法在全息图计算(线上计算)中只涉及到加法和乘法计算,避免了平方、开方、指数这些复杂运算,因此计算速度大大提高。

为了减少LUT的内存占用, Kim等^[10]提出了新型查找表(N-LUT)法。N-LUT法将一个三维物体沿轴向离散成多个二维平面,然后计算每个二维平面中心物点在全息图平面上复振幅分布,并存储在LUT中。在计算全息图时,可以根据物空间坐标中的相对位置关系来获得每个二维平面上的非中心物点在全息图平面上的复振幅分布。N-LUT法极大地降低了LUT法所需的内存占用量,但其存储的数据仍为多个二维矩阵,内存占用量依然很大,严重影响了数据读取速度。为解决这一问题, Pan等^[11]提出了一种分离查找表(S-LUT)法。S-LUT法通过近似将物点在全息图平面上的复振幅分布分解成水平和垂直方向调制因子的乘积。在线上计算时,只需通过调用调制因子便可计算出三维物体的全息图。S-LUT法减少了内存占用量,提高了全息图的计算速度,但其内存占用量仍会随着物体深度的增加而增加。随后, Jia等^[12]提出了压缩查找表(C-LUT)法。C-LUT法将深度信息从水平和垂直方向调制因子中分离出来使得离线存储数据不会随着物体深度的增加而增加。Gao等^[13]在此基础上提出了精准压缩查找表(AC-LUT)法。AC-LUT法通过额外的指数运算解决了C-LUT法中三维物体深度较大时重建图像中的畸变问题。Pi等^[14]在AC-LUT的基础上又提出了精准高压压缩查找表(AHC-LUT)法。AHC-LUT法成功将三维物体的离线数据压缩为两个一维向量,相比AC-LUT法,其内存占用量降低了98.64%,计算速度提升了11.62%。S-LUT

法、C-LUT 法、AC-LUT 法和 AHC-LUT 法的计算复杂度、内存占用量和畸变比较如表 1 所示。

表 1 不同方法的计算复杂度、内存占用量和畸变
Table 1 Computational complexity, memory usage, and distortion among different methods

Method	Online computational complexity	Offline computational complexity	Memory usage	Distortion of horizontal and vertical direction
S-LUT	$N_z(N_y(N_xq + pq) + pq)$	$(N_xp + N_yq)N_z$	$(N_xp + N_yq)N_zM \times 2$	—
C-LUT	$N_z(N_y(N_xq + pq) + pq)$	$N_xp + N_yq$	$(N_xp + N_yq)M \times 2$	$xz/d, yz/d$
AC-LUT	$N_z(N_y(N_xq + pq) + pq)$	$N_xp + N_yq$	$(N_xp + N_yq)M$	—
AHC-LUT	$N_z(N_y(N_xq + pq) + pq)$	$p + N_x\Delta x + q + N_y\Delta y$	$(p + N_x\Delta x + q + N_y\Delta y)M$	—

另一种加快速源法计算速度的思路是减小每个物点的计算区域。受此启发, Shimobaba 等^[15]提出了波前记录板(WRP)法。WRP 是一个虚拟平面, 放置在三维物体和全息图之间, 以协助点源法的加速计算。WRP 法包含两个步骤: 1) 使用点源法计算三维物体在 WRP 上的复振幅分布; 2) 利用平行平面间的衍射从 WRP 计算出三维物体的全息图。WRP 法的原理如图 2^[15]所示, 全息图平面上的复振幅分布可以表示为

$$H(x, y) = \mathcal{F}^{-1}\{\mathcal{F}[E(x, y)] \cdot \mathcal{F}[h(x, y)]\}, \quad (2)$$

式中: $E(x, y)$ 为三维物体在 WRP 上的复振幅分布; $h(x, y)$ 为脉冲响应函数; \mathcal{F} 和 \mathcal{F}^{-1} 分别为傅里叶变换和傅里叶逆变换。

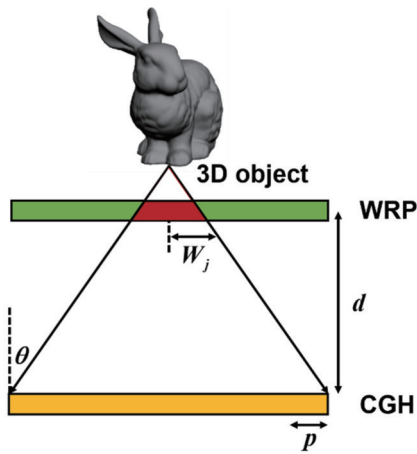


图 2 WRP 法示意图^[15]

Fig. 2 Schematic diagram of WRP method^[15]

由于 WRP 上记录了三维物体的振幅和相位信息, 因此从 WRP 到全息图的计算相当于直接从三维物体到全息图的计算。当 WRP 与三维物体距离比较近且三维物体深度范围较小时, 受最大衍射角的限制, 每个物点发出的球面波只覆盖 WRP 上的小部分区域。换句话说, 相比直接利用点源法计算三维物体的

全息图, WRP 上的计算量要小得多, 因此 WRP 法可以有效地提升计算速度。随后, Shimobaba 等^[16]又将 LUT 法应用在 WRP 的计算中, 进一步提升了 WRP 法的计算速度。此后, 学者们依次提出双 WRP^[17]、多 WRP^[18]、倾斜 WRP^[19] 等多种基于 WRP 的全息图计算方法以解决在应用 WRP 法时遇到的困难, 如对于大深度三维物体计算效率的下降、遮挡问题的剔除等。

2.2 面元法

面元法将三维物体离散为多边形的集合, 通常为三角形集合, 以利用计算机图形学中的照明、阴影、纹理等渲染信息添加方法来丰富场景的表现能力, 其原理如图 3^[20]所示。面元法的采样数量取决于三维物体的复杂程度以及建模方法的采样精度, 通常会达到数千个, 相比于点源法, 其基元数量减小了 2~3 个数量级。在面元法中, 离散面元与全息图平面一般来说是不平行的。因此, 面元法的核心问题是非平行平面之间的衍射计算。目前, 面元法大致可以分为传统面元法、传统解析面元法和三维仿射面元法。

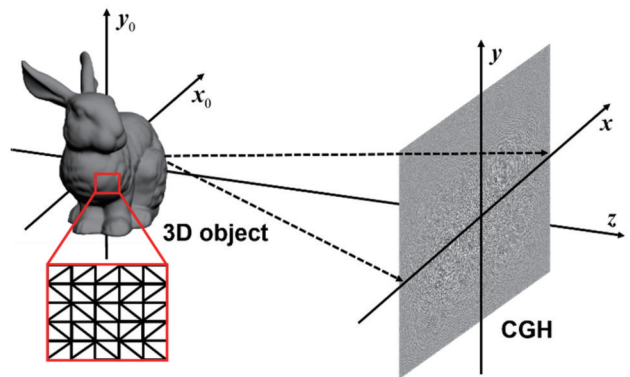


图 3 面元法示意图^[20]

Fig. 3 Schematic diagram of polygon-based method^[20]

传统面元法为每一个离散三角形建立倾斜局部坐标系和平行局部坐标系, 其中倾斜局部坐标系 $(\hat{x}, \hat{y}, \hat{z})$ 的 \hat{z} 轴与离散三角形的法向量平行, 其圆心通常选为

离散三角形的重心。平行局部坐标系 (x_0, y_0, z_0) 与全息图所在的全局坐标系 (x, y, z) 平行, 并与倾斜局部坐标系共享圆心。平行局部坐标系和倾斜局部坐标系可以通过三维旋转相互得到。传统面元法中, 对于每一个离散三角形, 首先在倾斜局部坐标系中对其进行

描绘, 然后进行快速傅里叶变换以得到该离散三角形在倾斜局部坐标系的频谱, 接着通过二维插值将该离散三角形在倾斜局部坐标系的频谱映射到平行局部坐标系, 最后通过角谱法计算该离散三角形在全息图平面上的复振幅分布, 其可表示为

$$H(x, y) = \iint e[\hat{u}(u, v), \hat{v}(u, v)] \exp\{j2\pi[\hat{u}(u, v)x + \hat{v}(u, v)y]\} \cdot J(u, v) \exp\left(jkz\sqrt{1 - \lambda_c^2 u^2 - \lambda_c^2 v^2}\right) dudv, \quad (3)$$

式中: u 和 v 分别为沿 x 和 y 方向的空间频率; \hat{u} 和 \hat{v} 分别为沿 \hat{x} 和 \hat{y} 方向的空间频率; $e(\hat{u}, \hat{v})$ 为全局坐标系下的三角形面元的频谱信息; $J(u, v)$ 为坐标变换 T^{-1} 下的雅可比行列式, 即从倾斜局部坐标系 $(\hat{x}, \hat{y}, \hat{z})$ 到平行局部坐标系 (x_0, y_0, z_0) 的变换可表示为 $k_0 = T^{-1}\hat{k}_0$ 。通过叠加所有离散三角形在全息图平面上的复振幅分布, 就可以计算出三维物体的全息图。传统面元法需要绘制每一个离散三角形, 并为其进行

一次快速傅里叶变换和一次二维线性插值, 因此可以灵活地为每一个离散三角形添加纹理、阴影、反射、散射等信息^[21]。然而, 以上操作都非常耗时, 这大大限制了传统面元法在实时全息三维显示中的应用。

传统解析面元法的核心思想是建立基元三角形与任意三角形之间的频谱关系^[22-23]。传统解析面元法首先定义一个基元三角形并求出其傅里叶频谱的解析表达式 e_Δ , 然后根据频谱的仿射变换公式即可得到任意三角形的解析频谱表达式 e_T , 其可表示为

$$e_T(\hat{u}, \hat{v}) = \frac{1}{|J_\Delta|} \exp\left\{\frac{j2\pi}{J_\Delta}[(a_{22}a_{13} - a_{12}a_{23})\hat{u} + (a_{11}a_{23} - a_{13}a_{21})\hat{v}]\right\} \times e_\Delta\left(\frac{a_{22}\hat{u} - a_{21}\hat{v}}{J_\Delta}, \frac{-a_{21}\hat{u} + a_{11}\hat{v}}{J_\Delta}\right), \quad (4)$$

式中: J_Δ 为仿射坐标变换引入的雅可比行列式。将式(4)代入式(3)即可得到任意三角形在全息图平面上的复振幅分布。之后将所有离散三角形在全息图平面上的复振幅分布进行叠加, 并通过设置每个离散三角形的权重来控制其振幅信息以获得三维物体的全息图。传统解析面元法避免了传统面元法中原始三角形的绘制、快速傅里叶变换和二维插值这些耗时计算, 但由于使用唯一的基元三角形, 其无法单独为每个离散三角形添加随机相位信息, 因此传统解析面元法计算生成的全息图通常被编码为振幅全息图或通过使用复振幅调制技术实现复振幅调制型全息显示。如果要将其编码为纯相位全息图, 需要对离散三角形进行细分, 但这会导致计算量的增加, 从而难以实现实时全息三维显示。

三维仿射传统面元法定义两个坐标系: 局部坐标系和全局坐标系。其中, 全息图位于全局坐标系, 且全息图法向量平行于全局坐标系的 z 轴。三维仿射传统面元法首先在局部坐标系中定义一个具有振幅和相位信息的基元三角形, 并计算其频谱, 然后应用伪逆矩阵求解三维仿射变换矩阵中的核心参数, 最后, 利用三维仿射变换矩阵中的核心参数, 直接获得任意三角形在全息图平面上的频谱^[24]。三维仿射传统面元法由于采用唯一的基元三角形, 因此省略了原始三角形的绘制以及快速傅里叶变换, 但其仍需二维线性插值以实现频谱映射。为了进一步加快计算速度, Pan 等^[25]提出了三维仿射解析面元法。三维仿射传统面元法在局部坐标系中定义一个具有振幅信息的基元三角形并解析

出其频谱, 然后通过三维仿射变换矩阵中的核心参数解析出任意三角形在全息图频谱面上的频谱值。由于基元三角形具有固定的振幅信息且不带有随机相位信息, 因此如果要将三维仿射解析面元法计算生成的全息图编码为纯相位全息图, 需要在计算时将任意三角形划分为多个相似的小三角形, 结合三维仿射变换计算特性, 求解每个小三角形频谱并增加随机相位。随后, Zhang 等^[26]通过一次三维旋转变换和一次二维仿射变换求解三维仿射变换矩阵, 避免了三维仿射面元法中伪逆矩阵的使用, 使计算结果更为精确。受到点源法中 LUT 法的启发, Wang 等^[27]将 LUT 法引入面元法中, 通过预先计算任意三角形在全息图平面上的解析频谱并将其存储在 LUT 中, 避免全息图计算过程中指数运算, 同时减少了乘法运算, 从而加快了计算速度。不同面元法计算时间对比结果如图 4^[27]所示, 其中传统解析面元法和三维仿射解析面元法在统计计算时间时没有对离散三角形进行细分。

2.3 层析法

层析法的基本思想是将三维物体分解为一系列平行于全息图的平面, 每个平面作为一个独立的计算单元, 然后利用平行平面之间的衍射计算来获得三维物体的全息图, 其原理如图 5 所示。层析法可以用多种衍射计算方法来实现, 包括角谱法、菲涅耳衍射法等。相比于菲涅耳衍射法, 角谱法没有进行傍轴近似, 重建精度较高。此外, 角谱法中物平面上的采样间隔不随着传播距离的改变而改变, 因此更适用于全息三维显示。基于角谱法的上述优点, Zhao 等^[28]提出了基于角

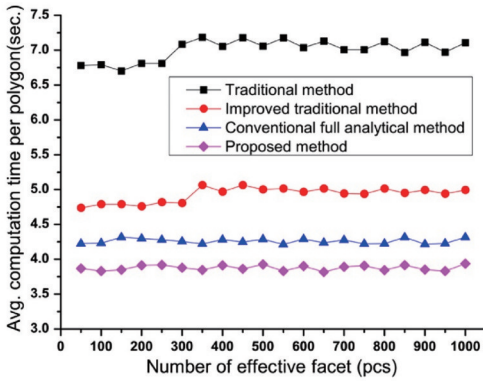


图 4 传统面元法、传统解析面元法、三维仿射传统面元法、三维仿射解析面元法计算时间对比^[27]

Fig. 4 Comparison of computation time among traditional polygon-based method, full analytical polygon-based method, traditional 3D affine transformation polygon-based method, and full analytical 3D affine transformation polygon-based method^[27]

谱的层析法。该方法对每层图像的振幅信息加入随机相位分布来模拟物体表面的漫反射,然后用角谱法计算每层图像在全息图平面上的复振幅分布并进行叠加以获得三维物体的全息图。整个过程可以表示为

$$H(x, y) = \sum_{i=1}^N \mathcal{F}^{-1} \left\{ \mathcal{F} \left[A_i(x, y) \right] \exp \left(jkz_i \sqrt{1 - \lambda^2 u^2 - \lambda^2 v^2} \right) \right\}, \quad (5)$$

式中: $A_i(x, y)$ 为第 i 层的复振幅分布; z_i 为第 i 层物平面到全息图平面的距离。

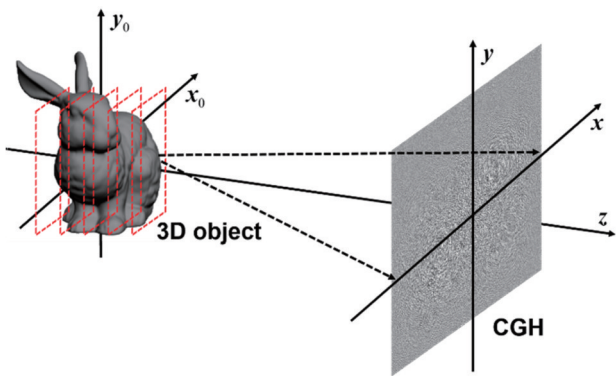


图 5 层析法示意图

Fig. 5 Schematic diagram of layer-based method

相比于点源法和面元法,层析法的基元数量最少,且在分层数量一定的前提下,计算量不会随着三维物体复杂程度的增加而增加,因此大大提高了复杂三维物体的计算速度。然而,不管采用哪种衍射计算方法来完成平行平面之间的衍射计算,二维快速傅里叶变换都是层析法中最基本的运算工具,这也在一定程度上限制了层析法的进一步加速。为了加快层析法的计算速度,Chang 等^[29]提出了一步层析法。该方法对全

息图进行非均匀采样,将其分解成不同采样率的若干部分。在计算全息图时,利用非均匀快速傅里叶变换算法直接完成从均匀采样的物平面到非均匀采样的全息图平面之间的衍射计算。此外, Jia 等^[30]提出了两层析法。该方法考虑到每层图像包含大面积的像素值为零的区域,在计算时首先将非零区域的信息传递到参考平面上,随后再利用菲涅耳衍射法完成从参考平面到全息图平面之间的衍射计算。由于每一层非零区域的面积相对较小,因此两步法大大降低了傅里叶变换的计算量。在计算参考平面的复振幅分布时,该方法还将二维快速傅里叶变换分解为两个一维快速傅里叶变换,并且不对零像素行或零像素列进行计算,从而进一步降低傅里叶变换的计算量。

3 硬件加速

高性能硬件平台的使用对于加快全息图的计算速度也很重要。传统的全息图算法是在中央处理器 (CPU) 上运行的, 尽管目前 CPU 的处理能力有所提高, 但受到 CPU 串行运算的限制, 在计算复杂三维场景的全息图时计算时间会大幅提升, 远远达不到实时全息三维显示的需求。近年来, 随着硬件设备性能的提升, 越来越多的研究人员利用高性能硬件设备来完成全息图的计算。图形处理器 (GPU) 因其良好的底层硬件架构, 高效的数据并行化处理效率, 以及较低的开发学习成本, 成为研究人员处理海量数据时的首选。目前, 研究人员主要是通过设计适配 GPU 的全息图快速生成算法和优化硬件系统架构来提升全息图的计算速度的。2006 年, Ahrenberg 等^[31]通过使用 GPU 计算集群实现了在 55 ms 完成 2048 个物点到 6400×3072 像点的全息图的计算。2009 年, Pan 等^[11]将 S-LUT 算法在多个 GPU 上运行, 最终在 0.5 s 计算了超 10^4 个物点在分辨率为 1024×768 的全息图上的复振幅分布。2014 年, Jackin 等^[32]解决了基于快速傅里叶变换方法的数据依赖问题, 成功利用多 GPU 实现了基于快速傅里叶变换方法的并行设计, 通过使用 4 块 GPU 和 4 块 CPU 混合分布式计算系统, 在 30 s 内完成了 10^8 物点到 9000×9000 像点的全息图的计算, 其系统架构如图 6^[32]所示。

2018 年, Kim 等^[33]提出了一种基于块并行的全息图并行计算方法。该方法通过分配共享内存和专用线程, 优化 GPU 在计算模式下的线程分配方式, 实现空闲区域最小化来增加计算时间和提高计算速度, 在 GeForce GTX 1080Ti 仅耗时 25.05 ms 便可以完成 10^4 个物点的全息图的计算。2020 年, Blender 等^[34]针对 3D 线型物体设计了一种适用于大规模并行计算架构的高效可计算近似算法, 并利用 RTX 3080 GPU 进行了计算加速, 相对于传统算法, 获得了 70 倍的速度提升。次年, Nishitsuji 等^[35]将波传播的积分计算与代数解决方案相结合, 提出了一种适用于 GPU 并能够实现

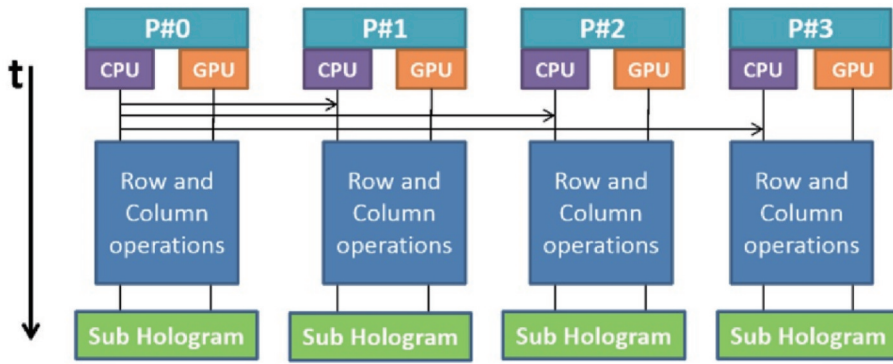


图 6 基于 CPU 和 GPU 的混合架构设计^[32]

Fig. 6 Hybrid architecture based on CPU and GPU^[32]

在不同深度层中计算任意线型对象的算法,该算法在 RTX 2080Ti GPU 上能够在 1.1 ms 内计算分辨率为

1920×1080 的全息图,不同方法计算时间的比较如图 7 所示。

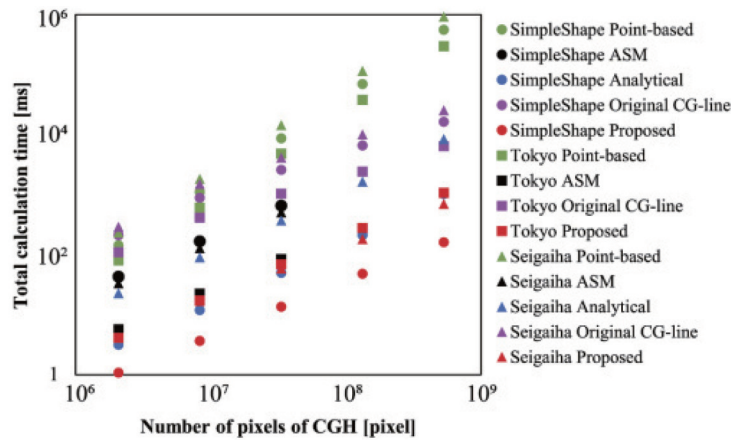


图 7 不同方法计算时间的比较^[35]

Fig. 7 Comparison of calculation time among different methods^[35]

4 全息图编码方法

空间光调制器因具备数字编码与动态刷新的能力而被广泛应用于计算全息三维显示系统中。当前大多数空间光调制器仅能实现单一的振幅或相位调制,为使全息图数据格式能够与空间光调制器相匹配,在完成全息图的计算之后还要将复振幅全息图编码为相应的振幅型全息图或相位型全息图。目前应用最广泛的空间光调制器为硅基液晶(LCOS)和数字微镜器件(DMD)。DMD是由多个小型反射镜面组成的高速光开关阵列,通过控制微镜阵列的翻转频率来实现对入射光场的振幅调制,其通常只能显示二进制全息图,刷新速度比LCOS快大约两个数量级。振幅型全息图可以比较容易地重建出目标图像,但其在计算过程中需要模拟物光波与参考光波的干涉并且只保留干涉场的振幅信息,因此其重建图像会受到零级光和共轭图像的干扰,光能利用率较低。LCOS主要通过液晶排列层中液晶的旋向和入射光的偏振态来实现对入射光场的相位调制,其可以显示灰度全息图,但刷新速度相对

较慢。相位型全息图的生成需在计算机中模拟记录物光波的衍射场,并将其振幅信息转化为常数。相位型全息图可以避免零级光和共轭图像,然而由于振幅信息的丢失,相位型全息图相当于是一个高通滤波器,仅能重建出被广泛扩散在全息图上的高频信息。随机相位可以使目标物体的信息充分地扩散在全息图上,有效地解决上述问题,并被广泛应用在具有宽视角的全息显示系统中。然而,随机相位会导致重建图像中产生严重的散斑噪声,影响全息显示质量。为避免随机相位引起的散斑噪声,在相位型全息图的计算中引入交替投影算法,其通过在不同平面之间投影迭代来更新相位全息图的分布,直到满足设计需求^[36-38]。最典型的交替投影算法是Gerchberg和Saxton^[38]在1972年提出的GS算法,其将物函数置于空域,将全息图置于频域,在迭代过程中分别在空域与频域施加相关约束条件,最终得到满足需求的相位型全息图,整个过程如图8^[38]所示。

近几年,非线性最小化算法被引入相位型全息图的编码中,其通过求解定义的损失函数对于相位全息

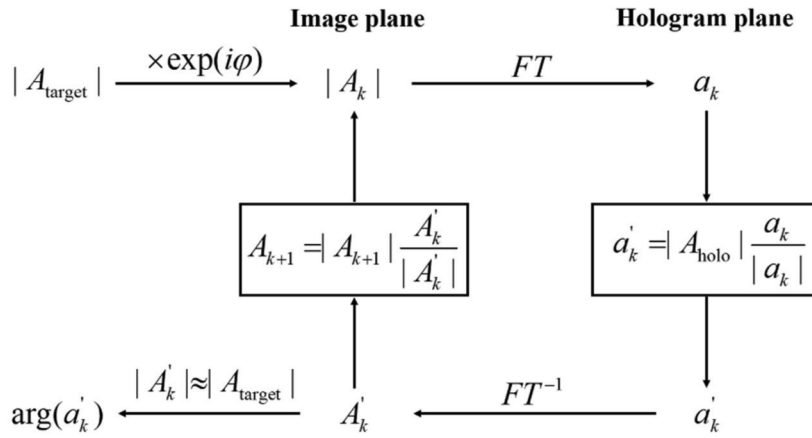


图 8 GS 算法流程图^[38]

Fig. 8 Schematic diagram of GS algorithm^[38]

图的导数,搜索损失函数梯度下降的方向并寻找相应的相位全息图分布^[39-40],非线性最小化类型算法在搜索过程中覆盖到的解函数范围略大于交替投影算法,因此在同等全息图参数与衍射重建模型下,其往往能够得到准确度更高、噪声更低的重建结果。

交替投影算法和非线性最小化算法都是通过像平面和全息图平面之间进行迭代来更新相位全息图的分布。不同于上述两种方法,误差扩散算法在全息图平面像素之间进行顺序迭代进而完成相位型全息图的编码^[41-42]。误差扩散法按照一定的扫描顺序依次扫描全息图平面上的每个像素点,然后计算每个像素点因振幅信息被移除而产生的误差,并将该误差按照一定的比例扩散至尚未扫描到的像素点。在整个迭代过程中,误差扩散算法不需要目标图像的信息。然而,无论是交替投影算法,非线性最小化算法还是误差扩散算法,都需要通过迭代来编码相位型全息图,因此难以实现实时全息三维显示。

相比于上述迭代编码方法,复振幅-相位编码方法因其高运算效率,被认为是极具潜力的实现实时全息三维显示的方法。任意复振幅分布在振幅被归一化后,都可以被分解为两个具有均匀振幅的纯相位分布,表示为

$$h = A \exp(j\theta) = \frac{\exp[j(\theta + \arccos A)]}{2} + \frac{\exp[j(\theta - \arccos A)]}{2}, \quad (6)$$

式中: $h = A \exp(j\theta)$ 为任意振幅被归一化的复振幅分布。

尽管空间光调制器无法实现复振幅调制,但被分解的两个具有均匀振幅的纯相位分布满足相位型空间光调制器的调制特性。受此启发,Kreis^[43]在2016年搭建基于双相位空间光调制器的干涉复振幅全息显示系

统。Kreis将迈克耳孙干涉仪中的反射镜替换为相位型空间光调制器,并将分解后的两个相位全息图分别加载至这两个空间光调制器上,实现了复振幅调制型全息显示。基于双相位空间光调制器的全息显示系统需要两个空间光调制器精确对准,同时对环境要求较为苛刻,难以实际应用。2012年,Song等^[44]将分解后的两幅相位全息图分别加载在同一个空间光调制器的不同区域,并利用余弦相位光栅的平移特性将两幅全息图进行干涉叠加,进而实现复振幅调制,大大减小了系统的复杂度。

上述方法虽然通过复振幅-相位编码实现了复振幅调制,并且避免了迭代过程,但需要借助两个空间光调制器或者辅助光学元件来实现全息三维显示,增加了系统的复杂性。除此之外,还可以通过将分解的两个相位全息图进行双相位全息图编码,从而使其适用于相位型空间光调制器,进而实现复振幅调制。双相位全息图按照编码方式的不同可以分为巨像素编码方法和单像素编码方法。前者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,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,如图 9(a)^[45]所示。巨像素编码方法既能够保持对复振幅全息图水平与垂直方向的采样间隔相等,又能通过不同方向的错位补偿,减小双相位全息图中的错位噪声。后者利用互补二维棋盘格对双相位分量进行采样,并将其组合成单像素结构的双相位全息图,如图 9(b)^[46]所示。单像素编码方法保证了从物体到全息图的运算过程中采样点数始终一致,极大地提升了运算效率。双相位全息图在重建时需要依靠滤波系统来滤除空间错位噪声,但由于包络函数分布特性,滤波系统也无法完全消除空间错位噪声。为了不影响编码效率,可以在编码过程中引入权重因子或包络函数来抑制空间错位噪声,进而提升双相位全息图的重建质量^[47-49]。

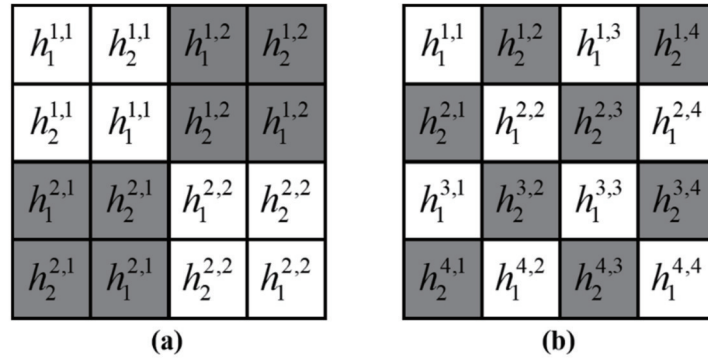


图 9 双相位编码方法。(a)巨像素编码方法^[45];(b)单像素编码方法^[46]

Fig. 9 Double-phase encoding method. (a) Macro pixel encoding method^[45]; (b) single pixel encoding method^[46]

5 深度学习

近几年,随着计算机性能的大幅提升,深度学习的概念被引入到计算全息中,为实现实时全息三维显示提供了新的思路。深度学习通过建立目标图像和全息图或者重建图像之间的物理映射关系,从而构建出神经网络的损失函数,进而实现卷积神经网络参数的快速更新,完成对神经网络的训练。当神经网络被训练完成后,可以通过该网络直接生成与空间光调制器调制方式相匹配的全息图,从而替代波前计算和波前编码两个步骤。有监督学习在训练过程中需要大规模的目标图像以及相应的全息图,通过建立目标图像与其对应全息图的损失函数关系,从而完成对神经网络的训练。当神经网络被训练完成后,便可以通过其快速计算生成满足需求的全息图。这种通过大规模数据学习来表征目标图像和全息图之间有效特征以及映射机制的方法,为实现实时全息三维显示提供了一种非常

有效的技术途径。Horisaki等^[50]使用多个随机相位图像及其传播的散斑状强度图像来训练神经网络,训练后的神经网络可快速计算生成手写数字图像的全息图,其网络架构和重建图像如图 10^[50]所示,计算速度相比传统的GS算法有3倍以上的提升。Lee等^[51]通过使用不同密度的点图像和圆形图像作为训练数据集,并以不同深度的多幅图像作为神经网络的输入进行训练,最终经过训练的神经网络可以仅耗时 234 ms 便可计算出实现多深度重建的全息图。Shi等^[52]使用 4000 对红绿蓝-深度图像和相应的全息图作为训练数据集,并通过可微分的波损失函数进行训练,经过训练得到的神经网络可以生成无散斑、逼真、高分辨率的全息图。Chang等^[53]开发了一个端到端的神经网络,其训练数据集包括 5000 个双目视图和左视图的全息图。由于双目图像可以通过双目相机捕捉,因此该神经网络可以直接从真实世界计算三维场景的全息图,其在 GeForce RTX2060 GPU 上仅需要 0.0150 s 左右便可计算出分

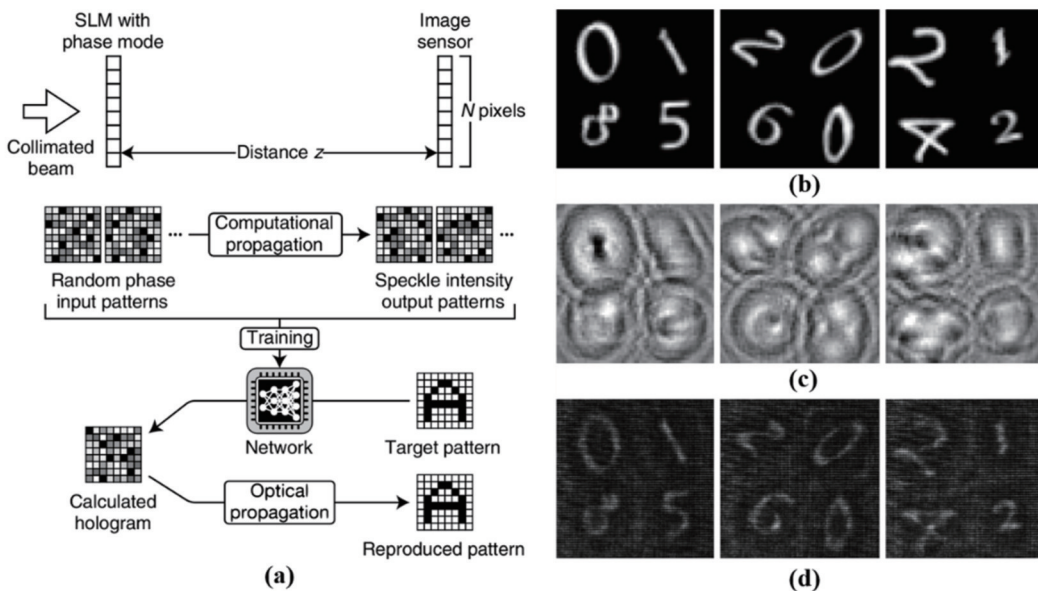


图 10 基于有监督学习的全息图生成算法。(a)原理图;(b)目标图像;(c)全息图;(d)重建图像^[50]

Fig. 10 Hologram generation algorithm based on supervised learning. (a) Schematic diagram; (b) target image; (c) holograms; (d) reconstructed images^[50]

辨率为 1920×1080 的复振幅全息图。随后, Chang 等^[54]再次设计了一种深度神经网络, 其仅依赖单个二维图像便可生成用于全息三维显示的全息图。值得说明的是, Chang 等提出的两个神经网络的输出均为复振幅全息图, 因此在进行光学重建时需要将其编码为双相位全息图以匹配空间光调制器的调制方式。

无监督学习在神经网络中引入了物理衍射传播模型, 因此其不需要训练数据集, 便可以直接生成全息图。Eybposh 等^[55]提出一种基于无监督深度学习的方法来生成计算全息图。该方法用经过训练的神经网络生成目标复振幅分布, 之后通过光场的反向传播计算出相应的相位全息图。Wu 等^[56]提出一种基于自动编

码器的神经网络, 即全息编码器。全息编码器通过将物理衍射传播模型结合到网络的训练中, 能够以无监督的方式学习到目标图像与全息图之间的物理映射关系。该全息编码器能够在 0.15 s 内生成高保真度分辨率为 4K 的全息图, 其网络架构和重建结果如图 11^[56]所示。Shui 等^[57]提出一种基于衍射模型的神经网络框架用来生成三维物体的全息图。该框架包含两个卷积神经网络, 第一个卷积神经网络将红绿蓝-深度图像转换为目标复振幅分布, 第二个卷积神经网络将全息图平面的复振幅分布转换为纯相位全息图。该网络框架可以在 0.17 s 内生成重建信噪比高达 25.08 dB 的计算全息图。

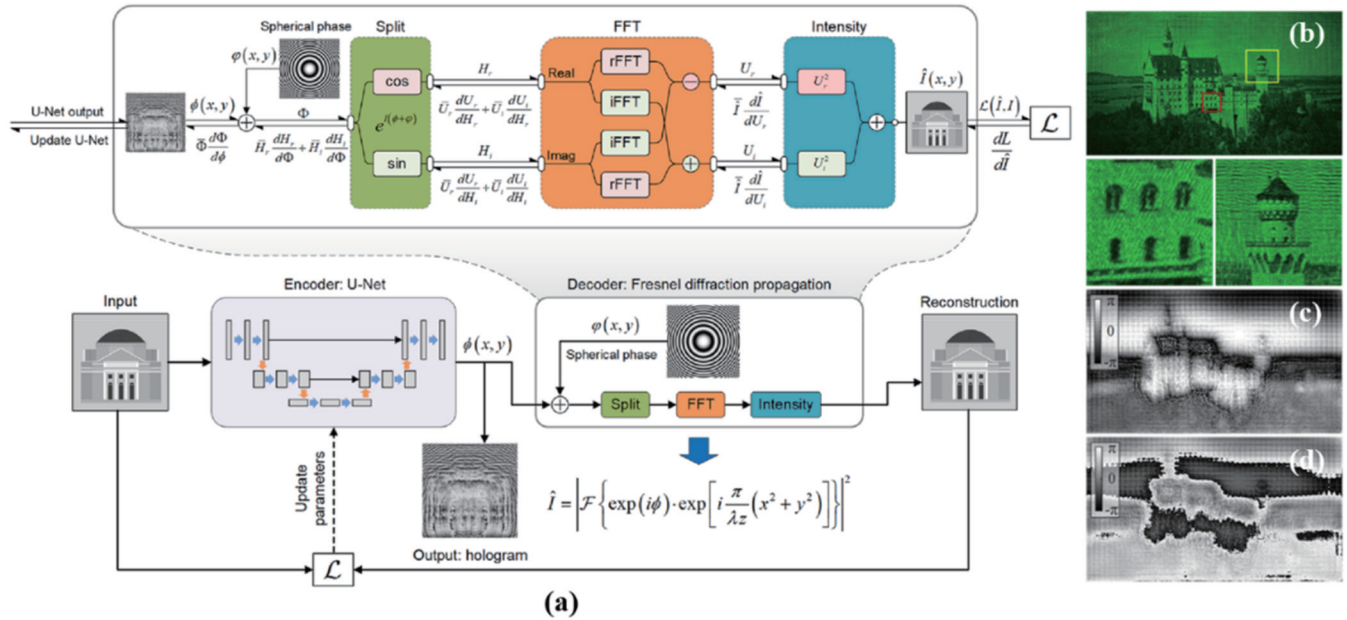


图 11 基于无监督学习的全息图生成算法。(a)全息编码器的结构;(b)重建强度分布;(c)全息图;(d)重建相位分布^[56]

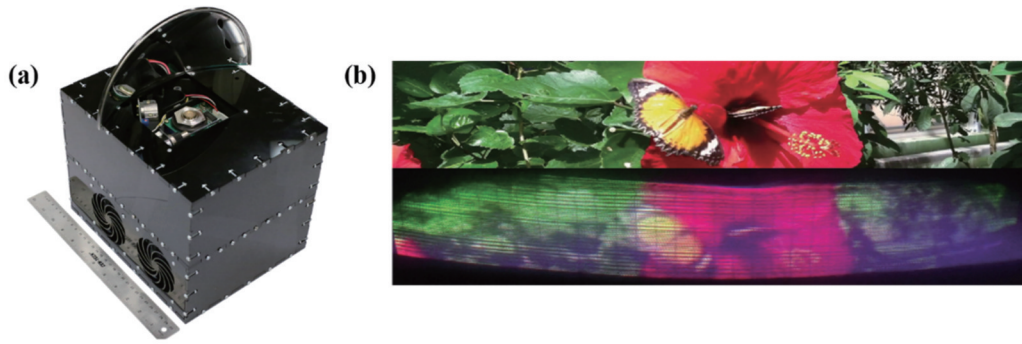
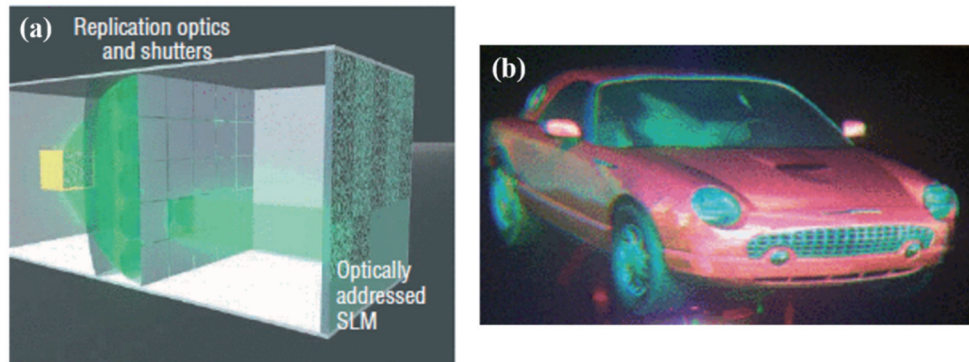
Fig. 11 Hologram generation algorithm based on unsupervised learning. (a) Architecture of the holo encoder; (b) reconstructed intensity distribution; (c) hologram; (d) reconstructed phase distribution^[56]

6 全息三维显示系统

美国麻省理工学院媒体实验室在 1989—2018 年间先后推出了五代以高速声光调制器为核心的 Mark 系列全息显示系统。Mark 系列全息显示系统的核心研究理念是利用声光调制器高时域带宽换取高空域带宽, 即全息图空间带宽积, 从而实现全息图拼接重建大尺寸大视场角三维场景。其中: Mark-I^[58]可以显示尺寸为 25 mm×25 mm×25 mm、视场角为 15°、刷新率为 20 Hz 的单色或彩色图像; Mark-II^[59]显示的彩色图像尺寸为 150 mm×75 mm×150 mm, 视场角可达 30°; Mark-III^[60]显示的彩色图像尺寸为 80 mm×60 mm×80 mm, 视场角为 24°。随后推出的 Mark-IV^[61]和 Mark-V^[62]对硬件结构进行进一步的改进并逐步降低全息显示器的成本和体积, 其中 Mark-V 的系统结构和显示效果如图 12^[62]所示。

从 2000 年开始, 英国剑桥大学高级光子和电子技术中心与英国 Qinetiq 公司利用电寻址液晶空间光调制器和光寻址双稳态液晶空间光调制器的拼接技术, 研制了一台名为 Active Tiling 的彩色全息视频显示系统^[63-64]。该系统采用 400 个 CPU 完成并行运算过程以及 4×4 的光寻址液晶空间光调制器拼接, 实现了刷新频率为 30 Hz、全息图尺寸为 136 mm×36 mm、像素总量超过 10⁸、显示宽度大于 300 mm 的全视差彩色三维图像, 其系统结构和显示效果如图 13^[63-64]所示。

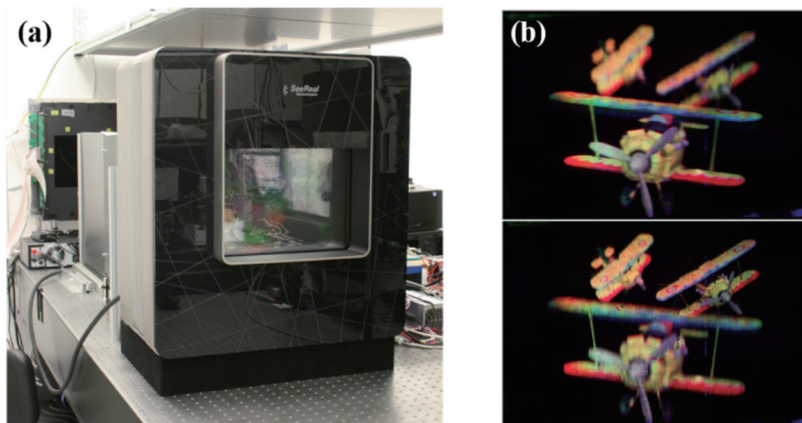
2007 年以来, 德国 SeeReal 公司推出一种新的全息显示系统以实现超大尺寸的全息三维显示^[65-66]。在该系统中, 人眼可以通过视窗观察到重建的三维场景, 当人眼移动时, 视窗也会通过人眼追踪系统移动到新的位置。2017 年该公司推出的最新系统利用全息光学元件替代了体积庞大的传统光学元件从而使显示系

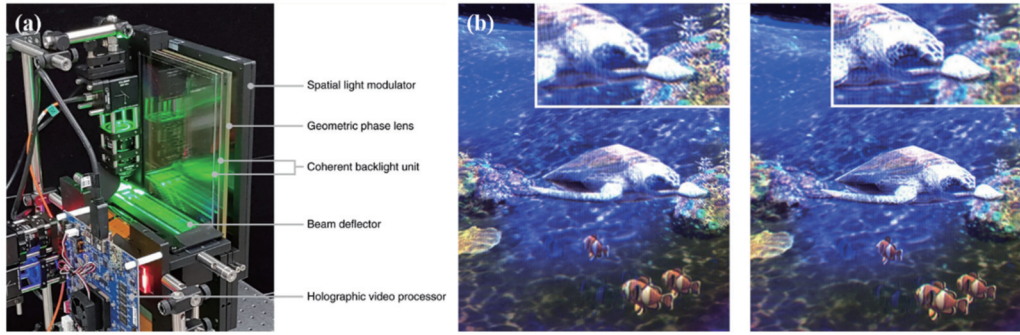
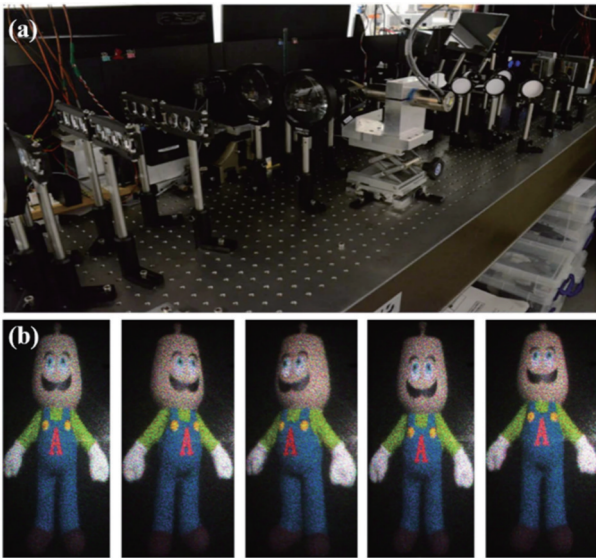
图 12 Mark-V 全息显示系统。(a) 系统结构; (b) 显示效果^[62]Fig. 12 Mark-V holographic display system. (a) System architecture; (b) display effect^[62]图 13 Active Tiling 全息显示系统。(a) 系统结构; (b) 显示效果^[63-64]Fig. 13 Active Tiling holographic display system. (a) System architecture; (b) display effect^[63-64]

统紧凑化,并通过逐像素对齐相位型空间光调制器和振幅型空间光调制器实现了复振幅调制,最终在一个 $300\text{ mm} \times 200\text{ mm}$ 的全息三维显示器上实现了彩色全息三维显示,其系统结构和显示效果如图 14^[67-68] 所示。

2020 年,三星先进技术研究所设计了一套基于相干背光单元和全息视频处理器的超薄全息显示系统^[69]。其中,相干背光单元使系统的有效空间带宽积扩大了 30 倍,全息视频处理器能够以 30 frame/s 的速

度计算 4K 超高清全息图。整个系统的厚度小于 10 cm,显示部件的厚度为 1 cm,可实现彩色实时全息三维显示,系统的结构和重建效果如图 15^[69] 所示。2022 年,英国剑桥大学光子器件与传感器研究中心和迪斯尼研究院合作提出了一种“全息砖”的概念,并采用两个动态全息砖的全光空间拼接搭建了一套动态彩色全息显示系统^[70]。该系统可以实现视场角为 40° 、帧率为 24 frame/s 的全彩色的动态全息显示,其显示系统和重建结果如图 16^[70] 所示。

图 14 大尺寸全息显示系统。(a) 系统结构; (b) 显示效果^[67-68]Fig. 14 Large size holographic display system. (a) System architecture; (b) display effect^[67-68]

图 15 超薄全息显示系统。(a)系统结构;(b)显示效果^[69]Fig. 15 Ultra-thin holographic display system. (a) System architecture; (b) display effect^[69]图 16 基于“全息砖”的彩色全息显示系统。(a)系统结构;(b)显示效果^[70]Fig. 16 Color holographic display system based on "Holobricks". (a) System architecture; (b) display effect^[70]

7 结束语

本文综述了以实时全息三维显示为目标的一系列创新工作。在全息图计算方面,点源法原理简单且应用广泛,但其离散基元过多,计算时间较长。研究人员从LUT和WRP两个方向对点源法进行了加速并取得了一定效果,同时利用加速硬件平台,能够在一定条件下实现实时全息三维显示。面元法比点源法的离散基元少,但其每个离散基元的计算相比点源法来说也更复杂。传统面元法可以对每个离散面元添加纹理、阴影、反射、散射等信息,但其在计算过程中需要对每个面元进行描绘、快速傅里叶变换和二维插值,非常耗时。传统解析面元法减少了传统面元法中的耗时操作,从而提高了计算速度,但其需要进行额外的纹理、散射等信息的添加,这同样非常耗时。三维仿射面元法可根据基元三角形直接获得任意三角形在全息图平面上的频谱,进一步加快了计算速度,然而其在计算时

间和面元的纹理信息之间仍存在着制约关系。目前虽然已经开发了许多基于物理模型的三维计算全息快速算法,但这些方法都有着自身的不足,且受到中央处理器串行运算的限制,在处理复杂三维场景时计算时间将会显著增加,远远达不到实时全息三维显示的需求。近年来,随着硬件设备性能的提升,将现有算法与高性能计算设备相结合逐渐成为实现实时全息三维显示的未来发展趋势^[71]。计算全息的波前编码方法与空间光调制器的调制能力紧密相关,复振幅编码方法相比交替投影算法和非线性最小化算法,其兼备了高运算效率与高重建精度,因此更适合应用于实时全息三维显示。深度学习可以灵活地更新和调整神经网络的架构和训练算法,具有极高的运算效率,为实现实时全息三维显示提供了新的思路。不过,深度学习相当于是一个黑箱子,我们很难获得一些关于它如何进行计算的见解,也很难保证某个卷积神经网络是否可以应用于所有目标场景。此外,以超表面为代表的新型调制器件的发展也为计算全息的波前编码提供了新角度与新手段^[72]。未来,计算全息三维显示的发展需要软硬并进,在设计更具实时性全息图算法和波前编码方法的同时也需要开发高性能运算平台和调制器件。随着计算全息算法和硬件设备的发展,相信在不久的将来,全息显示会真正走入我们的生活中。

参 考 文 献

- [1] Pi D P, Liu J, Wang Y T. Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display[J]. *Light: Science & Applications*, 2022, 11: 231.
- [2] Blinder D, Birbaum T, Ito T, et al. The state-of-the-art in computer generated holography for 3D display[J]. *Light: Advanced Manufacturing*, 2022, 3(3): 35.
- [3] Gabor D. A new microscopic principle[J]. *Nature*, 1948, 161 (4098): 777-778.
- [4] Leith E N, Upatnieks J. Reconstructed wavefronts and communication theory[J]. *Journal of the Optical Society of America*, 1962, 52(10): 1123-1130.
- [5] Leith E N, Upatnieks J. Wavefront reconstruction with continuous-tone objects[J]. *Journal of the Optical Society of America*, 1963, 53(12): 1377-1381.
- [6] Leith E N, Upatnieks J. Wavefront reconstruction with diffused

- illumination and three-dimensional objects[J]. *Journal of the Optical Society of America*, 1964, 54(11): 1295-1301.
- [7] Jiao S M, Zhuang Z Y, Zou W B. Fast computer generated hologram calculation with a mini look-up table incorporated with radial symmetric interpolation[J]. *Optics Express*, 2017, 25(1): 112-123.
- [8] Tsang P W M, Poon T C, Wu Y M. Review of fast methods for point-based computer-generated holography[J]. *Photonics Research*, 2018, 6(9): 837-846.
- [9] Lucente M E. Interactive computation of holograms using a look-up table[J]. *Journal of Electronic Imaging*, 1993, 2(1): 28-34.
- [10] Kim S C, Kim E S. Effective generation of digital holograms of three-dimensional objects using a novel look-up table method[J]. *Applied Optics*, 2008, 47(19): D55-D62.
- [11] Pan Y C, Xu X W, Solanki S, et al. Fast CGH computation using S-LUT on GPU[J]. *Optics Express*, 2009, 17(21): 18543-18555.
- [12] Jia J A, Wang Y T, Liu J A, et al. Reducing the memory usage for effective computer-generated hologram calculation using compressed look-up table in full-color holographic display[J]. *Applied Optics*, 2013, 52(7): 1404-1412.
- [13] Gao C, Liu J, Li X, et al. Accurate compressed look up table method for CGH in 3D holographic display[J]. *Optics Express*, 2015, 23(26): 33194-33204.
- [14] Pi D P, Liu J A, Kang R D, et al. Reducing the memory usage of computer-generated hologram calculation using accurate high-compressed look-up-table method in color 3D holographic display [J]. *Optics Express*, 2019, 27(20): 28410.
- [15] Shimobaba T, Masuda N, Ito T. Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane[J]. *Optics Letters*, 2009, 34(20): 3133-3135.
- [16] Shimobaba T, Nakayama H, Masuda N, et al. Rapid calculation algorithm of Fresnel computer-generated-hologram using look-up table and wavefront-recording plane methods for three-dimensional display[J]. *Optics Express*, 2010, 18(19): 19504-19509.
- [17] Phan A H, Piao M L, Gil S K, et al. Generation speed and reconstructed image quality enhancement of a long-depth object using double wavefront recording planes and a GPU[J]. *Applied Optics*, 2014, 53(22): 4817-4824.
- [18] Hasegawa N, Shimobaba T, Kakue T, et al. Acceleration of hologram generation by optimizing the arrangement of wavefront recording planes[J]. *Applied Optics*, 2016, 56(1): A97-A103.
- [19] Arai D, Shimobaba T, Murano K, et al. Acceleration of computer-generated holograms using tilted wavefront recording plane method[J]. *Optics Express*, 2015, 23(2): 1740-1747.
- [20] Zhang Y P, Fan H X, Wang F, et al. Polygon-based computer-generated holography: a review of fundamentals and recent progress[J]. *Applied Optics*, 2022, 61(5): B363-B374.
- [21] Pan Y J, Wang Y T, Liu J A, et al. Analytical brightness compensation algorithm for traditional polygon-based method in computer-generated holography[J]. *Applied Optics*, 2013, 52(18): 4391-4399.
- [22] Ahrenberg L, Benzie P, Magnor M, et al. Computer generated holograms from three dimensional meshes using an analytic light transport model[J]. *Applied Optics*, 2008, 47(10): 1567-1574.
- [23] Kim H, Hahn J, Lee B. Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography[J]. *Applied Optics*, 2008, 47(19): D117- D127.
- [24] Pan Y J, Wang Y T, Liu J A, et al. Fast polygon-based method for calculating computer-generated holograms in three-dimensional display[J]. *Applied Optics*, 2012, 52(1): A290-A299.
- [25] Pan Y J, Wang Y T, Liu J A, et al. Improved full analytical polygon-based method using Fourier analysis of the three-dimensional affine transformation[J]. *Applied Optics*, 2014, 53(7): 1354-1362.
- [26] Zhang Y P, Wang F, Poon T C, et al. Fast generation of full analytical polygon-based computer-generated holograms[J]. *Optics Express*, 2018, 26(15): 19206-19224.
- [27] Wang F, Shimobaba T, Zhang Y P, et al. Acceleration of polygon-based computer-generated holograms using look-up tables and reduction of the table size via principal component analysis[J]. *Optics Express*, 2021, 29(22): 35442-35455.
- [28] Zhao Y, Cao L C, Zhang H, et al. Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method[J]. *Optics Express*, 2015, 23(20): 25440-25449.
- [29] Chang C L, Xia J, Lei W. One step hologram calculation for multi-plane objects based on nonuniform sampling (Invited Paper) [J]. *Chinese Optics Letters*, 2014, 12(6): 060020.
- [30] Jia J A, Si J, Chu D P. Fast two-step layer-based method for computer generated hologram using sub-sparse 2D fast Fourier transform[J]. *Optics Express*, 2018, 26(13): 17487-17497.
- [31] Ahrenberg L, Benzie P, Magnor M, et al. Computer generated holography using parallel commodity graphics hardware[J]. *Optics Express*, 2006, 14(17): 7636-7641.
- [32] Jackin B J, Miyata H, Ohkawa T, et al. Distributed calculation method for large-pixel-number holograms by decomposition of object and hologram planes[J]. *Optics Letters*, 2014, 39(24): 6867-6870.
- [33] Kim D W, Lee Y H, Seo Y H. High-speed computer-generated hologram based on resource optimization for block-based parallel processing[J]. *Applied Optics*, 2018, 57(13): 3511-3518.
- [34] Blinder D, Nishitsuji T, Schelkens P. Real-time computation of 3D wireframes in computer-generated holography[J]. *IEEE Transactions on Image Processing*, 2021, 30: 9418-9428.
- [35] Nishitsuji T, Blinder D, Kakue T, et al. GPU-accelerated calculation of computer-generated holograms for line-drawn objects[J]. *Optics Express*, 2021, 29(9): 12849-12866.
- [36] Chang C L, 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.
- [37] Fienup J R. Iterative method applied to image reconstruction and to computer-generated holograms[J]. *Optical Engineering*, 1980, 19(3): 297-305.
- [38] Gerchberg R W, Saxton O W. A practical algorithm for the determination of phase from image and diffraction plane pictures [J]. *Optik*, 1972, 2(35): 237-246.
- [39] 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.
- [40] Zhang J Z, Pégard N, Zhong J S, et al. 3D computer-generated holography by non-convex optimization[J]. *Optica*, 2017, 4(10): 1306-1313.
- [41] 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.
- [42] Liu K X, He Z H, Cao L C. Pattern-adaptive error diffusion algorithm for improved phase-only hologram generation[J]. *Chinese Optics Letters*, 2021, 19(5): 050501.
- [43] Kreis T. 3-D display by referenceless phase holography[J]. *IEEE Transactions on Industrial Informatics*, 2016, 12(2): 685-693.
- [44] Song H, Sung G, Choi S, et al. Optimal synthesis of double-phase computer generated holograms using a phase-only spatial light modulator with grating filter[J]. *Optics Express*, 2012, 20(28): 29844-29853.
- [45] Arrizón V, Sánchez-de-la-Llave D. Double-phase holograms implemented with phase-only spatial light modulators: performance evaluation and improvement[J]. *Applied Optics*, 2002, 41(17): 3436-3447.
- [46] Mendoza-Yero O, Mínguez-Vega G, Lancis J. Encoding complex fields by using a phase-only optical element[J]. *Optics Letters*, 2014, 39(7): 1740-1743.

- [47] Kim Y K, Lee J S, Won Y H. Low-noise high-efficiency double-phase hologram by multiplying a weight factor[J]. *Optics Letters*, 2019, 44(15): 3649-3652.
- [48] Sui X M, He Z H, Jin G F, 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.
- [49] Sui X M, He Z H, Jin G F, et al. Spectral-envelope modulated double-phase method for computer-generated holography[J]. *Optics Express*, 2022, 30(17): 30552-30563.
- [50] Horisaki R, Takagi R, Tanida J. Deep-learning-generated holography[J]. *Applied Optics*, 2018, 57(14): 3859-3863.
- [51] Lee J, Jeong J, Cho J, et al. Deep neural network for multi-depth hologram generation and its training strategy[J]. *Optics Express*, 2020, 28(18): 27137-27154.
- [52] Shi L, Li B C, Kim C, et al. Towards real-time photorealistic 3D holography with deep neural networks[J]. *Nature*, 2021, 591(7849): 234-239.
- [53] Chang C L, Wang D, Zhu D C, et al. Deep-learning-based computer-generated hologram from a stereo image pair[J]. *Optics Letters*, 2022, 47(6): 1482-1485.
- [54] Chang C L, Zhu D C, Li J M, et al. Three-dimensional computer holography enabled from a single 2D image[J]. *Optics Letters*, 2022, 47(9): 2202-2205.
- [55] Eybposh M H, Caira N W, Atisa M, et al. DeepCGH: 3D computer-generated holography using deep learning[J]. *Optics Express*, 2020, 28(18): 26636-26650.
- [56] Wu J C, Liu K X, Sui X M, et al. High-speed computer-generated holography using an autoencoder-based deep neural network[J]. *Optics Letters*, 2021, 46(12): 2908-2911.
- [57] Shui X H, Zheng H D, Xia X, et al. Diffraction model-informed neural network for unsupervised layer-based computer-generated holography[J]. *Optics Express*, 2022, 30(25): 44814-44826.
- [58] St-Hilaire P, Benton S A, Lucente M E, et al. Electronic display system for computational holography[J]. *Proceedings of SPIE*, 1990, 1212: 174-182.
- [59] St-Hilaire P, Benton S A, Lucente M E, et al. Advances in holographic video[J]. *Proceedings of SPIE*, 1993, 1914: 188-196.
- [60] Smalley D E, Smithwick Q Y J, Bove V M Jr. Holographic video display based on guided-wave acousto-optic devices[J]. *Proceedings of SPIE*, 2007, 6488: 64880L.
- [61] Smalley D E, Smithwick Q Y J, Bove V M, et al. Anisotropic leaky-mode modulator for holographic video displays[J]. *Nature*, 2013, 498(7454): 313-317.
- [62] Henrie A, Codling J R, Gneiting S, et al. Hardware and software improvements to a low-cost horizontal parallax holographic video monitor[J]. *Applied Optics*, 2018, 57(1): A122-A133.
- [63] Slinger C, Cameron C, Stanley M. Computer-generated holography as a generic display technology[J]. *Computer*, 2005, 38(8): 46-53.
- [64] Stanley M, Conway P B, Coomber S D, et al. Novel electro-optic modulator system for the production of dynamic images from giga-pixel computer-generated holograms[J]. *Proceedings of SPIE*, 2000, 3956: 13-22.
- [65] Häussler R, Reichelt S, Leister N, et al. Large real-time holographic displays: from prototypes to a consumer product[J]. *Proceedings of SPIE*, 2009, 7237: 72370S.
- [66] Leister N, Schwerdtner A, Fütterer G, et al. Full-color interactive holographic projection system for large 3D scene reconstruction[J]. *Proceedings of SPIE*, 2008, 6911: 69110V.
- [67] Häussler R, Leister N, Stolle H. Large holographic 3D display for real-time computer-generated holography[J]. *Proceedings of SPIE*, 2017, 10335: 103350X.
- [68] Häussler R, Gritsai Y, Zschau E, et al. Large real-time holographic 3D displays: enabling components and results[J]. *Applied Optics*, 2017, 56(13): F45-F52.
- [69] An J, Won K, Kim Y, et al. Slim-panel holographic video display[J]. *Nature Communications*, 2020, 11: 5568.
- [70] Li J, Smithwick Q, Chu D P. Holobricks: modular coarse integral holographic displays[J]. *Light: Science & Applications*, 2022, 11: 57.
- [71] Zhang Y X, Liu J A, Li X, et al. Fast processing method to generate gigabyte computer generated holography for three-dimensional dynamic holographic display[J]. *Chinese Optics Letters*, 2016, 14(3): 030901.
- [72] Liu L X, Zhang X Q, Kenney M, et al. Broadband metasurfaces with simultaneous control of phase and amplitude[J]. *Advanced Materials*, 2014, 26(29): 5031-5036.

Research Progress of Real-Time Holographic 3D Display Technology

Liu Juan^{1,2,3*}, Pi Dapu^{1,2,3}, Wang Yongtian^{1,2,3}

¹*School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China;*

²*Beijing Engineering Research Center for Mixed Reality and Advanced Display Technology, Beijing 100081, China;*

³*Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education, Beijing 100081, China*

Abstract

Significance Holographic three-dimensional (3D) display technology can effectively reconstruct the wavefront of 3D objects and provide whole depth cues for human eyes, so it has become a research hotspot in the 3D display field. Compared with optical holography, computer-generated holography simulates the recording process of the hologram by computers and adopts the refreshable spatial light modulator instead of holographic recording material as the hologram-carrying media. Due to the above characteristics, computer-generated holography becomes an ideal technology to realize real-time holographic 3D displays and has a broad application prospect in military navigation, industrial manufacturing,

medical treatment, education, and entertainment fields. At present, the development of real-time holographic 3D displays is hindered by the huge data of 3D objects, the insufficient modulation capacity of spatial light modulators, and the low display degree of holographic 3D display systems. In order to overcome these problems, researchers have made many innovations from both algorithm and hardware aspects.

Progress We review the progress of real-time holographic 3D displays. Firstly, the basic principle and development history of holography are outlined. Next, the fast calculation methods of computer generated holograms (CGHs) and wavefront coding methods for current spatial light modulators are introduced in detail. Then, the contribution of deep learning to real-time holographic 3D displays is discussed, and some typical holographic display systems are introduced. Finally, the future development of real-time holographic 3D displays is prospected. The fast calculation methods can be classified into algorithm optimization and hardware acceleration. The algorithm optimization mainly simplifies the calculation complexity and reduces the redundant computation of traditional calculation methods, including point-based method, polygon-based method, and layer-based method. Hardware acceleration mainly speeds up the CGH calculation by designing fast calculation algorithms adapted to the hardware platform and optimizing hardware system architectures. The wavefront coding methods for current spatial light modulators can be mainly classified into iterative methods and non-iterative methods. Iterative methods solve the desired phase-only hologram by iterative calculation between the image plane and the hologram plane or pixels in the hologram plane, which are time-consuming. Non-iterative methods convert the diffracted complex wavefront to an intensity-constant distribution analytically. Compared with iterative methods, non-iterative methods are more efficient and suitable for real-time holographic 3D displays. In recent years, deep learning is also introduced into the computer-generated holography field. Deep learning completes the CGH calculation and wavefront coding through the trained neural network, which shows great potential for realizing real-time holographic 3D displays. Furthermore, with the development of algorithms, devices, and systems, the holographic display system is gradually developing towards large size, large field of view, and real-time color display.

Conclusions and Prospects Real-time holographic 3D display is the ultimate goal of the holographic 3D display. Although there is still a long way to go, it is believed that there is great potential for the further development of real-time holographic 3D displays in both software (algorithms) and hardware (devices and systems). It is expected that holographic 3D displays will eventually achieve real-time display and come gradually into the market and daily life, thus bringing revolutionary changes to our future life.

Key words holographic three-dimensional display; computer-generated holography; hologram; deep learning